(12) STANDARD PATENT (11) Application No. **AU 2007226231 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Novel isolated and purified strains of Chikungunya virus and polynucleotides and polypeptides sequences, diagnostic and immunogenical uses thereof

(51) International Patent Classification(s)
C12N 9/00 (2006.01) ***C07K 14/18*** (2006.01)

(21) Application No: **2007226231** (22) Date of Filing: **2007.03.15**

(87) WIPO No: **WO07/105111**

(30) Priority Data

(31) Number	(32) Date	(33) Country
2,545,597	**2006.04.04**	**CA**
2,538,898	**2006.03.15**	**CA**

(43) Publication Date: **2007.09.20**
(44) Accepted Journal Date: **2013.07.11**

(71) Applicant(s)
Institut Pasteur

(72) Inventor(s)
Despres, Philippe;Marechal, Valerie;Charneau, Pierre;Brehin, Anne-Claire;Souque, Philippe

(74) Agent / Attorney
Fisher Adams Kelly, GPO Box 1413, BRISBANE, QLD, 4001

(56) Related Art
Biosis Abstract Accession 1999:076525, Yuzhen et al, VIROLOGICA SINICA, 1998, 13(3):222-225

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 September 2007 (20.09.2007)

PCT

(10) International Publication Number
WO 2007/105111 A3

(51) International Patent Classification:
C07K 14/18 (2006.01)

(21) International Application Number:
PCT/IB2007/001716

(22) International Filing Date: 15 March 2007 (15.03.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

2,538,898	15 March 2006 (15.03.2006)	CA
2,545,597	4 April 2006 (04.04.2006)	CA

(71) Applicant *(for all designated States except US)*: **INSTITUT PASTEUR** [FR/FR]; 28 rue du Docteur Roux, F-75015 Paris (FR).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **DESPRES, Philippe** [FR/FR]; 18 Place de la Liberté, F-92250 La Garenne Colombes (FR). **BREHIN, Anne-Claire** [FR/FR]; 3 rue de Nemours, F-75011 Paris (FR). **MARECHAL, Valérie** [FR/FR]; 18 avenue des Chardons, F-94800 Villejuif (FR). **CHARNEAU, Pierre** [FR/FR]; 6 bis rue des Ecoles, F-75005 Paris (FR). **SOUQUE, Philippe** [FR/FR]; 48 rue Henri Guillaumet, F-78370 Plaisir (FR).

(74) Agent: **CABINET ORES**; 36 rue de St Pétersbourg, F-75008 Paris (FR).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
27 March 2008

WO 2007/105111 A3

(54) Title: NOVEL ISOLATED AND PURIFIED STRAINS OF CHIKUNGUNYA VIRUS AND POLYNUCLEOTIDES AND POLYPEPTIDES SEQUENCES, DIAGNOSTIC AND IMMUNOGENICAL USES THEREOF

(57) Abstract: The present invention concerns wild-strains of Chikungunya virus isolated from patients exhibiting severe forms of infection and stemming from a human arbovirosis epidemy. The present invention also concerns polypeptide sequences and fragment thereof derived from their genome, the polynucleotide encoding same and their use as diagnostic products, as vaccine and/or as immunogenic compositions.

NOVEL ISOLATED AND PURIFIED STRAINS OF CHIKUNGUNYA VIRUS AND POLYNUCLEOTIDES AND POLYPEPTIDES SEQUENCES, DIAGNOSTIC AND IMMUNOGENICAL USES THEREOF

FIELD OF THE INVENTION

The present invention concerns wild-strains of Chikungunya virus isolated from patients exhibiting severe forms of infection and stemming from a human arbovirosis epidemy. The present invention also concerns polypeptide sequences and fragment thereof derived from their genome, the polynucleotide encoding same and their use as diagnostic products, as vaccine and/or as immunogenic compositions.

BACKGROUND OF THE INVENTION

Chikungunya virus (CHIKV) is a mosquito-transmitted *Alphavirus* belonging to family *Togaviridae* [1,2]. It was isolated for the first time from a Tanzanian outbreak in 1952 [3]. It is responsible for an acute infection of abrupt onset, characterized by high fever, arthralgia, myalgia, headache and rash [4,5]. Poly-arthralgia, the pathognomonic sign of the disease, is very painful. Symptoms are generally self-limiting and last 1 to 10 days. However, arthralgia or arthritic symptoms may persist for months or years. In some patients, minor hemorrhagic signs such as epistaxis or gingivorrhagia have also been described.

CHIKV is geographically distributed in Africa, India and South East Asia. In Africa, the virus is maintained through a sylvatic transmission cycle between wild primates and mosquitoes such as *Aedes luteocephalus*, *Ae. furcifer* or *Ae. taylori* [4]. In Asia, CHIKV is mainly transmitted from human to human by *Ae. aegypti* and to a lesser extent by *Ae. albopictus* through an urban transmission cycle. Since the 1952 Tanzania outbreak, CHIKV has caused outbreaks in East Africa

(Tanzania, Uganda), in Austral Africa (Zimbabwe, South Africa), in West Africa (Senegal, Nigeria) and in Central Africa (Central African Republic, Democratic Republic of the Congo) [4]. The most recent epidemic re-emergence was documented in 1999-2000 in Kinshasa, where an estimated 50,000 persons were infected [6]. Since the first documented Asian outbreak in 1958 in Bangkok, Thailand, outbreaks have been documented in Thailand, Cambodia, Vietnam, Laos, Myanmar, Malasia, Philippines and Indonesia [4,5]. The most recent epidemic re-emergence was documented in 2001-2003 in Java after 20 years [7]. Either in Africa or Asia, the re-emergence was unpredictable, with intervals of 7-8 years to 20 years between consecutive epidemics.

Since the end of 2004, Chikungunya virus (CHIKV) has emerged in the islands of the south-western Indian Ocean. Between January and March 2005, more than 5,000 cases were reported in Comoros. Later in 2005, the virus has circulated in the other islands, *i.e* Mayotte, Seychelles, Réunion and Mauritius. Starting in December 2005, the rainy season gave rise to a renewed epidemic circulation of the virus. Between January 1st and March 1st, 2006, 2,553, 3,471, and 4,650 cases have been reported in Mauritius, Mayotte and Seychelles (March 12th, 2006). The most affected island is Reunion with an estimated 212,000 cases until March 12th, 2006 (total population: 770,000). More recently, circulation of the virus has been documented in Madagascar.

In Reunion Island, the first documented cases were patients coming 1 ng back from Comoros in March 2005. More than 3,000 cases were reported from March to June. The transmission was limited during the winter season of the southern hemisphere and a major upsurge has been observed since mid-December, with an estimated 210,000 cases between January and March 2006 [8]. Since March 2005, 85 patients with a confirmed CHIKV infection have developed severe clinical signs (meningoencephalitis or fulminant hepatitis) which justified hospitalization in an intensive care unit. Several cases of meningo-encephalitis and major algic syndrome have been associated with vertical transmission of the virus 9.

To date, two CHIKV complete nucleotide sequences have been determined, for the strains Ross (accession no: AF490259) and S27 [9], both isolated from patients during the 1952 Tanzania outbreak. Another complete nucleotide sequence has been determined for a strain isolated in *Ae. furcifer* during the Senegal 1983 outbreak (accession no AY726732). Khan and coworkers [9] showed that the S27 genome was similar in its structure to that of other alphaviruses and that O'nyong-nyong virus (ONN) was the closest relative to CHIKV. In addition, phylogenetic analyses based on partial E1 sequences from African and Asian isolates revealed the existence of three distinct CHIKV phylogroups, one containing all isolates from West Africa, one containing isolates from Asia, and one corresponding to Eastern, Central and Southern African isolates [10]. Strains isolated in 1999-2000 in the Democratic Republic of the Congo belonged to the latter phylogroup [6].

SUMMARY OF THE INVENTION

An aspect of the invention is to provide new diagnostic and immunologic tools against CHIK virus associated diseases, such as arbovirosis.

Such an aspect is particularly achieved by providing an isolated and purified wild strain of Chikungunya virus (CHIK) capable of *in vitro* infecting human cells; and its use for the detection of a CHIKV associated to an arbovirus, or for the preparation of a composition that prevents and/or treats an arbovirus.

Another aspect of the invention concerns an isolated and purified strain of CHIKV comprising at least one mutation in structural protein E1 and/or structural protein E2; and its use for the detection of a CHIKV associated to an arbovirus, or for the preparation of a composition that prevents and/or treats an arbovirus.

Another aspect of the invention concerns an isolated and purified polynucleotide

comprising all or part of the sequence of SEQ ID NOS: 1, 2, 3, 4, 5 or 6; and its use for the detection of a CHIKV associated to an arbovirus, or for the preparation of a composition that prevents and/or treats an arbovirus.

Another aspect of the invention concerns a fragment of the polynucleotide of the invention wherein it codes for the ectodomain of glycoprotein E2 or E1; and its use for the detection of a CHIKV associated to an arbovirus, or for the preparation of a composition that prevents and/or treats an arbovirus.

Other aspects of the invention concern a vector or plasmid comprising a polynucleotide or fragment contemplated by the present invention, and host cell comprising said vector or plasmid; and its use for the detection of a CHIKV associated to an arbovirus, or for the preparation of a composition that prevents and/or treats an arbovirus.

Yet another aspect of the invention concerns a purified polypeptide encoded by a polynucleotide or fragment of the invention; and its use for the detection of a CHIKV associated to an arbovirus, or for the preparation of a composition that prevents and/or treats an arbovirus.

A further aspect of the invention concerns a monoclonal or polyclonal antibody or fragment thereof that specifically binds to a polypeptide of the invention; and its use for the detection of a CHIKV associated to an arbovirus, or for the preparation of a composition that prevents and/or treats an arbovirus.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1: Localization of the E1 changes on the 3D structure modeled from the crystal structure of SFV E1 [43] [19].

A) Ribbon diagram of E1, with domain I colored red, domain II yellow and domain III, blue. Green tubes mark the disulfide bonds. The fusion peptide, at the tip of

the molecule (in domain II) is colored orange and labeled. The N-terminus and the C-terminus observed in the crystal (which is 30 aa upstream of the trans-membrane region) are also labeled. The 2 unique changes observed in the Indian Ocean isolates are indicated by stars and labeled: positions 226 (white) and 284 (magenta).

B) Partial representation (one octant, slightly extended) of the icosahedral E1 scaffold at the surface of the virion, viewed down a 5-fold symmetry axis. One E1 protomer is highlighted in colors, as in A); all the others are represented in grey. The location of some of the icosahedral symmetry axes are drawn as solid black symbols: pentagon for 5-fold axis, triangle for 3-fold axes, ellipse for 2-fold axes (which in the T=4 lattice of alphaviruses are coincident with quasi 6-fold axes). Open triangles indicate roughly the location of the E2 trimers that interact tightly with E1, covering domain II and the fusion peptide, and presenting the main antigenic sites. The open triangles mark also quasi 3-fold symmetry axes of the T=4 surface icosahedral lattice. A magenta ball marks the location of Glu 284, at an inter-E1 protomer contact site. This contact is propagated 240 times at the surface lattice (note all pink balls drawn on the grey protomers). Note that the fusion peptide, in orange, is pointing up and away from contacts with other E1 protomers. This is more easily seen at the periphery of the virion, where one of them is labeled (FP). In the virion, this region of E1 is not accessible, covered underneath the E2 molecule [19].

Figure 2: Phylogenetic relationships among chikungunya isolates based on partial E1 nucleotide sequences. Isolates from the Indian Ocean outbreak (Reunion, Seychelles, Mayotte, Mauritius, Madagascar) represent a distinct clade within a large East, Central and South African (ECSA) phylogroup. Bootstrap resampling values are indicated at major nodes. The branch leading to West-African phylogroup (of length approx. 15%) was shortened for convenience.

Figure 3: Proposed evolutionary scenario of chikungunya virus isolates from 1 the Indian Ocean outbreak. The scenario is based on six genome sequences

determined by direct sequencing of RT-PCR products obtained using RNA extracts as templates; the sequences thus correspond to consensus sequences (Cons. Seq.) of the possible mixture of coexisting genomes (quasispecies). Inset: number of cases of E1-226A and E1-226V at different time intervals in Reunion Island, based on partial E1 sequences. E1-226V was observed in consensus sequences 2, 3 and 4, and therefore most E1-226V isolates genotyped based on partial E1 sequences are likely related to these genotypes. However, the independent appearance of E1-226V in other genotypes cannot be excluded. Int.: intermediate sequence. The location, size and relative position of the Islands and the African border are indicative. Consensus sequence 1 was obtained from a Reunion patient who traveled back from Comoros in March 2005, and from a Reunion Island patient. Sequences 2 to 4 were sampled in Reunion Island; sequence 5 was sampled in the Seychelles.

Figure 4 shows the nucleotide sequence of the genome of a CHIK virus strain according to a preferred embodiment of the invention and more specifically for the preferred strain named 05.115 (SEQ ID NO:1).

Figure 5 shows the nucleotide sequence of the genome of a CHIK virus strain according to a preferred embodiment of the invention and more specifically for the preferred strain named 05.209 (SEQ ID NO:2).

Figure 6 shows the nucleotide sequence of the genome of a CHIK virus strain according to a preferred embodiment of the invention and more specifically for the preferred strain named 06.21 (SEQ ID NO:3).

Figure 7 shows the nucleotide sequence of the genome of a CHIK virus strain according to a preferred embodiment of the invention and more specifically for the preferred strain named 06.27 (SEQ ID NO:4).

Figure 8 shows the nucleotide sequence of the genome of a CHIK virus strain

according to a preferred embodiment of the invention and more specifically for the preferred strain named 06.49 (SEQ ID NO:5).

Figure 9 shows the nucleotide sequence of the genome of a CHIK virus strain according to a preferred embodiment of the invention and more specifically for the preferred strain named 05.61 (SEQ ID NO:6).

Figure 10 shows a nucleotide sequence of a fragment of a CHIK virus according to a preferred embodiment of the invention, and more specifically a fragment which codes for the ectodomain of the glycoprotein E2 of the preferred strain named 06.21 (SEQ ID NO:7).

Figure 11 shows a nucleotide sequence of a fragment of a CHIK virus according to a preferred embodiment of the invention, and more specifically a fragment which codes for the ectodomain of the glycoprotein E2 of the preferred strain named 06.27 (SEQ ID NO:8).

Figure 12 shows a nucleotide sequence of a fragment of a CHIK virus according to a preferred embodiment of the invention, and more specifically a fragment which codes for the ectodomain of the glycoprotein E2 of the preferred strain named 06.49 (SEQ ID NO:9).

Figure 13 shows a nucleotide sequence of a fragment of a CHIK virus according to a preferred embodiment of the invention, and more specifically a fragment which codes for the ectodomain of the glycoprotein E2 of the preferred strain named 06.115 (SEQ ID NO:10).

Figure 14 shows a nucleotide sequence of a fragment of a CHIK virus according to a preferred embodiment of the invention, and more specifically a fragment which codes for a soluble form of the glycoprotein E2 of the preferred strain named 06.21 (SEQ ID NO:11).

Figure 15 shows a nucleotide sequence of a fragment of a CHIK virus according to a preferred embodiment of the invention, and more specifically a fragment which codes for a soluble form of the glycoprotein E2 of the preferred strain named 06.27 (SEQ ID NO:12).

Figure 16 shows a nucleotide sequence of a fragment of a CHIK virus according to a preferred embodiment of the invention, and more specifically a fragment which codes for a soluble form of the glycoprotein E2 of the preferred strain named 06.49 (SEQ ID NO:13).

Figure 17 shows a nucleotide sequence of a fragment of a CHIK virus according to a preferred embodiment of the invention, and more specifically a fragment which codes for a soluble form of the glycoprotein E2 of the preferred strain named 06.115 (SEQ ID NO:14).

Figure 18 shows an amino acid sequence of a preferred CHIK virus polypeptide according to a preferred embodiment of the invention, and related more specifically to the ectodomain of the glycoprotein E2 of the preferred strain named 06.21 (SEQ ID NO:15).

Figure 19 shows an amino acid sequence of a preferred CHIK virus polypeptide according to a preferred embodiment of the invention, and related more specifically to the ectodomain of the glycoprotein E2 of the preferred strain named 06.27 (SEQ ID NO:16).

Figure 20 shows an amino acid sequence of a preferred CHIK virus polypeptide according to a preferred embodiment of the invention, and related more specifically to the ectodomain of the glycoprotein E2 of the preferred strain named 06.49 (SEQ ID NO:17).

Figure 21 shows an amino acid sequence of a preferred CHIK virus polypeptide according to a preferred embodiment of the invention, and related more

specifically to the ectodomain of the glycoprotein E2 of the preferred strain named 06.115 (SEQ ID NO:18).

Figure 22 shows an amino acid sequence of a preferred CHIK virus polypeptide according to a preferred embodiment of the invention, and related more specifically to a soluble form of the glycoprotein E2 of the preferred strain named 06.21 (SEQ ID NO:19).

Figure 23 shows an amino acid sequence of a preferred CHIK virus polypeptide according to a preferred embodiment of the invention, and related more specifically to a soluble form of the glycoprotein E2 of the preferred strain named 06.27 (SEQ ID NO:20).

Figure 24 shows an amino acid sequence of a preferred CHIK virus polypeptide according to a preferred embodiment of the invention, and related more specifically to a soluble form of the glycoprotein E2 of the preferred strain named 06.49 (SEQ ID NO:21).

Figure 25 shows an amino acid sequence of a preferred CHIK virus polypeptide according to a preferred embodiment of the invention, and related more specifically to a soluble form of the glycoprotein E2 of the preferred strain named 06.115 (SEQ ID NO:22).

Figure 26: Repeat Sequence Elements found in the 3'NTR region

A. Alignment of Repeat Sequence Elements found in the 3'NTR region of chikungunya virus genome. All sequences form conserved and stable stem-loop structures in which the less conserved nucleotides around position 20 constitute the loop. Three RSE are found in all chikungunya genomes. The first one (RSE1) is inserted before the internal poly-A sequence of S27 genome [9], whereas the two others are found downstream this motif.

B. Predicted secondary structure for RSE1 of isolate 05-115.

Figure 27: Focus phenotype of chikungunya viruses on AP61 cells by Focus Immunoassay

Mosquito AP61 cells in 24-well plates were infected with CHIK virus stocks grown on mosquito cells (virus titers 2-5 x 10^8 FFU. mL-1) at 0.0001 (top well) or 0.00001 (bottow well) multiplicity of infection. Infected cells were overlaid with CMC in Leibovitz L15 growth medium with 2% FBS for 2 days to allow focus development at 28°C. The cells were fixed with 3% PFA in PBS, permeabilized with Triton X-100 in PBS, and foci of CHIK virus replication were immunostained with mouse anti-CHIK HMAF (dilution 1 :2,000) and peroxidase-conjugated goat anti-mouse Ig (dilution 1 :100).

Figure 28: Viral preparation containing pE2. pE2 proteins detected by anti-CHIK antibodies.

Figure 29: Alignment of nucleotide sequences encoding soluble form of E2 glycoprotein (E2-1 to E2-361) from Indian Ocean CHIK virus strains -21, -27, -49 and -115.

Figure 30: Primer sequences (SEQ ID NO:79 and 80) used for the amplification and cloning of the soluble form of the E2 (E2-1 to E2-364) (N-terminal and C-terminal nucleic acid fragment: SEQ ID NO:81 and 82; N-terminal and C-terminal protein fragment: SEQ ID NO:83 and 84) from CHIK virus into the shuttle vector pMT2/BiP/V5-HisA.

Figure 31: SDS-PAGE showing CHIK-sE2 staining by Coomassie blue.

Figure 32: Immunoblot analysis of highly purified CHIK sE2 protein.

Figure 33: Construct of the TRIP vector expressing the secreted soluble form of the E2 glycoprotein (sE2) from Chikungunya virus La Réunion 05 strains.

Figure 34 shows the nucleotide sequence coding the secreted soluble form of

the E2 glycoprotein (sE2) into the TRIP vector (SEQ ID NO:85).

Figure 35: Immunofluorescent (IF) assay using anti-CHIK antibodies on TRIP/CHIK.sE2-transduced 293 cells.

Figure 36 shows a direct ELISA with 10^{-4} mL of enriched pE2 protein per well. Antigens were tested respectively with a mouse anti-DEN1 (dilution 1:1000), anti-WN (dilution 1:1000) and anti-CHIK (dilution 1:10 000).

Figure 37 shows an amino acid sequence of the ORF 2 (structural proteins) of the CHIK S27 strain (GenBank AF339485; SEQ ID NO: 23).

Figure 38 shows an amino acid sequence of the ORF 2 (structural proteins) of a preferred CHIK virus according to a preferred embodiment of the invention, namely the 05.61 strain (SEQ ID NO: 24).

Figure 39 shows an amino acid sequence of the ORF 2 (structural proteins) of a preferred CHIK virus according to a preferred embodiment of the invention, namely the 05.209 strain (SEQ ID NO: 25).

Figure 40 shows an amino acid sequence of the ORF 2 (structural proteins) of a preferred CHIK virus according to a preferred embodiment of the invention, namely the 05.115 strain (SEQ ID NO: 26).

Figure 41 shows an amino acid sequence of the ORF 2 (structural proteins) of a preferred CHIK virus according to a preferred embodiment of the invention, namely the 06.49 strain (SEQ ID NO: 27).

Figure 42 shows an amino acid sequence of the ORF 2 (structural proteins) of a preferred CHIK virus according to a preferred embodiment of the invention, namely the 06.27 strain (SEQ ID NO: 28).

Figure 43 shows an amino acid sequence of the ORF 2 (structural proteins) of a

preferred CHIK virus according to a preferred embodiment of the invention, namely the 06.21 strain (SEQ ID NO: 29).

Figure 44 shows an amino acid sequence of the ORF 1 (non-structural proteins) of a preferred CHIK virus according to a preferred embodiment of the invention, namely the 05.61 strain (SEQ ID NO: 30).

Figure 45 shows an amino acid sequence of the ORF 1 (non-structural proteins) of a preferred CHIK virus according to a preferred embodiment of the invention, namely the 05.209 strain (SEQ ID NO: 31).

Figure 46 shows an amino acid sequence of the ORF 1 (non-structural proteins) of a preferred CHIK virus according to a preferred embodiment of the invention, namely the 05.115 strain (SEQ ID NO: 32).

Figure 47 shows an amino acid sequence of the ORF 1 (non-structural proteins) of a preferred CHIK virus according to a preferred embodiment of the invention, namely the 06.49 strain (SEQ ID NO: 33).

Figure 48 shows an amino acid sequence of the ORF 1 (non-structural proteins) of a preferred CHIK virus according to a preferred embodiment of the invention, namely the 06.27 strain (SEQ ID NO: 34).

Figure 49 shows an amino acid sequence of the ORF 1 (non-structural proteins) of a preferred CHIK virus according to a preferred embodiment of the invention, namely the 06.21 strain (SEQ ID NO: 78).

Figure 50: Evaluation of anti-CHIK E2 Mab reactivity by ELISA.

Figure 51: Evaluation of anti-CHIK E2 MAb reactivity on CHIK virions by ELISA.

Figure 52: Immunofluorescence (IF) analysis of anti-CHIK E2 Mab reactivity on CHIKV-infected Vero cells.

Figure 53: Immunofluorescence (IF) analysis of anti-CHIK E2 Mab reactivity on TRIP/CHIK.sE2-transduced 293A cells.

Figure 54: Anti-CHIK E2 Mab binding on cell surface of CHIK virus-infected Vero cells by FACS analysis.

Figure 55: Western blot analysis of CHIKsE2 expression in TRIP/CHIK.sE2-transduced 293A cells.

DETAILED DESCRIPTION OF THE INVENTION

In the present study, the inventors determined the nearly complete nucleotide sequences of viruses isolated from six patients originating from Reunion and Seychelles Islands. The present invention allows to determine the genome structure as well as the unique molecular features of the Indian Ocean outbreak isolates, which distinguish them from other reported CHIKV and alphavirus sequences.

As one in the art may appreciate, the originality of the present invention is the identification of novel strains of the Chikungunya (CHIK) virus which are distinguished from CHIK virus of the prior art, and the use of these CHIK strains and the polypeptides and the polynucleotides encoding same derived from their genome in the diagnostic, prevention and/or treatment of arbovirosis.

According to a first aspect, the present invention concerns an isolated and purified wild strain of chikungunya virus (CHIKV) capable of *in vitro* infecting human cells. Preferably, the present invention concerns a wild strain of CHIK virus which exhibits the same characteristics than those selected from the group consisting of the isolates 05.115, 05.61, 05.209, 06.21, 06.27 and 06.49. According to a preferred embodiment, the strains that are within the scope of the

present invention are characterized in that their genome comprises at least one mutation when compared to the sequence of the genome of the CHIK virus strain S-27 (GenBank AF339485). Also within the scope of the invention, is any strain grown or obtained by cell culture from a sample of a preferred CHIK strain of the invention. The genome of the preferred strains according to the present invention comprises a sequence as shown in Figure 4, 5, 6, 7, 8 or 9 (SEQ ID NO: 1, 2, 3, 4, 5 or 6).

According to another aspect, the present invention provides an isolated and purified strain of chikungunya virus (CHIKV) comprising at least one mutation in structural protein E1 and/or in structural protein E2, and more particularly in their ectodomain region. According to a preferred embodiment, the strain of the invention is characterized by the fact that its genome comprises at least one mutation in the E2 protein at a position homologous to amino acid position 382, 399, 404, 485, 489, 506, 536, 624, 637, 669, 700 or 711 of SEQ ID NO: 23 (Fig. 37). More particularly, the mutation is preferably selected from the group consisting of G382K, I399M, G404E, N485T, A489T, L506M, I536T, S629N, T637M, A669T, S700T and V711A as shown in Table 6. According to another preferred embodiment, the strain of the invention is characterized by the fact that its genome comprises at least one mutation in the E1 protein at a position homologous to amino acid position 1035, 1078, 1093 or 1131 of SEQ ID NO: 23. More particularly, the mutation is preferably selected from the group consisting of A1035V, M1078V, D1093E and V1131A as shown in Table 6. Most preferably, the mutation in the E1 protein is A1035V.

As use herein, the expression "at a position homologous to an amino acid position" of a protein, refers to amino acid positions that are determined to correspond to one another based on sequence and/or structural alignments with a specified reference protein. For instance, in a position corresponding to an amino acid position of a CHIK virus structural protein set forth as SEQ ID NO: 1 can be determined empirically by aligning the sequences of amino acids set forth in SEQ ID NO: 1 with a particular CHIK virus structural protein. Homologous or

corresponding positions can be determined by such alignment by one of skill in the art using manual alignments or by using the numerous alignment programs available (for example, BLASTP). Homologous or corresponding positions also can be based on structural alignment, for example by using computers simulated alignments of protein structure. Recitation that amino acids of a polypeptide correspond to amino acids in a disclosed sequence refers to amino acids identified upon alignment of the polypeptide with the disclosed sequence to maximize identity or homology (where conserved amino acids are aligned) using a standard algorithm, such as the GAP algorithm. As used herein, "at a position homologous to" refers to a position of interest (i.e., base number or residue number) in a nucleic acid molecule or protein relative to the position in another reference nucleic acid molecule or protein. The position of interest to the position in another reference protein can be in, for example, an amino acid sequence from the same protein of another CHIK strain. Homologous positions can be determined by comparing and aligning sequences to maximize the number of matching nucleotides or residues, for instance, such that identity between the sequences is greater than 95%, preferably greater than 96%, more preferably greater than 97%, even more preferably greater than 98% and most preferably greater than 99%. The position of interest is then given the number assigned in the reference nucleic acid molecule.

Another aspect of the invention concerns an isolated and purified polynucleotide comprising all or part of the sequence as shown in Figures 4, 5, 6, 7, 8 or 9 (SEQ ID NO: 1, 2, 3, 4, 5 or 6).

Another aspect of the invention concerns a fragment of the polynucleotide of the invention characterized by the fact that it codes for the glycoprotein E1 or E2, and more preferably for their ectodomain region. Advantageously, the fragment of the invention when coding for the E2 ectodomain, comprises, or more preferably, consists of a nucleotide sequence as shown in Figures 10, 11, 12 or 13 (SEQ ID NO: 7, 8, 9 or 10).

Yet another aspect of the invention concerns a fragment of the polynucleotide of the invention characterized by the fact that it codes for a soluble form of glycoprotein E2. According to a preferred embodiment, the soluble fragment of glycoprotein E2 comprises or more preferably consists of a nucleotide sequence as shown in Figures 14, 15, 16 or 17 (SEQ ID NO. 11, 12, 13 or 14).

As one skilled in the art may appreciate, a fragment as contemplated by the present invention may be obtained by:

- use of restriction enzymes wherein their cleavage sites are present in the polynucleotide comprising said fragment;
- amplification with specific primers for said fragment;
- *in vitro* transcription; or
- chemical synthesis.

According to another aspect, the present invention is concerned with an isolated and purified polypeptide encoded by a polynucleotide or by a fragment of the invention. As used herein, the terms "polypeptide" and "protein" are used interchangeably to denote an amino acid polymer or a set of two or more interacting or bound amino acid polymers.

By "isolated" is meant, when referring to a polypeptide, that the indicated molecule is separate and discrete from the whole organism with which the molecule is found in nature or is present in the substantial absence of other biological macro-molecules of the same type. The term "isolated" with respect to a polynucleotide is a nucleic acid molecule devoid, in whole or part, of sequences normally associated with it in nature; or a sequence, as it exists in nature, but having heterologous sequences in association therewith; or a molecule disassociated from the chromosome.

Broadly defined, the terms "purified polypeptide" or "purified polynucleotide" refer to polypeptides or polynucleotides that are sufficiently free of other proteins or polynucleotides, or carbohydrates, and lipids with which they are naturally

associated. The polypeptide or polynucleotide may be purified by any process by which the protein or polynucleotide is separated from other elements or compounds on the basis for instance, of charge, molecular size, or binding affinity.

The preferred peptides of the invention comprise at least one amino acid substitution compared with the amino acid sequence of strain S-27 (GenBank AF339485) and are derived from the sequence of a protein coded by a fragment of the invention. Preferably, a purified polypeptide of the invention comprises all or part of the amino acid sequence of a CHIK virus ORF 1 or 2 contemplated by the present invention such as one defined in any one of SEQ ID NOS 24 to 29 (ORF 2) or of SEQ ID NOS 30 to 34 and 78 (ORF 1). More preferably, a purified polypeptide of the invention comprises all or part of the amino acid sequence of a glycoprotein E2 contemplated by the present invention such as one defined in any one of SEQ ID NOS 15 to 18 (Figures 18 to 21). Even more preferably, a purified polypeptide of the invention comprises all or part of the amino acid sequence of a soluble form of glycoprotein E2 contemplated by the present invention such as one defined in any one of SEQ ID NOS 19 to 22 (Figures 22 to 25).

The present invention is also concerned with a vector comprising a polynucleotide of the invention or a fragment of a polynucleotide of the invention. As used herein, the term "vector" refers to a polynucleotide construct designed for transduction/transfection of one or more cell types. Vectors may be, for example, "cloning vectors" which are designed for isolation, propagation and replication of inserted nucleotides, "expression vectors" which are designed for expression of a nucleotide sequence in a host cell, or a "viral vector" which is designed to result in the production of a recombinant virus or virus-like particle, or "shuttle vectors", which comprise the attributes of more than one type of vector. Preferred vector are those deposited at the CNCM (Collection Nationale de Cultures de Microorganismes), 28 rue du Docteur Roux, 75724 PARIS Cedex 15, France, on March 15, 2006 under accession numbers I-3587, I-3588, I-3589 and I-3590.

Another preferred vector contemplated by the present invention is the plasmid called TRIP-CHIK.sE2 which has been deposited at the CNCM (Collection Nationale de Cultures de Microorganismes), 28 rue du Docteur Roux, 75724 PARIS Cedex 15, France, on March 14, 2007, under accession number I-3733. Such a vector comprises a fragment which codes for a soluble form of the glycoprotein E2 of the invention. This preferred vector has been optimised for efficient production of the recombinant E2 protein into mammalian cells. As used herein, the term "optimised" means that the vector incorporates regulation sequences, such as a signal peptide sequence, in order to provide adequate expression of the desired encoded protein.

In a related aspect, the present invention provides a host cell comprising a vector as defined above. The term "host cell" refers to a cell that has a new combination of nucleic acid segments that are not covalently linked to each other in nature. A new combination of nucleic acid segments can be introduced into an organism using a wide array of nucleic acid manipulation techniques available to those skilled in the art. A host cell can be a single eukaryotic cell, or a single prokaryotic cell, or a mammalian cell. The host cell can harbor a vector that is extragenomic. An extragenomic nucleic acid vector does not insert into the cell's genome. A host cell can further harbor a vector or a portion thereof that is intragenomic. The term intragenomic defines a nucleic acid construct incorporated within the host cell's genome. A preferred host cell of the invention *E. coli* such as the one containing a vector of the invention and deposited at the CNCM (Collection Nationale de Cultures de Microorganismes), 28 rue du Docteur Roux, 75724 PARIS Cedex 15, France, on March 15, 2006 under accession numbers I-3587, I-3588, I-3589 and I-3590 and on March 14, 2007 under accession number I-3733.

The present invention is further concerned with a monoclonal antibody or polyclonal antibodies, or fragments thereof, that specifically bind to a polypeptide of the invention. As used herein, the term "specifically binds to" refers to antibodies that bind with a relatively high affinity to one or more epitopes of a protein of the invention, but which do not substantially recognize and bind to

molecules other than the one(s) of interest. As used herein, the term "relatively high affinity" means a binding affinity between the antibody and the protein of interest of at least 10^{-6} M, and preferably of at least about 10^{-7} M and even more preferably 10^{-8} M to 10^{-10} M. Determination of such affinity is preferably conducted under standard competitive binding immunoassay conditions which is common knowledge to one skilled in the art.

As used herein, the term "antibody" refers to a glycoprotein produced by lymphoid cells in response to a stimulation with an immunogen. Antibodies possess the ability to react *in vitro* and *in vivo* specifically and selectively with an antigenic determinant or epitope eliciting their production or with an antigenic determinant closely related to the homologous antigen. The term "antibody" is meant to encompass constructions using the binding (variable) region of such an antibody, and other antibody modifications. Thus, an antibody useful in the method of the invention may comprise a whole antibody, an antibody fragment, a polyfunctional antibody aggregate, or in general a substance comprising one or more specific binding sites from an antibody. The antibody fragment may be a fragment such as an Fv, Fab or $F(ab')_2$ fragment or a derivative thereof, such as a single chain Fv fragment. The antibody or antibody fragment may be non-recombinant, recombinant or humanized. The antibody may be of an immunoglobulin isotype, e.g., IgG, IgM, and so forth. In addition, an aggregate, polymer, derivative and conjugate of an immunoglobulin or a fragment thereof can be used where appropriate.

Another aspect of the invention is the use of an element selected from the group consisting of a strain, a polynucleotide, a fragment, a vector, a host cell, a polypeptide and an antibody of the invention for either the detection of a CHIKV associated to an arbovirosis, or for the preparation of a composition that prevents and/or treats an arbovirosis.

Another aspect of the present invention relates to a composition for treating and/or preventing an arbovirosis. The composition of the present invention

advantageously comprises at least one element selected from the group consisting of a strain, a polynucleotide, a fragment, a vector, a host cell, a polypeptide and an antibody of the invention. The composition of the invention may further comprise an acceptable carrier. In a related aspect, the invention provides a method for treating and/or preventing an arbovirosis. The method comprises the step of administering to a subject in need thereof a composition of the invention.

As used herein, the term "treating" refers to a process by which the development of an infection from a CHIKV is affected or completely eliminated. As used herein, the term "preventing" refers to a process by which the CHIKV infection is obstructed or delayed.

As used herein, the expression "an acceptable carrier" means a vehicle for containing the components (or elements) of the composition of the invention that can be administered to a animal host without adverse effects. Suitable carriers known in the art include, but are not limited to, gold particles, sterile water, saline, glucose, dextrose, or buffered solutions. Carriers may include auxiliary agents including, but not limited to, diluents, stabilizers (i. e., sugars and amino acids), preservatives, wetting agents, emulsifying agents, pH buffering agents, viscosity enhancing additives, colors and the like.

The amount of components of the composition of the invention is preferably a therapeutically effective amount. A therapeutically effective amount of components of the composition of the invention is the amount necessary to allow the same to perform their preventing and/or treating role against a CHIKV infection without causing overly negative effects in the host to which the composition is administered. The exact amount of components to be used and the composition to be administered will vary according to factors such as the mode of administration, as well as the other ingredients in the composition.

The composition of the invention may be given to a host (such as a human) through various routes of administration. For instance, the composition may be

administered in the form of sterile injectable preparations, such as sterile injectable aqueous or oleaginous suspensions. These suspensions may be formulated according to techniques known in the art using suitable dispersing or wetting agents and suspending agents. The sterile injectable preparations may also be sterile injectable solutions or suspensions in non-toxic parenterally-acceptable diluents or solvents. They may be given parenterally, for example intravenously, intramuscularly or sub-cutaneously by injection, by infusion or *per os*. Suitable dosages will vary, depending upon factors such as the amount of each of the components in the composition, the desired effect (short or long term), the route of administration, the age and the weight of the host to be treated. Any other methods well known in the art may be used for administering the composition of the invention.

Yet another aspect of the invention is the use of a composition as defined hereabove for the preparation of a medicament for treating and/or preventing an arbovirosis in a subject in need thereof.

Yet another aspect of the invention is to provide a kit for the detection of a CHIKV associated to an arbovirosis, comprising at least one element selected from the group consisting of a strain, a polynucleotide, a fragment, a vector, a host cell, a polypeptide and an antibody of the invention. Kits according to this embodiment of the invention may comprise packages, each containing one or more of the above mentioned elements (typically in concentrated form) which are required to perform the respective diagnostic tests.

EXAMPLES

The examples here below will highlight other characteristics and advantages of the present invention, and will serve to illustrate the scope of the use of the present invention and not to limit its scope. Modifications and variations may be made without departing from the spirit and the scope of the invention. Although it is possible to use other methods or products equivalent to those that are found

here below to test or to realize the present invention, the preferred material and methods are described.

EXAMPLE 1: Identification and characterization of CHIK viruses causing the Indian Ocean outbreak

The inventors (as sometimes referred therein as "we") report the nearly complete genome sequence of six selected clinical isolates, along with partial sequences of glycoprotein E1 from a total of 60 patients from Réunion, Seychelles, Mauritius, Madagascar and Mayotte Islands. The present results indicate that the outbreak was initiated by a strain related to East-African isolates, from which viral variants have evolved following a traceable microevolution history. Unique molecular features of the outbreak isolates were identified. Notably, in the region coding for the non-structural proteins, ten amino acid changes were found, three of which being located in alphavirus conserved positions of nsP2 (which contains helicase, protease and RNA triphosphatase activities) and of the polymerase nsP4. The sole isolate obtained from the cerebrospinal fluid of a patient showed unique changes in nsP1 (T301I), nsP2 (Y642N) and nsP3 (E460 deletion). In the structural protein region, two noteworthy changes (A226V and D284E) were observed in the membrane fusion glycoprotein E1. Homology 3D modelling allowed mapping of these two changes to regions that are important for virion assembly and for membrane fusion. Change E1-A226V was absent in the initial strains but was observed in >85% of subsequent viral sequences from Reunion, denoting evolutionary success possibly due to adaptation to the mosquito vector.

MATERIAL AND METHODS

Patients. The 60 patients for whom partial or complete CHIKV nucleotide sequences were determined originated from Reunion (N=43), Seychelles (N=3), Madagascar (N=7), Mayotte (N=4) and Mauritius (N=3). Characteristics of the patients and biological samples are listed in Table 1.

Virus isolation and RNA extraction. Viruses were isolated either from serum or cerebrospinal fluid (CSF) (Table 1). Briefly, C6-36 *Aedes albopictus* cells were inoculated with 1 ml of serum or CSF diluted 1:10 in L15 medium (Gibco). The cells were grown at 28°C in L15 supplemented with 5% foetal bovine serum and 10% tryptose-phosphate. Cells and supernatants were harvested after the first passage (5 days) and the second passage (7 days). The virus isolates were identified as CHIKV by indirect immunofluorescence, using CHIKV hyper immune ascitic fluid. In the case of isolates 05.115, 06.21, 06.27 and 06.49 whose genomes were sequenced, absence of yellow fever, dengue and West Nile viruses was confirmed by indirect immunofluorescence using specific sera. RNA was extracted using the QIAAmp Viral Minikit (Qiagen, France).

Nucleotide sequencing. Primers (Table 4) were designed based on the nucleotide sequence 20 of the S27 strain. RT-PCR was performed using the Titan One Tube RT-PCR kit (Roche, France). RT-PCR fragments were purified by ultrafiltration prior to sequencing (Millipore, France). Sequencing reactions were performed using the BigDye Terminator v1.1 cycle sequencing kit (Applied Biosystems, USA) and purified by ethanol precipitation. Sequence chromatograms were obtained on automated sequence analysers ABI3100 or ABI3700 (Applied Biosystems). All amplicons were sequenced on both strands.

Assembly of genome sequences and sequence analysis. Contig assembly was performed independently by distinct operators and software, using either BioNumerics version 4.5 (Applied-Maths, Sint-Martens-Latem, Belgium) or PhredPhrap / Consed [11]. Both analyses yielded exactly the same consensus sequence for all strains. A single contig of 11,601 nt was obtained for five isolates, whereas for strain 05.61, a sequence portion was missing, between S27 positions 5,246 to 5,649 (positions 390 to 524 of nsP3). Sequence alignments and computation of substitution tables were performed using programs BioNumerics, DNASP version 4.10 [12] and DAMBE version 4.2.13 [13]. Alignments of nucleotide and amino acid sequences against selected alphavirus sequences were performed with the ClustalW1.7 software [14]. Sequence

identities were computed with the Phylip package [15]. RNA secondary structure was predicted with the Vienna RNA secondary structure server [16]. Neighbor-joining trees were constructed using MEGA version 3.1 [17] with the Kimura-2 parameter corrections of multiple substitutions. Reliability of nodes was assessed by bootstrap resampling with 1,000 replicates. Amounts of synonymous substitutions per synonymous site (Ks) and of non synonymous substitutions per non synonymous site (Ka) were estimated using DNASP. RDP2 [18] was used to detect putative mosaic sequences.

3D structure modeling. The crystallographic structure of the ectodomain of the glycoprotein E1 of Semliki Forest Virus (SFV) at neutral pH [19]; Protein Data Bank code 2ALA) was used as a template to model and analyze the two amino acid mutations of the Indian Ocean isolates. Figure 2 was prepared using the program RIBBONS [20].

Detection of viral foci by immunological staining. *Aedes pseudoscutellaris* AP61 cells were grown in a 24-well tissue culture plates in Leibovitz L-15 growth medium with 10% heat inactivated fetal calf serum (FCS) for 24 h. Mosquito cell monolayers were washed once with Leibovitz L-15 and 0.2 ml Leibovitz L-15/2% FCS were added. Cells were infected with CHIK virus in 0.2 ml of Leibovitz L-15/2% FCS and incubated at 28°C for 1 h. Overlay medium consisting of 0.4 ml of Leibovitz L-15/2% FBS and carboxymethylcellulose (CMC) (1.6%) was then added and the tissue culture plates were incubated at 28°C for 2 days. Foci of infected cells were visualized by focus immunoassay (FIA). The cells were washed with PBS, fixed with 3% paraformaldehyde (PFA) in PBS for 20 min, and permeabilized with 0.5% Triton X-100 in PBS for 4 min at room temperature. The fixed cells were incubated for 20 min at 37°C with 1:2,000 dilution of hyperimmune mouse ascitic fluid (HMAF) directed against CHIKV. Goat anti-mouse IgG, horseradish peroxidase conjugated was used as the second antibody (1:100 dilution) at 37°C for 20 min. Foci were visualized with DAB·Peroxidase Substrate (Sigma).

1. Genome structure and molecular signatures of the Indian Ocean outbreak chikungunya viruses

Genome organization. We determined the nearly complete genome sequences of six CHIKV isolates (05.115, 05.61, 05.209, 06.21, 06.27 and 06.49) representing distinct geographic origins, time points and clinical forms (Table 1) of the Indian Ocean outbreak of chikungunya virus. 11,601 nucleotides were determined, corresponding to positions 52 (5'NTR) to 11,667 (3'NTR, end of third Repeat Sequence Element) in the nucleotide sequence of the 1952 Tanzanian isolate S27 (total length 11,826 nt). There were three insertion/deletion events between S27 and Réunion isolates, two of which were observed in the 3'NTR. First, the internal poly-A stretch of 14 nucleotides observed in S27 (11,440-11,443) and corresponding to a probable internal poly-A site [9] was replaced by a stretch of only 5 A in Indian Ocean isolates, similar to what was observed in other chikungunya viruses, e.g. the Ross strain (accession no.: AF490259). Second, one A was missing in Indian Ocean isolates in a 5-A stretch at S27 position 11,625. Finally, one codon was missing in isolate 06.27, corresponding to nsP3 codon 460, at which all other Indian Ocean isolates analyzed and available alphavirus sequences are GAA, coding for Glu.

The genome sequences of the six isolates presented therein was similar to those previously reported for alphaviruses [9, 21, 22]. Coding sequences consisted of two large open reading frames (ORF) of 7,422 nt and 3,744 nt encoding the non-structural polyprotein (2,474 amino-acids) and the structural polyprotein (1,248 amino-acids), respectively. The non structural polyprotein is the precursor of proteins nsP1 (535 aa), nsP2 (798 aa), nsP3 (530 aa) and nsP4 (611 aa), and the structural polyprotein is the precursor of proteins C (261 aa), p62 (487 aa, precursor to E3 - 64 aa – and E2 - 423 aa), 6K (61 aa), and E1 (439 aa). Cleavage sites characteristic of the alphavirus family in the non-structural and structural polyproteins were conserved. Glycosylation sites in E3, E2 and E1 were also conserved. A 65 nt junction sequence was identified between the stop codon (TAG, 7499-7501) of the non-structural ORF and the start codon (7567-7569) of the structural ORF. The 5' non-translated region (5'NTR) ended at

position 76. The 3'NTR region started at position 11,314 and contained three repeat sequence elements (RSE) with predicted secondary structures (Fig. 26) that were consistent with previous work [9].

Differences between Indian Ocean outbreak isolates and strain S27. Compared to strain S27, Réunion isolate 05.115 showed 28 aa changes (1.13%) in the non-structural proteins (Table 5, with the highest proportion in nsP3 (2.26%) and the lowest in nsP2 (0.6%). Ten out of 12 amino acid changes in nsP3 were concentrated between positions 326 and 524 (5.0% variation), similar to findings in ONN viruses [23]. One important difference with S27 was that the Indian Ocean isolates exhibited an opal stop codon (UGA) at nsP3 codon 524, instead of Arg (CGA) in S27. This opal codon was observed in related alphaviruses [9, 22, 23], and is believed to regulate the expression of nsP4, the putative RNA polymerase, by a read-through mechanism [21, 24].

Compared to S27, the structural proteins showed 21 (1.68%, for 05.115) to 22 (1.76%, for other isolates) amino-acid substitutions in Indian Ocean isolates (Table 6). Notably, envelope protein E2 showed the highest variation, with 14 (3.3%) aa changes, higher than envelope protein E1 (0.68%) and the capsid protein (0.38%). The ratio of rates of evolution of synonymous and non-synonymous sites (Ks/Ka) between S27 and 05.115 isolates was 11.0 for the whole polyprotein, whereas it was only 6.12 for protein E2, probably indicative of a positive selection in favor of amino-acid changes in this immunogenic protein. By comparison, Ks/Ka was 18.75 for the non-structural polyprotein.

Indian Ocean outbreak molecular signatures in non-structural proteins and phenotypic variation. Ten positions (excluding polymorphic positions) had aa that were unique to the non-structural proteins of outbreak isolates, when compared to other CHIKV sequences (Table 2). First, nsP2-54 was Asn in Indian Ocean isolates and in SFV, but was Ser in all other sequences. Second, nsP2-374 was Tyr in Indian Ocean isolates, but was His or Asn in other alphavirus sequences (Table 2). Third, position 500 in nsP4 was Leu in the Indian Ocean sequences

instead of Gln in the four other reported CHIKV sequences. Interestingly, this position, which is about 30 aa from the catalytic "GDD" motif, is a strictly conserved Glu in all other alphaviruses. The remaining seven changes took place in relatively variable regions.

Additional specific changes were observed in isolates 05.209 (S358P) and 06.27 (nsP1-T301I, nsP2-Y642N, and nsP3-460del). Notably, our phenotypic assays conducted in parallel showed differences for strain 06.27. Focus immunoassay showed that CHIKV stocks 05.115, 06.21, 06.27 and 06.49 formed mixtures of foci with different sizes on *Ae. Albopictus* C6/36 (data not shown) and *Ae. pseudoscuterallis* AP61 cells (Fig. 27). Interestingly, only isolate 06-27 formed medium foci, whereas others formed minutes and small foci. The particular phenotype of 06-27 could be linked to the observed aa differences in the non structural proteins, which are involved in the viral replication [21].

Indian Ocean molecular signatures in structural proteins and 3D modelling. When analyzing the aa sequences of the structural proteins, seven positions (four in E2, one in 6K and two in E1) were found to be unique to isolates from the Indian Ocean outbreak (Table 2). Two of these were located in the E2 ectodomain, with Thr 164 and Met 312 being identified in our isolates instead of Ala and Thr, respectively, in all other available CHIKV sequences (Table 2). The first of these two positions is variable in alphaviruses; it lies in a region defined previously as containing neutralizing epitopes [5, 25]. At position 312, Thr is present in other CHIKV, in ONNV and in SFV, but varies in other alphaviruses; it lies in a region identified as important for E1-E2 oligomerization [5, 25].

In E1, two crucial substitutions were observed, one at residue 284, specific to Indian Ocean isolates, and one at residue 284, present in 3 out of 6 Indian isolates (06.21, 06.27 and 06.49). Both mutations were mapped on the 3D structure (modeled from the crystal structure of SFV E1) in Fig. 1. Interestingly, residue 226 is Ala in all reported CHIKV sequences (Table 2), and was also Ala in the first of our Indian Ocean isolates sequenced here (05.61 and 05.115,

obtained at the beginning of the outbreak). All subsequent isolates (obtained from patients collected in November and December 2005) displayed a Val residue at this position. Although position 226 is relatively variable among alphaviruses, it was observed that a single mutation at this position (Pro to Ser) allowed SFV to adapt to growth in cholesterol-depleted insect cells [26, 27].

The other unique aa observed in E1 from Indian Ocean isolates was Glu 284. This is a highly conserved position in E1, which displays an Asp in the majority of alphaviruses or an Asn in SIN (Table 2). This amino acid is located at the interface between E1 protomers at the surface of the virion, participating in contacts that make up the icosahedral E1 scaffold (Fig. 1).

2. Phylogenetic analysis

Previous work based on E1 protein sequences showed strong phylogeographic structure of the chikungunya virus species [6, 10]. In order to determine the progenitor phylogroup from which the Indian Ocean outbreak isolates emerged, we compared a 1,044 nt region within the E1 coding sequence (positions 271 to 1314, *i.e.*, codons 91 to 438) from 63 biological specimens from 60 patients from Reunion, Seychelles, Madagascar, Mayotte and Comoros (Table 1) with 29 other available chikungunya sequences (Table 7). Phylogenetic analysis (Fig. 2) clearly demonstrated that the current Indian Ocean isolates represent a homogeneous clade within a broad group (group ECSA) comprising isolates from East, Central and South Africa (ECSA, Fig. 2). The isolates from an outbreak in Democratic Republic of the Congo [6] also formed a homogeneous clade within group ECSA. There was no ECSA group member showing a significantly closer relationship with the Indian Ocean isolates. Asian isolates were less related to Indian Ocean isolates and constituted the sister group of group ECSA, whereas West-African isolates were even more divergent. Inclusion of other alphaviruses, including the closest relative ONN, placed the root of the chikungunya isolates on the branch leading to the West-African phylogroup (data not shown).

Comparison of the sequences of Indian Ocean outbreak isolates to the S27 sequence revealed 316 (2.7%) nucleotide substitutions in isolate 05.115

(Table 8). The Asian clade Nagpur strain showed 5.1% average nucleotide divergence from 05.115, whereas the West-African clade Senegal strain 37997 displayed 15% difference (Table 8). Interestingly, the latter strain showed complete conservation of an 87 nucleotides portion (9,958 – 10,045, at the junction between structural proteins 6K and E1) with East-African and Indian Ocean outbreak isolates. Sequence identity in this portion may reflect a past event of genetic recombination between West-African and East/Central-African strains. Differently, we did not find statistical support (P > 7E-2) for sequence mosaicism or recombination since the split between S27 and Reunion isolates, although some genomic regions differed in their density of nucleotide polymorphisms.

3. Genotypic and phenotypic variation among Indian Ocean outbreak isolates and microevolutionary scenario

Specific aa changes in the non-structural proteins were observed in the isolates 05.209 (S358P) and 06.27 (nsP1-T301I, nsP2-Y642N, and nsP3-460del). In the structural proteins, change E1-A226V was observed in isolates 06.21, 06.27 and 06.49, and change E2-Q146R in the Seychelles isolate 05.209. In addition to these non-synonymous changes, there were 8 silent substitutions, observed in 05.209, 06.27 and 06.49 (Table 3).

A history of probable sequence evolution that occurred during the outbreak (Fig. 3) was deduced from the 14 amino-acid variations observed among the six complete genomes (Table 3). Isolate 05.61 was initially selected for genome analysis because it was isolated in March 2005, at the onset of the outbreak, from a Reunion patient returning 1 turning from Comoros Island, where the outbreak had been going on since January 2005. Remarkably, the isolates 05.61 and 05.115 (which was the second earliest isolate analyzed), the African isolate S27 and previous unrelated chikungunya isolates from Africa and Asia were identical at all 14 polymorphic sites. Therefore, the consensus sequence of isolates 05.61 and 05.115 (consensus sequence 1) likely represents the ancestral genotype of the Réunion outbreak. Distribution of the 14 polymorphisms suggested that this

founder gave rise to three consensus sequences that likely evolved in four steps. First, substitution at genome position 10,670 (causing the E1 A226V change) gave rise to consensus sequence 2, represented by the late-November 2005 isolate 06.21. Second, a G to A synonymous substitution at position 6,547 (nsP4) led to an intermediate sequence, which itself gave rise to two late sequences: consensus sequence 3 (isolate 06.27), following four additional substitutions and one codon deletion (Table 3), and consensus sequence 4 (06.49), which arose after three distinct synonymous substitutions (Table 3). A fifth consensus sequence was represented by the Seychelles isolate 05.209 alone, which exhibited four substitutions (two of them causing aa changes in nsP3 - S358P and in E2 - Q146R) compared to consensus sequence 1 (Fig. 3).

Since Reunion isolates had E1-226A at the beginning of the outbreak and E1-266V A at the beginning of the outbreak and E1-266V later in the epidemics, we compared residue 226 in 57 additional sequences (57 sequences from 54 sera and 3 CSF) from the Indian Ocean epidemic. Remarkably, the nature of E1-226 differed totally on Reunion Island before and after the winter season. Five sequences from patients sampled from March to June 2005 (including the sequence originating from a traveller back from Comoros) had E1-226A. Between September and end December 2005, 21 sequences showed E1-226V. Among 17 Reunion sequences from 2006, E1-226V was observed 12 times and E1-226A 5 times (Table 1). On Madagascar and Seychelles sequences, for which the samples were collected when the first clinical cases were suspected (i.e probably at the beginning of the outbreaks), only the E1-226 Ala was observed. On Mayotte 2006 sequences, only the E1-226 V was observed. On Mauritius 2006 sequences, both E1-226 Ala and Val were observed.

To date, only CHIKV laboratory strains, passaged many times on mosquito or mammalian cells, had been entirely sequenced [9]. We provide for the first time nearly complete nucleotide sequences of six clinical isolates passaged *in-vitro* only once or twice (see M&M section). The presence in infected patients of a mixed viral population, called quasispecies [31-33], with genotypes co-existing in

an equilibrium governed by a balance between mutation and natural selection. The presence in S27 of an Arg codon instead of the opal stop codon in Indian Ocean isolates is probably explained by numerous *in-vitro* passages of S27, as evolution of opal to Arg was observed experimentally in ONN viruses [23]. Whereas it may be advantageous for viral quasispecies to maintain the opal codon *in-vivo*, an Arg codon probably confers a selective advantage *in-vitro*, as observed for the closely related Semliki Forest virus [34]. Chikungunya virus quasispecies situation *in-vivo* could also explain the nsP1-T301I polymorphism observed for the LCR isolate 06.27. Indeed, it is likely that selection for a subset of genotypes harboring this change may be associated with invasion of the LCR [33]. These results underscore that the genome sequence of laboratory "reference" strains may not accurately reflect the natural situation, as the genotypic complexity of quasispecies *in-vivo* is subject to erosion by *in-vitro* selection. Since the Indian Ocean isolates sequenced here were subjected to *in-vitro* selection for only a few generations, they probably correspond more closely to the *in-vivo* genotypes than previously sequenced chikungunya strains.

The amino acid (aa) differences detected among the outbreak 1 isolates may relate to biological or pathogenic characteristics of the virus. Although our viral culture results are preliminary, they clearly show phenotypic differences between the unique isolate from CSF (06.27), isolated from a neonatal encephalopathy case, and three other isolates, associated with either the classical form of the disease or encephalopathy. The larger foci observed in culture with 06.27 could reflect a higher replication rate of the virus and be linked to the specific amino acid changes identified in nsP1, nsP2 and nsP3. Single amino-acid changes in nsP1, including a Thr/Ile change (residue 538 of Sindbis virus) [35,36] and a 18-nt deletion in nsP3 have previously been shown to affect neurovirulence in other alphaviruses [35-37]. However, in the absence of nsP1 structural data, it is difficult to predict the structural or functional impact of the I301T change observed in 06.27 isolate. It should also be noted that all the viral sequences determined from either the serum or the isolates from three neonatal encephalopathy cases and an adult meningo-encephalitis case had E1-226 Val. However, as this

genotype is observed also in classical forms of the disease, a potential link of E1-226 Val with neuropathogenesis needs further studies. Host factors have to be considered in the occurrence of neurological forms of the disease. For example, the blood-brain crossing may be favoured by young age or hypertension.

Unique molecular signatures of the Indian Ocean outbreak genomes were identified when they were compared to all other reported alphavirus sequences. These features represent interesting targets for future functional studies, as well as for epidemiological follow-up. One particularly interesting feature was the E1-226 Val residue (see above). Another interesting molecular signature of Indian Ocean outbreak genomes was E1-284 Asp. Although pseudo-atomic model of the scaffold used is of modest resolution (the resolution of the crystal structure is limited - approaching 3Å - and the model results of fitting this structure into a 9Å resolution cryo-electron microscopy reconstruction), it appears that the side-chain of Asp 284 interacts with the main chain of an adjacent E1 polypeptide in the virion. Indeed, it is in a position compatible with acceptance of a hydrogen bond from main chain amide 379 from the neighboring E1 protomer. Because the packing is very tight (see Fig. 1B), it is possible that the longer Glutamic acid side chain (which has an extra CH2 group compared to Asp or Asn) may introduce a slight distortion at the contact sites, an effect that is propagated by the icosahedral T=4 symmetry of the virion. Thus, a cooperative effect due to this change at position Asp 284 may play a role in either allowing a less efficient assembly of new particles in infected cells, or a more efficient particle disassembly process during invasion of a new cell, or a combination of both. This information 1 tion can guide new site-directed mutagenesis studies, using reverse genetics, to test the effect of the Asp/Glu replacement on the virus cycle.

Example 2: Identification and characterization of a soluble form of E2 (sE2) of the CHIK virus.

The TOPO/CHIK-21.pE2 (CNCM I-3587) plasmid containing the cDNA coding for the pE2 glycoprotein (E3+E2) from the CHIK 21 virus strain (Schuffenecker et al., *Plos Med.*, 3:1058, 2006) was used as a template for the amplification by PCR of the ectodomain sequence of the E2 envelope glycoprotein (Fig. 29). The ectodomain of gp-E2 (E2-1 to E2-364; 85% of E2) is strictly conserved among the CHIK-21, -27, 49 and 115 cell lines isolated in the Indian Ocean during the epidemic outbreak of 2005-06 (see Fig. 29). The soluble form of the sE2 corresponds to the gp-E2 ectodomain which is deleted at the carboxylic terminal of its transmembrane anchor region. It is of interest that the soluble form carries the main epitopes eliciting virus-neutralizing antibodies. The PCR primers are described in Figure 30 (SEQ ID NO:79 and 80): they allow the cloning of the sE2 sequence between the unique sites *Bg/II* and *NotI* of the pMT/BiP/V5-HisA vector (Invitrogen), on the one hand in a phase dependent on the BiP peptide signal at the N-terminal, and on the other hand in joining successive V5 (His)6 tags at its carboxylic terminal.

Drosophila S2 cells were transfected with the recombinant plasmid pMT/BiP/CHIK-sE2 in the presence of the plasmid coding for the blasticidin resistance gene. The S2/CHIK-sE2 stable cell line was obtained by successive passages in presence of blasticidin. The cell line was selected for its capacity to promote efficient secretion of the CHIK- sE2 virus following the activation of the metallothioneine promoter.

The S2/CHIK-sE2 cells in suspension were induced for the secretion of sE2 during 21 days in the presence of Cu_{2+}. The cellular supernatant is filtered at 0.22µM and concentrated for 16 hours on an affinity column of 5 ml HiTrap Chelating HP (Amersham Biosciences) with the help of a peristaltic pump. The CHIK sE2 protein is eluded from the affinity column in the presence of increasing concentrations of imidazole (50, 100 and 500 mM, pH 8). The CHIK sE2 protein is specifically eluded at a concentration of 500 mM imidazole (E_3 elution) from the

E_37 fraction (Fig. 31). The sE2 protein is detected as being highly purified in PAGESDS following Coomassie Blue coloration. The sE2 protein eluded in the E_39 fraction is specifically immunodetected by an ascite (HMAF) of a mouse hyperimmunized against the CHIK virus which was produced at the IMFH unit (Fig. 32). No cross-reactivity was observed with the anti-dengue (DEN) or anti-West Nile (WN) HMAHF and the monoclonal antibody 9D12 anti-DEN E. The soluble DEN sE proteins (DEN-3 and DEN-4) and WN sE purified from the supernatants of the S2 cellular clones induced according to the protocol described hereinabove are used as control viral antigens for the specificity of anti-CHIK murine antibodies.

Example 3: Construction of the TRIP vector expressing the soluble form of E2 (sE2) of the CHIK virus according to the present invention.

The gene coding the CHIK sE2 protein has been optimised by the Genecust firm so as to provide a synthetic DNA with an enriched G+C content in comparison to the cDNA obtained from the viral genomic RNA. The G+C rich codons (amino acids E2-1 to E2-364, soluble gp-E2 ectodomain, sE2) were fused to the signal peptide sequence of the human calreticuline (ssCRT) MLLSVPLLLLGLLGLAA (SEQ ID NO: 77) for translocation of the viral protein into the secretion pathway. The enzyme restriction sites *BamHi* in 5' and *Xhol* in 3' have been added at their respective ends of the sequences coding for the fusion ssCRT+sE2 protein.

The synthetic gene was cloned into the TRIP vector between the *BamHI* and *Xhol* sites under the transcription of the ieCMV promoter. The non-replicative and integrative TRIP/CHIK.sE2 plasmid thus produced was validated for the expression of the sE2 protein following transduction of 293 cells.

As shown in figure 33, the inventors have constructed a vector expressing the CHIK sE2. As mentioned above, the original CHIK sE2 sequence cloned into the TRIP vector has been modified for improving expression in mammalian cells (fig. 34).

Figure 35 shows mammalian cells, such as the 293 cells, transduced with the TRIP-CHIK.sE2 vector. The expressed sE2 protein has been revealed by IF with anti-CHIK antibodies.

Example 4: Production of recombinant protein sE2 and specific monoclonal antibodies.

The inventors have generated the stable inducible S2/CHIK.sE2 cell line which releases the soluble form of the envelope E2-glycoprotein (sE2) from Reunion CHIK virus strains. The inventors have also generated a stable cell line 293A/CHIK.sE2 which was transducted by the recombinant lentiviral vector TRIP/CHIK.sE2. A synthetic sE2 gene that was modified for optimal codon usage in mammalian cells had to be used in order to obtain efficient expression of CHIK virus sE2 in human fibroblastic 293A cells. The TRIP/CHIK.sE2 vector is currently assessed for its capability to induce protective immunity in a murine model of experimental infection. Viral suspension mainly enriched in CHIK pE2 (E2 precursor or E3E2) was obtained by solubilizing CHIK virions grown in mosquito cells with Triton X-100. Adult mice were hyperimmunized with CHIK pE2 in the presence of adjuvant in order to generate hybridoma directed against CHIK structural proteins. Anti-CHIK E2 monoclonal antibodies produced by mouse hybridoma were characterized by ELISA assay on highly purified CHIK virion and Western blot on secreted sE2 from stable cell line S2/CHIK.sE2. (Figs 50, 51 and Table 9). Fluorescent immunodetection assays of intracellular or surface viral antigens were also established on CHIK virus-infected VERO cells and stable 293A/CHIK.sE2 transduced cell line (Figs 52-55). Anti-CHIK.sE2 MAbs of the inventors find a potential use in developing early viral diagnosis of CHIK disease based on immunocapture of CHIK virions in viremic blood of patients, and as tools for immunological as well as virological studies.

Table 1. Characteristics of the patients

Patients No.	Island	Region or Island	Town or Locality	Sample (b)	Sampling Date	Clinical signs (c)	Virus Isolate No. (d)	E1-226 (e)
1	Réunion (Comoros) (a)	-	-	S	16-Mar-05	Classical	05.61 (G)	A (*)
2	Réunion	West	St Gilles Les Bains	S	11-Apr-05	Classical	05.55	A (*)
3	Réunion	South	Saint Pierre	S	2-May-05	Classical	05.107	A (*)
4	Réunion	West	Mare Sèche Cilaos	S	4-May-05	Classical	05.111	A (*)
5	Réunion	South	La Rivière St Louis	S	6-May-05	Classical	05.115 (G)	A (*)
6	Réunion	South	St Louis	CSF	7-Sep-05	Neonatal encephalopathy	05.223	V (**)
7	Réunion	South	La Rivière St Louis	S	11-Oct-05	Classical	06.55	V (**)
8	Réunion	South	St Louis	S	21-Oct-05	Classical	06.59	V (**)
9	Réunion	South	La Rivière St Louis	S	21-Oct-05	Classical	06.53	V (**)
10	Réunion	South	La Rivière St Louis	P	26-Oct-05	Classical	n.i.	V (**)
11	Réunion	South	St Joseph	P	9-Nov-05	Classical	n.i.	V (**)
12	Réunion	South	La Rivière St Louis	P	10-Nov-05	Classical	n.i.	V (**)
13	Réunion	South	St Louis	P	20-Nov-05	Classical	n.i.	V (**)
14	Réunion	South	La Rivière St Louis	P	21-Nov-05	Classical	n.i.	V (**)
15	Réunion	South	La Rivière St Louis	S	23-Nov-05	Classical	06.45	V (**)
16	Réunion	South	La Rivière St Louis (parents)	S	28-Nov-05	Neonatal ME	06.21 (G)	V (***)
17	Réunion	South	St Joseph	S	23-Nov-05	Classical	06.47	V (**)
18	Réunion	South	La Rivière St Louis	P	24-Nov-05	Classical	n.i.	V (**)
19	Réunion	South	Le Tampon	P	26-Nov-05	Classical	n.i.	V (**)
20	Réunion	South	Ravine des Cabris	P	25-Nov-05	Classical	n.i.	V (**)
21	Réunion	South	St Joseph (parents)	S	29-Nov-05	Neonatal ME	06.25	V (**)
21	Réunion	South	St Joseph (parents)	CSF	29-Nov-05	Neonatal ME	06.27 (G)	V (***)
22	Réunion	South	St Louis	S	2-Dec-05	Classical	06.49 (G)	V (***)

Table 1. Characteristics of the patients (con't)

Patients No.	Island	Region or Island	Town or Locality	Sample (b)	Sampling Date	Clinical signs (c)	Virus Isolate No. (d)	E1-226 (e)
23	Réunion	South	St Louis	P	8-Dec-05	Classical	n.i.	V (**)
24	Réunion	South	Ravine des Cabris	S	9-Dec-05	ME	06.17	V (**)
25	Réunion	South	St Louis	P	13-Dec-05	Classical	n.i.	V (**)
26	Réunion	South	St Pierre	P	2-Jan-06	Classical	n.i.	A (**)
27	Réunion	South	St Pierre	P	4-Jan-06	Algic symdrome	n.i.	A (**)
28	Réunion	East	St André	S	4-Jan-06	Classical	n.i.	A (**)
29	Réunion	South	St Louis	P	29-Dec-05	Algic syndrome	n.i.	V (**)
29	Réunion	South	St Louis	CSF	29-Dec-05	Algic syndrome	n.i.	V (**)
30	Réunion	South	La Rivière St Louis	P	29-Dec-05	Classical	n.i.	V (**)
31	Réunion	South	La Rivière St Louis	P	27-Dec-05	Classical	n.i.	V (**)
32	Réunion	South	St Pierre	P	27-Dec-05	Severe vesicular rash lower limbs	n.i.	V (**)
32	Réunion	South	St Pierre	L	28-Dec-05	Severe vesicular rash lower limbs	n.i.	V (**)
33	Réunion	South	Ravine des Cabris	P	4-Jan-06	n.d.	n.i.	A (**)
34	Réunion	South	St Joseph	P	3-Jan-06	Algic syndrome	n.i.	V (**)
35	Réunion	South	St Louis	P	2-Jan-06	Algic syndrome	n.i.	V (**)
36	Réunion	South	St Joseph	P	5-Jan-06	Algic syndrome	n.i.	To be determined

Table 1. Characteristics of the patients (con't)

Patients No.	Island	Region or Island	Town or Locality	Sample (b)	Sampling Date	Clinical signs (c)	Virus Isolate No. (d)	E1-226 (e)
37	Réunion	South	Ravine des Cabris	P	6-Jan-06	Classical	n.i.	V (**)
38	Réunion	South	St Louis	P	6-Jan-06	Algic syndrome	n.i.	V (**)
39	Réunion	West	Saint-Paul hospital	S	5-Jan-06	n.d.	n.i.	A (**)
40	Réunion	West	St Leu	S	19-Jan-06	Classical	n.i.	V (**)
41	Réunion	South	Les Avirons	S	30-Jan-06	Classical	06.97	V (**)
42	Réunion	East	St Benoît	S	3-Feb-06	Hepatitis	n.i.	V (**)
43	Réunion	n.d.	n.d.	S	22-Feb-06	Classical	n.i.	V (**)
44	Seychelles	Mahe island	Anse aux Pins	S	9-Aug-05	Classical	05.209 (G)	A (*)
45	Seychelles	Mahe island	Anse aux Pins	S	10-Aug-05	Classical	negative	A (**)
46	Seychelles	Mahe island	Anse aux Pins	S	10-Aug-05	Classical	negative	A (**)
47	Madagascar	East	Toamasina	S	1-Feb-06	Classical	06.103	A (**)
48	Madagascar	East	Toamasina	S	8-Feb-06	Classical	06.99	A (**)
49	Madagascar	East	Toamasina	S	9-Feb-06	Classical	06.101	A (**)

Table 1. Characteristics of the patients (con't)

Patients No.	Island	Region or Island	Town or Locality	Sample (b)	Sampling Date	Clinical signs (c)	Virus Isolate No. (d)	E1-226 (e)
50	Madagascar	East	Toamasina	S	15-Feb-06	Classical	n.i.	A (**)
51	Madagascar	North	Ampany	S	14-Feb-06	Classical	n.i.	A (**)
52	Madagascar	North	Djamandjary	S	14-Feb-06	Classical	n.i.	A (**)
53	Madagascar	North	Djamandjary	S	15-Feb-06	Classical	n.i.	A (**)
54	Mayotte	n.d.	n.d.	S	7-Feb-06	Classical	06.111	To be determined
55	Mayotte	n.d.	n.d.	S	11-Feb-06	Classical	n.i.	V (**)
56	Mayotte	n.d.	n.d.	S	13-Feb-06	Classical	n.i.	V (**)
57	Mayotte	n.d.	n.d.	S	13-Feb-06	Classical	n.i.	V (**)
58	Mauritius	n.d.	n.d.	S	12-Feb-06	Classical	06.93	V (*)
59	Mauritius	n.d.	n.d.	S	27-Feb-06	Classical	n.i.	A (**)
60	Mauritius	n.d.	n.d.	S	1-Mar-06	Classical	n.i.	A (**)

(a) This patent traveled back from Comoros and is believed to have been infected there.

(b) S: serum; P: plasma; CSF: cerebrospinal fluid

(c) ME: meningo-encephalitis

(d) Isolates labeled with (G) correspond to those for which the nearly complete genome sequence was established

(e) A: Alanine; V: Valine; (*) : Sequence determined from virus isolates; (**): Sequence determined from biological samples; (***): Sequence determined from both virus isolates and biological samples

n.i.: isolation of virus not intended; n.d.: not determined

Table 2. Relevant amino acid changes identified between Indian Ocean isolates *versus* a selection of *Alphavirus* sequences.

	Non-structural proteins														Structural proteins								
Protein	nsP1	nsP1	nsP2	nsP2	nsP2	nsP2	nsP3	nsP3	nsP3	nsP3	nsP3	nsP4	nsP4	nsP4	E2	E2	E2	E2	E2	6K	E1	E1	E1
Polypeptide position (a)	301	488	589	909	1177	1328	1550	1670	1691	1793	1804	1938	2117	2363	471	489	637	700	711	756	1035	1078	1093
Protein position (a)	301	488	54	374	642	793	217	337	358	460	471	75	254	500	146	164	312	375	386	8	226	269	284
05.115 (Genotype 1)	T	R	N	Y	Y	V	H	I	S	E	S	A	A	L	Q	T	M	T	A	I	A	V	E
06.21 (Genotype 2)	T	R	N	Y	Y	V	H	I	S	Nd	nd	A	A	L	Q	T	M	T	A	I	V	V	E
06.27 (Genotype 3)	I	R	N	Y	N	V	H	I	S	Del	S	A	A	L	Q	T	M	T	A	I	V	V	E

Table 2. Relevant amino acid changes identified between Indian Ocean isolates *versus* a selection of *Alphavirus* sequences. (con't)

	Non-structural proteins													Structural proteins							
06.49 (Genotype 4)	T	R	N	Y	Y	V	H	I	S	E	S	A	A	L	Q	T	M	T	A	I	V	V	E
05.209 (Genotype 5)	T	R	N	Y	Y	V	H	I	P	E	S	A	A	L	R	T	M	T	A	I	A	V	E
S27	T	Q	S	H	C	A	Y	T	S	L	P	T	T	Q	Q	A	T	S	V	I	A	M	D
Ross	T	Q	S	H	C	A	Y	T	S	L	P	T	T	Q	Q	A	T	S	V	V	A	M	D
37997 (West-African phylogroup)	T	K	S	H	Y	A	Y	T	S	P	P	T	T	Q	Q	A	T	S	V	A	A	V	D
Nagpur (Asian phylogroup)	nd	nd	nd	nd	nd	nd	nd	nd	nd	Nd	nd	nd	nd	nd	Q	A	T	S	G	V	A	M	D
ONNV	S	Q	S	N	H	A	Y	**	S	**	**	T	T	E	H	A*	T	S	L*	T*	A	V	D

Table 2. Relevant amino acid changes identified between Indian Ocean isolates *versus* a selection of *Alphavirus* sequences. (con't)

	Non-structural proteins														Structural proteins								
EEV	S	Q	S	N	E	R	N	**	**	**	**	I	V	E	E	G*	S	A	T*	D*	A	E	D
SFV	V	S	N	H	Y	A	L	**	S	**	**	T	T	E	H	V*	T	S	C*	A*	P	M	D
RRV	V	N	S	H	Y	G	S	**	**	**	**	V	T	E	H	D*	D	S	C*	A*	P	M	D
SINV	V	M	S	H	E	R	K	**	**	**	**	I	T	E	V	A*	V	T	V*	S*	A	V	N

(a) S27 reference numbering

*Variable position, ** Hypervariable position

ONV: o'nyong-nyong virus; SFV: Semliki Forest virus; RRV: Ross River virus; SINV: Sindbis virus; EEV: Eastern-Equine Encephalitis virus.

nd: not determined. Note that the opal stop codon observed in nsP3-524 of Indian Ocean outbreak isolates, but not in S27, is not represented in the Table.

Table 3. Polymorphisms observed among Indian Ocean isolates

Genome position (b)	978	1378	3605	4705	5147	5452	5453-5455	6547	7045	8978	9600	10670	11295	11421
Protein	nsP1	nsP1	nsP2	nsP3	nsP3	nsP3	nsP3	nsP4	nsP4	E2	E2	E1	E1	3'NTR
Protein position (for aa change)	301	-	642	-	358	-	460	-	-	146	-	226	-	-
05.115 (Genotype 1)	C	G	T	T	T	G	GAA (Glu)	G	C	A	T	C	G	C
05.61 (Genotype 1)	C	G	T	T	T	n.d.	n.d.	G	C	A	T	C	G	C
06.21 (Genotype 2)	C	G	T	T	T	G	GAA (Glu)	G	C	A	T	T (Ala→Val)	G	C
06.27 (Genotype 3)	T (Thr→Ile)	G	A (Tyr→Asn)	A	T	A	Deleted	A	C	A	T	T (Ala→Val)	G	C
06.49 (Genotype 4)	C	A	T	T	T	G	GAA (Glu)	A	C	A	C	T (Ala→Val)	G	T
05.209 (Genotype 5)	C	G	T	T	C (Ser→Pro)	G	GAA (Glu)	G	T	G (Gln→Arg)	T	C	A	C
S27 (a)	C	G	T	T	T	G	GAA (Glu)	G	C	A	T	C	G	C

(a) Only sites that are variable among Indian Ocean outbreak isolates are represented; those sites that were distinct between Indian Ocean outbreak isolates and other viruses are given as Supplementary Information

(b) S27 numbering

Table 4. Primers used for RT-PCR and sequencing

Fragment	Gene	Primer (a)	Sequence (5' to 3')	SEQ ID NO.
FG1	5'NC	18F	CACGTAGCCTACCAGTTTCTTA	35
	nsP1	871R	ATGGAACACCGATGGTAGGTG	36
FG2	nsP1	616F	AACCCCGTTCATGTACAATGC	37
	nsP1	1435R	CGGTACCACAAAGCTGTCAAAC	38
FG3	nsP1	1317F	CACTGACCTGCTGCTGTCTATG	39
	nsP2	2130R	AGTCCTGCAGCTTCTTCCTTC	40
FG4	nsP1	1412F	CGAGTTTGACAGCTTTGTGGTA	41
	nsP2	2227R	ATGACTGCAATTTTGTATGGGC	42
FG5	nsP2	1908F	CAATCTCGCCTGAAGACTTCC	43
	nsP2	2709R	TCCACTACAATCGGCTTGTTG	44
FG6	nsP2	2530F	GTGCGGCTTCTTCAATATGATG	45
	nsP2	3343R	TCCAGGCCTATTATCCCAGTG	46
FG7	nsP2	2577F	AACATCTGCACCCAAGTGTACC	47
	nsP2	3504R	GTCTCCTGTTGGCCGGTATAAT	48

Table 4. Primers used for RT-PCR and sequencing (con't)

Fragment	Gene	Primer (a)	Sequence (5' to 3')	SEQ ID NO.
FG8	nsP2	3332F	TAATAGGCCTGGAGGGAAGATG	49
	nsP3	4134R	CTACGCACTCTTCATCGTTCTT	50
FG9	nsP2	3885F	GAACGAGTCATCTGCGTATTGG	51
	nsP3	4725R	ATATCTCTGCCATATCCACTGC	52
FG10	nsP3	4458F	TCTTTACAGCCATGGACTCGAC	53
	nsP3	5273R	CGACAGGTACGGTGCTCATTAC	54
FG11	nsP3	5065F	TGTACAGGAAGCGAGTACGACC	55
	nsP4	5874R	TCTACTTTGCGCGACTGATACC	56
FG12	nsP4	5630F	ACGGACGACGAGTTACGACTAG	57
	nsP4	6380R	CCCAGTATTCTTGGTTGCATG	58
FG13	nsP4	6184F	AAAACAGCACGCTTACCACG	59
	nsP4	6936R	AACTTGAAGCGCGTACCTGTC	60

Table 4. Primers used for RT-PCR and sequencing (con't)

Fragment	Gene	Primer (a)	Sequence (5' to 3')	SEQ ID NO.
FG14	nsP4	6732F	TCATAGCCGCACACTTTAAGC	61
	nsP4	7495R	AGGACCGCCGTACAAAGTTAC	62
FG15	nsP4	7278F	GCAGGTGACGAACAAGATGAG	63
	C	8034R	CCGCTTAAAGGCCAATTTG	64
FG16	C	7910F	TCGAAGTCAAGCACGAAGG	65
	E2	8670R	GTCTGTCGCTTCATTTCTGATG	66
FG17	E3	8459F	TGCTTGAGGACAACGTCATGAG	67
	E2	9240R	TTTGTGATTGGTGACCGCG	68
FG18	E2	9093F	AGTCCGGCAACGTAAAGATCAC	69
	6K	9861R	AAAGGTTGCTGCTCGTTCCAC	70
FG19	E2	9648F	AGTTGTGTCAGTGGCCTCGTTC	71
	E1	10403R	TAAAGGACGCGGAGCTTAGCTG	72
FG20	E1	10145F	ACAAAACCGTCATCCCGTCTC	73
	E1	11158R	TGACTATGTGGTCCTTCGGAGG	74
FG21	E1	10959F	CAGCAAGAAAGGCAAGTGTGC	75
	3'NC	11770R	TTTGCCAATTATGGTATTCA	76

(a) The primer name indicates their position and direction on the nucleotide sequence of the S27 genome.

Table 5. Amino-acid changes observed between strain S27 and Indian Ocean outbreak strains in the non-structural proteins

nsP1 and nsP2

Protein	nsP1	nsP1	nsP1	nsP1	nsP1	nsP1	nsP1	nsP1	nsP2	nsP2	nsP2	nsP2	nsP2
Protein position	172	234	301	383	384	481	488	507	54	374	642	643	793
Polypeptide position	172	234	301	383	384	481	488	507	589	909	1177	1178	1328
S27	L	E	T	M	I	T	Q	L	S	H	C	S	A
05.115	V	K	T	L	L	I	R	R	N	Y	Y	N	V
05.61	V	K	T	L	L	I	R	R	N	Y	Y	N	V
06.21	V	K	T	L	L	I	R	R	N	Y	Y	N	V
06.27	V	K	T	L	L	I	R	R	N	Y	N	N	V
06.49	V	K	T	L	L	I	R	R	N	Y	Y	N	V
05.209	V	K	T	L	L	I	R	R	N	Y	Y	N	V

nsP3 and nsP4

Protein	nsP3	nsP3	nsP3	nsP3	nsP3	nsP3	nsP3	nsP3	nsP3	nsP3	nsP3	nsP3	nsP3	nsP3	nsP4	nsP4	nsP4	nsP4	nsP4	nsP4
Protein position	175	217	326	331	337	352	358	376	382	460	461	462	471	524	75	254	500	514	555	604
Polypeptide position	1508	1550	1659	1664	1670	1685	1691	1709	1715	1793	1794	1795	1804	1857	1938	2117	2363	2377	2418	2467
S27	V	Y	P	V	T	K	S	I	A	L	L	S	P	R	T	T	Q	I	V	V
05.115	I	H	S	A	I	E	S	T	T	E	P	N	S	STOP	A	A	L	T	I	I
05.61	I	H	S	A	I	E	S	T	T	nd	nd	nd	nd	nd	A	A	L	T	I	I
06.21	I	H	S	A	I	E	S	T	T	E	P	N	S	STOP	A	A	L	T	I	I
06.27	I	H	S	A	I	E	S	T	T	del	P	N	S	STOP	A	A	L	T	I	I
06.49	I	H	S	A	I	E	S	T	T	E	P	N	S	STOP	A	A	L	T	I	I
05.209	I	H	S	A	I	E	P	T	T	E	P	N	S	STOP	A	A	L	T	I	I

Grayed cells correspond to aa that were variable among Indian Ocean outbreak isolates

Table 6. Amino-acid changes observed in the structural genes among S27, Ross and Indian Ocean outbreak strains.

Protein	C	E3	E2	E2	E2	E2	E2	E2	E2	E2	E2	E2	E2	E2	E2	E2	E2	E2	E2	6K	6K	E1	E1	E1	E1
Cleaved protein AA position	63	23	57	74	79	1446	160	164	181	194	195	211	251	267	299	312	344	375	386	8	54	226	269	284	322
Polyprotein AA position	63	284	382	399	404	471	485	489	506	519	520	536	576	592	624	637	669	700	711	756	802	1035	1078	1093	1131
Genome nt position (a)	7754	8417	8710	8763	8777	8978	9020	9031	9082	9121	9125	9173	9293	9341	9437	9476	9571	9664	9698	9832	9970	10670	10798	10845	10958
S27	K	I	G	I	G	Q	N	A	L	S	Q	I	R	M	S	T	A	S	V	V	I	A	M	D	V
Ross	R	I	G	I	G	Q	N	A	L	G	R	I	P	R	S	T	A	S	V	V	I	A	M	D	V
05.115	R	T	K	M	E	Q	T	T	M	G	Q	T	R	R	N	M	T	T	A	I	V	A	V	E	A
05.61	R	T	K	M	E	Q	T	T	M	G	Q	T	R	R	N	M	T	T	A	I	V	A	V	E	A
06.21	R	T	K	M	E	Q	T	T	M	G	Q	T	R	R	N	M	T	T	A	I	V	V	V	E	A
06.27	R	T	K	M	E	Q	T	T	M	G	Q	T	R	R	N	M	T	T	A	I	V	V	V	E	A
06.49	R	T	K	M	E	Q	T	T	M	G	Q	T	R	R	N	M	T	T	A	I	V	V	V	E	A
05.209	R	T	K	M	E	R	T	T	M	G	Q	T	R	R	N	M	T	T	A	I	V	A	V	E	A

(a): When two nt positions were variable in the same codon, only the position of the upstream nt is given

Grayed cells correspond to amino acid changes among Indian Ocean outbreak isolates.

Table 7. Sequence used for the phylogenetic analysis of partial E1 sequences

Accession No.	Strain	Genomic domain	Strain Origin	Isolation Date	Phylogroup	Reference
AF192906	CAR 256	E1 partial	Central African Region	Unknown	Central Africa	1
AF192907	Ag41855	E1 partial	Uganda	1982	Central Africa	1
AY549583	ChikRCA	E1 partial	DRC (b)	1996	Central Africa	2
AF192903	AR 18211	E1 partial	South African Republic	1976	Central-East/South Africa	1
AF192904	SA H2123	E1 partial	South African Republic	1976	Central-East/South Africa	1
AF192905	Ross	E1 partial	Tanzanie	1953	Central-East/South Africa	1
AF490259	Ross	Complete genome	Tanzanie	1953	Central-East/South Africa	na
AF369024	S27	Complete genome	Tanzanie	1952	Central-East/South Africa	3
AY549576	DRC010	E1 partial	DRC (b)	2000	Central Africa	2
AY549577	DRC027	E1 partial	DRC (b)	2000	Central Africa	2
AY549579	DRC1719	E1 partial	DRC (b)	2000	Central Africa	2
AY549575	DRC007	E1 partial	DRC (b)	2000	Central Africa	2
AY549578	DRC1718	E1 partial	DRC (b)	2000	Central Africa	2
AY549581	DRC1725	E1 partial	DRC (b)	2000	Central Africa	2
AY549582	DRC1728	E1 partial	DRC (b)	2000	Central Africa	2
AY549580	DRC1720	E1 partial	DRC (b)	2000	Central Africa	2
AY549584	DRC1730	E1 partial	DRC (b)	2000	Central Africa	2
AF192896	644188	E1 partial	Thailand	1988	Asian	1
AF192899	3412/78	E1 partial	Thailand	1978	Asian	1
AF192894	RSU1	E1 partial	Indonesia	1985	Asian	1
AF192895	H15483	E1 partial	Philippines	1985	Asian	1
AF192898	1455/75	E1 partial	Thailand	1975	Asian	1
AF192901	Gibbs 63-263	E1 partial	India	1963	Asian	1
AF192902	PO731460	E1 partial	India	1973	Asian	1
AF192897	C-03295	E1 partial	Thailand	1995	Asian	1
AF192900	SV045196	E1 partial	Thailand	1996	Asian	1

Table 7. Sequence used for the phylogenetic analysis of partial E1 sequences. (con't)

Accession No.	Strain	Genomic domain	Strain Origin	Isolation Date	Phylogroup	Reference
L37661	Vaccine strain	polyprotein gene	Na	na	Asian	na
AF192892	37997	E1 partial	Na	na	West African	1
AY726732	37997	Complete genome	Senegal	1983	West African	4
AF192891	PM2951	E1 partial	Senegal	1966	West African	1
AF192893	IbH35 E1	E1 partial	Nigeria	1964	West African	1

References: (1) Powers et *al.*, Pastorino et al., 2004; (3) Khan et al., 2002; (4) Vanlandingham et al., 2005.

(b) Democratic Republic of the Congo

Table 8. Sequence percent similarity based on amino acids and nucleotides (in parentheses) for the structural (SP) and non-structural (NSP) proteins of selected Alphaviruses.

Virus	Strain	Accession No.	05.115/06.49	05.115	06.49
			NSP	SP	SP
CHIKV	05.115	To be submitted	100 (100)	100 (100)	-
	06.49	To be submitted	100 (99.97)	99.91 (99.95)	100 (100)
	S27	AF369024	98.79 (97.3)	98.47 (97.34)	98.38 (97.33)
	37997	AY726732	95.88 (85.5)	95.82 (84.87)	95.74 (84.81)
	Nagpur	AY424803	NA	97.18 (94.85)	97.10 (94.79)
	Vaccine	L37661	NA	96.92 (94.24)	96.83 (94.19)
ONNV	Gulu	M20303	85.90	87.30	87.22
SFV	42S RNA genome	X04129	70.55	65.20	65.20
RRV	NB5092	M20162	69.66	64.40	64.40
SINV	HRSP	J02363	59.25	47.40	47.31

CHIKV: chikungunya virus; ONNV: o'nyong-nyong virus; SFV: Semliki Forest virus; RRV: Ross River virus; SINV: Sindbis virus

NA: Not Available.

Table 9: List of biological assays performed to validate the reactivity of anti-CHIK E2 MAbs

BIOLOGICAL ASSAYS
ELISA on solubilized antigens from CHIK virions
ELISA on purified CHIK virions (La Réunion Isl.)
ELISA on purified CHIK virions (+ TX-100)
ELISA on purified CHIK virions (+ NP-40)
IF assay on CHIKV-infected VERO cells
FACS analysis on cell surface of CHIKV-infected VERO cells
Western blot on recombinant CHIK sE2 from S2 cells
IF assay on stable TRIP/CHIK.sE2-transduced 293A cell clone
Western blot on recombinant CHIK sE2 from TRIP/CHIK.sE2-transduced 293A cell clone

REFERENCES

1. Strauss EG, Strauss JH (1986) Structure and replication of the alphavirus genome. In Schlesinger S, Schlesinger MJ, editors. The *Togaviridae* and *Flaviviridae*. New York: Plenum Press. pp. 35-90.

2. Porterfield JH (1980) Antigenic characteristics and classification of the *Togaviridae*. In: Schlesinger R, editor. The Togaviruses. New York: Academic Press. pp. 13-46.

3. Ross RW (1956) The Newala epidemic. III. The virus: isolation, pathogenic properties and relationship to the epidemic. J Hyg 54: 177-191.

4. Jupp PG, McIntosh BM (1988) Chikungunya disease. In: editors MTP, editor. The Arboviruses: epidemiology and ecology. Boca Raton, Florida: CRC Press. pp. 137-13 157.

5. Johnston RE, Peters CJ (1996) Alphaviruses associated primarily with fever and polyarthritis. In: Fields BN, Knipe DM, Howley PM, editors. Fields Virology. pp. 16 843-898.

6. Pastorino B, Muyembe-Tamfum JJ, Bessaud M, Tock F, Tolou H, et al. (2004) Epidemic resurgence of Chikungunya virus in democratic Republic of the Congo: identification of a new central African strain. J Med Virol 74: 277-282.

7. Laras K, Sukri NC, Larasati RP, Bangs MJ, Kosim R, et al. (2005) Tracking the re-emergence of epidemic chikungunya virus in Indonesia. Trans R Soc Trop Med Hyg 99: 128-141.

8. Paquet C, Quatresous I, Solet JL, Sissoko D, Renault P (2006) Chikungunya outbreak in Reunion: epidemiology and surveillance, 2005 to early January 2006. Eurosurveillance weekly 11: 2.

9. Khan AH, Morita K, Parquet Md Mdel C, Hasebe F, Mathenge EG, et al. (2002) Complete nucleotide sequence of chikungunya virus and evidence for an internal polyadenylation site. J Gen Virol 83: 3075-3084.

10. Powers AM, Brault AC, Tesh RB, Weaver SC (2000) Re-emergence of Chikungunya and O'nyong-nyong viruses: evidence for distinct geographical lineages and distant evolutionary relationships. J Gen Virol 81: 471-479.

11. Gordon D AC, Green P. (1998) Consed: a graphical tool for sequence finishing. Genome Res 8: 195-202.

12. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA 2 polymorphism analyses by the coalescent and other methods. Bioinformatics 19: 2496-2497.

13. Xia X, Xie Z (2001) DAMBE: software package for data analysis in molecular biology and evolution. J Hered 92: 371-373.

14. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Research 22: 4673-4680.

15. Felsenstein J (1989) PHYLIP - Phylogeny Interferne Package (version 3.2). Cladistics 5: 164-166.

16. Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids Res 31: 3429-3431.

17. Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 5: 150-163.

18. Martin DP, Williamson C, Posada D (2005) RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21: 260-262.

19. Roussel A, Lescar J, Vaney MC, Wengler G, Wengler G, et al. (2006) Structure and interactions at the viral surface of the envelope protein E1 of semliki forest virus. Structure 14: 75-86.

20. Carson M (1987) Ribbon models of macromolecules. J Mol Graph 5: 103-106.

21. Strauss JH, Strauss EG (1994) The alphaviruses: gene expression, replication, and evolution. Microbiol Rev 58: 491-562.

22. Lavergne A, Thoisy BD, Lacoste V, Pascalis H, Pouliquen JF, et al. (2005) Mayaro virus: Complete nucleotide sequence and phylogenetic relationships with other alphaviruses. Virus Res in press.

23. Lanciotti RS, Ludwig ML, Rwaguma EB, Lutwama JJ, Kram TM, et al. (1998) Emergence of epidemic O'nyong-nyong fever in Uganda after a 35-year absence: genetic characterization of the virus. Virology 252: 258-268.

24. Strauss EG, Levinson R, Rice CM, Dalrymple J, Strauss JH (1988) Nonstructural proteins nsP3 and nsP4 of Ross River and O'Nyong-nyong viruses: sequence and comparison with those of other alphaviruses. Virology 164: 265-274.

25. Griffin DE (2001) Alphaviruses. In: Knipe DM, Howley PM, editors. Fields Virology. Philadelphia: Lippincott Williams & Wilkins. pp. 917-962.

26. Vashishtha M, Phalen T, Marquardt MT, Ryu JS, Ng AC, et al. (1998) A single point mutation controls the cholesterol dependence of Semliki Forest virus entry and exit. J Cell Biol 140: 91-99.

27. Ahn A, Schoepp RJ, Sternberg D, Kielian M (1999) Growth and stability of a cholesterol-ndependent Semliki Forest virus mutant in mosquitoes. Virology 262: 452-456.

28. Williams MC, Woodall JP, Corbet PS, Gillett JD (1965) O'nyong-Nyong Fever: An Epidemic Virus Disease In East Africa. 8. Virus Isolations From Anopheles Mosquitoes. Trans R Soc Trop Med Hyg 59: 300-306.

29. Weaver SC, Barrett AD (2004) Transmission cycles, host range, evolution and emergence of arboviral disease. Nat Rev Microbiol 2: 789-801.

30. Lu YE, Cassese T, Kielian M (1999) The cholesterol requirement for sindbis virus entry and exit and characterization of a spike protein region involved in cholesterol dependence. J Virol 73: 4272-4278.

31. Holland J, Spindler K, Horodyski F, Grabau E, Nichol S, et al. (1982) Rapid evolution of RNA genomes. Science 215: 1577-1585.

32. Domingo E, Holland JJ (1997) RNA virus mutations and fitness for survival. Annu Rev Microbiol 51: 151-178.

33. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R (2006) Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature 439: 344-348.

34. Kim KH, Rumenapf T, Strauss EG, Strauss JH (2004) Regulation of Semliki Forest virus RNA replication: a model for the control of alphavirus pathogenesis in invertebrate hosts. Virology 323: 153-163.

35. Heise C, Kirn DH (2000) Replication-selective adenoviruses as oncolytic agents. J Clin Invest 105: 847-851.

36. Heise MT, White LJ, Simpson DA, Leonard C, Bernard KA, et al. (2003) An attenuating mutation in nsP1 of the Sindbis-group virus S.A.AR86 accelerates non-structural protein processing and up-regulates viral 26S RNA synthesis. J Virol 77: 1149-1156.

37. Suthar MS, Shabman R, Madric K, Lambeth C, Heise MT (2005) Identification of adult mouse neurovirulence determinants of the Sindbis virus strain AR86. J Virol 79: 4219-4228.

38. Condon RJ, Rouse IL (1995) Acute symptoms and sequelae of Ross River virus infection in South-Western Australia: a follow-up study. Clin Diagn Virol 3: 273-284.

39. Selden SM, Cameron 1 ron AS (1996) Changing epidemiology of Ross River virus disease in South Australia. Med J Aust 165: 313-317.

40. Mazaud R, Salaün JJ, Montabone H, Goube P, Bazillio R (1971) Troubles neurologiques et sensoriels aigus dans la dengue et la fièvre à Chikungunya. Bull Soc Pathol Exot 64: 22-30.

41. Nimmannitya S, Halstead SB, Cohen SN, Margiotta MR (1969) Dengue and chikungunya virus infection in man in Thailand, 1962-1964. I. Observations on hospitalized patients with hemorrhagic fever. Am J Trop Med Hyg 18: 954-971.

42. Gratz NG (2004) Critical review of the vector status of *Aedes albopictus*. Med Vet Entomol 18: 215-227.

43. Lescar J, Roussel A, Wien MW, Navaza J, Fuller SD, et al. (2001) The Fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. Cell 105: 137-148.

CLAIMS

1. An isolated and purified wild strain 05.115 of chikungunya virus (CHIKV) capable of *in vitro* infecting human cells.

2. The strain according to claim 1, wherein its genome comprises a sequence as shown in Figure 4 (SEQ ID NO:1).

3. An isolated and purified polynucleotide comprising all or part of the sequence as shown in Figure 4, (SEQ ID NO: 1).

4. A fragment of the polynucleotide as defined in claim 3, wherein it codes for the ectodomain of glycoprotein E2.

5. A fragment of the polynucleotide as defined in claim 3, wherein it codes for a soluble form of glycoprotein E2.

6. A fragment of claim 3, wherein it codes for the ectodomain of glycoprotein E1.

7. A vector comprising a fragment as defined in any one of claims 3 to 6.

8. A host cell comprising a vector as defined in claim 7.

9. An isolated and purified polypeptide encoded by a polynucleotide as defined in claim 3, or by a fragment as defined in any one of claims 4 to 6.

10. An isolated and purified polypeptide selected in the group consisting of a purified polypeptide comprising all or part of the sequence as defined in Figure 40 (SEQ ID NO: 26).

11. Monoclonal or polyclonal antibodies, or fragment thereof, that specifically bind to a polypeptide as defined in any one of claims 9 and 10.

12. Use of an element selected from the group consisting of:

- a strain as defined in any one of claims 1 to 2;
- a polynucleotide as defined in claim 3;
- a fragment as defined in any one of claims 4 to 6;
- a vector as defined in claim 7;
- a host cell as defined in claim 8;
- a polypeptide as defined in any one of claims 9 or 10; and
- an antibody as defined in claim 11;

for the detection of a CHIKV associated to an arbovirosis.

13. Use of an element selected from the group consisting of:

- a strain as defined in any one of claims 1 to 2;
- a polynucleotide as defined in claim 3;
- a fragment as defined in any one of claims 4 to 6;
- a vector as defined in any one of claim 7;
- a host cell as defined in claim 8;
- a polypeptide as defined in any one of claims 9 or 10; and
- an antibody as defined in claim 11

for the preparation of a composition that prevents and/or treats an arbovirosis.

14. A composition comprising at least one element selected from the group consisting of:

- a strain as defined in any one of claims 1 to 2;
- a polynucleotide as defined in claim 3;
- a fragment as defined in any one of claims 4 to 6;
- a vector as defined in claim 7;
- a host cell as defined in claim 8;
- a polypeptide as defined in any one of claims 9 or 10; and

- an antibody as defined in claim 11;

for treating and/or preventing an arbovirosis.

15. Use of a composition as defined in claim 13 for the preparation of a medicament for treating and/or preventing an arbovirosis in a subject in need thereof.

16. A kit for the detection of a CHIKV associated to an arbovirosis, comprising at least one element selected from the group consisting of:

- a strain as defined in any one of claims 1 to 2;
- a polynucleotide as defined in claim 3;
- a fragment as defined in any one of claims 4 to 6;
- a vector as defined in claim 7;
- a host cell as defined in claim 8;
- a polypeptide as defined in any one of claims 9 or 10; and
- an antibody as defined in claim 11.

17. An isolated and purified wild strain 05.115 of chikungunya virus (CHIKV) substantially as herein described or a purified polypeptide selected in the group consisting of a purified polypeptide comprising all or part of the sequence as defined in Figure 40 (SEQ ID NO: 26) substantially as herein described.

18. An isolated and purified wild strain 05.115 of chikungunya virus (CHIKV) substantially as herein described with reference to the figures or a purified polypeptide selected in the group consisting of a purified polypeptide comprising all or part of the sequence as defined in Figure 40 (SEQ ID NO: 26) substantially as herein described with reference to the figures.

19. An isolated and purified wild strain 05.115 of chikungunya virus (CHIKV) substantially as herein described with reference to the examples or a purified polypeptide selected in the group consisting of a purified

polypeptide comprising all or part of the sequence as defined in Figure 40 (SEQ ID NO: 26) substantially as herein described with reference to the examples.

FIGURE 1

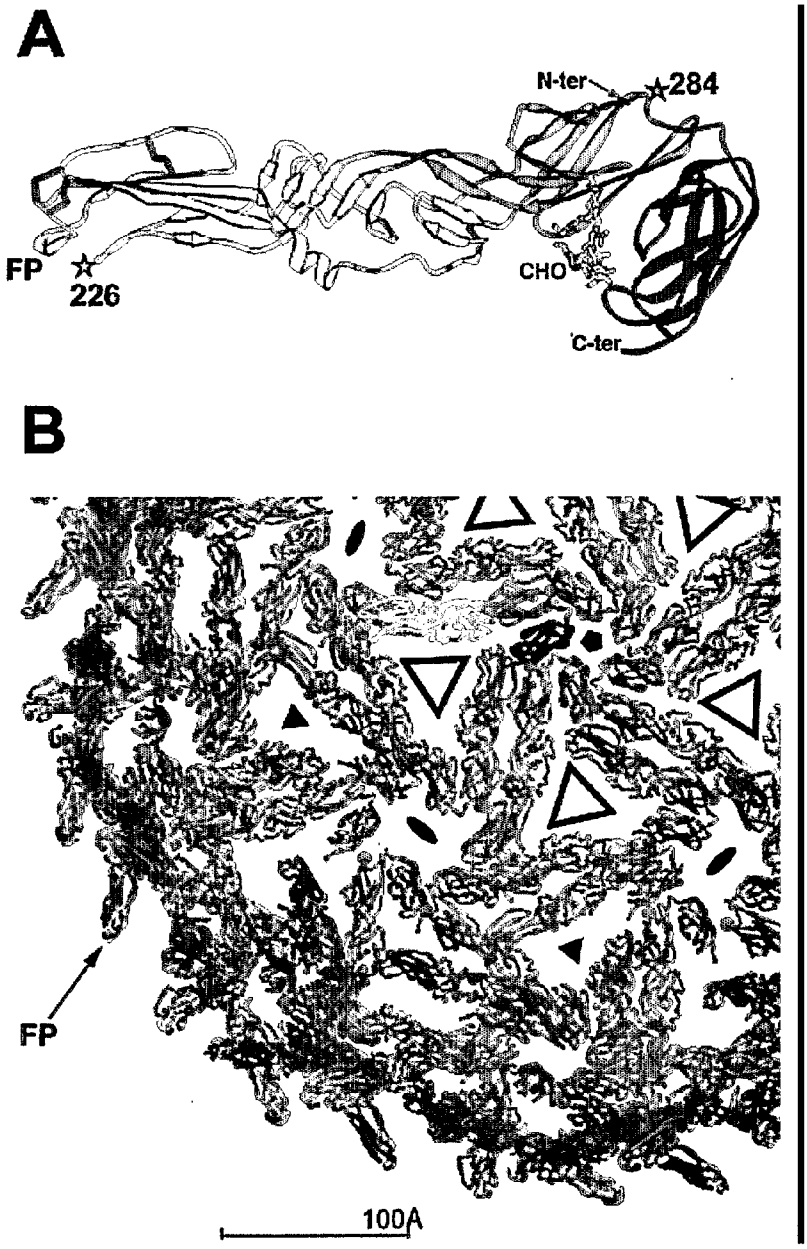


FIGURE 2

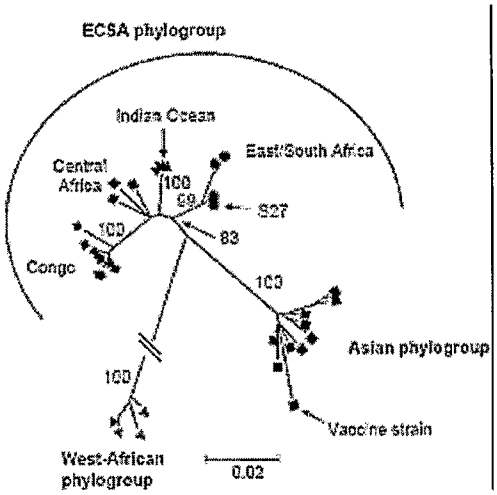

ECSA phylogroup
Indian Ocean
East/South Africa
Central Africa
100
99
S27
100
83
Congo
100
Asian phylogroup
100
Vaccine strain
West-African phylogroup
0.02

FIGURE 3

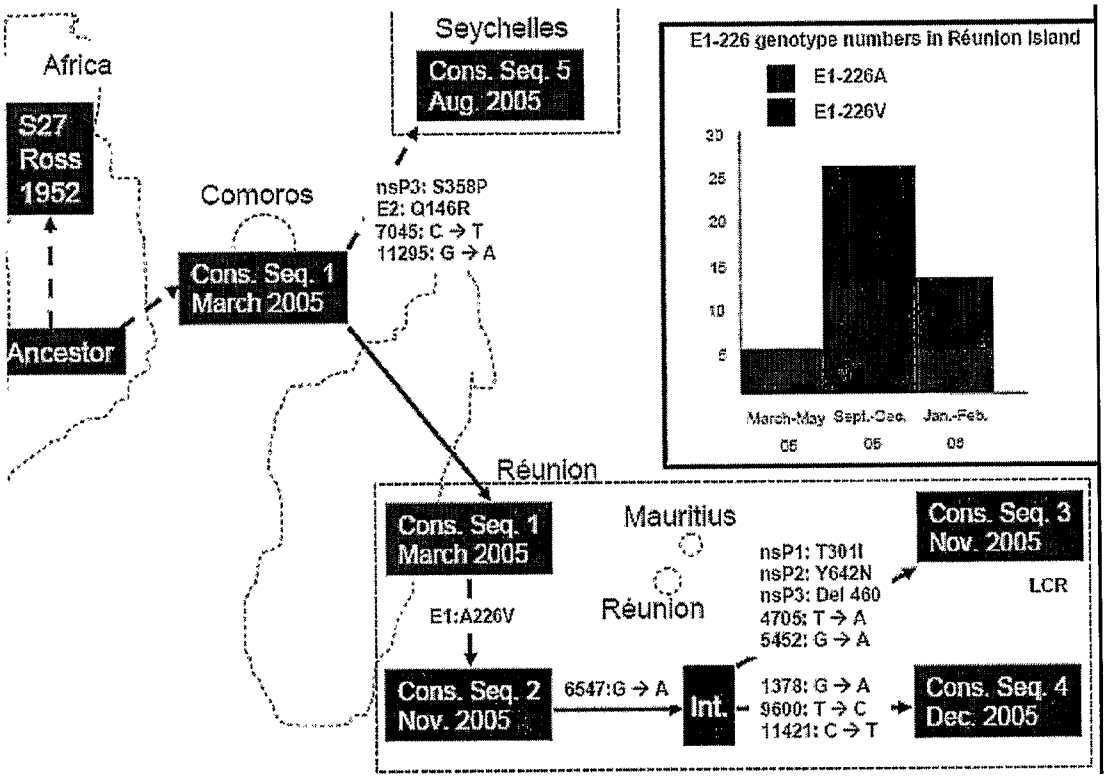

Africa
S27
Ross
1952
Ancestor
Comoros
Cons. Seq. 1
March 2005
Seychelles
Cons. Seq. 5
Aug. 2005
nsP3: S358P
E2: Q146R
7045: C → T
11295: G → A
E1-226 genotype numbers in Réunion Island
E1-226A
E1-226V
30
25
20
15
10
5
March-May 05
Sept.-Dec. 05
Jan.-Feb. 06
Réunion
Cons. Seq. 1
March 2005
E1:A226V
Cons. Seq. 2
Nov. 2005
6547:G → A
Mauritius
Réunion
Int.
nsP1: T301I
nsP2: Y642N
nsP3: Del 460
4705: T → A
5452: G → A
Cons. Seq. 3
Nov. 2005
LCR
1378: G → A
9600: T → C
11421: C → T
Cons. Seq. 4
Dec. 2005

FIGURE 4

```
caaagcaaga gattaataac ccatcatgga tcctgtgtac gtggacatag acgctgacag      60
cgccttttg aaggccctgc aacgtgcgta ccccatgttt gaggtggaac caaggcaggt     120
cacaccgaat gaccatgcta atgctagagc gttctcgcat ctagctataa aactaataga     180
gcaggaaatt gaccccgact caaccatcct ggatatcggc agtgcgccag caaggaggat     240
gatgtcggac aggaagtacc actgcgtctg cccgatgcgc agtgcggaag atcccgagag     300
actcgccaat tatgcgagaa agctagcatc tgccgcagga aaagtcctgg acagaaacat     360
ctctggaaag atcggggact tacaagcagt aatggccgtg ccagacacgg agacgccaac     420
attctgctta cacacagacg tctcatgtag acagagagca gacgtcgcta tataccaaga     480
cgtctatgct gtacacgcac ccacgtcgct ataccaccag gcgattaaag gggtccgagt     540
ggcgtactgg gttgggttcg acacaacccc gttcatgtac aatgccatgg cgggtgccta     600
cccctcatac tcgacaaact gggcagatga gcaggtactg aaggctaaga acataggatt     660
atgttcaaca gacctgacgg aaggtagacg aggcaagttg tctattatga gagggaaaaa     720
gctaaaaccg tgcgaccgtg tgctgttctc agtagggtca acgctctacc cggaaagccg     780
caagctactt aagagctggc acctgccatc ggtgttccat ttaaagggca aactcagctt     840
cacatgccgc tgtgatacag tggtttcgtg tgagggctac gtcgttaaga gaataacgat     900
gagcccaggc ctttatggaa aaaccacagg gtatgcggta acccaccacg cagacggatt     960
cctgatgtgc aagactaccg acacggttga cggcgaaaga gtgtcattct cggtgtgcac    1020
atacgtgccg gcgaccattt gtgatcaaat gaccggcatc cttgctacag aagtcacgcc    1080
ggaggatgca cagaagctgt tggtggggct gaaccagaga atagtggtta acggcagaac    1140
gcaacggaat acgaacacca tgaaaaatta tctgcttccc gtggtcgccc aagccttcag    1200
taagtgggca aaggagtgcc ggaaagacat ggaagatgaa aaactcctgg gggtcagaga    1260
aagaacactg acctgctgct gtctatgggc attcaagaag cagaaaacac acacggtcta    1320
caagaggcct gatacccagt caattcagaa ggttcaggcc gagtttgaca gctttgtggt    1380
accgagtctg tggtcgtccg ggttgtcaat ccctttgagg actagaatca aatggttgtt    1440
aagcaaggtg ccaaaaaccg acctgatccc atacagcgga gacgcccgag aagcccggga    1500
cgcagaaaaa gaagcagagg aagaacgaga agcagaactg actcgcgaag ccctaccacc    1560
tctacaggca gcacaggaag atgttcaggt cgaaatcgac gtggaacagc ttgaggacag    1620
agcgggcgca ggaataatag agactccgag aggagctatc aaagttactg cccaaccaac    1680
agaccacgtc gtgggagagt acctggtact ctccccgcag accgtactac gtagccagaa    1740
gctcagtctg attcacgctt tggcggagca agtgaagacg tgcacgcaca acggacgagc    1800
agggaggtat gcggtcgaag cgtacgacgg ccgagtccta gtgccctcag gctatgcaat    1860
ctcgcctgaa gacttccaga gtctaagcga aagcgcaacg atggtgtata acgaaagaga    1920
gttcgtaaac agaaagctac accatattgc gatgcacgga ccagccctga acaccgacga    1980
agagtcgtat gagctggtga gggcagagag gacagaacac gagtacgtct acgacgtgga    2040
tcagagaaga tgctgtaaga aggaagaagc cgcaggactg gtactggtgg gcgacttgac    2100
taatccgccc taccacgaat tcgcatatga agggctaaaa atccgccctg cctgcccata    2160
caaaattgca gtcataggag tcttcggagt accgggatct ggcaagtcag ctattatcaa    2220
gaacctagtt accaggcagg acctggtgac tagcggaaag aaagaaaact gccaagaaat    2280
caccaccgac gtgatgagac agagaggtct agagatatct gcacgtacgg ttgactcgct    2340
gctcttgaat ggatgcaaca gaccagtcga cgtgttgtac gtagacgagg cgtttgcgtg    2400
ccactctgga acgctacttg ctttgatcgc cttggtgaga ccaaggcaga aagttgtact    2460
ttgtggtgac ccgaagcagt gcggcttctt caatatgatg cagatgaaag tcaactataa    2520
tcacaacatc tgcacccaag tgtaccacaa aagtatctcc aggcggtgta cactgcctgt    2580
gaccgccatt gtgtcatcgt tgcattacga aggcaaaatg cgcactacga atgagtacaa    2640
caagccgatt gtagtggaca ctacaggctc aacaaaacct gaccctggag acctcgtgtt    2700
aacgtgcttc agagggtggg ttaaacaact gcaaattgac tatcgtggat acgaggtcat    2760
```

FIGURE 4 (con't)

```
gacagcagcc gcatcccaag ggttaaccag aaaaggagtt tacgcagtta gacaaaaagt    2820
taatgaaaac ccgctctatg catcaacgtc agagcacgtc aacgtactcc taacgcgtac    2880
ggaaggtaaa ctggtatgga agacactttc cggcgacccg tggataaaga cgctgcagaa    2940
cccaccgaaa ggaaacttca aagcaactat taaggagtgg gaggtggagc atgcatcaat    3000
aatggcgggc atctgcagtc accaaatgac cttcgataca ttccaaaata aagccaacgt    3060
ttgttgggct aagagcttgg tccctatcct cgaaacagcg gggataaaac taaatgatag    3120
gcagtggtct cagataattc aagccttcaa agaagacaaa gcatactcac ctgaagtagc    3180
cctgaatgaa atatgtacgc gcatgtatgg ggtggatcta gacagcgggc tattttctaa    3240
accgttggtg tctgtgtatt acgcggataa ccactgggat aataggcctg gagggaaaat    3300
gttcggattt aaccccgagg cagcatccat tctagaaaga aagtatccat tcacaaaagg    3360
gaagtggaac atcaacaagc agatctgcgt gactaccagg aggatagaag actttaaccc    3420
taccaccaac atcataccgg ccaacaggag actaccacac tcattagtgg ccgaacaccg    3480
cccagtaaaa ggggaaagaa tggaatggct ggttaacaag ataaacggcc accacgtgct    3540
cctggtcagt ggctataacc ttgcactgcc tactaagaga gtcacttggg tagcgccgtt    3600
aggtgtccgc ggagcggact acacatacaa cctagagttg ggtctgccag caacgcttgg    3660
taggtatgac ctagtggtca taaacatcca cacacctttt cgcatacacc attaccaaca    3720
gtgcgtcgac cacgcaatga aactgcaaat gctcgggggt gactcattga gactgctcaa    3780
accgggcggc tctctattga tcagagcata tggttacgca gatagaacca gtgaacgagt    3840
catctgcgta ttgggacgca agtttagatc gtctagagcg ttgaaaccac catgtgtcac    3900
cagcaacact gagatgtttt tcctattcag caactttgac aatggcagaa ggaatttcac    3960
aactcatgtc atgaacaatc aactgaatgc agccttcgta ggacaggtca cccgagcagg    4020
atgtgcaccg tcgtaccggg taaaacgcat ggacatcgcg aagaacgatg aagagtgcgt    4080
agtcaacgcc gctaaccctc gcgggttacc gggtgacggt gtttgcaagg cagtatacaa    4140
aaaatggccg gagtccttta agaacagtgc aacaccagtg ggaaccgcaa aaacagttat    4200
gtgcggtacg tatccagtaa tccacgctgt tggaccaaac ttctctaatt attcggagtc    4260
tgaaggggac cgggaattgg cagctgccta tcgagaagtc gcaaaggaag taactaggct    4320
gggagtaaat agtgtagcta tacctctcct ctccacaggt gtatactcag gagggaaaga    4380
caggctgacc cagtcactga accacctctt tacagccatg gactcgacgg atgcagacgt    4440
ggtcatctac tgccgcgaca aagaatggga gaagaaaata tctgaggcca tacagatgcg    4500
gacccaagta gagctgctgg atgagcacat ctccatagac tgcgatattg ttcgcgtgca    4560
ccctgacagc agcttggcag gcagaaaagg atacagcacc acggaaggcg cactgtactc    4620
atatctagaa gggacccgtt ttcatcagac ggctgtggat atggcggaga tacatactat    4680
gtggccaaag caaacagagg ccaatgagca agtctgccta tatgccctgg gggaaagtat    4740
tgaatcgatc aggcagaaat gcccggtgga tgatgcagac gcatcatctc cccccaaaac    4800
tgtcccgtgc ctttgccgtt acgctatgac tccagaacgc gtcacccggc ttcgcatgaa    4860
ccacgtcaca agcataattg tgtgttcttc gtttcccctc ccaaagtaca aaatagaagg    4920
agtgcaaaaa gtcaaatgct ctaaggtaat gctatttgac cacaacgtgc catcgcgcgt    4980
aagtccaagg gaatatagat cttcccagga gtctgcacag gaggcgagta caatcacgtc    5040
actgacgcat agtcaattcg acctaagcgt tgatggcgag atactgcccg tcccgtcaga    5100
cctggatgct gacgccccag ccctagaacc agcactagac gacggggcga cacacacgct    5160
gccatccaca accggaaacc ttgcggccgt gtctgattgg gtaatgagca ccgtacctgt    5220
cgcgccgccc agaagaaggc gagggagaaa cctgactgtg acatgtgacg agagagaagg    5280
gaatataaca cccatggcta gcgtccgatt ctttagggca gagctgtgtc cggtcgtaca    5340
agaaacagcg gagacgcgtg acacagcaat gtctcttcag gcaccaccga gtaccgccac    5400
ggaaccgaat catccgccga tctccttcgg agcatcaagc gagacgttcc ccattacatt    5460
tggggacttc aacgaaggag aaatcgaaag cttgtcttct gagctactaa ctttcggaga    5520
```

FIGURE 4 (con't)

cttcttacca ggagaagtgg atgacttgac agacagcgac tggtccacgt gctcagacac 5580
ggacgacgag ttatgactag acagggcagg tgggtatata ttctcgtcgg acaccggtcc 5640
aggtcattta caacagaagt cagtacgcca gtcagtgctg ccggtgaaca ccctggagga 5700
agtccacgag gagaagtgtt acccacctaa gctggatgaa gcaaaggagc aactattact 5760
taagaaactc caggagagtg catccatggc caacagaagc aggtatcagt cgcgcaaagt 5820
agaaaacatg aaagcagcaa tcatccagag actaaagaga ggctgtagac tatacttaat 5880
gtcagagacc ccaaaagtcc ctacttaccg gactacatat ccggcgcctg tgtactcgcc 5940
tccgatcaac gtccgattgt ccaatcccga gtccgcagtg gcagcatgca atgagttctt 6000
agctagaaac tatccaactg tctcatcata ccaaattacc gacgagtatg atgcatatct 6060
agacatggtg gacgggtcgg agagttgcct ggaccgagcg acattcaatc cgtcaaaact 6120
caggagctac ccgaaacagc acgcttacca cgcgccctcc atcagaagcg ctgtaccgtc 6180
cccattccag aacacactac agaatgtact ggcagcagcc acgaaaagaa actgcaacgt 6240
cacacagatg agggaattac ccactttgga ctcagcagta ttcaacgtgg agtgtttcaa 6300
aaaattcgca tgcaaccaag aatactggga agaatttgct gccagcccta ttaggataac 6360
aactgagaat ttagcaacct atgttactaa actaaaaggg ccaaaagcag cagcgctatt 6420
cgcaaaaacc cataatctac tgccactaca ggaagtacca atggataggt tcacagtaga 6480
tatgaaaagg gacgtgaagg tgactcctgg tacaaagcat acagaggaaa gacctaaggt 6540
gcaggttata caggcggctg aacccttggc gacagcatac ctatgtggga ttcacagaga 6600
gctggttagg aggctgaacg ccgtcctcct acccaatgta catacactat ttgacatgtc 6660
tgccgaggat ttcgatgcca tcatagccgc acactttaag ccaggagaca ctgttttgga 6720
aacggacata gcctcctttg ataagagcca agatgattca cttgcgctta ctgctttgat 6780
gctgttagag gatttagggg tggatcactc cctgctggac ttgatagagg ctgctttcgg 6840
agagatttcc agctgtcacc taccgacagg tacgcgcttc aagttcggcg ccatgatgaa 6900
atcaggtatg ttcctaactc tgttcgtcaa cacattgtta aacatcacca tcgccagccg 6960
agtgctggaa gatcgtctga caaaatccgc gtgcgcggcc ttcatcggcg acgacaacat 7020
aatacatgga gtcgtctccg atgaattgat ggcagccaga tgtgccactt ggatgaacat 7080
ggaagtgaag atcatagatg cagttgtatc cttgaaagcc ccttactttt gtggagggtt 7140
tatactgcac gatactgtga caggaacagc ttgcagagtg gcagacccgc taaaaaggct 7200
ttttaaactg ggcaaaccgc tagcggcagg tgacgaacaa gatgaagata gaagacgagc 7260
gctggctgac gaagtgatca gatggcaacg aacagggcta attgatgagc tggagaaagc 7320
ggtatactct aggtacgaag tgcagggtat atcagttgtg gtaatgtcca tggccacctt 7380
tgcaagctcc agatccaact tcgagaagct cagaggaccc gtcataactt tgtacggcgg 7440
tcctaaatag gtacgcacta cagctaccta ttttgcagaa gccgacagca agtatctaaa 7500
cactaatcag ctacaatgga gttcatccca acccaaactt tttacaatag gaggtaccag 7560
cctcgaccct ggactccgcg ccctactatc caagtcatca ggcccagacc gcgccctcag 7620
aggcaagctg ggcaacttgc ccagctgatc tcagcagtta ataaactgac aatgcgcgcg 7680
gtaccccaac agaagccacg caggaatcgg aagaataaga agcaaaagca aaaacaacag 7740
gcgccacaaa acaacacaaa tcaaaagaag cagccaccta aaaagaaacc ggctcaaaag 7800
aaaaagaagc cgggccgcag agagaggatg tgcatgaaaa tcgaaaatga ttgtattttc 7860
gaagtcaagc acgaaggtaa ggtaacaggt tacgcgtgcc tggtgggggа caaagtaatg 7920
aaaccagcac acgtaaaggg gaccatcgat aacgcggacc tggccaaact ggcctttaag 7980
cggtcatcta agtatgacct tgaatgcgcg cagatacccg tgcacatgaa gtccgacgct 8040
tcgaagttca cccatgagaa accggagggg tactacaact ggcaccacgg agcagtacag 8100
tactcaggag gccggttcac catccctaca ggtgctggca aaccagggga cagcggcaga 8160
ccgatcttcg acaacaaggg acgcgtggtg gccatagtct taggaggagc taatgaagga 8220
gcccgtacag ccctctcggt ggtgacctgg aataaagaca ttgtcactaa aatcaccccc 8280
gaggggggcg aagagtggag tcttgccatc ccagttatgt gcctgttggc aaacaccacg 8340
ttcccctgct cccagccccc ttgcacgccc tgctgctacg aaaaggaacc ggaggaaacc 8400

FIGURE 4 (con't)

ctacgcatgc ttgaggacaa cgtcatgaga cctgggtact atcagctgct acaagcatcc 8460
ttaacatgtt ctccccaccg ccagcgacgc agcaccaagg acaacttcaa tgtctataaa 8520
gccacaagac catacttagc tcactgtccc gactgtggag aagggcactc gtgccatagt 8580
cccgtagcac tagaacgcat cagaaatgaa gcgacagacg ggacgctgaa aatccaggtc 8640
tccttgcaaa tcggaataaa gacggatgac agccacgatt ggaccaagct gcgttatatg 8700
gacaaccaca tgccagcaga cgcagagagg gcggggctat ttgtaagaac atcagcaccg 8760
tgtacgatta ctggaacaat gggacacttc atcctggccc gatgtccaaa aggggaaact 8820
ctgacggtgg gattcactga cagtaggaag attagtcact catgtacgca cccatttcac 8880
cacgaccctc ctgtgatagg tcgggaaaaa ttccattccc gaccgcagca cggtaaagag 8940
ctaccttgca gcacgtacgt gcagagcacc gccgcaacta ccgaggagat agaggtacac 9000
atgcccccag acacccctga tcgcacatta atgtcacaac agtccggcaa cgtaaagatc 9060
acagtcaatg gccagacggt gcggtacaag tgtaattgcg gtggctcaaa tgaaggacta 9120
acaactacag acaaagtgat taataactgc aaggttgatc aatgtcatgc cgcggtcacc 9180
aatcacaaaa agtggcagta taactcccct ctggtcccgc gtaatgctga acttggggac 9240
cgaaaaggaa aaattcacat cccgtttccg ctggcaaatg taacatgcag ggtgcctaaa 9300
gcaaggaacc ccaccgtgac gtacgggaaa aaccaagtca tcatgctact gtatcctgac 9360
cacccaacac tcctgtccta ccggaatatg ggagaagaac caaactatca agaagagtgg 9420
gtgatgcata agaaggaagt cgtgctaacc gtgccgactg aagggctcga ggtcacgtgg 9480
ggcaacaacg agccgtataa gtattggccg cagttatcta caaacggtac agcccatggc 9540
cacccgcatg agataattct gtattattat gagctgtacc ccactatgac tgtagtagtt 9600
gtgtcagtgg ccacgttcat actcctgtcg atggtgggta tggcagcggg gatgtgcatg 9660
tgtgcacgac gcagatgcat cacaccgtat gaactgacac caggagctac cgtccctttc 9720
ctgcttagcc taatatgctg catcagaaca gctaaagcgg ccacatacca agaggctgcg 9780
atatacctgt ggaacgagca gcaacctttg ttttggctac aagcccttat tccgctggca 9840
gccctgattg ttctatgcaa ctgtctgaga ctcttaccat gctgctgtaa aacgttggct 9900
tttttagccg taatgagcgt cggtgcccac actgtgagcg cgtacgaaca cgtaacagtg 9960
atcccgaaca cggtgggagt accgtataag actctagtca atagacctgg ctacagcccc 10020
atggtattgg agatggaact actgtcagtc actttggagc caacactatc gcttgattac 10080
atcacgtgcg agtacaaaac cgtcatcccg tctccgtacg tgaagtgctg cggtacagca 10140
gagtgcaagg acaaaaacct acctgactac agctgtaagg tcttcaccgg cgtctaccca 10200
tttatgtggg gcggcgccta ctgcttctgc gacgctgaaa acacgcagtt gagcgaagca 10260
cacgtggaga agtccgaatc atgcaaaaca gaatttgcat cagcatacag ggctcatacc 10320
gcatctgcat cagctaagct ccgcgtcctt taccaaggaa ataacatcac tgtaactgcc 10380
tatgcaaacg gcgaccatgc cgtcacagtt aaggacgcca aattcattgt ggggccaatg 10440
tcttcagcct ggacaccttt cgacaacaaa attgtggtgt acaaaggtga cgtctataac 10500
atggactacc cgccctttgg cgcaggaaga ccaggacaat ttggcgatat ccaaagtcgc 10560
acacctgaga gtaaagacgt ctatgctaat acacaactgg tactgcagag accggctgcg 10620
ggtacggtac acgtgccata ctctcaggca ccatctggct ttaagtattg gctaaaagaa 10680
cgcggggcgt cgctgcagca cacagcacca tttggctgcc aaatagcaac aaacccggta 10740
agagcggtga actgcgccgt agggaacatg cccatctcca tcgacatacc ggaagcggcc 10800
ttcactaggg tcgtcgacgc gccctcttta acggacatgt cgtgcgaggt accagcctgc 10860
acccattcct cagactttgg gggcgtcgcc attattaaat atgcagccag caagaaaggc 10920
aagtgtgcgg tgcattcgat gactaacgcc gtcactattc gggaagctga gatagaagtt 10980
gaagggaatt ctcagctgca aatctctttc tcgacggcct tagccagcgc cgaattccgc 11040
gtacaagtct gttctacaca agtacactgt gcagccgagt gccacccccc gaaggaccac 11100
atagtcaact acccggcgtc acataccacc ctcggggtcc aggacatctc cgctacggcg 11160
atgtcatggg tgcagaagat cacgggaggt gtgggactgg ttgttgctgt tgccgcactg 11220
attctaatcg tggtgctatg cgtgtcgttc agcaggcact aacttgacaa ttaagtatga 11280

FIGURE 4 (con't)

```
aggtatatgt gtcccctaag agacacactg tacatagcaa ataatctata gatcaaaggg   11340
ctacgcaacc cctgaatagt aacaaaatac aaaatcacta aaaattataa aaacagaaaa   11400
atacataaat aggtatacgt gtcccctaag agacacattg tatgtaggtg ataagtatag   11460
atcaaagggc cgaataaccc ctgaatagta acaaaatatg aaaatcaata aaaatcataa   11520
aatagaaaaa ccataaacag aagtagttca aagggctata aaacccctga atagtaacaa   11580
aacataaaat taataaaaat c                                             11601
```

FIGURE 5

caaagcaaga gattaataac ccatcatgga tcctgtgtac gtggacatag acgctgacag 60
cgccttttg aaggccctgc aacgtgcgta ccccatgttt gaggtggaac caaggcaggt 120
cacaccgaat gaccatgcta atgctagagc gttctcgcat ctagctataa aactaataga 180
gcaggaaatt gaccccgact caaccatcct ggatatcggc agtgcgccag caaggaggat 240
gatgtcggac aggaagtacc actgcgtctg cccgatgcgc agtgcggaag atcccgagag 300
actcgccaat tatgcgagaa agctagcatc tgccgcagga aaagtcctgg acagaaacat 360
ctctggaaag atcggggact tacaagcagt aatggccgtg ccagacacgg agacgccaac 420
attctgctta cacacagacg tctcatgtag acagagagca gacgtcgcta tataccaaga 480
cgtctatgct gtacacgcac ccacgtcgct ataccaccag gcgattaaag gggtccgagt 540
ggcgtactgg gttgggttcg acacaacccc gttcatgtac aatgccatgg cgggtgccta 600
cccctcatac tcgacaaact gggcagatga gcaggtactg aaggctaaga acataggatt 660
atgttcaaca gacctgacgg aaggtagacg aggcaagttg tctattatga gagggaaaaa 720
gctaaaaccg tgcgaccgtg tgctgttctc agtagggtca acgctctacc cggaaagccg 780
caagctactt aagagctggc acctgccatc ggtgttccat ttaaagggca aactcagctt 840
cacatgccgc tgtgatacag tggtttcgtg tgagggctac gtcgttaaga gaataacgat 900
gagcccaggc ctttatggaa aaaccacagg gtatgcggta acccaccacg cagacggatt 960
cctgatgtgc aagactaccg acacggttga cggcgaaaga gtgtcattct cggtgtgcac 1020
atacgtgccg gcgaccattt gtgatcaaat gaccggcatc cttgctacag aagtcacgcc 1080
ggaggatgca cagaagctgt tggtggggct gaaccagaga atagtggtta acggcagaac 1140
gcaacggaat acgaacacca tgaaaaatta tctgcttccc gtggtcgccc aagccttcag 1200
taagtgggca aaggagtgcc ggaaagacat ggaagatgaa aaactcctgg gggtcagaga 1260
aagaacactg acctgctgct gtctatgggc attcaagaag cagaaaacac acacggtcta 1320
caagaggcct gatacccagt caattcagaa ggttcaggcc gagtttgaca gctttgtggt 1380
accgagtctg tggtcgtccg ggttgtcaat ccctttgagg actagaatca aatggttgtt 1440
aagcaaggtg ccaaaaaccg acctgatccc atacagcgga gacgcccgag aagcccggga 1500
cgcagaaaaa gaagcagagg aagaacgaga agcagaactg actcgcgaag ccctaccacc 1560
tctacaggca gcacaggaag atgttcaggt cgaaatcgac gtggaacagc ttgaggacag 1620
agcgggcgca ggaataatag agactccgag aggagctatc aaagttactg cccaaccaac 1680
agaccacgtc gtgggagagt acctggtact ctccccgcag accgtactac gtagccagaa 1740
gctcagtctg attcacgctt tggcggagca agtgaagacg tgcacgcaca acggacgagc 1800
agggaggtat gcggtcgaag cgtacgacgg ccgagtccta gtgccctcag gctatgcaat 1860
ctcgcctgaa gacttccaga gtctaagcga aagcgcaacg atggtgtata acgaaagaga 1920
gttcgtaaac agaaagctac accatattgc gatgcacgga ccagccctga acaccgacga 1980
agagtcgtat gagctggtga gggcagagag gacagaacac gagtacgtct acgacgtgga 2040
tcagagaaga tgctgtaaga aggaagaagc cgcaggactg gtactggtgg gcgacttgac 2100
taatccgccc taccacgaat tcgcatatga agggctaaaa atccgccctg cctgcccata 2160
caaaattgca gtcataggag tcttcggagt accgggatct ggcaagtcag ctattatcaa 2220
gaacctagtt accaggcagg acctggtgac tagcggaaag aaagaaaact gccaagaaat 2280
caccaccgac gtgatgagac agagaggtct agagatatct gcacgtacgg ttgactcgct 2340
gctcttgaat ggatgcaaca gaccagtcga cgtgttgtac gtagacgagg cgtttgcgtg 2400
ccactctgga acgctacttg ctttgatcgc cttggtgaga ccaaggcaga aagttgtact 2460
ttgtggtgac ccgaagcagt gcggcttctt caatatgatg cagatgaaag tcaactataa 2520
tcacaacatc tgcacccaag tgtaccacaa aagtatctcc aggcggtgta cactgcctgt 2580
gaccgccatt gtgtcatcgt tgcattacga aggcaaaatg cgcactacga atgagtacaa 2640
caagccgatt gtagtggaca ctacaggctc aacaaaacct gaccctggag acctcgtgtt 2700
aacgtgcttc agagggtggg ttaaacaact gcaaattgac tatcgtggat acgaggtcat 2760
gacagcagcc gcatcccaag ggttaaccag aaaaggagtt tacgcagtta gacaaaaagt 2820

FIGURE 5 (con't)

taatgaaaac ccgctctatg catcaacgtc agagcacgtc aacgtactcc taacgcgtac 2880
ggaaggtaaa ctggtatgga agacactttc cggcgacccg tggataaaga cgctgcagaa 2940
cccaccgaaa ggaaacttca aagcaactat taaggagtgg gaggtggagc atgcatcaat 3000
aatggcgggc atctgcagtc accaaatgac cttcgataca ttccaaaata aagccaacgt 3060
ttgttgggct aagagcttgg tccctatcct cgaaacagcg gggataaaac taaatgatag 3120
gcagtggtct cagataattc aagccttcaa agaagacaaa gcatactcac ctgaagtagc 3180
cctgaatgaa atatgtacgc gcatgtatgg ggtggatcta gacagcgggc tattttctaa 3240
accgttggtg tctgtgtatt acgcggataa ccactgggat aataggcctg gagggaaaat 3300
gttcggattt aaccccgagg cagcatccat tctagaaaga aagtatccat tcacaaaagg 3360
gaagtggaac atcaacaagc agatctgcgt gactaccagg aggatagaag actttaaccc 3420
taccaccaac atcataccgg ccaacaggag actaccacac tcattagtgg ccgaacaccg 3480
cccagtaaaa ggggaaagaa tggaatggct ggttaacaag ataaacggcc accacgtgct 3540
cctggtcagt ggctataacc ttgcactgcc tactaagaga gtcacttggg tagcgccgtt 3600
aggtgtccgc ggagcggact acacatacaa cctagagttg ggtctgccag caacgcttgg 3660
taggtatgac ctagtggtca taaacatcca cacacctttt cgcatacacc attaccaaca 3720
gtgcgtcgac cacgcaatga aactgcaaat gctcgggggt gactcattga gactgctcaa 3780
accgggcggc tctctattga tcagagcata tggttacgca gatagaacca gtgaacgagt 3840
catctgcgta ttgggacgca agtttagatc gtctagagcg ttgaaaccac catgtgtcac 3900
cagcaacact gagatgtttt tcctattcag caactttgac aatggcagaa ggaatttcac 3960
aactcatgtc atgaacaatc aactgaatgc agccttcgta ggacaggtca cccgagcagg 4020
atgtgcaccg tcgtaccggg taaaacgcat ggacatcgcg aagaacgatg aagagtgcgt 4080
agtcaacgcc gctaaccctc gcgggttacc gggtgacggt gtttgcaagg cagtatacaa 4140
aaaatggccg gagtccttta agaacagtgc aacaccagtg ggaaccgcaa aaacagttat 4200
gtgcggtacg tatccagtaa tccacgctgt tggaccaaac ttctctaatt attcggagtc 4260
tgaaggggac cgggaattgg cagctgccta tcgagaagtc gcaaaggaag taactaggct 4320
gggagtaaat agtgtagcta tacctctcct ctccacaggt gtatactcag gagggaaaga 4380
caggctgacc cagtcactga accacctctt tacagccatg gactcgacgg atgcagacgt 4440
ggtcatctac tgccgcgaca aagaatggga gaagaaaata tctgaggcca tacagatgcg 4500
gacccaagta gagctgctgg atgagcacat ctccatagac tgcgatattg ttcgcgtgca 4560
ccctgacagc agcttggcag gcagaaaagg atacagcacc acggaaggcg cactgtactc 4620
atatctagaa gggacccgtt ttcatcagac ggctgtggat atggcggaga tacatactat 4680
gtggccaaag caaacagagg ccaatgagca agtctgccta tatgccctgg gggaaagtat 4740
tgaatcgatc aggcagaaat gcccggtgga tgatgcagac gcatcatctc cccccaaaac 4800
tgtcccgtgc ctttgccgtt acgctatgac tccagaacgc gtcacccggc ttcgcatgaa 4860
ccacgtcaca agcataattg tgtgttcttc gtttcccctc ccaaagtaca aaatagaagg 4920
agtgcaaaaa gtcaaatgct ctaaggtaat gctatttgac cacaacgtgc catcgcgcgt 4980
aagtccaagg gaatatagat cttcccagga gtctgcacag gaggcgagta caatcacgtc 5040
actgacgcat agtcaattcg acctaagcgt tgatggcgag atactgcccg tcccgccaga 5100
cctggatgct gacgccccag ccctagaacc agcactagac gacggggcga cacacacgct 5160
gccatccaca accggaaacc ttgcggccgt gtctgattgg gtaatgagca ccgtacctgt 5220
cgcgccgccc agaagaaggc gagggagaaa cctgactgtg acatgtgacg agagagaagg 5280
gaatataaca cccatggcta gcgtccgatt ctttagggca gagctgtgtc cggtcgtaca 5340
agaaacagcg gagacgcgtg acacagcaat gtctcttcag gcaccaccga gtaccgccac 5400
ggaacgaat catccgccga tctccttcgg agcatcaagc gagacgttcc ccattacatt 5460
tggggacttc aacgaaggag aaatcgaaag cttgtcttct gagctactaa ctttcggaga 5520
cttcttacca ggagaagtgg atgacttgac agacagcgac tggtccacgt gctcagacac 5580
ggacgacgag ttatgactag acagggcagg tgggtatata ttctcgtcgg acaccggtcc 5640

FIGURE 5 (con't)

```
aggtcattta caacagaagt cagtacgcca gtcagtgctg ccggtgaaca ccctggagga    5700
agtccacgag gagaagtgtt acccacctaa gctggatgaa gcaaaggagc aactattact    5760
taagaaactc caggagagtg catccatggc caacagaagc aggtatcagt cgcgcaaagt    5820
agaaaacatg aaagcagcaa tcatccagag actaaagaga ggctgtagac tatacttaat    5880
gtcagagacc ccaaaagtcc ctacttaccg gactacatat ccggcgcctg tgtactcgcc    5940
tccgatcaac gtccgattgt ccaatcccga gtccgcagtg gcagcatgca atgagttctt    6000
agctagaaac tatccaactg tctcatcata ccaaattacc gacgagtatg atgcatatct    6060
agacatggtg gacgggtcgg agagttgcct ggaccgagcg acattcaatc cgtcaaaact    6120
caggagctac ccgaaacagc acgcttacca cgcgccctcc atcagaagcg ctgtaccgtc    6180
cccattccag aacacactac agaatgtact ggcagcagcc acgaaaagaa actgcaacgt    6240
cacacagatg agggaattac ccactttgga ctcagcagta ttcaacgtgg agtgtttcaa    6300
aaaattcgca tgcaaccaag aatactggga agaatttgct gccagcccta ttaggataac    6360
aactgagaat ttagcaacct atgttactaa actaaaaggg ccaaaagcag cagcgctatt    6420
cgcaaaaacc cataatctac tgccactaca ggaagtacca atggataggt tcacagtaga    6480
tatgaaaagg gacgtgaagg tgactcctgg tacaaagcat acagaggaaa gacctaaggt    6540
gcaggttata caggcggctg aacccttggc gacagcatac ctatgtggga ttcacagaga    6600
gctggttagg aggctgaacg ccgtcctcct acccaatgta catacactat ttgacatgtc    6660
tgccgaggat ttcgatgcca tcatagccgc acactttaag ccaggagaca ctgttttgga    6720
aacggacata gcctcctttg ataagagcca agatgattca cttgcgctta ctgctttgat    6780
gctgttagag gatttagggg tggatcactc cctgctggac ttgatagagg ctgctttcgg    6840
agagatttcc agctgtcacc taccgacagg tacgcgcttc aagttcggcg ccatgatgaa    6900
atcaggtatg ttcctaactc tgttcgtcaa cacattgtta aacatcacca tcgccagccg    6960
agtgctggaa gatcgtctga caaaatccgc gtgtgcggcc ttcatcggcg acgacaacat    7020
aatacatgga gtcgtctccg atgaattgat ggcagccaga tgtgccactt ggatgaacat    7080
ggaagtgaag atcatagatg cagttgtatc cttgaaagcc ccttactttt gtggagggtt    7140
tatactgcac gatactgtga caggaacagc ttgcagagtg gcagacccgc taaaaaggct    7200
ttttaaactg ggcaaaccgc tagcggcagg tgacgaacaa gatgaagata gaagacgagc    7260
gctggctgac gaagtgatca gatggcaacg aacagggcta attgatgagc tggagaaagc    7320
ggtatactct aggtacgaag tgcagggtat atcagttgtg gtaatgtcca tggccacctt    7380
tgcaagctcc agatccaact tcgagaagct cagaggaccc gtcataactt tgtacggcgg    7440
tcctaaatag gtacgcacta cagctaccta ttttgcagaa gccgacagca agtatctaaa    7500
cactaatcag ctacaatgga gttcatccca acccaaactt tttacaatag gaggtaccag    7560
cctcgaccct ggactccgcg ccctactatc caagtcatca ggcccagacc gcgccctcag    7620
aggcaagctg ggcaacttgc ccagctgatc tcagcagtta ataaactgac aatgcgcgcg    7680
gtaccccaac agaagccacg caggaatcgg aagaataaga agcaaaagca aaaacaacag    7740
gcgccacaaa acaacacaaa tcaaaagaag cagccaccta aaaagaaacc ggctcaaaag    7800
aaaaagaagc cgggccgcag agagaggatg tgcatgaaaa tcgaaaatga ttgtatttc     7860
gaagtcaagc acgaaggtaa ggtaacaggt tacgcgtgcc tggtgggga caaagtaatg     7920
aaaccagcac acgtaaaggg gaccatcgat aacgcggacc tggccaaact ggcctttaag    7980
cggtcatcta agtatgacct tgaatgcgcg cagatacccg tgcacatgaa gtccgacgct    8040
tcgaagttca cccatgagaa accggagggg tactacaact ggcaccacgg agcagtacag    8100
tactcaggag gccggttcac catccctaca ggtgctggca aaccagggga cagcggcaga    8160
ccgatcttcg acaacaaggg acgcgtggtg gccatagtct taggaggagc taatgaagga    8220
gcccgtacag ccctctcggt ggtgacctgg aataaagaca ttgtcactaa aatcaccccc    8280
gaggggggcg aagagtggag tcttgccatc ccagttatgt gcctgttggc aaacaccacg    8340
ttcccctgct cccagccccc ttgcacgccc tgctgctacg aaaaggaacc ggaggaaacc    8400
ctacgcatgc ttgaggacaa cgtcatgaga cctgggtact atcagctgct acaagcatcc    8460
ttaacatgtt ctccccaccg ccagcgacgc agcaccaagg acaacttcaa tgtctataaa    8520
```

FIGURE 5 (con't)

```
gccacaagac catacttagc tcactgtccc gactgtggag aagggcactc gtgccatagt   8580
cccgtagcac tagaacgcat cagaaatgaa gcgacagacg ggacgctgaa aatccaggtc   8640
tccttgcaaa tcggaataaa gacggatgac agccacgatt ggaccaagct gcgttatatg   8700
gacaaccaca tgccagcaga cgcagagagg gcggggctat ttgtaagaac atcagcaccg   8760
tgtacgatta ctggaacaat gggacacttc atcctggccc gatgtccaaa aggggaaact   8820
ctgacggtgg gattcactga cagtaggaag attagtcact catgtacgca cccatttcac   8880
cacgaccctc ctgtgatagg tcgggaaaaa ttccattccc gaccgcggca cggtaaagag   8940
ctaccttgca gcacgtacgt gcagagcacc gccgcaacta ccgaggagat agaggtacac   9000
atgcccccag acacccctga tcgcacatta atgtcacaac agtccggcaa cgtaaagatc   9060
acagtcaatg gccagacggt gcggtacaag tgtaattgcg gtggctcaaa tgaaggacta   9120
acaactacag acaaagtgat taataactgc aaggttgatc aatgtcatgc cgcggtcacc   9180
aatcacaaaa agtggcagta taactcccct ctggtcccgc gtaatgctga acttggggac   9240
cgaaaaggaa aaattcacat cccgtttccg ctggcaaatg taacatgcag ggtgcctaaa   9300
gcaaggaacc ccaccgtgac gtacgggaaa aaccaagtca tcatgctact gtatcctgac   9360
cacccaacac tcctgtccta ccggaatatg ggagaagaac caaactatca agaagagtgg   9420
gtgatgcata agaaggaagt cgtgctaacc gtgccgactg aagggctcga ggtcacgtgg   9480
ggcaacaacg agccgtataa gtattggccg cagttatcta caaacggtac agcccatggc   9540
cacccgcatg agataattct gtattattat gagctgtacc ccactatgac tgtagtagtt   9600
gtgtcagtgg ccacgttcat actcctgtcg atggtgggta tggcagcggg gatgtgcatg   9660
tgtgcacgac gcagatgcat cacaccgtat gaactgacac caggagctac cgtccctttc   9720
ctgcttagcc taatatgctg catcagaaca gctaaagcgg ccacatacca agaggctgcg   9780
atatacctgt ggaacgagca gcaacctttg ttttggctac aagcccttat tccgctggca   9840
gccctgattg ttctatgcaa ctgtctgaga ctcttaccat gctgctgtaa aacgttggct   9900
tttttagccg taatgagcgt cggtgcccac actgtgagcg cgtacgaaca cgtaacagtg   9960
atcccgaaca cggtgggagt accgtataag actctagtca atagacctgg ctacagcccc  10020
atggtattgg agatggaact actgtcagtc actttggagc caacactatc gcttgattac  10080
atcacgtgcg agtacaaaac cgtcatcccg tctccgtacg tgaagtgctg cggtacagca  10140
gagtgcaagg acaaaaacct acctgactac agctgtaagg tcttcaccgg cgtctaccca  10200
tttatgtggg gcggcgccta ctgcttctgc gacgctgaaa acacgcagtt gagcgaagca  10260
cacgtggaga agtccgaatc atgcaaaaca gaatttgcat cagcatacag ggctcatacc  10320
gcatctgcat cagctaagct ccgcgtcctt taccaaggaa ataacatcac tgtaactgcc  10380
tatgcaaacg gcgaccatgc cgtcacagtt aaggacgcca aattcattgt ggggccaatg  10440
tcttcagcct ggacaccttt cgacaacaaa attgtggtgt acaaaggtga cgtctataac  10500
atggactacc cgccctttgg cgcaggaaga ccaggacaat ttggcgatat ccaaagtcgc  10560
acacctgaga gtaaagacgt ctatgctaat acacaactgg tactgcagag accggctgcg  10620
ggtacggtac acgtgccata ctctcaggca ccatctggct ttaagtattg gctaaaagaa  10680
cgcggggcgt cgctgcagca cacagcacca tttggctgcc aaatagcaac aaacccggta  10740
agagcggtga actgcgccgt agggaacatg cccatctcca tcgacatacc ggaagcggcc  10800
ttcactaggg tcgtcgacgc gccctcttta acggacatgt cgtgcgaggt accagcctgc  10860
acccattcct cagactttgg gggcgtcgcc attattaaat atgcagccag caagaaaggc  10920
aagtgtgcgg tgcattcgat gactaacgcc gtcactattc gggaagctga gatagaagtt  10980
gaagggaatt ctcagctgca aatctctttc tcgacggcct tagccagcgc cgaattccgc  11040
gtacaagtct gttctacaca agtacactgt gcagccgagt gccacccccc gaaggaccac  11100
atagtcaact acccggcgtc acataccacc ctcggggtcc aggacatctc cgctacggcg  11160
atgtcatggg tgcagaagat cacgggaggt gtgggactgg ttgttgctgt tgccgcactg  11220
attctaatcg tggtgctatg cgtatcgttc agcaggcact aacttgacaa ttaagtatga  11280
aggtatatgt gtcccctaag agacacactg tacatagcaa ataatctata gatcaaaggg  11340
ctacgcaacc cctgaatagt aacaaaatac aaaatcacta aaaattataa aaacagaaaa  11400
```

FIGURE 5 (con't)

```
atacataaat aggtatacgt gtcccctaag agacacattg tatgtaggtg ataagtatag   11460
atcaaagggc cgaataaccc ctgaatagta acaaaatatg aaaatcaata aaaatcataa   11520
aatagaaaaa ccataaacag aagtagttca aagggctata aaacccctga atagtaacaa   11580
aacataaaat taataaaaat c                                             11601
```

FIGURE 6

```
caaagcaaga gattaataac ccatcatgga tcctgtgtac gtggacatag acgctgacag      60
cgccttttg aaggccctgc aacgtgcgta ccccatgttt gaggtggaac caaggcaggt      120
cacaccgaat gaccatgcta atgctagagc gttctcgcat ctagctataa aactaataga     180
gcaggaaatt gaccccgact caaccatcct ggatatcggc agtgcgccag caaggaggat      240
gatgtcggac aggaagtacc actgcgtctg cccgatgcgc agtgcggaag atcccgagag      300
actcgccaat tatgcgagaa agctagcatc tgccgcagga aaagtcctgg acagaaacat      360
ctctggaaag atcggggact tacaagcagt aatggccgtg ccagacacgg agacgccaac      420
attctgctta cacacagacg tctcatgtag acagagagca gacgtcgcta tataccaaga      480
cgtctatgct gtacacgcac ccacgtcgct ataccaccag gcgattaaag gggtccgagt      540
ggcgtactgg gttgggttcg acacaacccc gttcatgtac aatgccatgg cgggtgccta      600
cccctcatac tcgacaaact gggcagatga gcaggtactg aaggctaaga acataggatt      660
atgttcaaca gacctgacgg aaggtagacg aggcaagttg tctattatga gagggaaaaa      720
gctaaaaccg tgcgaccgtg tgctgttctc agtagggtca acgctctacc cggaaagccg      780
caagctactt aagagctggc acctgccatc ggtgttccat ttaaagggca aactcagctt      840
cacatgccgc tgtgatacag tggtttcgtg tgagggctac gtcgttaaga gaataacgat      900
gagcccaggc ctttatggaa aaaccacagg gtatgcggta acccaccacg cagacggatt      960
cctgatgtgc aagactaccg acacggttga cggcgaaaga gtgtcattct cggtgtgcac     1020
atacgtgccg gcgaccattt gtgatcaaat gaccggcatc cttgctacag aagtcacgcc     1080
ggaggatgca cagaagctgt tggtggggct gaaccagaga atagtggtta acggcagaac     1140
gcaacggaat acgaacacca tgaaaaatta tctgcttccc gtggtcgccc aagccttcag     1200
taagtgggca aaggagtgcc ggaaagacat ggaagatgaa aaactcctgg gggtcagaga     1260
aagaacactg acctgctgct gtctatgggc attcaagaag cagaaaacac acacggtcta     1320
caagaggcct gatacccagt caattcagaa ggttcaggcc gagtttgaca gctttgtggt     1380
accgagtctg tggtcgtccg ggttgtcaat ccctttgagg actagaatca aatggttgtt     1440
aagcaaggtg ccaaaaaccg acctgatccc atacagcgga gacgcccgag aagcccggga     1500
cgcagaaaaa gaagcagagg aagaacgaga agcagaactg actcgcgaag ccctaccacc     1560
tctacaggca gcacaggaag atgttcaggt cgaaatcgac gtggaacagc ttgaggacag     1620
agcgggcgca ggaataatag agactccgag aggagctatc aaagttactg cccaaccaac     1680
agaccacgtc gtgggagagt acctggtact ctccccgcag accgtactac gtagccagaa     1740
gctcagtctg attcacgctt tggcggagca agtgaagacg tgcacgcaca acggacgagc     1800
agggaggtat gcggtcgaag cgtacgacgg ccgagtccta gtgccctcag gctatgcaat     1860
ctcgcctgaa gacttccaga gtctaagcga aagcgcaacg atggtgtata acgaaagaga     1920
gttcgtaaac agaaagctac accatattgc gatgcacgga ccagccctga acaccgacga     1980
agagtcgtat gagctggtga gggcagagag gacagaacac gagtacgtct acgacgtgga     2040
tcagagaaga tgctgtaaga aggaagaagc cgcaggactg gtactggtgg gcgacttgac     2100
taatccgccc taccacgaat tcgcatatga agggctaaaa atccgccctg cctgcccata     2160
caaaattgca gtcataggag tcttcggagt accgggatct ggcaagtcag ctattatcaa     2220
gaacctagtt accaggcagg acctggtgac tagcggaaag aaagaaaact gccaagaaat     2280
caccaccgac gtgatgagac agagaggtct agagatatct gcacgtacgg ttgactcgct     2340
gctcttgaat ggatgcaaca gaccagtcga cgtgttgtac gtagacgagg cgtttgcgtg     2400
ccactctgga acgctacttg ctttgatcgc cttggtgaga ccaaggcaga aagttgtact     2460
ttgtggtgac ccgaagcagt gcggcttctt caatatgatg cagatgaaag tcaactataa     2520
tcacaacatc tgcacccaag tgtaccacaa aagtatctcc aggcggtgta cactgcctgt     2580
gaccgccatt gtgtcatcgt tgcattacga aggcaaaatg cgcactacga atgagtacaa     2640
caagccgatt gtagtggaca ctacaggctc aacaaaacct gaccctggag acctcgtgtt     2700
aacgtgcttc agagggtggg ttaaacaact gcaaattgac tatcgtggat acgaggtcat     2760
gacagcagcc gcatcccaag ggttaaccag aaaaggagtt tacgcagtta gacaaaaagt     2820
```

FIGURE 6 (con't)

taatgaaaac ccgctctatg catcaacgtc agagcacgtc aacgtactcc taacgcgtac 2880
ggaaggtaaa ctggtatgga agacactttc cggcgacccg tggataaaga cgctgcagaa 2940
cccaccgaaa ggaaacttca aagcaactat taaggagtgg gaggtggagc atgcatcaat 3000
aatggcgggc atctgcagtc accaaatgac cttcgataca ttccaaaata aagccaacgt 3060
ttgttgggct aagagcttgg tccctatcct cgaaacagcg gggataaaac taaatgatag 3120
gcagtggtct cagataattc aagccttcaa agaagacaaa gcatactcac ctgaagtagc 3180
cctgaatgaa atatgtacgc gcatgtatgg ggtggatcta gacagcgggc tattttctaa 3240
accgttggtg tctgtgtatt acgcggataa ccactgggat aataggcctg gagggaaaat 3300
gttcggattt aaccccgagg cagcatccat tctagaaaga aagtatccat tcacaaaagg 3360
gaagtggaac atcaacaagc agatctgcgt gactaccagg aggatagaag actttaaccc 3420
taccaccaac atcataccgg ccaacaggag actaccacac tcattagtgg ccgaacaccg 3480
cccagtaaaa ggggaaagaa tggaatggct ggttaacaag ataaacggcc accacgtgct 3540
cctggtcagt ggctataacc ttgcactgcc tactaagaga gtcacttggg tagcgccgtt 3600
aggtgtccgc ggagcggact acacatacaa cctagagttg ggtctgccag caacgcttgg 3660
taggtatgac ctagtggtca taaacatcca cacacctttt cgcatacacc attaccaaca 3720
gtgcgtcgac cacgcaatga aactgcaaat gctcgggggt gactcattga gactgctcaa 3780
accgggcggc tctctattga tcagagcata tggttacgca gatagaacca gtgaacgagt 3840
catctgcgta ttgggacgca agtttagatc gtctagagcg ttgaaaccac catgtgtcac 3900
cagcaacact gagatgtttt tcctattcag caactttgac aatggcagaa ggaatttcac 3960
aactcatgtc atgaacaatc aactgaatgc agccttcgta ggacaggtca cccgagcagg 4020
atgtgcaccg tcgtaccggg taaaacgcat ggacatcgcg aagaacgatg aagagtgcgt 4080
agtcaacgcc gctaaccctc gcgggttacc gggtgacggt gtttgcaagg cagtatacaa 4140
aaaatggccg gagtccttta agaacagtgc aacaccagtg ggaaccgcaa aaacagttat 4200
gtgcggtacg tatccagtaa tccacgctgt tggaccaaac ttctctaatt attcggagtc 4260
tgaaggggac cgggaattgg cagctgccta tcgagaagtc gcaaaggaag taactaggct 4320
gggagtaaat agtgtagcta tacctctcct ctccacaggt gtatactcag gagggaaaga 4380
caggctgacc cagtcactga accacctctt tacagccatg gactcgacgg atgcagacgt 4440
ggtcatctac tgccgcgaca aagaatggga gaagaaaata tctgaggcca tacagatgcg 4500
gacccaagta gagctgctgg atgagcacat ctccatagac tgcgatattg ttcgcgtgca 4560
ccctgacagc agcttggcag gcagaaaagg atacagcacc acggaaggcg cactgtactc 4620
atatctagaa gggacccgtt ttcatcagac ggctgtggat atggcggaga tacatactat 4680
gtggccaaag caaacagagg ccaatgagca agtctgccta tatgccctgg gggaaagtat 4740
tgaatcgatc aggcagaaat gcccggtgga tgatgcagac gcatcatctc cccccaaaac 4800
tgtcccgtgc ctttgccgtt acgctatgac tccagaacgc gtcacccggc ttcgcatgaa 4860
ccacgtcaca agcataattg tgtgttcttc gtttcccctc ccaaagtaca aaatagaagg 4920
agtgcaaaaa gtcaaatgct ctaaggtaat gctatttgac cacaacgtgc catcgcgcgt 4980
aagtccaagg gaatatagat cttcccagga gtctgcacag gaggcgagta caatcacgtc 5040
actgacgcat agtcaattcg acctaagcgt tgatggcgag atactgcccg tcccgtcaga 5100
cctggatgct gacgccccag ccctagaacc agcactagac gacggggcga cacacacgct 5160
gccatccaca accggaaacc ttgcggccgt gtctgattgg gtaatgagca ccgtacctgt 5220
cgcgccgccc agaagaaggc gagggagaaa cctgactgtg acatgtgacg agagagaagg 5280
gaatataaca cccatggcta gcgtccgatt ctttagggca gagctgtgtc cggtcgtaca 5340
agaaacagcg gagacgcgtg acacagcaat gtctcttcag gcaccaccga gtaccgccac 5400
ggaaccgaat catccgccga tctccttcgg agcatcaagc gagacgttcc ccattacatt 5460
tggggacttc aacgaaggag aaatcgaaag cttgtcttct gagctactaa ctttcggaga 5520
cttcttacca ggagaagtgg atgacttgac agacagcgac tggtccacgt gctcagacac 5580
ggacgacgag ttatgactag acagggcagg tgggtatata ttctcgtcgg acaccggtcc 5640
aggtcattta caacagaagt cagtacgcca gtcagtgctg ccggtgaaca ccctggagga 5700

FIGURE 6 (con't)

agtccacgag gagaagtgtt acccacctaa gctggatgaa gcaaaggagc aactattact 5760
taagaaactc caggagagtg catccatggc caacagaagc aggtatcagt cgcgcaaagt 5820
agaaaacatg aaagcagcaa tcatccagag actaaagaga ggctgtagac tatacttaat 5880
gtcagagacc ccaaaagtcc ctacttaccg gactacatat ccggcgcctg tgtactcgcc 5940
tccgatcaac gtccgattgt ccaatcccga gtccgcagtg gcagcatgca atgagttctt 6000
agctagaaac tatccaactg tctcatcata ccaaattacc gacgagtatg atgcatatct 6060
agacatggtg gacgggtcgg agagttgcct ggaccgagcg acattcaatc cgtcaaaact 6120
caggagctac ccgaaacagc acgcttacca cgcgccctcc atcagaagcg ctgtaccgtc 6180
cccattccag aacacactac agaatgtact ggcagcagcc acgaaaagaa actgcaacgt 6240
cacacagatg agggaattac ccactttgga ctcagcagta ttcaacgtgg agtgtttcaa 6300
aaaattcgca tgcaaccaag aatactggga agaatttgct gccagcccta ttaggataac 6360
aactgagaat ttagcaacct atgttactaa actaaaaggg ccaaaagcag cagcgctatt 6420
cgcaaaaacc cataatctac tgccactaca ggaagtacca atggataggt tcacagtaga 6480
tatgaaaagg gacgtgaagg tgactcctgg tacaaagcat acagaggaaa gacctaaggt 6540
gcaggttata caggcggctg aacccttggc gacagcatac ctatgtggga ttcacagaga 6600
gctggttagg aggctgaacg ccgtcctcct acccaatgta catacactat ttgacatgtc 6660
tgccgaggat ttcgatgcca tcatagccgc acactttaag ccaggagaca ctgttttgga 6720
aacggacata gcctcctttg ataagagcca agatgattca cttgcgctta ctgctttgat 6780
gctgttagag gatttagggg tggatcactc cctgctggac ttgatagagg ctgctttcgg 6840
agagatttcc agctgtcacc taccgacagg tacgcgcttc aagttcggcg ccatgatgaa 6900
atcaggtatg ttcctaactc tgttcgtcaa cacattgtta aacatcacca tcgccagccg 6960
agtgctggaa gatcgtctga caaaatccgc gtgcgcggcc ttcatcggcg acgacaacat 7020
aatacatgga gtcgtctccg atgaattgat ggcagccaga tgtgccactt ggatgaacat 7080
ggaagtgaag atcatagatg cagttgtatc cttgaaagcc ccttactttt gtggagggtt 7140
tatactgcac gatactgtga caggaacagc ttgcagagtg gcagacccgc taaaaaggct 7200
ttttaaactg ggcaaaccgc tagcggcagg tgacgaacaa gatgaagata gaagacgagc 7260
gctggctgac gaagtgatca gatggcaacg aacagggcta attgatgagc tggagaaagc 7320
ggtatactct aggtacgaag tgcagggtat atcagttgtg gtaatgtcca tggccacctt 7380
tgcaagctcc agatccaact tcgagaagct cagaggaccc gtcataactt tgtacggcgg 7440
tcctaaatag gtacgcacta cagctaccta ttttgcagaa gccgacagca agtatctaaa 7500
cactaatcag ctacaatgga gttcatccca acccaaactt tttacaatag gaggtaccag 7560
cctcgaccct ggactccgcg ccctactatc caagtcatca ggcccagacc gcgccctcag 7620
aggcaagctg ggcaacttgc ccagctgatc tcagcagtta ataaactgac aatgcgcgcg 7680
gtaccccaac agaagccacg caggaatcgg aagaataaga agcaaaagca aaaacaacag 7740
gcgccacaaa acaacacaaa tcaaaagaag cagccaccta aaaagaaacc ggctcaaaag 7800
aaaaagaagc cgggccgcag agagaggatg tgcatgaaaa tcgaaaatga ttgtattttc 7860
gaagtcaagc acgaaggtaa ggtaacaggt tacgcgtgcc tggtggggga caaagtaatg 7920
aaaccagcac acgtaaaggg gaccatcgat aacgcggacc tggccaaact ggcctttaag 7980
cggtcatcta agtatgacct tgaatgcgcg cagatacccg tgcacatgaa gtccgacgct 8040
tcgaagttca cccatgagaa accggagggg tactacaact ggcaccacgg agcagtacag 8100
tactcaggag gccggttcac catccctaca ggtgctggca aaccagggga cagcggcaga 8160
ccgatcttcg acaacaaggg acgcgtggtg gccatagtct taggaggagc taatgaagga 8220
gcccgtacag ccctctcggt ggtgacctgg aataaagaca ttgtcactaa aatcaccccc 8280
gaggggggccg aagagtggag tcttgccatc ccagttatgt gcctgttggc aaacaccacg 8340
ttcccctgct cccagccccc ttgcacgccc tgctgctacg aaaaggaacc ggaggaaacc 8400
ctacgcatgc ttgaggacaa cgtcatgaga cctgggtact atcagctgct acaagcatcc 8460
ttaacatgtt ctccccaccg ccagcgacgc agcaccaagg acaacttcaa tgtctataaa 8520
gccacaagac catacttagc tcactgtccc gactgtggag aagggcactc gtgccatagt 8580

FIGURE 6 (con't)

cccgtagcac tagaacgcat cagaaatgaa gcgacagacg ggacgctgaa aatccaggtc 8640
tccttgcaaa tcggaataaa gacggatgac agccacgatt ggaccaagct gcgttatatg 8700
gacaaccaca tgccagcaga cgcagagagg gcggggctat ttgtaagaac atcagcaccg 8760
tgtacgatta ctggaacaat gggacacttc atcctggccc gatgtccaaa aggggaaact 8820
ctgacggtgg gattcactga cagtaggaag attagtcact catgtacgca cccatttcac 8880
cacgaccctc ctgtgatagg tcgggaaaaa ttccattccc gaccgcagca cggtaaagag 8940
ctaccttgca gcacgtacgt gcagagcacc gccgcaacta ccgaggagat agaggtacac 9000
atgcccccag acacccctga tcgcacatta atgtcacaac agtccggcaa cgtaaagatc 9060
acagtcaatg gccagacggt gcggtacaag tgtaattgcg gtggctcaaa tgaaggacta 9120
acaactacag acaaagtgat taataactgc aaggttgatc aatgtcatgc cgcggtcacc 9180
aatcacaaaa agtggcagta taactccccc ctggtcccgc gtaatgctga acttggggac 9240
cgaaaaggaa aaattcacat cccgtttccg ctggcaaatg taacatgcag ggtgcctaaa 9300
gcaaggaacc ccaccgtgac gtacgggaaa aaccaagtca tcatgctact gtatcctgac 9360
cacccaacac tcctgtccta ccggaatatg ggagaagaac caaactatca agaagagtgg 9420
gtgatgcata agaaggaagt cgtgctaacc gtgccgactg aagggctcga ggtcacgtgg 9480
ggcaacaacg agccgtataa gtattggccg cagttatcta caaacggtac agcccatggc 9540
cacccgcatg agataattct gtattattat gagctgtacc ccactatgac tgtagtagtt 9600
gtgtcagtgg ccacgttcat actcctgtcg atggtgggta tggcagcggg gatgtgcatg 9660
tgtgcacgac gcagatgcat cacaccgtat gaactgacac caggagctac cgtccctttc 9720
ctgcttagcc taatatgctg catcagaaca gctaaagcgg ccacatacca agaggctgcg 9780
atatacctgt ggaacgagca gcaacctttg ttttggctac aagcccttat tccgctggca 9840
gccctgattg ttctatgcaa ctgtctgaga ctcttaccat gctgctgtaa aacgttggct 9900
tttttagccg taatgagcgt cggtgcccac actgtgagcg cgtacgaaca cgtaacagtg 9960
atcccgaaca cggtgggagt accgtataag actctagtca atagacctgg ctacagcccc 10020
atggtattgg agatggaact actgtcagtc actttggagc caacactatc gcttgattac 10080
atcacgtgcg agtacaaaac cgtcatcccg tctccgtacg tgaagtgctg cggtacagca 10140
gagtgcaagg acaaaaacct acctgactac agctgtaagg tcttcaccgg cgtctaccca 10200
tttatgtggg gcggcgccta ctgcttctgc gacgctgaaa acacgcagtt gagcgaagca 10260
cacgtggaga agtccgaatc atgcaaaaca gaatttgcat cagcatacag ggctcatacc 10320
gcatctgcat cagctaagct ccgcgtcctt taccaaggaa ataacatcac tgtaactgcc 10380
tatgcaaacg gcgaccatgc cgtcacagtt aaggacgcca aattcattgt ggggccaatg 10440
tcttcagcct ggacaccttt cgacaacaaa attgtggtgt acaaaggtga cgtctataac 10500
atggactacc cgccctttgg cgcaggaaga ccaggacaat ttggcgatat ccaaagtcgc 10560
acacctgaga gtaaagacgt ctatgctaat acacaactgg tactgcagag accggctgtg 10620
ggtacggtac acgtgccata ctctcaggca ccatctggct ttaagtattg gctaaaagaa 10680
cgcggggcgt cgctgcagca cacagcacca tttggctgcc aaatagcaac aaacccggta 10740
agagcggtga actgcgccgt agggaacatg cccatctcca tcgacatacc ggaagcggcc 10800
ttcactaggg tcgtcgacgc gccctcttta acggacatgt cgtgcgaggt accagcctgc 10860
acccattcct cagactttgg gggcgtcgcc attattaaat atgcagccag caagaaaggc 10920
aagtgtgcgg tgcattcgat gactaacgcc gtcactattc gggaagctga gatagaagtt 10980
gaagggaatt ctcagctgca aatctctttc tcgacggcct tagccagcgc cgaattccgc 11040
gtacaagtct gttctacaca agtacactgt gcagccgagt gccacccccc gaaggaccac 11100
atagtcaact acccggcgtc acataccacc ctcggggtcc aggacatctc cgctacggcg 11160
atgtcatggg tgcagaagat cacgggaggt gtgggactgg ttgttgctgt tgccgcactg 11220
attctaatcg tggtgctatg cgtgtcgttc agcaggcact aacttgacaa ttaagtatga 11280
aggtatatgt gtcccctaag agacacactg tacatagcaa ataatctata gatcaaaggg 11340
ctacgcaacc cctgaatagt aacaaaatac aaaatcacta aaaattataa aaacagaaaa 11400
atacataaat aggtatacgt gtcccctaag agacacattg tatgtaggtg ataagtatag 11460

FIGURE 6 (con't)

```
atcaaagggc cgaataaccc ctgaatagta acaaaatatg aaaatcaata aaaatcataa   11520
aatagaaaaa ccataaacag aagtagttca aagggctata aaacccctga atagtaacaa   11580
aacataaaat taataaaaat c                                             11601
```

FIGURE 7

```
caaagcaaga gattaataac ccatcatgga tcctgtgtac gtggacatag acgctgacag     60
cgccttttg aaggccctgc aacgtgcgta ccccatgttt gaggtggaac caaggcaggt     120
cacaccgaat gaccatgcta atgctagagc gttctcgcat ctagctataa aactaataga     180
gcaggaaatt gaccccgact caaccatcct ggatatcggc agtgcgccag caaggaggat     240
gatgtcggac aggaagtacc actgcgtctg cccgatgcgc agtgcggaag atcccgagag     300
actcgccaat tatgcgagaa agctagcatc tgccgcagga aaagtcctgg acagaaacat     360
ctctggaaag atcggggact tacaagcagt aatggccgtg ccagacacgg agacgccaac     420
attctgctta cacacagacg tctcatgtag acagagagca gacgtcgcta tataccaaga     480
cgtctatgct gtacacgcac ccacgtcgct ataccaccag gcgattaaag gggtccgagt     540
ggcgtactgg gttgggttcg acacaacccc gttcatgtac aatgccatgg cgggtgccta     600
cccctcatac tcgacaaact gggcagatga gcaggtactg aaggctaaga acataggatt     660
atgttcaaca gacctgacgg aaggtagacg aggcaagttg tctattatga gagggaaaaa     720
gctaaaaccg tgcgaccgtg tgctgttctc agtagggtca acgctctacc cggaaagccg     780
caagctactt aagagctggc acctgccatc ggtgttccat ttaaagggca aactcagctt     840
cacatgccgc tgtgatacag tggtttcgtg tgagggctac gtcgttaaga gaataacgat     900
gagcccaggc ctttatggaa aaaccatagg gtatgcggta acccaccacg cagacggatt     960
cctgatgtgc aagactaccg acacggttga cggcgaaaga gtgtcattct cggtgtgcac    1020
atacgtgccg gcgaccattt gtgatcaaat gaccggcatc cttgctacag aagtcacgcc    1080
ggaggatgca cagaagctgt tggtggggct gaaccagaga atagtggtta acggcagaac    1140
gcaacggaat acgaacacca tgaaaaatta tctgcttccc gtggtcgccc aagccttcag    1200
taagtgggca aaggagtgcc ggaaagacat ggaagatgaa aaactcctgg gggtcagaga    1260
aagaacactg acctgctgct gtctatgggc attcaagaag cagaaaacac acacggtcta    1320
caagaggcct gatacccagt caattcagaa ggttcaggcc gagtttgaca gctttgtggt    1380
accgagtctg tggtcgtccg ggttgtcaat ccctttgagg actagaatca aatggttgtt    1440
aagcaaggtg ccaaaaaccg acctgatccc atacagcgga gacgcccgag aagcccggga    1500
cgcagaaaaa gaagcagagg aagaacgaga agcagaactg actcgcgaag ccctaccacc    1560
tctacaggca gcacaggaag atgttcaggt cgaaatcgac gtggaacagc ttgaggacag    1620
agcgggcgca ggaataatag agactccgag aggagctatc aaagttactg cccaaccaac    1680
agaccacgtc gtgggagagt acctggtact ctccccgcag accgtactac gtagccagaa    1740
gctcagtctg attcacgctt tggcggagca agtgaagacg tgcacgcaca acggacgagc    1800
agggaggtat gcggtcgaag cgtacgacgg ccgagtccta gtgccctcag gctatgcaat    1860
ctcgcctgaa gacttccaga gtctaagcga aagcgcaacg atggtgtata acgaaagaga    1920
gttcgtaaac agaaagctac accatattgc gatgcacgga ccagccctga acaccgacga    1980
agagtcgtat gagctggtga gggcagagag gacagaacac gagtacgtct acgacgtgga    2040
tcagagaaga tgctgtaaga aggaagaagc cgcaggactg gtactggtgg gcgacttgac    2100
taatccgccc taccacgaat tcgcatatga agggctaaaa atccgccctg cctgcccata    2160
caaaattgca gtcataggag tcttcggagt accgggatct ggcaagtcag ctattatcaa    2220
gaacctagtt accaggcagg acctggtgac tagcggaaag aaagaaaact gccaagaaat    2280
caccaccgac gtgatgagac agagaggtct agagatatct gcacgtacgg ttgactcgct    2340
gctcttgaat ggatgcaaca gaccagtcga cgtgttgtac gtagacgagg cgtttgcgtg    2400
ccactctgga acgctacttg ctttgatcgc cttggtgaga ccaaggcaga aagttgtact    2460
ttgtggtgac ccgaagcagt gcggcttctt caatatgatg cagatgaaag tcaactataa    2520
tcacaacatc tgcacccaag tgtaccacaa aagtatctcc aggcggtgta cactgcctgt    2580
gaccgccatt gtgtcatcgt tgcattacga aggcaaaatg cgcactacga atgagtacaa    2640
caagccgatt gtagtggaca ctacaggctc aacaaaacct gaccctggag acctcgtgtt    2700
aacgtgcttc agagggtggg ttaaacaact gcaaattgac tatcgtggat acgaggtcat    2760
```

FIGURE 7 (con't)

gacagcagcc gcatcccaag ggttaaccag aaaaggagtt tacgcagtta gacaaaaagt 2820
taatgaaaac ccgctctatg catcaacgtc agagcacgtc aacgtactcc taacgcgtac 2880
ggaaggtaaa ctggtatgga agacactttc cggcgacccg tggataaaga cgctgcagaa 2940
cccaccgaaa ggaaacttca aagcaactat taaggagtgg gaggtggagc atgcatcaat 3000
aatggcgggc atctgcagtc accaaatgac cttcgataca ttccaaaata aagccaacgt 3060
ttgttgggct aagagcttgg tccctatcct cgaaacagcg gggataaaac taaatgatag 3120
gcagtggtct cagataattc aagccttcaa agaagacaaa gcatactcac ctgaagtagc 3180
cctgaatgaa atatgtacgc gcatgtatgg ggtggatcta gacagcgggc tattttctaa 3240
accgttggtg tctgtgtatt acgcggataa ccactgggat aataggcctg gagggaaaat 3300
gttcggattt aaccccgagg cagcatccat tctagaaaga aagtatccat tcacaaaagg 3360
gaagtggaac atcaacaagc agatctgcgt gactaccagg aggatagaag actttaaccc 3420
taccaccaac atcataccgg ccaacaggag actaccacac tcattagtgg ccgaacaccg 3480
cccagtaaaa ggggaaagaa tggaatggct ggttaacaag ataaacggcc accacgtgct 3540
cctggtcagt ggcaataacc ttgcactgcc tactaagaga gtcacttggg tagcgccgtt 3600
aggtgtccgc ggagcggact acacatacaa cctagagttg ggtctgccag caacgcttgg 3660
taggtatgac ctagtggtca taaacatcca cacacctttt cgcatacacc attaccaaca 3720
gtgcgtcgac cacgcaatga aactgcaaat gctcgggggt gactcattga gactgctcaa 3780
accgggcggc tctctattga tcagagcata tggttacgca gatagaacca gtgaacgagt 3840
catctgcgta ttgggacgca agtttagatc gtctagagcg ttgaaaccac catgtgtcac 3900
cagcaacact gagatgtttt tcctattcag caactttgac aatggcagaa ggaatttcac 3960
aactcatgtc atgaacaatc aactgaatgc agccttcgta ggacaggtca cccgagcagg 4020
atgtgcaccg tcgtaccggg taaaacgcat ggacatcgcg aagaacgatg aagagtgcgt 4080
agtcaacgcc gctaaccctc gcgggttacc gggtgacggt gtttgcaagg cagtatacaa 4140
aaaatggccg gagtccttta agaacagtgc aacaccagtg ggaaccgcaa aaacagttat 4200
gtgcggtacg tatccagtaa tccacgctgt tggaccaaac ttctctaatt attcggagtc 4260
tgaaggggac cgggaattgg cagctgccta tcgagaagtc gcaaaggaag taactaggct 4320
gggagtaaat agtgtagcta tacctctcct ctccacaggt gtatactcag gagggaaaga 4380
caggctgacc cagtcactga accacctctt tacagccatg gactcgacgg atgcagacgt 4440
ggtcatctac tgccgcgaca aagaatggga gaagaaaata tctgaggcca tacagatgcg 4500
gacccaagta gagctgctgg atgagcacat ctccatagac tgcgatattg ttcgcgtgca 4560
ccctgacagc agcttggcag gcagaaaagg atacagcacc acggaaggcg cactgtactc 4620
atatctagaa gggacccgtt ttcatcagac ggcagtggat atggcggaga tacatactat 4680
gtggccaaag caaacagagg ccaatgagca agtctgccta tatgccctgg gggaaagtat 4740
tgaatcgatc aggcagaaat gcccggtgga tgatgcagac gcatcatctc cccccaaaac 4800
tgtcccgtgc ctttgccgtt acgctatgac tccagaacgc gtcacccggc ttcgcatgaa 4860
ccacgtcaca agcataattg tgtgttcttc gtttcccctc ccaaagtaca aaatagaagg 4920
agtgcaaaaa gtcaaatgct ctaaggtaat gctatttgac cacaacgtgc catcgcgcgt 4980
aagtccaagg gaatatagat cttcccagga gtctgcacag gaggcgagta caatcacgtc 5040
actgacgcat agtcaattcg acctaagcgt tgatggcgag atactgcccg tcccgtcaga 5100
cctggatgct gacgccccag ccctagaacc agcactagac gacggggcga cacacacgct 5160
gccatccaca accggaaacc ttgcggccgt gtctgattgg gtaatgagca ccgtacctgt 5220
cgcgccgccc agaagaaggc gagggagaaa cctgactgtg acatgtgacg agagagaagg 5280
gaatataaca cccatggcta gcgtccgatt ctttagggca gagctgtgtc cggtcgtaca 5340
agaaacagcg gagacgcgtg acacagcaat gtctcttcag gcaccaccga gtaccgccac 5400
accgaatcat ccgccgatct ccttcggagc atcaagcgag acgttcccca ttacatttgg 5460
ggacttcaac gaaggagaaa tcgaaagctt gtcttctgag ctactaactt tcggagactt 5520
cttaccagga gaagtggatg acttgacaga cagcgactgg tccacgtgct cagacacgga 5580
cgacgagtta tgactagaca gggcaggtgg gtatatattc tcgtcggaca ccggtccagg 5640

FIGURE 7 (con't)

```
tcatttacaa cagaagtcag tacgccagtc agtgctgccg gtgaacaccc tggaggaagt   5700
ccacgaggag aagtgttacc cacctaagct ggatgaagca aaggagcaac tattacttaa   5760
gaaactccag gagagtgcat ccatggccaa cagaagcagg tatcagtcgc gcaaagtaga   5820
aaacatgaaa gcagcaatca tccagagact aaagagaggc tgtagactat acttaatgtc   5880
agagacccca aaagtcccta cttaccggac tacatatccg gcgcctgtgt actcgcctcc   5940
gatcaacgtc cgattgtcca atcccgagtc cgcagtggca gcatgcaatg agttcttagc   6000
tagaaactat ccaactgtct catcatacca aattaccgac gagtatgatg catatctaga   6060
catggtggac gggtcggaga gttgcctgga ccgagcgaca ttcaatccgt caaaactcag   6120
gagctacccg aaacagcacg cttaccacgc gccctccatc agaagcgctg taccgtcccc   6180
attccagaac acactacaga atgtactggc agcagccacg aaaagaaact gcaacgtcac   6240
acagatgagg gaattaccca ctttggactc agcagtattc aacgtggagt gtttcaaaaa   6300
attcgcatgc aaccaagaat actgggaaga atttgctgcc agccctatta ggataacaac   6360
tgagaattta gcaacctatg ttactaaact aaaagggcca aaagcagcag cgctattcgc   6420
aaaaacccat aatctactgc cactacagga agtaccaatg gataggttca cagtagatat   6480
gaaaagggac gtaaaggtga ctcctggtac aaagcataca gaggaaagac ctaaggtgca   6540
ggttatacag gcggctgaac ccttggcgac agcataccta tgtgggattc acagagagct   6600
ggttaggagg ctgaacgccg tcctcctacc caatgtacat acactatttg acatgtctgc   6660
cgaggatttc gatgccatca tagccgcaca ctttaagcca ggagacactg ttttggaaac   6720
ggacatagcc tcctttgata agagccaaga tgattcactt gcgcttactg ctttgatgct   6780
gttagaggat ttaggggtgg atcactccct gctggacttg atagaggctg ctttcggaga   6840
gatttccagc tgtcacctac cgacaggtac gcgcttcaag ttcggcgcca tgatgaaatc   6900
aggtatgttc ctaactctgt tcgtcaacac attgttaaac atcaccatcg ccagccgagt   6960
gctggaagat cgtctgacaa aatccgcgtg cgcggccttc atcggcgacg acaacataat   7020
acatggagtc gtctccgatg aattgatggc agccagatgt gccacttgga tgaacatgga   7080
agtgaagatc atagatgcag ttgtatcctt gaaagcccct tacttttgtg gagggtttat   7140
actgcacgat actgtgacag gaacagcttg cagagtggca gacccgctaa aaaggctttt   7200
taaactgggc aaaccgctag cggcaggtga cgaacaagat gaagatagaa gacgagcgct   7260
ggctgacgaa gtgatcagat ggcaacgaac agggctaatt gatgagctgg agaaagcggt   7320
atactctagg tacgaagtgc agggtatatc agttgtggta atgtccatgg ccacctttgc   7380
aagctccaga tccaacttcg agaagctcag aggacccgtc ataactttgt acggcggtcc   7440
taaataggta cgcactacag ctacctattt tgcagaagcc gacagcaagt atctaaacac   7500
taatcagcta caatggagtt catcccaacc caaacttttt acaataggag gtaccagcct   7560
cgaccctgga ctccgcgccc tactatccaa gtcatcaggc ccagaccgcg ccctcagagg   7620
caagctgggc aacttgccca gctgatctca gcagttaata aactgacaat gcgcgcggta   7680
ccccaacaga agccacgcag gaatcggaag aataagaagc aaaagcaaaa acaacaggcg   7740
ccacaaaaca acacaaatca aaagaagcag ccacctaaaa agaaaccggc tcaaaagaaa   7800
aagaagccgg gccgcagaga gaggatgtgc atgaaaatcg aaaatgattg tattttcgaa   7860
gtcaagcacg aaggtaaggt aacaggttac gcgtgcctgg tgggggacaa agtaatgaaa   7920
ccagcacacg taaaggggac catcgataac gcggacctgg ccaaactggc ctttaagcgg   7980
tcatctaagt atgaccttga atgcgcgcag atacccgtgc acatgaagtc cgacgcttcg   8040
aagttcaccc atgagaaacc ggaggggtac tacaactggc accacggagc agtacagtac   8100
tcaggaggcc ggttcaccat ccctacaggt gctggcaaac caggggacag cggcagaccg   8160
atcttcgaca acaagggacg cgtggtggcc atagtcttag gaggagctaa tgaaggagcc   8220
cgtacagccc tctcggtggt gacctggaat aaagacattg tcactaaaat caccccccgag  8280
ggggccgaag agtggagtct tgccatccca gttatgtgcc tgttggcaaa caccacgttc   8340
ccctgctccc agcccccttg cacgccctgc tgctacgaaa aggaaccgga ggaaacccta   8400
cgcatgcttg aggacaacgt catgagacct gggtactatc agctgctaca agcatcctta   8460
acatgttctc cccaccgcca gcgacgcagc accaaggaca acttcaatgt ctataaagcc   8520
```

FIGURE 7 (con't)

acaagaccat acttagctca ctgtcccgac tgtggagaag ggcactcgtg ccatagtccc 8580
gtagcactag aacgcatcag aaatgaagcg acagacggga cgctgaaaat ccaggtctcc 8640
ttgcaaatcg gaataaagac ggatgacagc cacgattgga ccaagctgcg ttatatggac 8700
aaccacatgc cagcagacgc agagagggcg gggctatttg taagaacatc agcaccgtgt 8760
acgattactg gaacaatggg acacttcatc ctggcccgat gtccaaaagg ggaaactctg 8820
acggtgggat tcactgacag taggaagatt agtcactcat gtacgcaccc atttcaccac 8880
gaccctcctg tgataggtcg ggaaaaattc cattcccgac cgcagcacgg taaagagcta 8940
ccttgcagca cgtacgtgca gagcaccgcc gcaactaccg aggagataga ggtacacatg 9000
cccccagaca cccctgatcg cacattaatg tcacaacagt ccggcaacgt aaagatcaca 9060
gtcaatggcc agacggtgcg gtacaagtgt aattgcggtg gctcaaatga aggactaaca 9120
actacagaca aagtgattaa taactgcaag gttgatcaat gtcatgccgc ggtcaccaat 9180
cacaaaaagt ggcagtataa ctccctctg gtcccgcgta atgctgaact tggggaccga 9240
aaaggaaaaa ttcacatccc gtttccgctg gcaaatgtaa catgcagggt gcctaaagca 9300
aggaacccca ccgtgacgta cgggaaaaac caagtcatca tgctactgta tcctgaccac 9360
ccaacactcc tgtcctaccg gaatatggga gaagaaccaa actatcaaga agagtgggtg 9420
atgcataaga aggaagtcgt gctaaccgtg ccgactgaag ggctcgaggt cacgtggggc 9480
aacaacgagc cgtataagta ttggccgcag ttatctacaa acggtacagc ccatggccac 9540
ccgcatgaga taattctgta ttattatgag ctgtacccca ctatgactgt agtagttgtg 9600
tcagtggcca cgttcatact cctgtcgatg gtgggtatgg cagcggggat gtgcatgtgt 9660
gcacgacgca gatgcatcac accgtatgaa ctgacaccag gagctaccgt ccctttcctg 9720
cttagcctaa tatgctgcat cagaacagct aaagcggcca cataccaaga ggctgcgata 9780
tacctgtgga acgagcagca acctttgttt tggctacaag cccttattcc gctggcagcc 9840
ctgattgttc tatgcaactg tctgagactc ttaccatgct gctgtaaaac gttggctttt 9900
ttagccgtaa tgagcgtcgg tgcccacact gtgagcgcgt acgaacacgt aacagtgatc 9960
ccgaacacgg tgggagtacc gtataagact ctagtcaata gacctggcta cagccccatg 10020
gtattggaga tggaactact gtcagtcact ttggagccaa cactatcgct tgattacatc 10080
acgtgcgagt acaaaaccgt catcccgtct ccgtacgtga agtgctgcgg tacagcagag 10140
tgcaaggaca aaaacctacc tgactacagc tgtaaggtct tcaccggcgt ctacccattt 10200
atgtggggcg gcgcctactg cttctgcgac gctgaaaaca cgcagttgag cgaagcacac 10260
gtggagaagt ccgaatcatg caaaacagaa tttgcatcag catacagggc tcataccgca 10320
tctgcatcag ctaagctccg cgtcctttac caaggaaata acatcactgt aactgcctat 10380
gcaaacggcg accatgccgt cacagttaag gacgccaaat tcattgtggg gccaatgtct 10440
tcagcctgga cacctttcga caacaaaatt gtggtgtaca aaggtgacgt ctataacatg 10500
gactacccgc cctttggcgc aggaagacca ggacaatttg gcgatatcca aagtcgcaca 10560
cctgagagta aagacgtcta tgctaataca caactggtac tgcagagacc ggctgtgggt 10620
acggtacacg tgccatactc tcaggcacca tctggcttta agtattggct aaaagaacgc 10680
ggggcgtcgc tgcagcacac agcaccattt ggctgccaaa tagcaacaaa cccggtaaga 10740
gcggtgaact gcgccgtagg gaacatgccc atctccatcg acataccgga agcggccttc 10800
actagggtcg tcgacgcgcc ctctttaacg gacatgtcgt gcgaggtacc agcctgcacc 10860
cattcctcag actttggggg cgtcgccatt attaaatatg cagccagcaa gaaaggcaag 10920
tgtgcggtgc attcgatgac taacgccgtc actattcggg aagctgagat agaagttgaa 10980
gggaattctc agctgcaaat ctctttctcg acggccttag ccagcgccga attccgcgta 11040
caagtctgtt ctacacaagt acactgtgca gccgagtgcc acccccgaa ggaccacata 11100
gtcaactacc cggcgtcaca taccaccctc ggggtccagg acatctccgc tacggcgatg 11160
tcatgggtgc agaagatcac gggaggtgtg ggactggttg ttgctgttgc cgcactgatt 11220
ctaatcgtgg tgctatgcgt gtcgttcagc aggcactaac ttgacaatta agtatgaagg 11280
tatatgtgtc ccctaagaga cacactgtac atagcaaata atctatagat caaagggcta 11340
cgcaacccct gaatagtaac aaaatacaaa atcactaaaa attataaaaa cagaaaaata 11400

FIGURE 7 (con't)

```
cataaatagg tatacgtgtc ccctaagaga cacattgtat gtaggtgata agtatagatc   11460
aaagggccga ataacccctg aatagtaaca aaatatgaaa atcaataaaa atcataaaat   11520
agaaaaacca taaacagaag tagttcaaag ggctataaaa cccctgaata gtaacaaaac   11580
ataaaattaa taaaaatc                                                11598
```

FIGURE 8

```
caaagcaaga gattaataac ccatcatgga tcctgtgtac gtggacatag acgctgacag      60
cgccttttg aaggccctgc aacgtgcgta ccccatgttt gaggtggaac caaggcaggt     120
cacaccgaat gaccatgcta atgctagagc gttctcgcat ctagctataa aactaataga     180
gcaggaaatt gaccccgact caaccatcct ggatatcggc agtgcgccag caaggaggat     240
gatgtcggac aggaagtacc actgcgtctg cccgatgcgc agtgcggaag atcccgagag     300
actcgccaat tatgcgagaa agctagcatc tgccgcagga aaagtcctgg acagaaacat     360
ctctggaaag atcggggact tacaagcagt aatggccgtg ccagacacgg agacgccaac     420
attctgctta cacacagacg tctcatgtag acagagagca gacgtcgcta tataccaaga     480
cgtctatgct gtacacgcac ccacgtcgct ataccaccag gcgattaaag gggtccgagt     540
ggcgtactgg gttgggttcg acacaacccc gttcatgtac aatgccatgg cgggtgccta     600
cccctcatac tcgacaaact gggcagatga gcaggtactg aaggctaaga acataggatt     660
atgttcaaca gacctgacgg aaggtagacg aggcaagttg tctattatga gagggaaaaa     720
gctaaaaccg tgcgaccgtg tgctgttctc agtagggtca acgctctacc cggaaagccg     780
caagctactt aagagctggc acctgccatc ggtgttccat ttaaagggca aactcagctt     840
cacatgccgc tgtgatacag tggtttcgtg tgagggctac gtcgttaaga gaataacgat     900
gagcccaggc ctttatggaa aaaccacagg gtatgcggta acccaccacg cagacggatt     960
cctgatgtgc aagactaccg acacggttga cggcgaaaga gtgtcattct cggtgtgcac    1020
atacgtgccg gcgaccattt gtgatcaaat gaccggcatc cttgctacag aagtcacgcc    1080
ggaggatgca cagaagctgt tggtggggct gaaccagaga atagtggtta acggcagaac    1140
gcaacggaat acgaacacca tgaaaaatta tctgcttccc gtggtcgccc aagccttcag    1200
taagtgggca aaggagtgcc ggaaagacat ggaagatgaa aaactcctgg gggtcagaga    1260
aagaacactg acctgctgct gtctatgggc attcaagaag cagaaaacac acacggtcta    1320
caagagacct gatacccagt caattcagaa ggttcaggcc gagtttgaca gctttgtggt    1380
accgagtctg tggtcgtccg ggttgtcaat ccctttgagg actagaatca aatggttgtt    1440
aagcaaggtg ccaaaaaccg acctgatccc atacagcgga gacgcccgag aagcccggga    1500
cgcagaaaaa gaagcagagg aagaacgaga agcagaactg actcgcgaag ccctaccacc    1560
tctacaggca gcacaggaag atgttcaggt cgaaatcgac gtggaacagc ttgaggacag    1620
agcgggcgca ggaataatag agactccgag aggagctatc aaagttactg cccaaccaac    1680
agaccacgtc gtgggagagt acctggtact ctccccgcag accgtactac gtagccagaa    1740
gctcagtctg attcacgctt tggcggagca agtgaagacg tgcacgcaca acggacgagc    1800
agggaggtat gcggtcgaag cgtacgacgg ccgagtccta gtgccctcag gctatgcaat    1860
ctcgcctgaa gacttccaga gtctaagcga aagcgcaacg atggtgtata acgaaagaga    1920
gttcgtaaac agaaagctac accatattgc gatgcacgga ccagccctga acaccgacga    1980
agagtcgtat gagctggtga gggcagagag gacagaacac gagtacgtct acgacgtgga    2040
tcagagaaga tgctgtaaga aggaagaagc cgcaggactg gtactggtgg gcgacttgac    2100
taatccgccc taccacgaat tcgcatatga agggctaaaa atccgccctg cctgcccata    2160
caaaattgca gtcataggag tcttcggagt accgggatct ggcaagtcag ctattatcaa    2220
gaacctagtt accaggcagg acctggtgac tagcggaaag aaagaaaact gccaagaaat    2280
caccaccgac gtgatgagac agagaggtct agagatatct gcacgtacgg ttgactcgct    2340
gctcttgaat ggatgcaaca gaccagtcga cgtgttgtac gtagacgagg cgtttgcgtg    2400
ccactctgga acgctacttg ctttgatcgc cttggtgaga ccaaggcaga aagttgtact    2460
ttgtggtgac ccgaagcagt gcggcttctt caatatgatg cagatgaaag tcaactataa    2520
tcacaacatc tgcacccaag tgtaccacaa aagtatctcc aggcggtgta cactgcctgt    2580
gaccgccatt gtgtcatcgt tgcattacga aggcaaaatg cgcactacga atgagtacaa    2640
caagccgatt gtagtggaca ctacaggctc aacaaaacct gaccctggag acctcgtgtt    2700
aacgtgcttc agagggtggg ttaaacaact gcaaattgac tatcgtggat acgaggtcat    2760
gacagcagcc gcatcccaag ggttaaccag aaaaggagtt tacgcagtta gacaaaaagt    2820
```

FIGURE 8 (con't)

taatgaaaac ccgctctatg catcaacgtc agagcacgtc aacgtactcc taacgcgtac 2880
ggaaggtaaa ctggtatgga agacactttc cggcgacccg tggataaaga cgctgcagaa 2940
cccaccgaaa ggaaacttca aagcaactat taaggagtgg gaggtggagc atgcatcaat 3000
aatggcgggc atctgcagtc accaaatgac cttcgataca ttccaaaata aagccaacgt 3060
ttgttgggct aagagcttgg tccctatcct cgaaacagcg gggataaaac taaatgatag 3120
gcagtggtct cagataattc aagccttcaa agaagacaaa gcatactcac ctgaagtagc 3180
cctgaatgaa atatgtacgc gcatgtatgg ggtggatcta gacagcgggc tattttctaa 3240
accgttggtg tctgtgtatt acgcggataa ccactgggat aataggcctg gagggaaaat 3300
gttcggattt aaccccgagg cagcatccat tctagaaaga aagtatccat tcacaaaagg 3360
gaagtggaac atcaacaagc agatctgcgt gactaccagg aggatagaag actttaaccc 3420
taccaccaac atcataccgg ccaacaggag actaccacac tcattagtgg ccgaacaccg 3480
cccagtaaaa ggggaaagaa tggaatggct ggttaacaag ataaacggcc accacgtgct 3540
cctggtcagt ggctataacc ttgcactgcc tactaagaga gtcacttggg tagcgccgtt 3600
aggtgtccgc ggagcggact acacatacaa cctagagttg ggtctgccag caacgcttgg 3660
taggtatgac ctagtggtca taaacatcca cacacctttt cgcatacacc attaccaaca 3720
gtgcgtcgac cacgcaatga aactgcaaat gctcgggggt gactcattga gactgctcaa 3780
accgggcggc tctctattga tcagagcata tggttacgca gatagaacca gtgaacgagt 3840
catctgcgta ttgggacgca agtttagatc gtctagagcg ttgaaaccac catgtgtcac 3900
cagcaacact gagatgtttt tcctattcag caactttgac aatggcagaa ggaatttcac 3960
aactcatgtc atgaacaatc aactgaatgc agccttcgta ggacaggtca cccgagcagg 4020
atgtgcaccg tcgtaccggg taaaacgcat ggacatcgcg aagaacgatg aagagtgcgt 4080
agtcaacgcc gctaaccctc gcgggttacc gggtgacggt gtttgcaagg cagtatacaa 4140
aaaatggccg gagtccttta agaacagtgc aacaccagtg ggaaccgcaa aaacagttat 4200
gtgcggtacg tatccagtaa tccacgctgt tggaccaaac ttctctaatt attcggagtc 4260
tgaaggggac cgggaattgg cagctgccta tcgagaagtc gcaaaggaag taactaggct 4320
gggagtaaat agtgtagcta tacctctcct ctccacaggt gtatactcag gagggaaaga 4380
caggctgacc cagtcactga accacctctt tacagccatg gactcgacgg atgcagacgt 4440
ggtcatctac tgccgcgaca aagaatggga gaagaaaata tctgaggcca tacagatgcg 4500
gacccaagta gagctgctgg atgagcacat ctccatagac tgcgatattg ttcgcgtgca 4560
ccctgacagc agcttggcag gcagaaaagg atacagcacc acggaaggcg cactgtactc 4620
atatctagaa gggacccgtt ttcatcagac ggctgtggat atggcggaga tacatactat 4680
gtggccaaag caaacagagg ccaatgagca agtctgccta tatgccctgg gggaaagtat 4740
tgaatcgatc aggcagaaat gcccggtgga tgatgcagac gcatcatctc cccccaaaac 4800
tgtcccgtgc ctttgccgtt acgctatgac tccagaacgc gtcacccggc ttcgcatgaa 4860
ccacgtcaca agcataattg tgtgttcttc gtttcccctc ccaaagtaca aaatagaagg 4920
agtgcaaaaa gtcaaatgct ctaaggtaat gctatttgac cacaacgtgc catcgcgcgt 4980
aagtccaagg gaatatagat cttcccagga gtctgcacag gaggcgagta caatcacgtc 5040
actgacgcat agtcaattcg acctaagcgt tgatggcgag atactgcccg tcccgtcaga 5100
cctggatgct gacgccccag ccctagaacc agcactagac gacggggcga cacacacgct 5160
gccatccaca accggaaacc ttgcggccgt gtctgattgg gtaatgagca ccgtacctgt 5220
cgcgccgccc agaagaaggc gagggagaaa cctgactgtg acatgtgacg agagagaagg 5280
gaatataaca cccatggcta gcgtccgatt ctttagggca gagctgtgtc cggtcgtaca 5340
agaaacagcg gagacgcgtg acacagcaat gtctcttcag gcaccaccga gtaccgccac 5400
ggaaccgaat catccgccga tctccttcgg agcatcaagc gagacgttcc ccattacatt 5460
tggggacttc aacgaaggag aaatcgaaag cttgtcttct gagctactaa ctttcggaga 5520
cttcttacca ggagaagtgg atgacttgac agacagcgac tggtccacgt gctcagacac 5580
ggacgacgag ttatgactag acagggcagg tgggtatata ttctcgtcgg acaccggtcc 5640

FIGURE 8 (con't)

```
aggtcattta caacagaagt cagtacgcca gtcagtgctg ccggtgaaca ccctggagga  5700
agtccacgag gagaagtgtt acccacctaa gctggatgaa gcaaaggagc aactattact  5760
taagaaactc caggagagtg catccatggc caacagaagc aggtatcagt cgcgcaaagt  5820
agaaaacatg aaagcagcaa tcatccagag actaaagaga ggctgtagac tatacttaat  5880
gtcagagacc ccaaaagtcc ctacttaccg gactacatat ccggcgcctg tgtactcgcc  5940
tccgatcaac gtccgattgt ccaatcccga gtccgcagtg gcagcatgca atgagttctt  6000
agctagaaac tatccaactg tctcatcata ccaaattacc gacgagtatg atgcatatct  6060
agacatggtg gacgggtcgg agagttgcct ggaccgagcg acattcaatc cgtcaaaact  6120
caggagctac ccgaaacagc acgcttacca cgcgccctcc atcagaagcg ctgtaccgtc  6180
cccattccag aacacactac agaatgtact ggcagcagcc acgaaaagaa actgcaacgt  6240
cacacagatg agggaattac ccactttgga ctcagcagta ttcaacgtgg agtgtttcaa  6300
aaaattcgca tgcaaccaag aatactggga agaatttgct gccagcccta ttaggataac  6360
aactgagaat ttagcaacct atgttactaa actaaaaggg ccaaaagcag cagcgctatt  6420
cgcaaaaacc cataatctac tgccactaca ggaagtacca atggataggt tcacagtaga  6480
tatgaaaagg gacgtaaagg tgactcctgg tacaaagcat acagaggaaa gacctaaggt  6540
gcaggttata caggcggctg aacccttggc gacagcatac ctatgtggga ttcacagaga  6600
gctggttagg aggctgaacg ccgtcctcct acccaatgta catacactat ttgacatgtc  6660
tgccgaggat ttcgatgcca tcatagccgc acactttaag ccaggagaca ctgttttgga  6720
aacggacata gcctcctttg ataagagcca agatgattca cttgcgctta ctgctttgat  6780
gctgttagag gatttagggg tggatcactc cctgctggac ttgatagagg ctgctttcgg  6840
agagatttcc agctgtcacc taccgacagg tacgcgcttc aagttcggcg ccatgatgaa  6900
atcaggtatg ttcctaactc tgttcgtcaa cacattgtta aacatcacca tcgccagccg  6960
agtgctggaa gatcgtctga caaaatccgc gtgcgcggcc ttcatcggcg acgacaacat  7020
aatacatgga gtcgtctccg atgaattgat ggcagccaga tgtgccactt ggatgaacat  7080
ggaagtgaag atcatagatg cagttgtatc cttgaaagcc ccttactttt gtggagggtt  7140
tatactgcac gatactgtga caggaacagc ttgcagagtg gcagacccgc taaaaaggct  7200
ttttaaactg ggcaaaccgc tagcggcagg tgacgaacaa gatgaagata gaagacgagc  7260
gctggctgac gaagtgatca gatggcaacg aacagggcta attgatgagc tggagaaagc  7320
ggtatactct aggtacgaag tgcagggtat atcagttgtg gtaatgtcca tggccacctt  7380
tgcaagctcc agatccaact tcgagaagct cagaggaccc gtcataactt tgtacggcgg  7440
tcctaaatag gtacgcacta cagctaccta ttttgcagaa gccgacagca agtatctaaa  7500
cactaatcag ctacaatgga gttcatccca acccaaactt tttacaatag gaggtaccag  7560
cctcgaccct ggactccgcg ccctactatc caagtcatca ggcccagacc gcgccctcag  7620
aggcaagctg ggcaacttgc ccagctgatc tcagcagtta ataaactgac aatgcgcgcg  7680
gtaccccaac agaagccacg caggaatcgg aagaataaga agcaaaagca aaaacaacag  7740
gcgccacaaa acaacacaaa tcaaaagaag cagccaccta aaaagaaacc ggctcaaaag  7800
aaaaagaagc cgggccgcag agagaggatg tgcatgaaaa tcgaaaatga ttgtattttc  7860
gaagtcaagc acgaaggtaa ggtaacaggt tacgcgtgcc tggtgggga caaagtaatg  7920
aaaccagcac acgtaaaggg gaccatcgat aacgcggacc tggccaaact ggcctttaag  7980
cggtcatcta agtatgacct tgaatgcgcg cagatacccg tgcacatgaa gtccgacgct  8040
tcgaagttca cccatgagaa accggagggg tactacaact ggcaccacgg agcagtacag  8100
tactcaggag gccggttcac catccctaca ggtgctggca aaccagggga cagcggcaga  8160
ccgatcttcg acaacaaggg acgcgtggtg gccatagtct taggaggagc taatgaagga  8220
gcccgtacag ccctctcggt ggtgacctgg aataaagaca ttgtcactaa aatcaccccc  8280
gaggggccg aagagtggag tcttgccatc ccagttatgt gcctgttggc aaacaccacg  8340
ttcccctgct cccagccccc ttgcacgccc tgctgctacg aaaaggaacc ggaggaaacc  8400
ctacgcatgc ttgaggacaa cgtcatgaga cctgggtact atcagctgct acaagcatcc  8460
```

FIGURE 8 (con't)

```
ttaacatgtt ctccccaccg ccagcgacgc agcaccaagg acaacttcaa tgtctataaa    8520
gccacaagac catacttagc tcactgtccc gactgtggag aagggcactc gtgccatagt    8580
cccgtagcac tagaacgcat cagaaatgaa gcgacagacg ggacgctgaa aatccaggtc    8640
tccttgcaaa tcggaataaa gacggatgac agccacgatt ggaccaagct gcgttatatg    8700
gacaaccaca tgccagcaga cgcagagagg gcggggctat ttgtaagaac atcagcaccg    8760
tgtacgatta ctggaacaat gggacacttc atcctggccc gatgtccaaa aggggaaact    8820
ctgacggtgg gattcactga cagtaggaag attagtcact catgtacgca cccatttcac    8880
cacgaccctc ctgtgatagg tcgggaaaaa ttccattccc gaccgcagca cggtaaagag    8940
ctaccttgca gcacgtacgt gcagagcacc gccgcaacta ccgaggagat agaggtacac    9000
atgccccccag acacccctga tcgcacatta atgtcacaac agtccggcaa cgtaaagatc    9060
acagtcaatg gccagacggt gcggtacaag tgtaattgcg gtggctcaaa tgaaggacta    9120
acaactacag acaaagtgat taataactgc aaggttgatc aatgtcatgc cgcggtcacc    9180
aatcacaaaa agtggcagta taactcccct ctggtcccgc gtaatgctga acttggggac    9240
cgaaaaggaa aaattcacat cccgtttccg ctggcaaatg taacatgcag ggtgcctaaa    9300
gcaaggaacc ccaccgtgac gtacgggaaa aaccaagtca tcatgctact gtatcctgac    9360
cacccaacac tcctgtccta ccggaatatg ggagaagaac caaactatca agaagagtgg    9420
gtgatgcata agaaggaagt cgtgctaacc gtgccgactg aagggctcga ggtcacgtgg    9480
ggcaacaacg agccgtataa gtattggccg cagttatcta caaacggtac agcccatggc    9540
cacccgcacg agataattct gtattattat gagctgtacc ccactatgac tgtagtagtt    9600
gtgtcagtgg ccacgttcat actcctgtcg atggtgggta tggcagcggg gatgtgcatg    9660
tgtgcacgac gcagatgcat cacaccgtat gaactgacac caggagctac cgtccctttc    9720
ctgcttagcc taatatgctg catcagaaca gctaaagcgg ccacatacca agaggctgcg    9780
atatacctgt ggaacgagca gcaacctttg ttttggctac aagcccttat tccgctggca    9840
gccctgattg ttctatgcaa ctgtctgaga ctcttaccat gctgctgtaa aacgttggct    9900
tttttagccg taatgagcgt cggtgcccac actgtgagcg cgtacgaaca cgtaacagtg    9960
atcccgaaca cggtgggagt accgtataag actctagtca atagacctgg ctacagcccc   10020
atggtattgg agatggaact actgtcagtc actttggagc caacactatc gcttgattac   10080
atcacgtgcg agtacaaaac cgtcatcccg tctccgtacg tgaagtgctg cggtacagca   10140
gagtgcaagg acaaaaacct acctgactac agctgtaagg tcttcaccgg cgtctaccca   10200
tttatgtggg gcggcgccta ctgcttctgc gacgctgaaa acacgcagtt gagcgaagca   10260
cacgtggaga agtccgaatc atgcaaaaca gaatttgcat cagcatacag ggctcatacc   10320
gcatctgcat cagctaagct ccgcgtcctt taccaaggaa ataacatcac tgtaactgcc   10380
tatgcaaacg gcgaccatgc cgtcacagtt aaggacgcca aattcattgt ggggccaatg   10440
tcttcagcct ggacaccttt cgacaacaaa attgtggtgt acaaaggtga cgtctataac   10500
atggactacc cgcccctttgg cgcaggaaga ccaggacaat ttggcgatat ccaaagtcgc   10560
acacctgaga gtaaagacgt ctatgctaat acacaactgg tactgcagag accggctgtg   10620
ggtacggtac acgtgccata ctctcaggca ccatctggct ttaagtattg gctaaaagaa   10680
cgcggggcgt cgctgcagca cacagcacca tttggctgcc aaatagcaac aaacccggta   10740
agagcggtga actgcgccgt agggaacatg cccatctcca tcgacatacc ggaagcggcc   10800
ttcactaggg tcgtcgacgc gccctcttta acggacatgt cgtgcgaggt accagcctgc   10860
acccattcct cagactttgg gggcgtcgcc attattaaat atgcagccag caagaaaggc   10920
aagtgtgcgg tgcattcgat gactaacgcc gtcactattc gggaagctga gatagaagtt   10980
gaagggaatt ctcagctgca aatctctttc tcgacggcct tagccagcgc cgaattccgc   11040
gtacaagtct gttctacaca agtacactgt gcagccgagt gccaccccgc gaaggaccac   11100
atagtcaact acccggcgtc acataccacc ctcggggtcc aggacatctc cgctacggcg   11160
atgtcatggg tgcagaagat cacgggaggt gtgggactgg ttgttgctgt tgccgcactg   11220
attctaatcg tggtgctatg cgtgtcgttc agcaggcact aacttgacaa ttaagtatga   11280
```

FIGURE 8 (con't)

```
aggtatatgt gtcccctaag agacacactg tacatagcaa ataatctata gatcaaaggg   11340
ctacgcaacc cctgaatagt aacaaaatat aaaatcacta aaaattataa aaacagaaaa   11400
atacataaat aggtatacgt gtcccctaag agacacattg tatgtaggtg ataagtatag   11460
atcaaagggc cgaataaccc ctgaatagta acaaaatatg aaaatcaata aaaatcataa   11520
aatagaaaaa ccataaacag aagtagttca aagggctata aaacccctga atagtaacaa   11580
aacataaaat taataaaaat c                                             11601
```

FIGURE 9

```
cagtttctta ctgctctact ctgcaaagca agagattaat aacccatcat ggatcctgtg      60
tacgtggaca tagacgctga cagcgccttt ttgaaggccc tgcaacgtgc gtaccccatg     120
tttgaggtgg aaccaaggca ggtcacaccg aatgaccatg ctaatgctag agcgttctcg     180
catctagcta taaaactaat agagcaggaa attgaccccg actcaaccat cctggatatc     240
ggcagtgcgc cagcaaggag gatgatgtcg gacaggaagt accactgcgt ctgcccgatg     300
cgcagtgcgg aagatcccga gagactcgcc aattatgcga gaaagctagc atctgccgca     360
ggaaaagtcc tggacagaaa catctctgga aagatcgggg acttacaagc agtaatggcc     420
gtgccagaca cggagacgcc aacattctgc ttacacacag acgtctcatg tagacagaga     480
gcagacgtcg ctatatacca agacgtctat gctgtacacg cacccacgtc gctataccac     540
caggcgatta aaggggtccg agtggcgtac tgggttgggt tcgacacaac cccgttcatg     600
tacaatgcca tggcgggtgc ctacccctca tactcgacaa actgggcaga tgagcaggta     660
ctgaaggcta agaacatagg attatgttca acagacctga cggaaggtag acgaggcaag     720
ttgtctatta tgagagggaa aaagctaaaa ccgtgcgacc gtgtgctgtt ctcagtaggg     780
tcaacgctct acccggaaag ccgcaagcta cttaagagct ggcacctgcc atcggtgttc     840
catttaaagg gcaaactcag cttcacatgc cgctgtgata cagtggtttc gtgtgagggc     900
tacgtcgtta agagaataac gatgagccca ggcctttatg gaaaaaccac agggtatgcg     960
gtaacccacc acgcagacgg attcctgatg tgcaagacta ccgacacggt tgacggcgaa    1020
agagtgtcat tctcggtgtg cacatacgtg ccggcgacca tttgtgatca aatgaccggc    1080
atccttgcta cagaagtcac gccggaggat gcacagaagc tgttggtggg gctgaaccag    1140
agaatagtgg ttaacggcag aacgcaacgg aatacgaaca ccatgaaaaa ttatctgctt    1200
cccgtggtcg cccaagcctt cagtaagtgg gcaaaggagt gccggaaaga catggaagat    1260
gaaaaactcc tgggggtcag agaaagaaca ctgacctgct gctgtctatg ggcattcaag    1320
aagcagaaaa cacacacggt ctacaagagg cctgatacgc agtcaattca gaaggttcag    1380
gccgagtttg acagctttgt ggtaccgagt ctgtggtcgt ccgggttgtc aatccctttg    1440
aggactagaa tcaaatggtt gttaagcaag gtgccaaaaa ccgacctgat cccatacagc    1500
ggagacgccc gagaagcccg ggacgcagaa aaagaagcag aggaagaacg agaagcagaa    1560
ctgactcgcg aagccctacc acctctacag gcagcacagg aagatgttca ggtcgaaatc    1620
gacgtggaac agcttgagga cagagcgggc gcaggaataa tagagactcc gagaggagct    1680
atcaaagtta ctgcccaacc aacagaccac gtcgtgggag agtacctggt actctccccg    1740
cagaccgtac tacgtagcca gaagctcagt ctgattcacg ctttggcgga gcaagtgaag    1800
acgtgcacgc acaacggacg agcagggagg tatgcggtcg aagcgtacga cggccgagtc    1860
ctagtgccct caggctatgc aatctcgcct gaagacttcc agagtctaag cgaaagcgca    1920
acgatggtgt ataacgaaag agagttcgta aacagaaagc tacaccatat tgcgatgcac    1980
ggaccagccc tgaacaccga cgaagagtcg tatgagctgg tgagggcaga gaggacagaa    2040
cacgagtacg tctacgacgt ggatcagaga agatgctgta agaaggaaga agccgcagga    2100
ctggtactgg tgggcgactt gactaatccg ccctaccacg aattcgcata tgaagggcta    2160
aaaatccgcc ctgcctgccc atacaaaatt gcagtcatag gagtcttcgg agtaccggga    2220
tctggcaagt cagctattat caagaaccta gttaccaggc aggacctggt gactagcgga    2280
aagaaagaaa actgccaaga aatcaccacc gacgtgatga gacagagagg tctagagata    2340
tctgcacgta cggttgactc gctgctcttg aatggatgca acagaccagt cgacgtgttg    2400
tacgtagacg aggcgtttgc gtgccactct ggaacgctac ttgctttgat cgccttggtg    2460
agaccaaggc agaaagttgt actttgtggt gacccgaagc agtgcggctt cttcaatatg    2520
atgcagatga aagtcaacta taatcacaac atctgcaccc aagtgtacca caaaagtatc    2580
tccaggcggt gtacactgcc tgtgaccgcc attgtgtcat cgttgcatta cgaaggcaaa    2640
atgcgcacta cgaatgagta caacaagccg attgtagtgg acactacagg ctcaacaaaa    2700
cctgaccctg gagacctcgt gttaacgtgc ttcagagggt gggttaaaca actgcaaatt    2760
gactatcgtg gatacgaggt catgacagca gccgcatccc aagggttaac cagaaaagga    2820
```

FIGURE 9 (con't)

```
gtttacgcag ttagacaaaa agttaatgaa aacccgctct atgcatcaac gtcagagcac    2880
gtcaacgtac tcctaacgcg tacggaaggt aaactggtat ggaagacact ttccggcgac    2940
ccgtggataa agacgctgca gaacccaccg aaaggaaact tcaaagcaac tattaaggag    3000
tgggaggtgg agcatgcatc aataatggcg ggcatctgca gtcaccaaat gaccttcgat    3060
acattccaaa ataaagccaa cgtttgttgg gctaagagct tggtccctat cctcgaaaca    3120
gcggggataa aactaaatga taggcagtgg tctcagataa ttcaagcctt caaagaagac    3180
aaagcatact cacctgaagt agccctgaat gaaatatgta cgcgcatgta tggggtggat    3240
ctagacagcg ggctattttc taaaccgttg gtgtctgtgt attacgcgga taaccactgg    3300
gataataggc ctggagggaa aatgttcgga tttaaccccg aggcagcatc cattctagaa    3360
agaaagtatc cattcacaaa agggaagtgg aacatcaaca agcagatctg cgtgactacc    3420
aggaggatag aagactttaa ccctaccacc aacatcatac cggccaacag gagactacca    3480
cactcattag tggccgaaca ccgcccagta aaaggggaaa gaatggaatg gctggttaac    3540
aagataaacg gccaccacgt gctcctggtc agtggctata accttgcact gcctactaag    3600
agagtcactt gggtagcgcc gttaggtgtc cgcggagcgg actacacata caacctagag    3660
ttgggtctgc cagcaacgct tggtaggtat gacctagtgg tcataaacat ccacacacct    3720
tttcgcatac accattacca acagtgcgtc gaccacgcaa tgaaactgca aatgctcggg    3780
ggtgactcat tgagactgct caaaccgggc ggctctctat tgatcagagc atatggttac    3840
gcagatagaa ccagtgaacg agtcatctgc gtattgggac gcaagtttag atcgtctaga    3900
gcgttgaaac caccatgtgt caccagcaac actgagatgt ttttcctatt cagcaacttt    3960
gacaatggca gaaggaattt cacaactcat gtcatgaaca atcaactgaa tgcagccttc    4020
gtaggacagg tcacccgagc aggatgtgca ccgtcgtacc gggtaaaacg catggacatc    4080
gcgaagaacg atgaagagtg cgtagtcaac gccgctaacc ctcgcgggtt accgggtgac    4140
ggtgtttgca aggcagtata caaaaaatgg ccggagtcct ttaagaacag tgcaacacca    4200
gtgggaaccg caaaaacagt tatgtgcggt acgtatccag taatccacgc tgttggacca    4260
aacttctcta attattcgga gtctgaaggg gaccgggaat tggcagctgc ctatcgagaa    4320
gtcgcaaagg aagtaactag gctgggagta aatagtgtag ctatacctct cctctccaca    4380
ggtgtatact caggagggaa agacaggctg acccagtcac tgaaccacct ctttacagcc    4440
atggactcga cggatgcaga cgtggtcatc tactgccgcg acaaagaatg ggagaagaaa    4500
atatctgagg ccatacagat gcggacccaa gtagagctgc tggatgagca catctccata    4560
gactgcgata ttgttcgcgt gcaccctgac agcagcttgg caggcagaaa aggatacagc    4620
accacggaag gcgcactgta ctcatatcta gaagggaccc gttttcatca gacggctgtg    4680
gatatggcgg agatacatac tatgtggcca aagcaaacag aggccaatga gcaagtctgc    4740
ctatatgccc tgggggaaag tattgaatcg atcaggcaga aatgcccggt ggatgatgca    4800
gacgcatcat ctccccccaa aactgtcccg tgcctttgcc gttacgctat gactccagaa    4860
cgcgtcaccc ggcttcgcat gaaccacgtc acaagcataa ttgtgtgttc ttcgtttccc    4920
ctcccaaagt acaaaataga aggagtgcaa aaagtcaaat gctctaaggt aatgctattt    4980
gaccacaacg tgccatcgcg cgtaagtcca agggaatata gatcttccca ggagtctgca    5040
caggaggcga gtacaatcac gtcactgacg catagtcaat tcgacctaag cgttgatggc    5100
gagatactgc ccgtcccgtc agacctggat gctgacgccc cagccctaga accagcacta    5160
gacgacgggg cgacacacac gctgccatcc acaaccggaa accttgcggc cgtgtctaga    5220
cagggcaggt gggtatatat tctcgtcgga caccggtcca ggtcatttac aacagaagtc    5280
agtacgccag tcagtgctgc cggtgaacac cctggaggaa gtccacgagg agaagtgtta    5340
cccacctaag ctggatgaag caaaggagca actattactt aagaaactcc aggagagtgc    5400
atccatggcc aacagaagca ggtatcagtc gcgcaaagta gaaaacatga aagcagcaat    5460
catccagaga ctaaagagag gctgtagact atacttaatg tcagagaccc caaaagtccc    5520
tacttaccgg actacatatc cggcgcctgt gtactcgcct ccgatcaacg tccgattgtc    5580
caatcccgag tccgcagtgg cagcatgcaa tgagttctta gctagaaact atccaactgt    5640
```

FIGURE 9 (con't)

```
ctcatcatac caaattaccg acgagtatga tgcatatcta gacatggtgg acgggtcgga  5700
gagttgcctg gaccgagcga cattcaatcc gtcaaaactc aggagctacc cgaaacagca  5760
cgcttaccac gcgccctcca tcagaagcgc tgtaccgtcc ccattccaga acacactaca  5820
gaatgtactg gcagcagcca cgaaaagaaa ctgcaacgtc acacagatga gggaattacc  5880
cactttggac tcagcagtat tcaacgtgga gtgtttcaaa aaattcgcat gcaaccaaga  5940
atactgggaa gaatttgctg ccagccctat taggataaca actgagaatt tagcaaccta  6000
tgttactaaa ctaaaagggc caaaagcagc agcgctattc gcaaaaaccc ataatctact  6060
gccactacag gaagtaccaa tggataggtt cacagtagat atgaaaaggg acgtgaaggt  6120
gactcctggt acaaagcata cagaggaaag acctaaggtg caggttatac aggcggctga  6180
acccttggcg acagcatacc tatgtgggat tcacagagag ctggttagga ggctgaacgc  6240
cgtcctccta cccaatgtac atacactatt tgacatgtct gccgaggatt tcgatgccat  6300
catagccgca cactttaagc caggagacac tgttttggaa acggacatag cctcctttga  6360
taagagccaa gatgattcac ttgcgcttac tgctttgatg ctgttagagg atttaggggt  6420
ggatcactcc ctgctggact tgatagaggc tgctttcgga gagatttcca gctgtcacct  6480
accgacaggt acgcgcttca agttcggcgc catgatgaaa tcaggtatgt tcctaactct  6540
gttcgtcaac acattgttaa acatcaccat cgccagccga gtgctggaag atcgtctgac  6600
aaaatccgcg tgcgcggcct tcatcggcga cgacaacata atacatggag tcgtctccga  6660
tgaattgatg gcagccagat gtgccacttg gatgaacatg gaagtgaaga tcatagatgc  6720
agttgtatcc ttgaaagccc cttacttttg tggagggttt atactgcacg atactgtgac  6780
aggaacagct tgcagagtgg cagacccgct aaaaaggctt tttaaactgg gcaaaccgct  6840
agcggcaggt gacgaacaag atgaagatag aagacgagcg ctggctgacg aagtgatcag  6900
atggcaacga acagggctaa ttgatgagct ggagaaagcg gtatactcta ggtacgaagt  6960
gcagggtata tcagttgtgg taatgtccat ggccaccttt gcaagctcca gatccaactt  7020
cgagaagctc agaggacccg tcataacttt gtacggcggt cctaaatagg tacgcactac  7080
agctacctat tttgcagaag ccgacagcaa gtatctaaac actaatcagc tacaatggag  7140
ttcatcccaa cccaaacttt ttacaatagg aggtaccagc ctcgaccctg gactccgcgc  7200
cctactatcc aagtcatcag gcccagaccg cgccctcaga ggcaagctgg gcaacttgcc  7260
cagctgatct cagcagttaa taaactgaca atgcgcgcgg taccccaaca gaagccacgc  7320
aggaatcgga agaataagaa gcaaaagcaa aaacaacagg cgccacaaaa caacacaaat  7380
caaaagaagc agccacctaa aaagaaaccg gctcaaaaga aaaagaagcc gggccgcaga  7440
gagaggatgt gcatgaaaat cgaaaatgat tgtattttcg aagtcaagca cgaaggtaag  7500
gtaacaggtt acgcgtgcct ggtgggggac aaagtaatga aaccagcaca cgtaaagggg  7560
accatcgata acgcggacct ggccaaactg gcctttaagc ggtcatctaa gtatgacctt  7620
gaatgcgcgc agatacccgt gcacatgaag tccgacgctt cgaagttcac ccatgagaaa  7680
ccggaggggt actacaactg gcaccacgga gcagtacagt actcaggagg ccggttcacc  7740
atccctacag gtgctggcaa accaggggac agcggcagac cgatcttcga caacaaggga  7800
cgcgtggtgg ccatagtctt aggaggagct aatgaaggag cccgtacagc cctctcggtg  7860
gtgacctgga ataaagacat tgtcactaaa atcacccccg agggggccga agagtggagt  7920
cttgccatcc cagttatgtg cctgttggca aacaccacgt tcccctgctc ccagcccccc  7980
tgcacgccct gctgctacga aaaggaaccg gaggaaaccc tacgcatgct tgaggacaac  8040
gtcatgagac ctgggtacta tcagctgcta caagcatcct taacatgttc tccccaccgc  8100
cagcgacgca gcaccaagga caacttcaat gtctataaag ccacaagacc atacttagct  8160
cactgtcccg actgtggaga agggcactcg tgccatagtc ccgtagcact agaacgcatc  8220
agaaatgaag cgacagacgg gacgctgaaa atccaggtct ccttgcaaat cggaataaag  8280
acggatgaca gccacgattg gaccaagctg cgttatatgg acaaccacat gccagcagac  8340
gcagagaggg cggggctatt tgtaagaaca tcagcaccgt gtacgattac tggaacaatg  8400
ggacacttca tcctggcccg atgtccaaaa ggggaaactc tgacggtggg attcactgac  8460
```

FIGURE 9 (con't)

```
agtaggaaga ttagtcactc atgtacgcac ccatttcacc acgaccctcc tgtgataggt    8520
cgggaaaaat tccattcccg accgcagcac ggtaaagagc taccttgcag cacgtacgtg    8580
cagagcaccg ccgcaactac cgaggagata gaggtacaca tgcccccaga caccctgat    8640
cgcacattaa tgtcacaaca gtccggcaac gtaaagatca cagtcaatgg ccagacggtg    8700
cggtacaagt gtaattgcgg tggctcaaat gaaggactaa caactacaga caaagtgatt    8760
aataactgca aggttgatca atgtcatgcc gcggtcacca atcacaaaaa gtggcagtat    8820
aactcccctc tggtcccgcg taatgctgaa cttggggacc gaaaaggaaa aattcacatc    8880
ccgtttccgc tggcaaatgt aacatgcagg gtgcctaaag caaggaaccc caccgtgacg    8940
tacgggaaaa accaagtcat catgctactg tatcctgacc acccaacact cctgtcctac    9000
cggaatatgg gagaagaacc aaactatcaa gaagagtggg tgatgcataa gaaggaagtc    9060
gtgctaaccg tgccgactga agggctcgag gtcacgtggg gcaacaacga gccgtataag    9120
tattggccgc agttatctac aaacggtaca gcccatggcc acccgcatga gataattctg    9180
tattattatg agctgtaccc cactatgact gtagtagttg tgtcagtggc cacgttcata    9240
ctcctgtcga tggtgggtat ggcagcgggg atgtgcatgt gtgcacgacg cagatgcatc    9300
acaccgtatg aactgacacc aggagctacc gtccctttcc tgcttagcct aatatgctgc    9360
atcagaacag ctaaagcggc cacataccaa gaggctgcga tatacctgtg gaacgagcag    9420
caacctttgt tttggctaca agcccttatt ccgctggcag ccctgattgt tctatgcaac    9480
tgtctgagac tcttaccatg ctgctgtaaa acgttggctt ttttagccgt aatgagcgtc    9540
ggtgcccaca ctgtgagcgc gtacgaacac gtaacagtga tcccgaacac ggtgggagta    9600
ccgtataaga ctctagtcaa tagacctggc tacagcccca tggtattgga gatggaacta    9660
ctgtcagtca ctttggagcc aacactatcg cttgattaca tcacgtgcga gtacaaaacc    9720
gtcatcccgt ctccgtacgt gaagtgctgc ggtacagcag agtgcaagga caaaaaccta    9780
cctgactaca gctgtaaggt cttcaccggc gtctacccat ttatgtgggg cggcgcctac    9840
tgcttctgcg acgctgaaaa cacgcagttg agcgaagcac acgtggagaa gtccgaatca    9900
tgcaaaacag aatttgcatc agcatacagg gctcataccg catctgcatc agctaagctc    9960
cgcgtccttt accaaggaaa taacatcact gtaactgcct atgcaaacgg cgaccatgcc   10020
gtcacagtta aggacgccaa attcattgtg gggccaatgt cttcagcctg gacacctttc   10080
gacaacaaaa ttgtggtgta caaaggtgac gtctataaca tggactaccc gcccctttggc   10140
gcaggaagac caggacaatt tggcgatatc caaagtcgca cacctgagag taaagacgtc   10200
tatgctaata cacaactggt actgcagaga ccggctgcgg gtacggtaca cgtgccatac   10260
tctcaggcac catctggctt taagtattgg ctaaaagaac gcggggcgtc gctgcagcac   10320
acagcaccat ttggctgcca aatagcaaca aacccggtaa gagcggtgaa ctgcgccgta   10380
gggaacatgc ccatctccat cgacataccg gaagcggcct tcactagggt cgtcgacgcg   10440
ccctctttaa cggacatgtc gtgcgaggta ccagcctgca cccattcctc agactttggg   10500
ggcgtcgcca ttattaaata tgcagccagc aagaaaggca agtgtgcggt gcattcgatg   10560
actaacgccg tcactattcg ggaagctgag atagaagttg aagggaattc tcagctgcaa   10620
atctctttct cgacggcctt agccagcgcc gaattccgcg tacaagtctg ttctacacaa   10680
gtacactgtg cagccgagtg ccacccccgg aaggaccaca tagtcaacta cccggcgtca   10740
cataccaccc tcggggtcca ggacatctcc gctacggcga tgtcatgggt gcagaagatc   10800
acgggaggtg tgggactggt tgttgctgtt gccgcactga ttctaatcgt ggtgctatgc   10860
gtgtcgttca gcaggcacta acttgacaat taagtatgaa ggtatatgtg tcccctaaga   10920
gacacactgt acatagcaaa taatctatag atcaaagggc tacgcaaccc ctgaatagta   10980
acaaaatata aaatcactaa aaattataaa aacagaaaaa tacataaata ggtatacgtg   11040
tcccctaaga gacacattgt atgtaggtga taagtataga tcaaagggcc gaataacccc   11100
tgaatagtaa caaaatatga aaatcaataa aaatcataaa atagaaaaac cataaacaga   11160
agtagttcaa agggctataa aacccctgaa tagtaacaaa acataaaatt aataaaaatc   11220
aaatgaatac catatgg                                                  11237
```

FIGURE 10

```
atacccgtgc acatgaagtc cgacgcttcg aagttcaccc atgagaaacc ggaggggtac      60
tacaactggc accacggagc agtacagtac tcaggaggcc ggttcaccat ccctacaggt     120
gctggcaaac caggggacag cggcagaccg atcttcgaca acaagggacg cgtggtggcc     180
atagtcttag gaggagctaa tgaaggagcc cgtacagccc tctcggtggt gacctggaat     240
aaagacattg tcactaaaat caccccccgag ggggccgaag agtggagtct tgccatccca    300
gttatgtgcc tgttggcaaa caccacgttc ccctgctccc agccccccttg cacgccctgc    360
tgctacgaaa aggaaccgga ggaaacccta cgcatgcttg aggacaacgt catgagacct     420
gggtactatc agctgctaca agcatcctta acatgttctc cccaccgcca gcgacgcagc     480
accaaggaca acttcaatgt ctataaagcc acaagaccat acttagctca ctgtcccgac     540
tgtggagaag ggcactcgtg ccatagtccc gtagcactag aacgcatcag aaatgaagcg     600
acagacggga cgctgaaaat ccaggtctcc ttgcaaatcg gaataaagac ggatgacagc     660
cacgattgga ccaagctgcg ttatatggac aaccacatgc cagcagacgc agagagggcg     720
gggctatttg taagaacatc agcaccgtgt acgattactg gaacaatggg acacttcatc     780
ctggcccgat gtccaaaagg ggaaactctg acggtgggat tcactgacag taggaagatt     840
agtcactcat gtacgcaccc atttcaccac gaccctcctg tgataggtcg ggaaaaattc     900
cattcccgac cgcagcacgg taaagagcta ccttgcagca cgtacgtgca gagcaccgcc     960
gcaactaccg aggagataga ggtacacatg cccccagaca cccctgatcg cacattaatg    1020
tcacaacagt ccggcaacgt aaagatcaca gtcaatggcc agacggtgcg gtacaagtgt    1080
aattgcggtg gctcaaatga aggactaaca actacagaca aagtgattaa taactgcaag    1140
gttgatcaat gtcatgccgc ggtcaccaat cacaaaaagt ggcagtataa ctccccctctg   1200
gtcccgcgta atgctgaact tggggaccga aaaggaaaaa ttcacatccc gtttccgctg    1260
gcaaatgtaa catgcagggt gcctaaagca aggaacccca ccgtgacgta cgggaaaaac    1320
caagtcatca tgctactgta tcctgaccac ccaacactcc tgtcctaccg gaatatggga    1380
gaagaaccaa actatcaaga agagtgggtg atgcataaga aggaagtcgt gctaaccgtg    1440
ccgactgaag ggctcgaggt cacgtggggc aacaacgagc cgtataagta ttggccgcag    1500
ttatctacaa acggtacagc ccatggccac ccgcatgaga taattctgta ttattatgag    1560
ctgtacccca ctatgactgt agtagttgtg tcagtggcca cgttcatact cctgtcgatg    1620
gtgggtatgg cagcggggat gtgcatgcgt gcacgacgca gatgcatcac accgtatgaa    1680
ctgacaccag gagctaccgt ccctttcctg cttagcctaa tatgctgcat cagaacagct    1740
aaagcggcca cataccaaga ggctgcgata tacctgtgga acgagcagca accttta       1797
```

FIGURE 11

```
atacccgtgc acatgaagtc cgacgcttcg aagttcaccc atgagaaacc ggaggggtac      60
tacaactggc accacggagc agtacagtac tcaggaggcc ggttcaccat ccctacaggt     120
gctggcaaac caggggacag cggcagaccg atcttcgaca acaagggacg cgtggtggcc     180
atagtcttag gaggagctaa tgaaggagcc cgtacagccc tctcggtggt gacctggaat     240
aaagacattg tcactaaaat caccccccgag ggggccgaag agtggagtct tgccatccca     300
gttatgtgcc tgttggcaaa caccacgttc ccctgctccc agcccccttg cacgccctgc     360
tgctacgaaa aggaaccgga ggaaacccta cgcatgcttg aggacaacgt catgagacct     420
gggtactatc agctgctaca agcatcctta acatgttctc cccaccgcca gcgacgcagc     480
accaaggaca acttcaatgt ctataaagcc acaagaccat acttagctca ctgtcccgac     540
tgtggagaag ggcactcgtg ccatagtccc gtagcactag aacgcatcag aaatgaagcg     600
acagacggga cgctgaaaat ccaggtctcc ttgcaaatcg gaataaagac ggatgacagc     660
cacgattgga ccaagctgcg ttatatggac aaccacatgc cagcagacgc agagagggcg     720
gggctatttg taagaacatc agcaccgtgt acgattactg gaacaatggg acacttcatc     780
ctggcccgat gtccaaaagg ggaaactctg acggtgggat tcactgacag taggaagatt     840
agtcactcat gtacgcaccc atttcaccac gaccctcctg tgataggtcg ggaaaaattc     900
cattcccgac cgcagcacgg taaagagcta ccttgcagca cgtacgtgca gagcaccgcc     960
gcaactaccg aggagataga ggtacacatg cccccagaca cccctgatcg cacattaatg    1020
tcacaacagt ccggcaacgt aaagatcaca gtcaatggcc agacggtgcg gtacaagtgt    1080
aattgcggtg gctcaaatga aggactaaca actacagaca aagtgattaa taactgcaag    1140
gttgatcaat gtcatgccgc ggtcaccaat cacaaaaagt ggcagtataa ctcccctctg    1200
gtcccgcgta atgctgaact tggggaccga aaaggaaaaa ttcacatccc gtttccgctg    1260
gcaaatgtaa catgcagggt gcctaaagca aggaacccca ccgtgacgta cgggaaaaac    1320
caagtcatca tgctactgta tcctgaccac ccaacactcc tgtcctaccg gaatatggga    1380
gaagaaccaa actatcaaga agagtgggtg atgcataaga aggaagtcgt gctaaccgtg    1440
ccgactgaag ggctcgaggt cacgtggggc aacaacgagc cgtataagta ttggccgcag    1500
ttatctacaa acggtacagc ccatggccac ccgcatgaga taattctgta ttattatgag    1560
ctgtacccca ctatgactgt agtagttgtg tcagtggcca cgttcatact cctgtcgatg    1620
gtgggtatgg cagcggggat gtgcatgtgt gcacgacgca gatgcatcac accgtatgaa    1680
ctgacaccag gagctaccgt ccctttcctg cttagcctaa tatgctgcat cagaacagct    1740
aaagcggcca cataccaaga ggctgcgata tacctgtgga acgagcagca accttta       1797
```

FIGURE 12

```
atacccgtgc acatgaagtc cgacgcttcg aagttcaccc atgagaaacc ggaggggtac      60
tacaactggc accacggagc agtacagtac tcaggaggcc ggttcaccat ccctacaggt     120
gctggcaaac caggggacag cggcagaccg atcttcgaca acaagggacg cgtggtggcc     180
atagtcttag gaggagctaa tgaaggagcc cgtacagccc tctcggtggt gacctggaat     240
aaagacattg tcactaaaat caccccccgag ggggccgaag agtggagtct tgccatccca    300
gttatgtgcc tgttggcaaa caccacgttc ccctgctccc agcccccttg cacgccctgc     360
tgctacgaaa aggaaccgga ggaaacccta cgcatgcttg aggacaacgt catgagacct     420
gggtactatc agctgctaca agcatcctta acatgttctc cccaccgcca gcgacgcagc     480
accaaggaca acttcaatgt ctataaagcc acaagaccat acttagctca ctgtcccgac     540
tgtggagaag ggcactcgtg ccatagtccc gtagcactag aacgcatcag aaatgaagcg     600
acagacggga cgctgaaaat ccaggtctcc ttgcaaatcg gaataaagac ggatgacagc     660
cacgattgga ccaagctgcg ttatatggac aaccacatgc cagcagacgc agagagggcg     720
gggctatttg taagaacatc agcaccgtgt acgattactg gaacaatggg acacttcatc     780
ctggcccgat gtccaaaagg ggaaactctg acggtgggat tcactgacag taggaagatt     840
agtcactcat gtacgcaccc atttcaccac gaccctcctg tgataggtcg ggaaaaattc     900
cattcccgac cgcagcacgg taaagagcta ccttgcagca cgtacgtgca gagcaccgcc     960
gcaactaccg aggagataga ggtacacatg cccccagaca cccctgatcg cacattaatg    1020
tcacaacagt ccggcaacgt aaagatcaca gtcaatggcc agacggtgcg gtacaagtgt    1080
aattgcggtg gctcaaatga aggactaaca actacagaca aagtgattaa taactgcaag    1140
gttgatcaat gtcatgccgc ggtcaccaat cacaaaaagt ggcagtataa ctcccctctg    1200
gtcccgcgta atgctgaact tggggaccga aaaggaaaaa ttcacatccc gtttccgctg    1260
gcaaatgtaa catgcagggt gcctaaagca aggaacccca ccgtgacgta cgggaaaaac    1320
caagtcatca tgctactgta tcctgaccac ccaacactcc tgtcctaccg gaatatggga    1380
gaagaaccaa actatcaaga agagtgggtg atgcataaga aggaagtcgt gctaaccgtg    1440
ccgactgaag ggctcgaggt cacgtggggc aacaacgagc cgtataagta ttggccgcag    1500
ttatctacaa acggtacagc ccatggccac ccgcacgaga taattctgta ttattatgag    1560
ctgtacccca ctatgactgt agtagttgtg tcagtggcca cgttcatact cctgtcgatg    1620
gtgggtatgg cagcggggat gtgcatgtgt gcacgacgca gatgcatcac accgtatgaa    1680
ctgacaccag gagctaccgt cccttcctg cttagcctaa tatgctgcat cagaacagct     1740
aaagcggcca cataccaaga ggctgcgata tacctgtgga acgagcagca accttta       1797
```

FIGURE 13

```
atacccgtgc acatgaagtc cgacgcttcg aagttcaccc atgagaaacc ggaggggtac      60
tacaactggc accacggagc agtacagtac tcaggaggcc ggttcaccat ccctacaggt     120
gctggcaaac caggggacag cggcagaccg atcttcgaca acaagggacg cgtggtggcc     180
atagtcttag gaggagctaa tgaaggagcc cgtacagccc tctcggtggt gacctggaat     240
aaagacattg tcactaaaat caccccccgag ggggccgaag agtggagtct tgccatccca    300
gttatgtgcc tgttggcaaa caccacgttc ccctgctccc agcccccttg cacgccctgc     360
tgctacgaaa aggaaccgga ggaaacccta cgcatgcttg aggacaacgt catgagacct     420
gggtactatc agctgctaca agcatcctta acatgttctc cccaccgcca gcgacgcagc     480
accaaggaca acttcaatgt ctataaagcc acaagaccat acttagctca ctgtcccgac     540
tgtggagaag ggcactcgtg ccatagtccc gtagcactag aacgcatcag aaatgaagcg     600
acagacggga cgctgaaaat ccaggtctcc ttgcaaatcg gaataaagac ggatgacagc     660
cacgattgga ccaagctgcg ttatatggac aaccacatgc cagcagacgc agagagggcg     720
gggctatttg taagaacatc agcaccgtgt acgattactg gaacaatggg acacttcatc     780
ctggcccgat gtccaaaagg ggaaactctg acggtgggat tcactgacag taggaagatt     840
agtcactcat gtacgcaccc atttcaccac gaccctcctg tgataggtcg ggaaaaattc     900
cattcccgac cgcagcacgg taaagagata ccttgcagca cgtacgtgca gagcaccgcc     960
gcaactaccg aggagataga ggtacacatg cccccagaca cccctgatcg cacattaatg    1020
tcacaacagt ccggcaacgt aaagatcaca gtcaatggcc agacggtgcg gtacaagtgt    1080
aattgcggtg gctcaaatga aggactaaca actacagaca aagtgattaa taactgcaag    1140
gttgatcaat gtcatgccgc ggtcaccaat cacaaaaagt ggcagtataa ctcccctctg    1200
gtcccgcgta atgctgaact tggggaccga aaaggaaaaa ttcacatccc gtttccgctg    1260
gcaaatgtaa catgcagggt gcctaaagca aggaacccca ccgtgacgta cgggaaaaac    1320
caagtcatca tgctactgta tcctgaccac ccaacactcc tgtcctaccg gaatatggga    1380
gaagaaccaa actatcaaga agagtgggtg atgcataaga aggaagtcgt gctaaccgtg    1440
ccgactgaag ggctcgaggt cacgtggggc aacaacgagc cgtataagta ttggccgcag    1500
ttatctacaa acggtacagc ccatggccac ccgcatgaga taattctgta ttattatgag    1560
ctgtacccca ctatgactgt agtagttgtg tcagtggcca cgttcatact cctgtcgatg    1620
gtgggtatgg cagcggggat gtgcatgtgt gcacgacgca gatgcatcac accgtatgaa    1680
ctgacaccag gagctaccgt cccttcctg cttagcctaa tatgctgcat cagaacagct     1740
aaagcggcca cataccaaga ggctgcgata tacctgtgga acgagcagca accttta       1797
```

FIGURE 14

```
gacaacttca atgtctataa agccacaaga ccatacttag ctcactgtcc cgactgtgga      60
gaagggcact cgtgccatag tcccgtagca ctagaacgca tcagaaatga agcgacagac     120
gggacgctga aaatccaggt ctccttgcaa atcggaataa agacggatga cagccacgat     180
tggaccaagc tgcgttatat ggacaaccac atgccagcag acgcagagag ggcggggcta     240
tttgtaagaa catcagcacc gtgtacgatt actggaacaa tgggacactt catcctggcc     300
cgatgtccaa aaggggaaac tctgacggtg ggattcactg acagtaggaa gattagtcac     360
tcatgtacgc acccatttca ccacgaccct cctgtgatag gtcgggaaaa attccattcc     420
cgaccgcagc acggtaaaga gctaccttgc agcacgtacg tgcagagcac cgccgcaact     480
accgaggaga tagaggtaca catgccccca gacacccctg atcgcacatt aatgtcacaa     540
cagtccggca acgtaaagat cacagtcaat ggccagacgg tgcggtacaa gtgtaattgc     600
ggtggctcaa atgaaggact aacaactaca gacaaagtga ttaataactg caaggttgat     660
caatgtcatg ccgcggtcac caatcacaaa aagtggcagt ataactcccc tctggtcccg     720
cgtaatgctg aacttgggga ccgaaaagga aaaattcaca tcccgtttcc gctggcaaat     780
gtaacatgca gggtgcctaa agcaaggaac cccaccgtga cgtacgggaa aaaccaagtc     840
atcatgctac tgtatcctga ccacccaaca ctcctgtcct accggaatat gggagaagaa     900
ccaaactatc aagaagagtg ggtgatgcat aagaaggaag tcgtgctaac cgtgccgact     960
gaagggctcg aggtcacgtg gggcaacaac gagccgtata agtattggcc gcagttatct    1020
acaaacggta cagcccatgg ccacccgcat gagataattc tgtattatta tgagctgtac    1080
cccactatga ct                                                       1092
```

FIGURE 15

```
gacaacttca atgtctataa agccacaaga ccatacttag ctcactgtcc cgactgtgga      60
gaagggcact cgtgccatag tcccgtagca ctagaacgca tcagaaatga agcgacagac     120
gggacgctga aaatccaggt ctccttgcaa atcggaataa agacggatga cagccacgat     180
tggaccaagc tgcgttatat ggacaaccac atgccagcag acgcagagag ggcggggcta     240
tttgtaagaa catcagcacc gtgtacgatt actggaacaa tgggacactt catcctggcc     300
cgatgtccaa aaggggaaac tctgacggtg ggattcactg acagtaggaa gattagtcac     360
tcatgtacgc acccatttca ccacgaccct cctgtgatag gtcgggaaaa attccattcc     420
cgaccgcagc acggtaaaga gctaccttgc agcacgtacg tgcagagcac cgccgcaact     480
accgaggaga tagaggtaca catgccccca gacacccctg atcgcacatt aatgtcacaa     540
cagtccggca acgtaaagat cacagtcaat ggccagacgg tgcggtacaa gtgtaattgc     600
ggtggctcaa atgaaggact aacaactaca gacaaagtga ttaataactg caaggttgat     660
caatgtcatg ccgcggtcac caatcacaaa aagtggcagt ataactcccc tctggtcccg     720
cgtaatgctg aacttgggga ccgaaaagga aaaattcaca tcccgtttcc gctggcaaat     780
gtaacatgca gggtgcctaa agcaaggaac cccaccgtga cgtacgggaa aaaccaagtc     840
atcatgctac tgtatcctga ccacccaaca ctcctgtcct accggaatat gggagaagaa     900
ccaaactatc aagaagagtg ggtgatgcat aagaaggaag tcgtgctaac cgtgccgact     960
gaagggctcg aggtcacgtg gggcaacaac gagccgtata agtattggcc gcagttatct    1020
acaaacggta cagcccatgg ccacccgcat gagataattc tgtattatta tgagctgtac    1080
cccactatga ct                                                       1092
```

FIGURE 16

```
gacaacttca atgtctataa agccacaaga ccatacttag ctcactgtcc cgactgtgga      60
gaagggcact cgtgccatag tcccgtagca ctagaacgca tcagaaatga agcgacagac     120
gggacgctga aaatccaggt ctccttgcaa atcggaataa agacggatga cagccacgat     180
tggaccaagc tgcgttatat ggacaaccac atgccagcag acgcagagag ggcggggcta     240
tttgtaagaa catcagcacc gtgtacgatt actggaacaa tgggacactt catcctggcc     300
cgatgtccaa aaggggaaac tctgacggtg ggattcactg acagtaggaa gattagtcac     360
tcatgtacgc acccatttca ccacgaccct cctgtgatag gtcgggaaaa attccattcc     420
cgaccgcagc acggtaaaga gctaccttgc agcacgtacg tgcagagcac cgccgcaact     480
accgaggaga tagaggtaca catgccccca gacacccctg atcgcacatt aatgtcacaa     540
cagtccggca acgtaaagat cacagtcaat ggccagacgg tgcggtacaa gtgtaattgc     600
ggtggctcaa atgaaggact aacaactaca gacaaagtga ttaataactg caaggttgat     660
caatgtcatg ccgcggtcac caatcacaaa aagtggcagt ataactcccc tctggtcccg     720
cgtaatgctg aacttgggga ccgaaaagga aaaattcaca tcccgtttcc gctggcaaat     780
gtaacatgca gggtgcctaa agcaaggaac cccaccgtga cgtacgggaa aaaccaagtc     840
atcatgctac tgtatcctga ccacccaaca ctcctgtcct accggaatat gggagaagaa     900
ccaaactatc aagaagagtg ggtgatgcat aagaaggaag tcgtgctaac cgtgccgact     960
gaagggctcg aggtcacgtg gggcaacaac gagccgtata agtattggcc gcagttatct    1020
acaaacggta cagcccatgg ccacccgcac gagataattc tgtattatta tgagctgtac    1080
cccactatga ct                                                       1092
```

FIGURE 17

```
gacaacttca atgtctataa agccacaaga ccatacttag ctcactgtcc cgactgtgga      60
gaagggcact cgtgccatag tcccgtagca ctagaacgca tcagaaatga agcgacagac     120
gggacgctga aaatccaggt ctccttgcaa atcggaataa agacggatga cagccacgat     180
tggaccaagc tgcgttatat ggacaaccac atgccagcag acgcagagag ggcggggcta     240
tttgtaagaa catcagcacc gtgtacgatt actggaacaa tgggacactt catcctggcc     300
cgatgtccaa aaggggaaac tctgacggtg ggattcactg acagtaggaa gattagtcac     360
tcatgtacgc acccatttca ccacgaccct cctgtgatag gtcgggaaaa attccattcc     420
cgaccgcagc acggtaaaga gataccttgc agcacgtacg tgcagagcac cgccgcaact     480
accgaggaga tagaggtaca catgccccca gacacccctg atcgcacatt aatgtcacaa     540
cagtccggca acgtaaagat cacagtcaat ggccagacgg tgcggtacaa gtgtaattgc     600
ggtggctcaa atgaaggact aacaactaca gacaaagtga ttaataactg caaggttgat     660
caatgtcatg ccgcggtcac caatcacaaa aagtggcagt ataactcccc tctggtcccg     720
cgtaatgctg aacttgggga ccgaaaagga aaaattcaca tcccgtttcc gctggcaaat     780
gtaacatgca gggtgcctaa agcaaggaac cccaccgtga cgtacgggaa aaaccaagtc     840
atcatgctac tgtatcctga ccacccaaca ctcctgtcct accggaatat gggagaagaa     900
ccaaactatc aagaagagtg ggtgatgcat aagaaggaag tcgtgctaac cgtgccgact     960
gaagggctcg aggtcacgtg gggcaacaac gagccgtata agtattggcc gcagttatct    1020
acaaacggta cagcccatgg ccacccgcat gagataattc tgtattatta tgagctgtac    1080
cccactatga ct                                                       1092
```

FIGURE 18

```
IPVHMKSDAS KFTHEKPEGY YNWHHGAVQY SGGRFTIPTG AGKPGDSGRP IFDNKGRVVA    60
IVLGGANEGA RTALSVVTWN KDIVTKITPE GAEEWSLAIP VMCLLANTTF PCSQPPCTPC   120
CYEKEPEETL RMLEDNVMRP GYYQLLQASL TCSPHRQRRS TKDNFNVYKA TRPYLAHCPD   180
CGEGHSCHSP VALERIRNEA TDGTLKIQVS LQIGIKTDDS HDWTKLRYMD NHMPADAERA   240
GLFVRTSAPC TITGTMGHFI LARCPKGETL TVGFTDSRKI SHSCTHPFHH DPPVIGREKF   300
HSRPQHGKEL PCSTYVQSTA ATTEEIEVHM PPDTPDRTLM SQQSGNVKIT VNGQTVRYKC   360
NCGGSNEGLT TTDKVINNCK VDQCHAAVTN HKKWQYNSPL VPRNAELGDR KGKIHIPFPL   420
ANVTCRVPKA RNPTVTYGKN QVIMLLYPDH PTLLSYRNMG EEPNYQEEWV MHKKEVVLTV   480
PTEGLEVTWG NNEPYKYWPQ LSTNGTAHGH PHEIILYYYE LYPTMTVVVV SVATFILLSM   540
VGMAAGMCMR ARRRCITPYE LTPGATVPFL LSLICCIRTA KAATYQEAAI YLWNEQQPL    599
```

FIGURE 19

```
IPVHMKSDAS KFTHEKPEGY YNWHHGAVQY SGGRFTIPTG AGKPGDSGRP IFDNKGRVVA    60
IVLGGANEGA RTALSVVTWN KDIVTKITPE GAEEWSLAIP VMCLLANTTF PCSQPPCTPC   120
CYEKEPEETL RMLEDNVMRP GYYQLLQASL TCSPHRQRRS TKDNFNVYKA TRPYLAHCPD   180
CGEGHSCHSP VALERIRNEA TDGTLKIQVS LQIGIKTDDS HDWTKLRYMD NHMPADAERA   240
GLFVRTSAPC TITGTMGHFI LARCPKGETL TVGFTDSRKI SHSCTHPFHH DPPVIGREKF   300
HSRPQHGKEL PCSTYVQSTA ATTEEIEVHM PPDTPDRTLM SQQSGNVKIT VNGQTVRYKC   360
NCGGSNEGLT TTDKVINNCK VDQCHAAVTN HKKWQYNSPL VPRNAELGDR KGKIHIPFPL   420
ANVTCRVPKA RNPTVTYGKN QVIMLLYPDH PTLLSYRNMG EEPNYQEEWV MHKKEVVLTV   480
PTEGLEVTWG NNEPYKYWPQ LSTNGTAHGH PHEIILYYYE LYPTMTVVVV SVATFILLSM   540
VGMAAGMCMC ARRRCITPYE LTPGATVPFL LSLICCIRTA KAATYQEAAI YLWNEQQPL    599
```

FIGURE 20

```
IPVHMKSDAS KFTHEKPEGY YNWHHGAVQY SGGRFTIPTG AGKPGDSGRP IFDNKGRVVA    60
IVLGGANEGA RTALSVVTWN KDIVTKITPE GAEEWSLAIP VMCLLANTTF PCSQPPCTPC   120
CYEKEPEETL RMLEDNVMRP GYYQLLQASL TCSPHRQRRS TKDNFNVYKA TRPYLAHCPD   180
CGEGHSCHSP VALERIRNEA TDGTLKIQVS LQIGIKTDDS HDWTKLRYMD NHMPADAERA   240
GLFVRTSAPC TITGTMGHFI LARCPKGETL TVGFTDSRKI SHSCTHPFHH DPPVIGREKF   300
HSRPQHGKEL PCSTYVQSTA ATTEEIEVHM PPDTPDRTLM SQQSGNVKIT VNGQTVRYKC   360
NCGGSNEGLT TTDKVINNCK VDQCHAAVTN HKKWQYNSPL VPRNAELGDR KGKIHIPFPL   420
ANVTCRVPKA RNPTVTYGKN QVIMLLYPDH PTLLSYRNMG EEPNYQEEWV MHKKEVVLTV   480
PTEGLEVTWG NNEPYKYWPQ LSTNGTAHGH PHEIILYYYE LYPTMTVVVV SVATFILLSM   540
VGMAAGMCMC ARRRCITPYE LTPGATVPFL LSLICCIRTA KAATYQEAAI YLWNEQQPL    599
```

FIGURE 21

```
IPVHMKSDAS KFTHEKPEGY YNWHHGAVQY SGGRFTIPTG AGKPGDSGRP IFDNKGRVVA     60
IVLGGANEGA RTALSVVTWN KDIVTKITPE GAEEWSLAIP VMCLLANTTF PCSQPPCTPC    120
CYEKEPEETL RMLEDNVMRP GYYQLLQASL TCSPHRQRRS TKDNFNVYKA TRPYLAHCPD    180
CGEGHSCHSP VALERIRNEA TDGTLKIQVS LQIGIKTDDS HDWTKLRYMD NHMPADAERA    240
GLFVRTSAPC TITGTMGHFI LARCPKGETL TVGFTDSRKI SHSCTHPFHH DPPVIGREKF    300
HSRPQHGKEI PCSTYVQSTA ATTEEIEVHM PPDTPDRTLM SQQSGNVKIT VNGQTVRYKC    360
NCGGSNEGLT TTDKVINNCK VDQCHAAVTN HKKWQYNSPL VPRNAELGDR KGKIHIPFPL    420
ANVTCRVPKA RNPTVTYGKN QVIMLLYPDH PTLLSYRNMG EEPNYQEEWV MHKKEVVLTV    480
PTEGLEVTWG NNEPYKYWPQ LSTNGTAHGH PHEIILYYYE LYPTMTVVVV SVATFILLSM    540
VGMAAGMCMC ARRRCITPYE LTPGATVPFL LSLICCIRTA KAATYQEAAI YLWNEQQPL     599
```

FIGURE 22

```
DNFNVYKATR PYLAHCPDCG EGHSCHSPVA LERIRNEATD GTLKIQVSLQ IGIKTDDSHD     60
WTKLRYMDNH MPADAERAGL FVRTSAPCTI TGTMGHFILA RCPKGETLTV GFTDSRKISH    120
SCTHPFHHDP PVIGREKFHS RPQHGKELPC STYVQSTAAT TEEIEVHMPP DTPDRTLMSQ    180
QSGNVKITVN GQTVRYKCNC GGSNEGLTTT DKVINNCKVD QCHAAVTNHK KWQYNSPLVP    240
RNAELGDRKG KIHIPFPLAN VTCRVPKARN PTVTYGKNQV IMLLYPDHPT LLSYRNMGEE    300
PNYQEEWVMH KKEVVLTVPT EGLEVTWGNN EPYKYWPQLS TNGTAHGHPH EIILYYYELY    360
PTMT                                                                 364
```

FIGURE 23

```
DNFNVYKATR PYLAHCPDCG EGHSCHSPVA LERIRNEATD GTLKIQVSLQ IGIKTDDSHD     60
WTKLRYMDNH MPADAERAGL FVRTSAPCTI TGTMGHFILA RCPKGETLTV GFTDSRKISH    120
SCTHPFHHDP PVIGREKFHS RPQHGKELPC STYVQSTAAT TEEIEVHMPP DTPDRTLMSQ    180
QSGNVKITVN GQTVRYKCNC GGSNEGLTTT DKVINNCKVD QCHAAVTNHK KWQYNSPLVP    240
RNAELGDRKG KIHIPFPLAN VTCRVPKARN PTVTYGKNQV IMLLYPDHPT LLSYRNMGEE    300
PNYQEEWVMH KKEVVLTVPT EGLEVTWGNN EPYKYWPQLS TNGTAHGHPH EIILYYYELY    360
PTMT                                                                 364
```

FIGURE 24

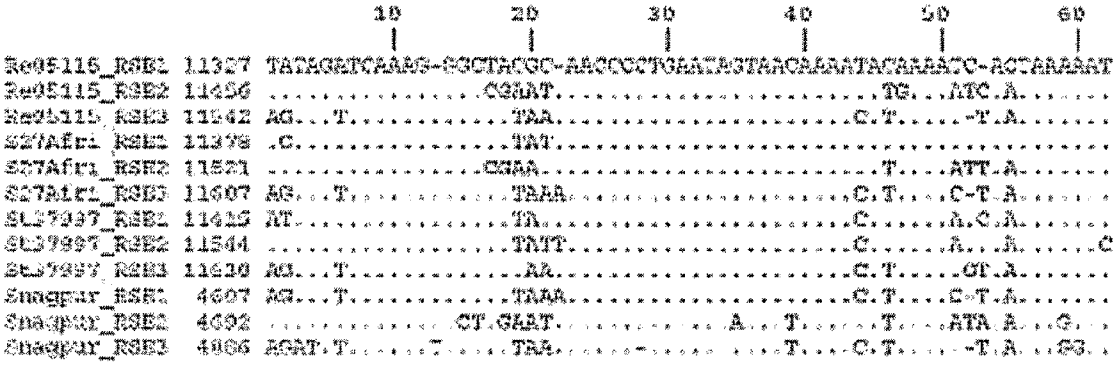
```
DNFNVYKATR PYLAHCPDCG EGHSCHSPVA LERIRNEATD GTLKIQVSLQ IGIKTDDSHD    60
WTKLRYMDNH MPADAERAGL FVRTSAPCTI TGTMGHFILA RCPKGETLTV GFTDSRKISH   120
SCTHPFHHDP PVIGREKFHS RPQHGKELPC STYVQSTAAT TEEIEVHMPP DTPDRTLMSQ   180
QSGNVKITVN GQTVRYKCNC GGSNEGLTTT DKVINNCKVD QCHAAVTNHK KWQYNSPLVP   240
RNAELGDRKG KIHIPFPLAN VTCRVPKARN PTVTYGKNQV IMLLYPDHPT LLSYRNMGEE   300
PNYQEEWVMH KKEVVLTVPT EGLEVTWGNN EPYKYWPQLS TNGTAHGHPH EIILYYYELY   360
PTMT                                                                364
```

FIGURE 25

```
DNFNVYKATR PYLAHCPDCG EGHSCHSPVA LERIRNEATD GTLKIQVSLQ IGIKTDDSHD    60
WTKLRYMDNH MPADAERAGL FVRTSAPCTI TGTMGHFILA RCPKGETLTV GFTDSRKISH   120
SCTHPFHHDP PVIGREKFHS RPQHGKEIPC STYVQSTAAT TEEIEVHMPP DTPDRTLMSQ   180
QSGNVKITVN GQTVRYKCNC GGSNEGLTTT DKVINNCKVD QCHAAVTNHK KWQYNSPLVP   240
RNAELGDRKG KIHIPFPLAN VTCRVPKARN PTVTYGKNQV IMLLYPDHPT LLSYRNMGEE   300
PNYQEEWVMH KKEVVLTVPT EGLEVTWGNN EPYKYWPQLS TNGTAHGHPH EIILYYYELY   360
PTMT                                                                364
```

FIGURE 26

A


10
20
30
40
50
60
Re05115_RSE1 11327 TATAGGTCAAAG-GGCTACGC-AACCCCTGAATAGTAACAAAATACAAAATC-ACTAAAAAT
Re05115_RSE2 11456 .....CGAAT.....TG...ATC.A.....
Re05115_RSE3 11542 AG...T.....TAA.....C.T.....-T.A.....
S27Afr1_RSE1 11378 .C.....TAT.....
S27Afr1_RSE2 11521 .....CGAA.....T....ATT.A.....
S27Afr1_RSE3 11607 AG...T.....TAAA.....C.T....C-T.A.....
SL37997_RSE1 11413 AT.....TA.....C.....A.C.A.....
SL37997_RSE2 11541 .....TATT.....C.....A...A.....C
SL37997_RSE3 11630 AG...T.....AA.....C.T.....GT.A.....
Snagpur_RSE1 4607 AG...T.....TAAA.....C.T....C-T.A.....
Snagpur_RSE2 4692 .....CT.GAAT.....A..T.....T....ATA.A...G...
Snagpur_RSE3 4806 AGAT.T.....T.....TAA.....T...C.T.....-T.A...GG..

B

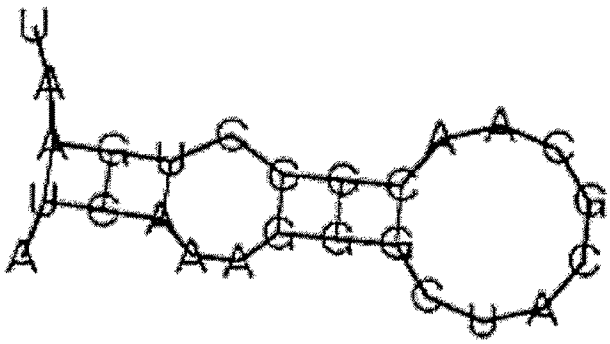


FIGURE 27

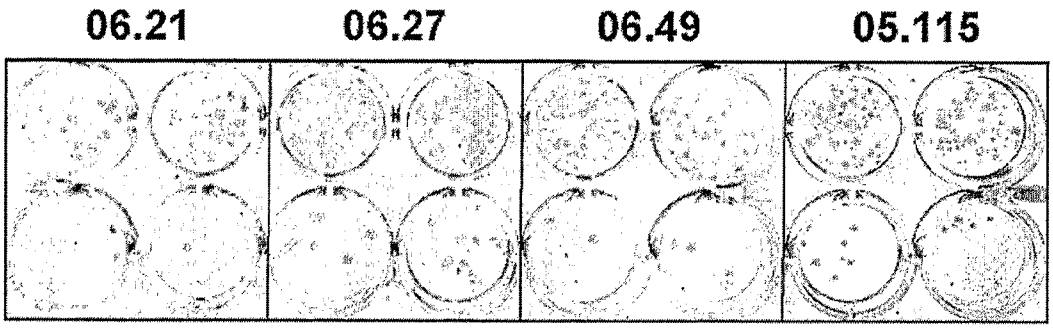


FIGURE 28

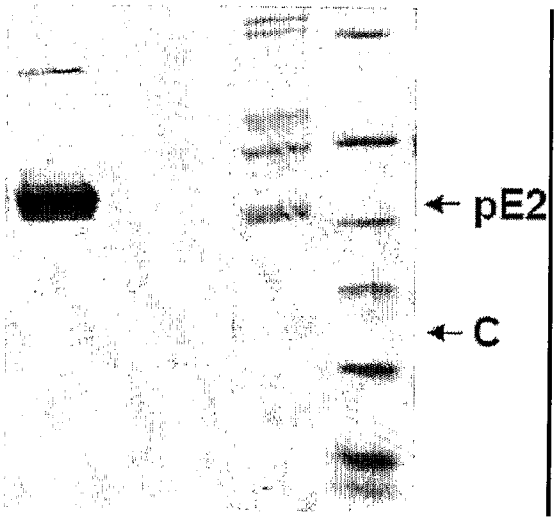


FIGURE 29

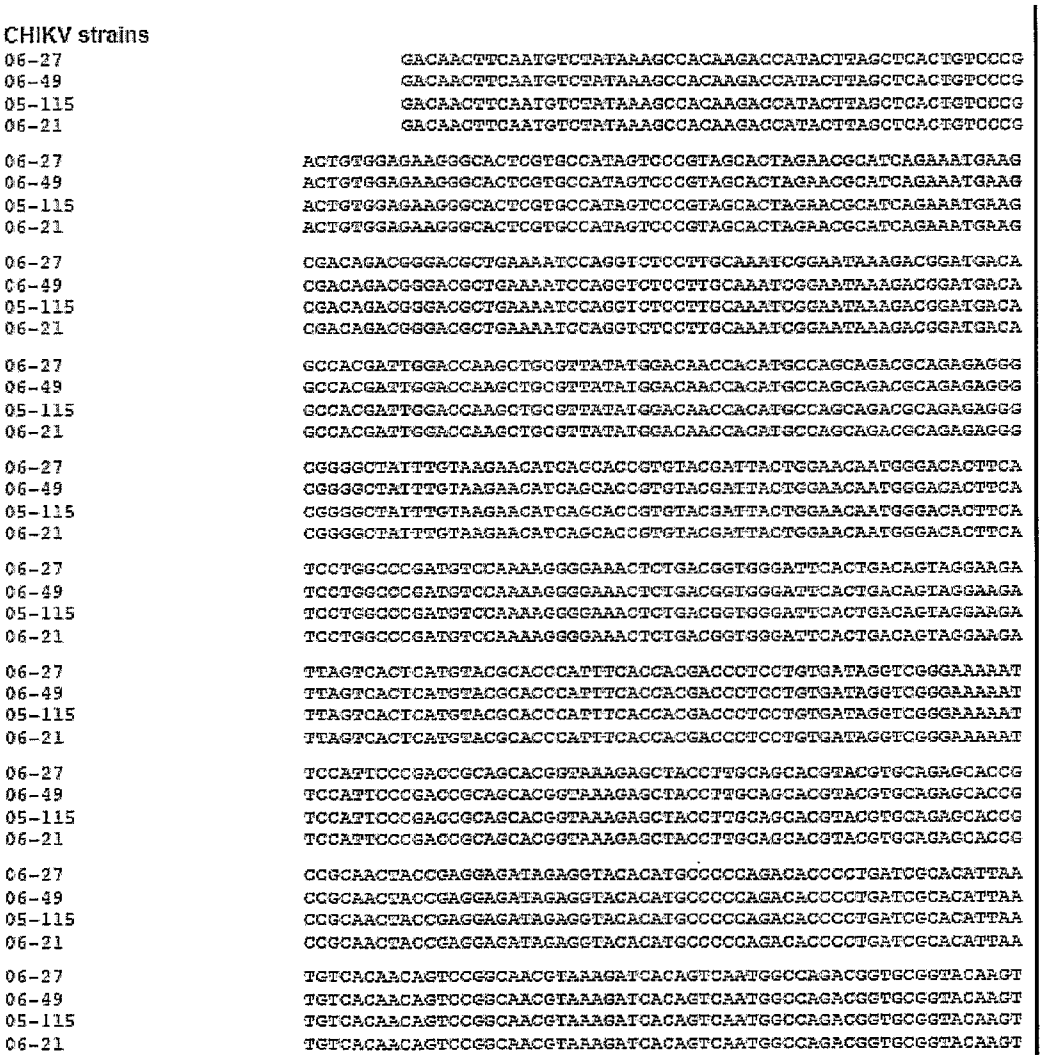

CHIKV strains
06-27 GACAACTTCAATGTCTATAAAGCCACAAGACCATACTTAGCTCACTGTCCCG
06-49 GACAACTTCAATGTCTATAAAGCCACAAGACCATACTTAGCTCACTGTCCCG
05-115 GACAACTTCAATGTCTATAAAGCCACAAGACCATACTTAGCTCACTGTCCCG
06-21 GACAACTTCAATGTCTATAAAGCCACAAGACCATACTTAGCTCACTGTCCCG
06-27 ACTGTGGAGAAGGGCACTCGTGCCATAGTCCCGTAGCACTAGAACGCATCAGAAATGAAG
06-49 ACTGTGGAGAAGGGCACTCGTGCCATAGTCCCGTAGCACTAGAACGCATCAGAAATGAAG
05-115 ACTGTGGAGAAGGGCACTCGTGCCATAGTCCCGTAGCACTAGAACGCATCAGAAATGAAG
06-21 ACTGTGGAGAAGGGCACTCGTGCCATAGTCCCGTAGCACTAGAACGCATCAGAAATGAAG
06-27 CGACAGACGGGACGCTGAAAATCCAGGTCTCCTTGCAAATCGGAATAAAGACGGATGACA
06-49 CGACAGACGGGACGCTGAAAATCCAGGTCTCCTTGCAAATCGGAATAAAGACGGATGACA
05-115 CGACAGACGGGACGCTGAAAATCCAGGTCTCCTTGCAAATCGGAATAAAGACGGATGACA
06-21 CGACAGACGGGACGCTGAAAATCCAGGTCTCCTTGCAAATCGGAATAAAGACGGATGACA
06-27 GCCACGATTGGACCAAGCTGCGTTATATGGACAACCACATGCCAGCAGACGCAGAGAGGG
06-49 GCCACGATTGGACCAAGCTGCGTTATATGGACAACCACATGCCAGCAGACGCAGAGAGGG
05-115 GCCACGATTGGACCAAGCTGCGTTATATGGACAACCACATGCCAGCAGACGCAGAGAGGG
06-21 GCCACGATTGGACCAAGCTGCGTTATATGGACAACCACATGCCAGCAGACGCAGAGAGGG
06-27 CGGGGCTATTTGTAAGAACATCAGCACCGTGTACGATTACTGGAAGAATGGGACACTTCA
06-49 CGGGGCTATTTGTAAGAACATCAGCACCGTGTACGATTACTGGAAGAATGGGACACTTCA
05-115 CGGGGCTATTTGTAAGAACATCAGCACCGTGTACGATTACTGGAACAATGGGACACTTCA
06-21 CGGGGCTATTTGTAAGAACATCAGCACCGTGTACGATTACTGGAACAATGGGACACTTCA
06-27 TCCTGGCCCGATGTCCAAAAGGGGAAACTCTGACGGTGGGATTCACTGACAGTAGGAAGA
06-49 TCCTGGCCCGATGTCCAAAAGGGGAAACTCTGACGGTGGGATTCACTGACAGTAGGAAGA
05-115 TCCTGGCCCGATGTCCAAAAGGGGAAACTCTGACGGTGGGATTCACTGACAGTAGGAAGA
06-21 TCCTGGCCCGATGTCCAAAAGGGGAAACTCTGACGGTGGGATTCACTGACAGTAGGAAGA
06-27 TTAGTCACTCATGTACGCACCCATTTCACCACGACCCTCCTGTGATAGGTCGGGAAAAAT
06-49 TTAGTCACTCATGTACGCACCCATTTCACCACGACCCTCCTGTGATAGGTCGGGAAAAAT
05-115 TTAGTCACTCATGTACGCACCCATTTCACCACGACCCTCCTGTGATAGGTCGGGAAAAAT
06-21 TTAGTCACTCATGTACGCACCCATTTCACCACGACCCTCCTGTGATAGGTCGGGAAAAAT
06-27 TCCATTCCCGACCGCAGCACGGTAAAGAGCTACCTTGCAGCACGTACGTGCAGAGCACCG
06-49 TCCATTCCCGACCGCAGCACGGTAAAGAGCTACCTTGCAGCACGTACGTGCAGAGCACCG
05-115 TCCATTCCCGACCGCAGCACGGTAAAGAGCTACCTTGCAGCACGTACGTGCAGAGCACCG
06-21 TCCATTCCCGACCGCAGCACGGTAAAGAGCTACCTTGCAGCACGTACGTGCAGAGCACCG
06-27 CCGCAACTACCGAGGAGATAGAGGTACACATGCCCCCAGACACCCCTGATCGCACATTAA
06-49 CCGCAACTACCGAGGAGATAGAGGTACACATGCCCCCAGACACCCCTGATCGCACATTAA
05-115 CCGCAACTACCGAGGAGATAGAGGTACACATGCCCCCAGACACCCCTGATCGCACATTAA
06-21 CCGCAACTACCGAGGAGATAGAGGTACACATGCCCCCAGACACCCCTGATCGCACATTAA
06-27 TGTCACAACAGTCCGGCAACGTAAAGATCACAGTCAATGGCCAGACGGTGCGGTACAAGT
06-49 TGTCACAACAGTCCGGCAACGTAAAGATCACAGTCAATGGCCAGACGGTGCGGTACAAGT
05-115 TGTCACAACAGTCCGGCAACGTAAAGATCACAGTCAATGGCCAGACGGTGCGGTACAAGT
06-21 TGTCACAACAGTCCGGCAACGTAAAGATCACAGTCAATGGCCAGACGGTGCGGTACAAGT

FIGURE 29 (con't)

```
06-27   GTAATTGCGGTGGCTCAAATGAAGGACTAACAACTACAGACAAAGTGATTAATAACTGCA
06-49   GTAATTGCGGTGGCTCAAATGAAGGACTAACAACTACAGACAAAGTGATTAATAACTGCA
05-115  GTAATTGCGGTGGCTCAAATGAAGGACTAACAACTACAGACAAAGTGATTAATAACTGCA
06-21   GTAATTGCGGTGGCTCAAATGAAGGACTAACAACTACAGACAAAGTGATTAATAACTGCA

06-27   AGGTTGATCAATGTCATGCCGCGGTCACCAATCACAAAAAGTGGCAGTATAACTCCCCTC
06-49   AGGTTGATCAATGTCATGCCGCGGTCACCAATCACAAAAAGTGGCAGTATAACTCCCCTC
05-115  AGGTTGATCAATGTCATGCCGCGGTCACCAATCACAAAAAGTGGCAGTATAACTCCCCTC
06-21   AGGTTGATCAATGTCATGCCGCGGTCACCAATCACAAAAAGTGGCAGTATAACTCCCCTC

06-27   TGGTCCCGCGTAATGCTGAACTTGGGGACCGAAAAGGAAAAATTCACATCCCGTTTCCGC
06-49   TGGTCCCGCGTAATGCTGAACTTGGGGACCGAAAAGGAAAAATTCACATCCCGTTTCCGC
05-115  TGGTCCCGCGTAATGCTGAACTTGGGGACCGAAAAGGAAAAATTCACATCCCGTTTCCGC
06-21   TGGTCCCGCGTAATGCTGAACTTGGGGACCGAAAAGGAAAAATTCACATCCCGTTTCCGC

06-27   TGGCAAATGTAACATGCAGGGTGCCTAAAGCAAGGAACCCCACCGTGACGTACGGGAAAA
06-49   TGGCAAATGTAACATGCAGGGTGCCTAAAGCAAGGAACCCCACCGTGACGTACGGGAAAA
05-115  TGGCAAATGTAACATGCAGGGTGCCTAAAGCAAGGAACCCCACCGTGACGTACGGGAAAA
06-21   TGGCAAATGTAACATGCAGGGTGCCTAAAGCAAGGAACCCCACCGTGACGTACGGGAAAA

06-27   ACCAAGTCATCATGCTACTGTATCCTGACCACCCAACACTCCTGTCCTACCGGAATATGG
06-49   ACCAAGTCATCATGCTACTGTATCCTGACCACCCAACACTCCTGTCCTACCGGAATATGG
05-115  ACCAAGTCATCATGCTACTGTATCCTGACCACCCAACACTCCTGTCCTACCGGAATATGG
06-21   ACCAAGTCATCATGCTACTGTATCCTGACCACCCAACACTCCTGTCCTACCGGAATATGG

06-27   GAGAAGAACCAAACTATCAAGAAGAGTGGGTGATGCATAAGAAGGAAGTCGTGCTAACCG
06-49   GAGAAGAACCAAACTATCAAGAAGAGTGGGTGATGCATAAGAAGGAAGTCGTGCTAACCG
05-115  GAGAAGAACCAAACTATCAAGAAGAGTGGGTGATGCATAAGAAGGAAGTCGTGCTAACCG
06-21   GAGAAGAACCAAACTATCAAGAAGAGTGGGTGATGCATAAGAAGGAAGTCGTGCTAACCG

06-27   TGCCGACTGAAGGGCTCGAGGTCACGTGGGGCAACAACGAGCCGTATAAGTATTGGCCGC
06-49   TGCCGACTGAAGGGCTCGAGGTCACGTGGGGCAACAACGAGCCGTATAAGTATTGGCCGC
05-115  TGCCGACTGAAGGGCTCGAGGTCACGTGGGGCAACAACGAGCCGTATAAGTATTGGCCGC
06-21   TGCCGACTGAAGGGCTCGAGGTCACGTGGGGCAACAACGAGCCGTATAAGTATTGGCCGC

06-27   AGTTATCTACAAACGGTACAGCCCATGGCCACCCGCATGAGATAATTCTGTATTATTATG
06-49   AGTTATCTACAAACGGTACAGCCCATGGCCACCCGCACGAGATAATTCTGTATTATTATG
05-115  AGTTATCTACAAACGGTACAGCCCATGGCCACCCGCATGAGATAATTCTGTATTATTATG
06-21   AGTTATCTACAAACGGTACAGCCCATGGCCACCCGCATGAGATAATTCTGTATTATTATG

06-27   AGCTGTACCCCACTATGACT
06-49   AGCTGTACCCCACTATGACT
05-115  AGCTGTACCCCACTATGACT
05-21   AGCTGTACCCCACTATGACT
```

FIGURE 30

Primer for the N-terminal sequence of CHIK E2-gp :

```
E2 furin cleavage site   D¹   N   F   N   V   Y   K   A   T   R   P   Y   L   A   H   C   P
5'-cgc agc acc aag/gac aac ttc aat gtc tat aaa gcc aca aga cca tac cta gct cac tgt cca-3'


5'-aaaaaagatctgacaacttcaatgtctataaagccacaagacc-3'          primer 43-mer   A+T:27   G+C:15
      BglII
```

Primer for the C-terminal sequence of the CHIK E2 ectodomain:

```
    P   H   E   I   I   L   Y   Y   Y   E   L   Y   P   T   M   T*** V   V   V   V   S
5'- ccg cat gag ata atc ttg tac tat tat gag ctg tac cct act atg act gta gta gtt gtg tca-3'


5'- ccg cat gag ata atc ttg tac tat tat gag ctg tac cct act atg acg cgg ccg caa aaa-3'
                                                                  NotI


5'-tttttgcggccgcgtcatagtagggtacagctcataatagtacaag-3'        primer 46-mer   A+T :25   G+C:21
       Not I
```

FIGURE 31

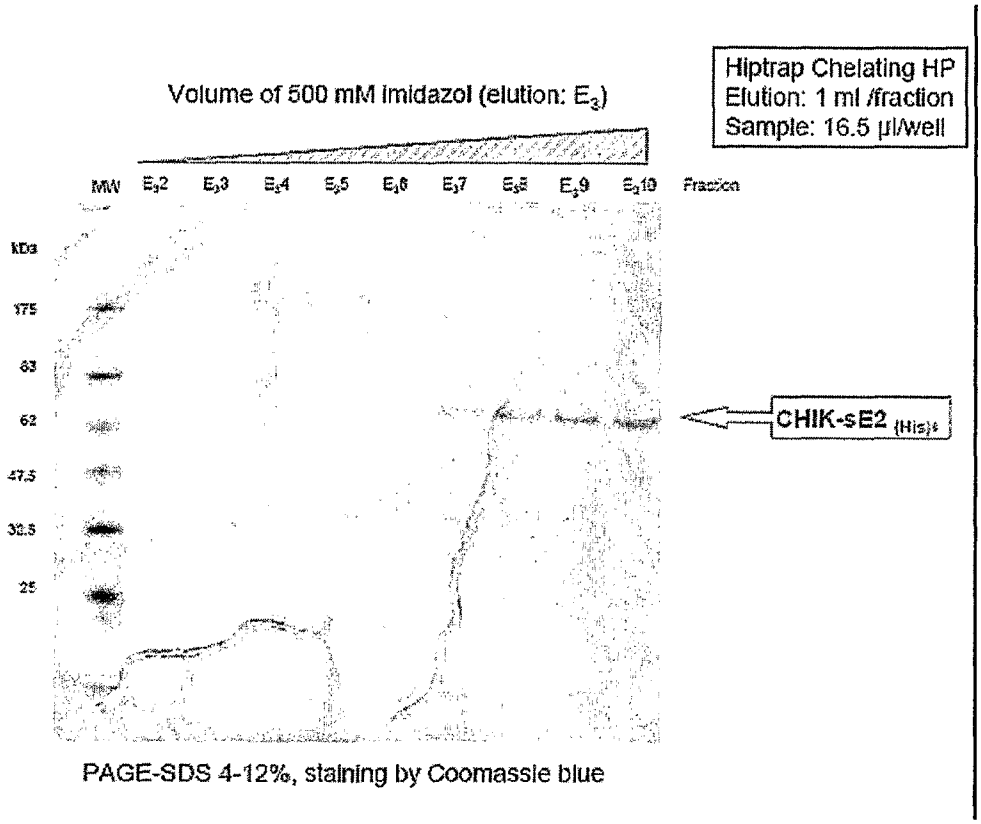

Hiptrap Chelating HP
Elution: 1 ml /fraction
Sample: 16.5 µl/well
Volume of 500 mM imidazol (elution: E3)
MW
E2
E3
E4
E5
E6
E7
E8
E9
E10
Fraction
kDa
175
83
62
47.5
32.5
25
CHIK-sE2 (His)6
PAGE-SDS 4-12%, staining by Coomassie blue

FIGURE 32

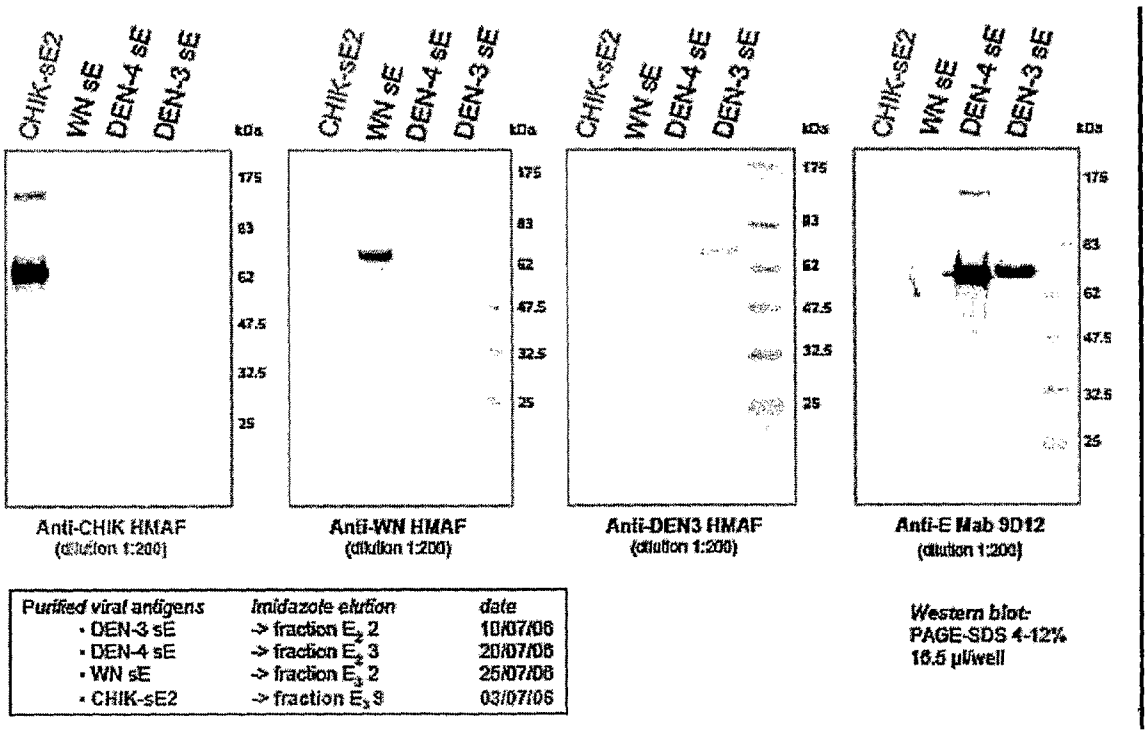

CHIK-sE2
WN sE
DEN-4 sE
DEN-3 sE
kDa
175
83
62
47.5
32.5
25
Anti-CHIK HMAF
(dilution 1:200)
Anti-WN HMAF
(dilution 1:200)
Anti-DEN3 HMAF
(dilution 1:200)
Anti-E Mab 9D12
(dilution 1:200)
Purified viral antigens
Imidazole elution
date
• DEN-3 sE
-> fraction E2 2
10/07/06
• DEN-4 sE
-> fraction E2 3
20/07/06
• WN sE
-> fraction E2 2
25/07/06
• CHIK-sE2
-> fraction E2 3
03/07/06
Western blot:
PAGE-SDS 4-12%
16.5 µl/well

FIGURE 33

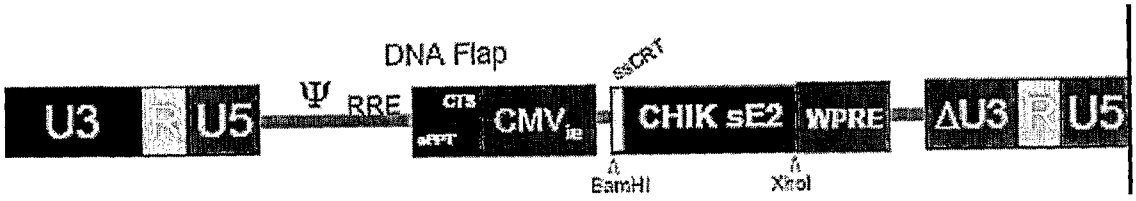

DNA Flap
ssCRT
U3
R
U5
Ψ
RRE
CTS
cPPT
CMV
IE
CHIK sE2
WPRE
ΔU3
R
U5
BamHI
XhoI

FIGURE 34

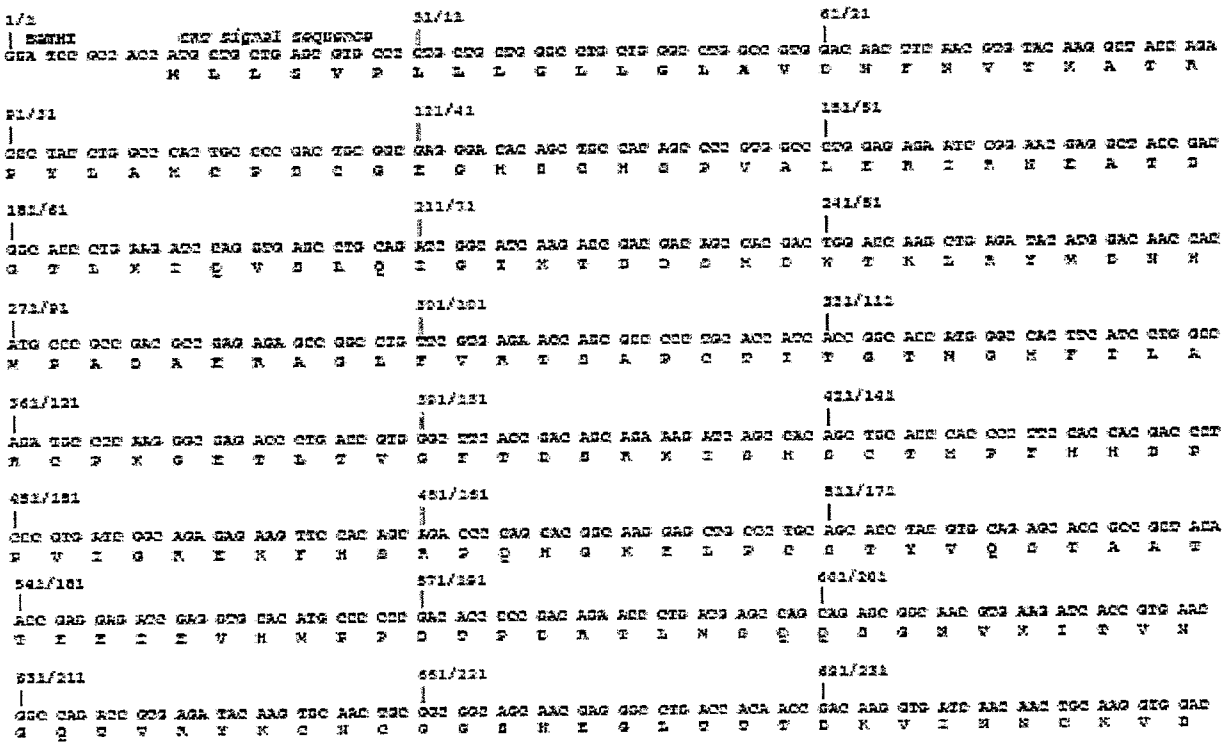

1/1
51/11
61/21
EcoRI
CH2 signal sequence
91/31
121/41
131/51
181/61
211/71
241/81
271/91
301/101
321/111
361/121
391/131
421/141
451/151
481/161
511/171
541/181
571/191
601/201
631/211
661/221
691/231

FIGURE 34(con't)

```
721/241                                 751/251                                 781/261
|                                       |                                       |
CAG TGC CAC GCT GCT GTG ACC AAC CAC AAG AAG TGG CAG TAC AAC AGC CCC CTG GTG CCC AGA AAC GCC GAG CTG GGC GAC AGA AAG GGC
Q   C   H   A   A   V   T   N   H   K   K   W   Q   Y   N   S   P   L   V   P   R   N   A   E   L   G   D   R   K   G

811/271                                 841/281                                 871/291
|                                       |                                       |
AAG ATC CAC ACC CCC TTC CCC CTG GCC AAC GTG ACC TGC AGA GTG CCC AAG GCC AGA AAC CCC ACC GTG ACC TAC GGC AAG AAC CAG GTG
K   I   H   T   P   F   P   L   A   N   V   T   C   R   V   P   K   A   R   N   P   T   V   T   Y   G   K   N   Q   V

901/301                                 931/311                                 961/321
|                                       |                                       |
ATC ATG CTG CTG TAC CCC GAC CAC CCC ACC CTG CTG AGC TAC AGA AAC ATG GGC GAG GAG CCC AAC TAC CAG GAG GAG TGG GTG ATG CAC
I   M   L   L   Y   P   D   H   P   T   L   L   S   Y   R   N   M   G   E   E   P   N   Y   Q   E   E   W   V   M   H

991/331                                 1021/341                                1051/351
|                                       |                                       |
AAG AAG GAG GTG GTG CTG ACC GTG CCC ACC GAG GGC CTG GAG GTG ACC TGG GGC AAC AAC GAG CCC TAC AAG TAC TGG CCC CAG CTG AGC
K   K   E   V   V   L   T   V   P   T   E   G   L   E   V   T   W   G   N   N   E   P   Y   K   Y   W   P   Q   L   S

1081/361                                1111/371                                1141/381
|                                       |                                       |           XhoI
ACC AAC GGC ACC GCC CAC GGA CAC CCC CAC GAG ATC ATC CTG TAC TAC TAC GAG CTG TAC CCC ACC ATG ACC TAA CTC GAG
T   N   G   T   A   H   G   H   P   H   E   I   I   L   Y   Y   Y   E   L   Y   P   T   M   T   *
```

FIGURE 35

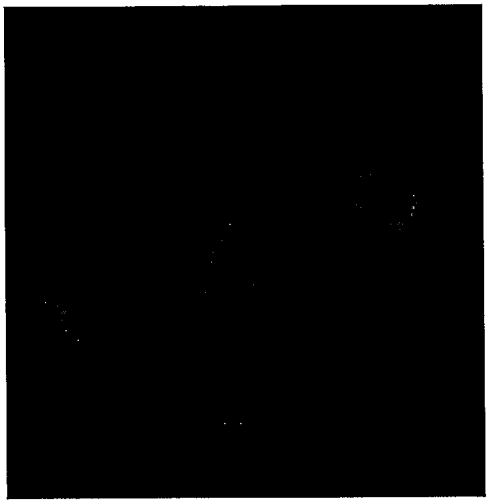


FIGURE 36

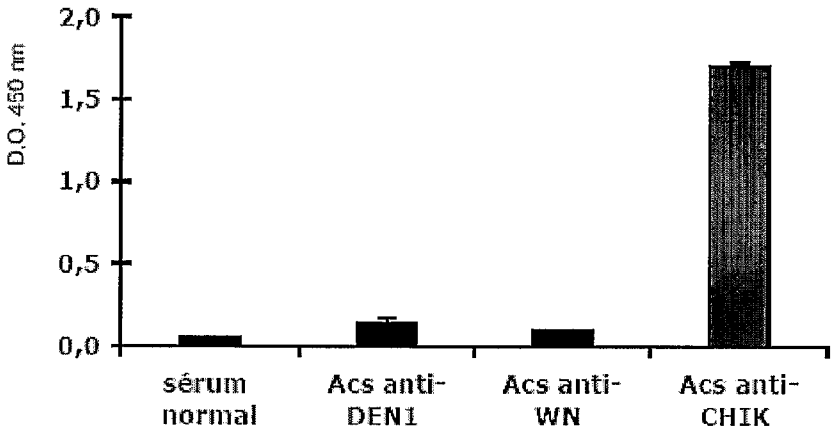

2,0
1,5
1,0
0,5
0,0
D.O. 450 nm
sérum normal
Acs anti-DEN1
Acs anti-WN
Acs anti-CHIK

FIGURE 37

```
MEFIPTQTFY NRRYQPRPWT PRPTIQVIRP RPRPQRQAGQ LAQLISAVNK LTMRAVPQQK      60
PRRNRKNKKQ KQKQQAPQNN TNQKKQPPKK KPAQKKKKPG RRERMCMKIE NDCIFEVKHE     120
GKVTGYACLV GDKVMKPAHV KGTIDNADLA KLAFKRSSKY DLECAQIPVH MKSDASKFTH     180
EKPEGYYNWH HGAVQYSGGR FTIPTGAGKP GDSGRPIFDN KGRVVAIVLG GANEGARTAL     240
SVVTWNKDIV TKITPEGAEE WSLAIPVMCL LANTTFPCSQ PPCIPCCYEK EPEETLRMLE     300
DNVMRPGYYQ LLQASLTCSP HRQRRSTKDN FNVYKATRPY LAHCPDCGEG HSCHSPVALE     360
RIRNEATDGT LKIQVSLQIG IGTDDSHDWT KLRYMDNHIP ADAGRAGLFV RTSAPCTITG     420
TMGHFILARC PKGETLTVGF TDSRKISHSC THPFHHDPPV IGREKFHSRP QHGKELPCST     480
YVQSNAATAE EIEVHMPPDT PDRTLLSQQS GNVKITVNGR TVRYKCNCGG SNEGLITTDK     540
VINNCKVDQC HAAVTNHKKW QYNSPLVPRN AELGDRKGKI HIPFPLANVT CMVPKARNPT     600
VTYGKNQVIM LLYPDHPTLL SYRSMGEEPN YQEEWVTHKK EVVLTVPTEG LEVTWGNNEP     660
YKYWPQLSAN GTAHGHPHEI ILYYYELYPT MTVVVVSVAS FILLSMVGMA VGMCMCARRR     720
CITPYELTPG ATVPFLLSLI CCIRTAKAAT YQEAAVYLWN EQQPLFWLQA LIPLAALIVL     780
CNCLRLLPCC CKTLAFLAVM SIGAHTVSAY EHVTVIPNTV GVPYKTLVNR PGYSPMVLEM     840
ELLSVTLEPT LSLDYITCEY KTVIPSPYVK CCGTAECKDK NLPDYSCKVF TGVYPFMWGG     900
AYCFCDAENT QLSEAHVEKS ESCKTEFASA YRAHTASASA KLRVLYQGNN ITVTAYANGD     960
HAVTVKDAKF IVGPMSSAWT PFDNKIVVYK GDVYNMDYPP FGAGRPGQFG DIQSRTPESK    1020
DVYANTQLVL QRPAAGTVHV PYSQAPSGFK YWLKERGASL QHTAPFGCQI ATNPVRAMNC    1080
AVGNMPISID IPDAAFTRVV DAPSLTDMSC EVPACTHSSD FGGVAIIKYA VSKKGKCAVH    1140
SMTNAVTIRE AEIEVEGNSQ LQISFSTALA SAEFRVQVCS TQVHCAAECH PPKDHIVNYP    1200
ASHTTLGVQD ISATAMSWVQ KITGGVGLVV AVAALILIVV LCVSFSRH                 1248
```

FIGURE 38

```
MEFIPTQTFY NRRYQPRPWT PRPTIQVIRP RPRPQRQAGQ LAQLISAVNK LTMRAVPQQK      60
PRRNRKNKKQ KQKQQAPQNN TNQKKQPPKK KPAQKKKKPG RRERMCMKIE NDCIFEVKHE     120
GKVTGYACLV GDKVMKPAHV KGTIDNADLA KLAFKRSSKY DLECAQIPVH MKSDASKFTH     180
EKPEGYYNWH HGAVQYSGGR FTIPTGAGKP GDSGRPIFDN KGRVVAIVLG GANEGARTAL     240
SVVTWNKDIV TKITPEGAEE WSLAIPVMCL LANTTFPCSQ PPCTPCCYEK EPEETLRMLE     300
DNVMRPGYYQ LLQASLTCSP HRQRRSTKDN FNVYKATRPY LAHCPDCGEG HSCHSPVALE     360
RIRNEATDGT LKIQVSLQIG IKTDDSHDWT KLRYMDNHMP ADAERAGLFV RTSAPCTITG     420
TMGHFILARC PKGETLTVGF TDSRKISHSC THPFHHDPPV IGREKFHSRP QHGKELPCST     480
YVQSTAATTE EIEVHMPPDT PDRTLMSQQS GNVKITVNGQ TVRYKCNCGG SNEGLTTTDK     540
VINNCKVDQC HAAVTNHKKW QYNSPLVPRN AELGDRKGKI HIPFPLANVT CRVPKARNPT     600
VTYGKNQVIM LLYPDHPTLL SYRNMGEEPN YQEEWVMHKK EVVLTVPTEG LEVTWGNNEP     660
YKYWPQLSTN GTAHGHPHEI ILYYYELYPT MTVVVVSVAT FILLSMVGMA AGMCMCARRR     720
CITPYELTPG ATVPFLLSLI CCIRTAKAAT YQEAAIYLWN EQQPLFWLQA LIPLAALIVL     780
CNCLRLLPCC CKTLAFLAVM SVGAHTVSAY EHVTVIPNTV GVPYKTLVNR PGYSPMVLEM     840
ELLSVTLEPT LSLDYITCEY KTVIPSPYVK CCGTAECKDK NLPDYSCKVF TGVYPFMWGG     900
AYCFCDAENT QLSEAHVEKS ESCKTEFASA YRAHTASASA KLRVLYQGNN ITVTAYANGD     960
HAVTVKDAKF IVGPMSSAWT PFDNKIVVYK GDVYNMDYPP FGAGRPGQFG DIQSRTPESK    1020
DVYANTQLVL QRPAAGTVHV PYSQAPSGFK YWLKERGASL QHTAPFGCQI ATNPVRAVNC    1080
AVGNMPISID IPEAAFTRVV DAPSLTDMSC EVPACTHSSD FGGVAIIKYA ASKKGKCAVH    1140
SMTNAVTIRE AEIEVEGNSQ LQISFSTALA SAEFRVQVCS TQVHCAAECH PPKDHIVNYP    1200
ASHTTLGVQD ISATAMSWVQ KITGGVGLVV AVAALILIVV LCVSFSRH                 1248
```

FIGURE 39

```
MEFIPTQTFY NRRYQPRPWT PRPTIQVIRP RPRPQRQAGQ LAQLISAVNK LTMRAVPQQK      60
PRRNRKNKKQ KQKQQAPQNN TNQKKQPPKK KPAQKKKKPG RRERMCMKIE NDCIFEVKHE     120
GKVTGYACLV GDKVMKPAHV KGTIDNADLA KLAFKRSSKY DLECAQIPVH MKSDASKFTH     180
EKPEGYYNWH HGAVQYSGGR FTIPTGAGKP GDSGRPIFDN KGRVVAIVLG GANEGARTAL     240
SVVTWNKDIV TKITPEGAEE WSLAIPVMCL LANTTFPCSQ PPCTPCCYEK EPEETLRMLE     300
DNVMRPGYYQ LLQASLTCSP HRQRRSTKDN FNVYKATRPY LAHCPDCGEG HSCHSPVALE     360
RIRNEATDGT LKIQVSLQIG IKTDDSHDWT KLRYMDNHMP ADAERAGLFV RTSAPCTITG     420
TMGHFILARC PKGETLTVGF TDSRKISHSC THPFHHDPPV IGREKFHSRP RHGKELPCST     480
YVQSTAATTE EIEVHMPPDT PDRTLMSQQS GNVKITVNGQ TVRYKCNCGG SNEGLTTTDK     540
VINNCKVDQC HAAVTNHKKW QYNSPLVPRN AELGDRKGKI HIPFPLANVT CRVPKARNPT     600
VTYGKNQVIM LLYPDHPTLL SYRNMGEEPN YQEEWVMHKK EVVLTVPTEG LEVTWGNNEP     660
YKYWPQLSTN GTAHGHPHEI ILYYYELYPT MTVVVVSVAT FILLSMVGMA AGMCMCARRR     720
CITPYELTPG ATVPFLLSLI CCIRTAKAAT YQEAAIYLWN EQQPLFWLQA LIPLAALIVL     780
CNCLRLLPCC CKTLAFLAVM SVGAHTVSAY EHVTVIPNTV GVPYKTLVNR PGYSPMVLEM     840
ELLSVTLEPT LSLDYITCEY KTVIPSPYVK CCGTAECKDK NLPDYSCKVF TGVYPFMWGG     900
AYCFCDAENT QLSEAHVEKS ESCKTEFASA YRAHTASASA KLRVLYQGNN ITVTAYANGD     960
HAVTVKDAKF IVGPMSSAWT PFDNKIVVYK GDVYNMDYPP FGAGRPGQFG DIQSRTPESK    1020
DVYANTQLVL QRPAAGTVHV PYSQAPSGFK YWLKERGASL QHTAPFGCQI ATNPVRAVNC    1080
AVGNMPISID IPEAAFTRVV DAPSLTDMSC EVPACTHSSD FGGVAIIKYA ASKKGKCAVH    1140
SMTNAVTIRE AEIEVEGNSQ LQISFSTALA SAEFRVQVCS TQVHCAAECH PPKDHIVNYP    1200
ASHTTLGVQD ISATAMSWVQ KITGGVGLVV AVAALILIVV LCVSFSRH                 1248
```

FIGURE 40

```
MEFIPTQTFY NRRYQPRPWT PRPTIQVIRP RPRPQRQAGQ LAQLISAVNK LTMRAVPQQK      60
PRRNRKNKKQ KQKQQAPQNN TNQKKQPPKK KPAQKKKKPG RRERMCMKIE NDCIFEVKHE     120
GKVTGYACLV GDKVMKPAHV KGTIDNADLA KLAFKRSSKY DLECAQIPVH MKSDASKFTH     180
EKPEGYYNWH HGAVQYSGGR FTIPTGAGKP GDSGRPIFDN KGRVVAIVLG GANEGARTAL     240
SVVTWNKDIV TKITPEGAEE WSLAIPVMCL LANTTFPCSQ PPCTPCCYEK EPEETLRMLE     300
DNVMRPGYYQ LLQASLTCSP HRQRRSTKDN FNVYKATRPY LAHCPDCGEG HSCHSPVALE     360
RIRNEATDGT LKIQVSLQIG IKTDDSHDWT KLRYMDNHMP ADAERAGLFV RTSAPCTITG     420
TMGHFILARC PKGETLTVGF TDSRKISHSC THPFHHDPPV IGREKFHSRP QHGKELPCST     480
YVQSTAATTE EIEVHMPPDT PDRTLMSQQS GNVKITVNGQ TVRYKCNCGG SNEGLTTTDK     540
VINNCKVDQC HAAVTNHKKW QYNSPLVPRN AELGDRKGKI HIPFPLANVT CRVPKARNPT     600
VTYGKNQVIM LLYPDHPTLL SYRNMGEEPN YQEEWVMHKK EVVLTVPTEG LEVTWGNNEP     660
YKYWPQLSTN GTAHGHPHEI ILYYYELYPT MTVVVVSVAT FILLSMVGMA AGMCMCARRR     720
CITPYELTPG ATVPFLLSLI CCIRTAKAAT YQEAAIYLWN EQQPLFWLQA LIPLAALIVL     780
CNCLRLLPCC CKTLAFLAVM SVGAHTVSAY EHVTVIPNTV GVPYKTLVNR PGYSPMVLEM     840
ELLSVTLEPT LSLDYITCEY KTVIPSPYVK CCGTAECKDK NLPDYSCKVF TGVYPFMWGG     900
AYCFCDAENT QLSEAHVEKS ESCKTEFASA YRAHTASASA KLRVLYQGNN ITVTAYANGD     960
HAVTVKDAKF IVGPMSSAWT PFDNKIVVYK GDVYNMDYPP FGAGRPGQFG DIQSRTPESK    1020
DVYANTQLVL QRPAAGTVHV PYSQAPSGFK YWLKERGASL QHTAPFGCQI ATNPVRAVNC    1080
AVGNMPISID IPEAAFTRVV DAPSLTDMSC EVPACTHSSD FGGVAIIKYA ASKKGKCAVH    1140
SMTNAVTIRE AEIEVEGNSQ LQISFSTALA SAEFRVQVCS TQVHCAAECH PPKDHIVNYP    1200
ASHTTLGVQD ISATAMSWVQ KITGGVGLVV AVAALILIVV LCVSFSRH                 1248
```

FIGURE 41

```
MEFIPTQTFY NRRYQPRPWT PRPTIQVIRP RPRPQRQAGQ LAQLISAVNK LTMRAVPQQK      60
PRRNRKNKKQ KQKQQAPQNN TNQKKQPPKK KPAQKKKKPG RRERMCMKIE NDCIFEVKHE     120
GKVTGYACLV GDKVMKPAHV KGTIDNADLA KLAFKRSSKY DLECAQIPVH MKSDASKFTH     180
EKPEGYYNWH HGAVQYSGGR FTIPTGAGKP GDSGRPIFDN KGRVVAIVLG GANEGARTAL     240
SVVTWNKDIV TKITPEGAEE WSLAIPVMCL LANTTFPCSQ PPCTPCCYEK EPEETLRMLE     300
DNVMRPGYYQ LLQASLTCSP HRQRRSTKDN FNVYKATRPY LAHCPDCGEG HSCHSPVALE     360
RIRNEATDGT LKIQVSLQIG IKTDDSHDWT KLRYMDNHMP ADAERAGLFV RTSAPCTITG     420
TMGHFILARC PKGETLTVGF TDSRKISHSC THPFHHDPPV IGREKFHSRP QHGKELPCST     480
YVQSTAATTE EIEVHMPPDT PDRTLMSQQS GNVKITVNGQ TVRYKCNCGG SNEGLTTTDK     540
VINNCKVDQC HAAVTNHKKW QYNSPLVPRN AELGDRKGKI HIPFPLANVT CRVPKARNPT     600
VTYGKNQVIM LLYPDHPTLL SYRNMGEEPN YQEEWVMHKK EVVLTVPTEG LEVTWGNNEP     660
YKYWPQLSTN GTAHGHPHEI ILYYYELYPT MTVVVVSVAT FILLSMVGMA AGMCMCARRR     720
CITPYELTPG ATVPFLLSLI CCIRTAKAAT YQEAAIYLWN EQQPLFWLQA LIPLAALIVL     780
CNCLRLLPCC CKTLAFLAVM SVGAHTVSAY EHVTVIPNTV GVPYKTLVNR PGYSPMVLEM     840
ELLSVTLEPT LSLDYITCEY KTVIPSPYVK CCGTAECKDK NLPDYSCKVF TGVYPFMWGG     900
AYCFCDAENT QLSEAHVEKS ESCKTEFASA YRAHTASASA KLRVLYQGNN ITVTAYANGD     960
HAVTVKDAKF IVGPMSSAWT PFDNKIVVYK GDVYNMDYPP FGAGRPGQFG DIQSRTPESK    1020
DVYANTQLVL QRPAVGTVHV PYSQAPSGFK YWLKERGASL QHTAPFGCQI ATNPVRAVNC    1080
AVGNMPISID IPEAAFTRVV DAPSLTDMSC EVPACTHSSD FGGVAIIKYA ASKKGKCAVH    1140
SMTNAVTIRE AEIEVEGNSQ LQISFSTALA SAEFRVQVCS TQVHCAAECH PPKDHIVNYP    1200
ASHTTLGVQD ISATAMSWVQ KITGGVGLVV AVAALILIVV LCVSFSRH                 1248
```

FIGURE 42

```
MEFIPTQTFY NRRYQPRPWT PRPTIQVIRP RPRPQRQAGQ LAQLISAVNK LTMRAVPQQK      60
PRRNRKNKKQ KQKQQAPQNN TNQKKQPPKK KPAQKKKKPG RRERMCMKIE NDCIFEVKHE     120
GKVTGYACLV GDKVMKPAHV KGTIDNADLA KLAFKRSSKY DLECAQIPVH MKSDASKFTH     180
EKPEGYYNWH HGAVQYSGGR FTIPTGAGKP GDSGRPIFDN KGRVVAIVLG GANEGARTAL     240
SVVTWNKDIV TKITPEGAEE WSLAIPVMCL LANTTFPCSQ PPCTPCCYEK EPEETLRMLE     300
DNVMRPGYYQ LLQASLTCSP HRQRRSTKDN FNVYKATRPY LAHCPDCGEG HSCHSPVALE     360
RIRNEATDGT LKIQVSLQIG IKTDDSHDWT KLRYMDNHMP ADAERAGLFV RTSAPCTITG     420
TMGHFILARC PKGETLTVGF TDSRKISHSC THPFHHDPPV IGREKFHSRP QHGKELPCST     480
YVQSTAATTE EIEVHMPPDT PDRTLMSQQS GNVKITVNGQ TVRYKCNCGG SNEGLTTTDK     540
VINNCKVDQC HAAVTNHKKW QYNSPLVPRN AELGDRKGKI HIPFPLANVT CRVPKARNPT     600
VTYGKNQVIM LLYPDHPTLL SYRNMGEEPN YQEEWVMHKK EVVLTVPTEG LEVTWGNNEP     660
YKYWPQLSTN GTAHGHPHEI ILYYYELYPT MTVVVVSVAT FILLSMVGMA AGMCMCARRR     720
CITPYELTPG ATVPFLLSLI CCIRTAKAAT YQEAAIYLWN EQQPLFWLQA LIPLAALIVL     780
CNCLRLLPCC CKTLAFLAVM SVGAHTVSAY EHVTVIPNTV GVPYKTLVNR PGYSPMVLEM     840
ELLSVTLEPT LSLDYITCEY KTVIPSPYVK CCGTAECKDK NLPDYSCKVF TGVYPFMWGG     900
AYCFCDAENT QLSEAHVEKS ESCKTEFASA YRAHTASASA KLRVLYQGNN ITVTAYANGD     960
HAVTVKDAKF IVGPMSSAWT PFDNKIVVYK GDVYNMDYPP FGAGRPGQFG DIQSRTPESK    1020
DVYANTQLVL QRPAVGTVHV PYSQAPSGFK YWLKERGASL QHTAPFGCQI ATNPVRAVNC    1080
AVGNMPISID IPEAAFTRVV DAPSLTDMSC EVPACTHSSD FGGVAIIKYA ASKKGKCAVH    1140
SMTNAVTIRE AEIEVEGNSQ LQISFSTALA SAEFRVQVCS TQVHCAAECH PPKDHIVNYP    1200
ASHTTLGVQD ISATAMSWVQ KITGGVGLVV AVAALILIVV LCVSFSRH                 1248
```

FIGURE 43

```
MEFIPTQTFY NRRYQPRPWT PRPTIQVIRP RPRPQRQAGQ LAQLISAVNK LTMRAVPQQK      60
PRRNRKNKKQ KQKQQAPQNN TNQKKQPPKK KPAQKKKKPG RRERMCMKIE NDCIFEVKHE     120
GKVTGYACLV GDKVMKPAHV KGTIDNADLA KLAFKRSSKY DLECAQIPVH MKSDASKFTH     180
EKPEGYYNWH HGAVQYSGGR FTIPTGAGKP GDSGRPIFDN KGRVVAIVLG GANEGARTAL     240
SVVTWNKDIV TKITPEGAEE WSLAIPVMCL LANTTFPCSQ PPCTPCCYEK EPEETLRMLE     300
DNVMRPGYYQ LLQASLTCSP HRQRRSTKDN FNVYKATRPY LAHCPDCGEG HSCHSPVALE     360
RIRNEATDGT LKIQVSLQIG IKTDDSHDWT KLRYMDNHMP ADAERAGLFV RTSAPCTITG     420
TMGHFILARC PKGETLTVGF TDSRKISHSC THPFHHDPPV IGREKFHSRP QHGKELPCST     480
YVQSTAATTE EIEVHMPPDT PDRTLMSQQS GNVKITVNGQ TVRYKCNCGG SNEGLTTTDK     540
VINNCKVDQC HAAVTNHKKW QYNSPLVPRN AELGDRKGKI HIPFPLANVT CRVPKARNPT     600
VTYGKNQVIM LLYPDHPTLL SYRNMGEEPN YQEEWVMHKK EVVLTVPTEG LEVTWGNNEP     660
YKYWPQLSTN GTAHGHPHEI ILYYYELYPT MTVVVVSVAT FILLSMVGMA AGMCMCARRR     720
CITPYELTPG ATVPFLLSLI CCIRTAKAAT YQEAAIYLWN EQQPLFWLQA LIPLAALIVL     780
CNCLRLLPCC CKTLAFLAVM SVGAHTVSAY EHVTVIPNTV GVPYKTLVNR PGYSPMVLEM     840
ELLSVTLEPT LSLDYITCEY KTVIPSPYVK CCGTAECKDK NLPDYSCKVF TGVYPFMWGG     900
AYCFCDAENT QLSEAHVEKS ESCKTEFASA YRAHTASASA KLRVLYQGNN ITVTAYANGD     960
HAVTVKDAKF IVGPMSSAWT PFDNKIVVYK GDVYNMDYPP FGAGRPGQFG DIQSRTPESK    1020
DVYANTQLVL QRPAVGTVHV PYSQAPSGFK YWLKERGASL QHTAPFGCQI ATNPVRAVNC    1080
AVGNMPISID IPEAAFTRVV DAPSLTDMSC EVPACTHSSD FGGVAIIKYA ASKKGKCAVH    1140
SMTNAVTIRE AEIEVEGNSQ LQISFSTALA SAEFRVQVCS TQVHCAAECH PPKDHIVNYP    1200
ASHTTLGVQD ISATAMSWVQ KITGGVGLVV AVAALILIVV LCVSFSRH                 1248
```

FIGURE 44

```
MDPVYVDIDA DSAFLKALQR AYPMFEVEPR QVTPNDHANA RAFSHLAIKL IEQEIDPDST      60
ILDIGSAPAR RMMSDRKYHC VCPMRSAEDP ERLANYARKL ASAAGKVLDR NISGKIGDLQ     120
AVMAVPDTET PTFCLHTDVS CRQRADVAIY QDVYAVHAPT SLYHQAIKGV RVAYWVGFDT     180
TPFMYNAMAG AYPSYSTNWA DEQVLKAKNI GLCSTDLTEG RRGKLSIMRG KKLKPCDRVL     240
FSVGSTLYPE SRKLLKSWHL PSVFHLKGKL SFTCRCDTVV SCEGYVVKRI TMSPGLYGKT     300
TGYAVTHHAD GFLMCKTTDT VDGERVSFSV CTYVPATICD QMTGILATEV TPEDAQKLLV     360
GLNQRIVVNG RTQRNTNTMK NYLLPVVAQA FSKWAKECRK DMEDEKLLGV RERTLTCCCL     420
WAFKKQKTHT VYKRPDTQSI QKVQAEFDSF VVPSLWSSGL SIPLRTRIKW LLSKVPKTDL     480
IPYSGDAREA RDAEKEAEEE REAELTREAL PPLQAAQEDV QVEIDVEQLE DRAGAGIIET     540
PRGAIKVTAQ PTDHVVGEYL VLSPQTVLRS QKLSLIHALA EQVKTCTHNG RAGRYAVEAY     600
DGRVLVPSGY AISPEDFQSL SESATMVYNE REFVNRKLHH IAMHGPALNT DEESYELVRA     660
ERTEHEYVYD VDQRRCCKKE EAAGLVLVGD LTNPPYHEFA YEGLKIRPAC PYKIAVIGVF     720
GVPGSGKSAI IKNLVTRQDL VTSGKKENCQ EITTDVMRQR GLEISARTVD SLLLNGCNRP     780
VDVLYVDEAF ACHSGTLLAL IALVRPRQKV VLCGDPKQCG FFNMMQMKVN YNHNICTQVY     840
HKSISRRCTL PVTAIVSSLH YEGKMRTTNE YNKPIVVDTT GSTKPDPGDL VLTCFRGWVK     900
QLQIDYRGYE VMTAAASQGL TRKGVYAVRQ KVNENPLYAS TSEHVNVLLT RTEGKLVWKT     960
LSGDPWIKTL QNPPKGNFKA TIKEWEVEHA SIMAGICSHQ MTFDTFQNKA NVCWAKSLVP    1020
ILETAGIKLN DRQWSQIIQA FKEDKAYSPE VALNEICTRM YGVDLDSGLF SKPLVSVYYA    1080
DNHWDNRPGG KMFGFNPEAA SILERKYPFT KGKWNINKQI CVTTRRIEDF NPTTNIIPAN    1140
RRLPHSLVAE HRPVKGERME WLVNKINGHH VLLVSGYNLA LPTKRVTWVA PLGVRGADYT    1200
YNLELGLPAT LGRYDLVVIN IHTPFRIHHY QQCVDHAMKL QMLGGDSLRL LKPGGSLLIR    1260
AYGYADRTSE RVICVLGRKF RSSRALKPPC VTSNTEMFFL FSNFDNGRRN FTTHVMNNQL    1320
NAAFVGQVTR AGCAPSYRVK RMDIAKNDEE CVVNAANPRG LPGDGVCKAV YKKWPESFKN    1380
SATPVGTAKT VMCGTYPVIH AVGPNFSNYS ESEGDRELAA AYREVAKEVT RLGVNSVAIP    1440
LLSTGVYSGG KDRLTQSLNH LFTAMDSTDA DVVIYCRDKE WEKKISEAIQ MRTQVELLDE    1500
HISIDCDIVR VHPDSSLAGR KGYSTTEGAL YSYLEGTRFH QTAVDMAEIH TMWPKQTEAN    1560
EQVCLYALGE SIESIRQKCP VDDADASSPP KTVPCLCRYA MTPERVTRLR MNHVTSIIVC    1620
SSFPLPKYKI EGVQKVKCSK VMLFDHNVPS RVSPREYRSS QESAQEASTI TSLTHSQFDL    1680
SVDGEILPVP SDLLDADAPAL EPALDDGATH TLPSTTGNLA AVSDWVMSTV PVAPPRRRRG    1740
RNLTVTCDER EGNITPMASV RFFRAELCPV VQETAETRDT AMSLQAPPST ATEPNHPPIS    1800
FGASSETFPI TFGDFNEGEI ESLSSELLTF GDFLPGEVDD LTDSDWSTCS DTDDELLDRA    1860
GGYIFSSDTG PGHLQQKSVR QSVLPVNTLE EVHEEKCYPP KLDEAKEQLL LKKLQESASM    1920
ANRSRYQSRK VENMKAAIIQ RLKRGCRLYL MSETPKVPTY RTTYPAPVYS PPINVRLSNP    1980
ESAVAACNEF LARNYPTVSS YQITDEYDAY LDMVDGSESC LDRATFNPSK LRSYPKQHAY    2040
HAPSIRSAVP SPFQNTLQNV LAAATKRNCN VTQMRELPTL DSAVFNVECF KKFACNQEYW    2100
EEFAASPIRI TTENLATYVT KLKGPKAAAL FAKTHNLLPL QEVPMDRFTV DMKRDVKVTP    2160
GTKHTEERPK VQVIQAAEPL ATAYLCGIHR ELVRRLNAVL LPNVHTLFDM SAEDFDAIIA    2220
AHFKPGDTVL ETDIASFDKS QDDSLALTAL MLLEDLGVDH SLLDLIEAAF GEISSCHLPT    2280
GTRFKFGAMM KSGMFLTLFV NTLLNITIAS RVLEDRLTKS ACAAFIGDDN IIHGVVSDEL    2340
MAARCATWMN MEVKIIDAVV SLKAPYFCGG FILHDTVTGT ACRVADPLKR LFKLGKPLAA    2400
GDEQDEDRRR ALADEVIRWQ RTGLIDELEK AVYSRYEVQG ISVVVMSMAT FASSRSNFEK    2460
LRGPVITLYG GPK                                                       2473
```

FIGURE 45

```
MDPVYVDIDA DSAFLKALQR AYPMFEVEPR QVTPNDHANA RAFSHLAIKL IEQEIDPDST      60
ILDIGSAPAR RMMSDRKYHC VCPMRSAEDP ERLANYARKL ASAAGKVLDR NISGKIGDLQ     120
AVMAVPDTET PTFCLHTDVS CRQRADVAIY QDVYAVHAPT SLYHQAIKGV RVAYWVGFDT     180
TPFMYNAMAG AYPSYSTNWA DEQVLKAKNI GLCSTDLTEG RRGKLSIMRG KKLKPCDRVL     240
FSVGSTLYPE SRKLLKSWHL PSVFHLKGKL SFTCRCDTVV SCEGYVVKRI TMSPGLYGKT     300
TGYAVTHHAD GFLMCKTTDT VDGERVSFSV CTYVPATICD QMTGILATEV TPEDAQKLLV     360
GLNQRIVVNG RTQRNINTMK NYLLPVVAQA FSKWAKECRK DMEDEKLLGV RERTLTCCCL     420
WAFKKQKTHT VYKRPDTQSI QKVQAEFDSF VVPSLWSSGL SIPLRTRIKW LLSKVPKTDL     480
IPYSGDAREA RDAEKEAEEE REAELTREAL PPLQAAQEDV QVEIDVEQLE DRAGAGIIET     540
PRGAIKVTAQ PTDHVVGEYL VLSPQTVLRS QKLSLIHALA EQVKTCTHNG RAGRYAVEAY     600
DGRVLVPSGY AISPEDFQSL SESATMVYNE REFVNRKLHH IAMHGPALNT DEESYELVRA     660
ERTEHEYVYD VDQRRCCKKE EAAGLVLVGD LTNPPYHEFA YEGLKIRPAC PYKIAVIGVF     720
GVPGSGKSAI IKNLVTRQDL VTSGKKENCQ EITTDVMRQR GLEISARTVD SLLLNGCNRP     780
VDVLYVDEAF ACHSGTLLAL IALVRPRQKV VLCGDPKQCG FFNMMQMKVN YNHNICTQVY     840
HKSISRRCTL PVTAIVSSLH YEGKMRTTNE YNKPIVVDTT GSTKPDPGDL VLTCFRGWVK     900
QLQIDYRGYE VMTAAASQGL TRKGVYAVRQ KVNENPLYAS TSEHVNVLLT RTEGKLVWKT     960
LSGDPWIKTL QNPPKGNFKA TIKEWEVEHA SIMAGICSHQ MTFDTFQNKA NVCWAKSLVP    1020
ILETAGIKLN DRQWSQIIQA FKEDKAYSPE VALNEICTRM YGVDLDSGLF SKPLVSVYYA    1080
DNHWDNRPGG KMFGFNPEAA SILERKYPFT KGKWNINKQI CVTTRRIEDF NPTTNIIPAN    1140
RRLPHSLVAE HRPVKGERME WLVNKINGHH VLLVSGYNLA LPTKRVTWVA PLGVRGADYT    1200
YNLELGLPAT LGRYDLVVIN IHTPFRIHHY QQCVDHAMKL QMLGGDSLRL LKPGGSLLIR    1260
AYGYADRTSE RVICVLGRKF RSSRALKPPC VTSNTEMFFL FSNFDNGRRN FTTHVMNNQL    1320
NAAFVGQVTR AGCAPSYRVK RMDIAKNDEE CVVNAANPRG LPGDGVCKAV YKKWPESFKN    1380
SATPVGTAKT VMCGTYPVIH AVGPNFSNYS ESEGDRELAA AYREVAKEVT RLGVNSVAIP    1440
LLSTGVYSGG KDRLTQSLNH LFTAMDSTDA DVVIYCRDKE WEKKISEAIQ MRTQVELLDE    1500
HISIDCDIVR VHPDSSLAGR KGYSTTEGAL YSYLEGTRFH QTAVDMAEIH TMWPKQTEAN    1560
EQVCLYALGE SIESIRQKCP VDDADASSPP KTVPCLCRYA MTPERVTRLR MNHVTSIIVC    1620
SSFPLPKYKI EGVQKVKCSK VMLFDHNVPS RVSPREYRSS QESAQEASTI TSLTHSQFDL    1680
SVDGEILPVP PDLDADAPAL EPALDDGATH TLPSTTGNLA AVSDWVMSTV PVAPPRRRRG    1740
RNLTVTCDER EGNITPMASV RFFRAELCPV VQETAETRDT AMSLQAPPST ATEPNHPPIS    1800
FGASSETFPI TFGDFNEGEI ESLSSELLTF GDFLPGEVDD LTDSDWSTCS DTDDELLDRA    1860
GGYIFSSDTG PGHLQQKSVR QSVLPVNTLE EVHEEKCYPP KLDEAKEQLL LKKLQESASM    1920
ANRSRYQSRK VENMKAAIIQ RLKRGCRLYL MSETPKVPTY RTTYPAPVYS PPINVRLSNP    1980
ESAVAACNEF LARNYPTVSS YQITDEYDAY LDMVDGSESC LDRATFNPSK LRSYPKQHAY    2040
HAPSIRSAVP SPFQNTLQNV LAAATKRNCN VTQMRELPTL DSAVFNVECF KKFACNQEYW    2100
EEFAASPIRI TTENLATYVT KLKGPKAAAL FAKTHNLLPL QEVPMDRFTV DMKRDVKVTP    2160
GTKHTEERPK VQVIQAAEPL ATAYLCGIHR ELVRRLNAVL LPNVHTLFDM SAEDFDAIIA    2220
AHFKPGDTVL ETDIASFDKS QDDSLALTAL MLLEDLGVDH SLLDLIEAAF GEISSCHLPT    2280
GTRFKFGAMM KSGMFLTLFV NTLLNITIAS RVLEDRLTKS ACAAFIGDDN IIHGVVSDEL    2340
MAARCATWMN MEVKIIDAVV SLKAPYFCGG FILHDTVTGT ACRVADPLKR LFKLGKPLAA    2400
GDEQDEDRRR ALADEVIRWQ RTGLIDELEK AVYSRYEVQG ISVVVMSMAT FASSRSNFEK    2460
LRGPVITLYG GPK                                                       2473
```

FIGURE 46

```
MDPVYVDIDA DSAFLKALQR AYPMFEVEPR QVTPNDHANA RAFSHLAIKL IEQEIDPDST      60
ILDIGSAPAR RMMSDRKYHC VCPMRSAEDP ERLANYARKL ASAAGKVLDR NISGKIGDLQ     120
AVMAVPDTET PTFCLHTDVS CRQRADVAIY QDVYAVHAPT SLYHQAIKGV RVAYWVGFDT     180
TPFMYNAMAG AYPSYSTNWA DEQVLKAKNI GLCSTDLTEG RRGKLSIMRG KKLKPCDRVL     240
FSVGSTLYPE SRKLLKSWHL PSVFHLKGKL SFTCRCDTVV SCEGYVVKRI TMSPGLYGKT     300
TGYAVTHHAD GFLMCKTTDT VDGERVSFSV CTYVPATICD QMTGILATEV TPEDAQKLLV     360
GLNQRIVVNG RTQRNTNTMK NYLLPVVAQA FSKWAKECRK DMEDEKLLGV RERTLTCCCL     420
WAFKKQKTHT VYKRPDTQSI QKVQAEFDSF VVPSLWSSGL SIPLRTRIKW LLSKVPKTDL     480
IPYSGDAREA RDAEKEAEEE REAELTREAL PPLQAAQEDV QVEIDVEQLE DRAGAGIIET     540
PRGAIKVTAQ PTDHVVGEYL VLSPQTVLRS QKLSLIHALA EQVKTCTHNG RAGRYAVEAY     600
DGRVLVPSGY AISPEDFQSL SESATMVYNE REFVNRKLHH IAMHGPALNT DEESYELVRA     660
ERTEHEYVYD VDQRRCCKKE EAAGLVLVGD LTNPPYHEFA YEGLKIRPAC PYKIAVIGVF     720
GVPGSGKSAI IKNLVTRQDL VTSGKKENCQ EITTDVMRQR GLEISARTVD SLLLNGCNRP     780
VDVLYVDEAF ACHSGTLLAL IALVRPRQKV VLCGDPKQCG FFNMMQMKVN YNHNICTQVY     840
HKSISRRCTL PVTAIVSSLH YEGKMRTTNE YNKPIVVDTT GSTKPDPGDL VLTCFRGWVK     900
QLQIDYRGYE VMTAAASQGL TRKGVYAVRQ KVNENPLYAS TSEHVNVLLT RTEGKLVWKT     960
LSGDPWIKTL QNPPKGNFKA TIKEWEVEHA SIMAGICSHQ MTFDTFQNKA NVCWAKSLVP    1020
ILETAGIKLN DRQWSQIIQA FKEDKAYSPE VALNEICTRM YGVDLDSGLF SKPLVSVYYA    1080
DNHWDNRPGG KMFGFNPEAA SILERKYPFT KGKWNINKQI CVTTRRIEDF NPTTNIIPAN    1140
RRLPHSLVAE HRPVKGERME WLVNKINGHH VLLVSGYNLA LPTKRVTWVA PLGVRGADYT    1200
YNLELGLPAT LGRYDLVVIN IHTPFRIHHY QQCVDHAMKL QMLGGDSLRL LKPGGSLLIR    1260
AYGYADRTSE RVICVLGRKF RSSRALKPPC VTSNTEMFFL FSNFDNGRRN FTTHVMNNQL    1320
NAAFVGQVTR AGCAPSYRVK RMDIAKNDEE CVVNAANPRG LPGDGVCKAV YKKWPESFKN    1380
SATPVGTAKT VMCGTYPVIH AVGPNFSNYS ESEGDRELAA AYREVAKEVT RLGVNSVAIP    1440
LLSTGVYSGG KDRLTQSLNH LFTAMDSTDA DVVIYCRDKE WEKKISEAIQ MRTQVELLDE    1500
HISIDCDIVR VHPDSSLAGR KGYSTTEGAL YSYLEGTRFH QTAVDMAEIH TMWPKQTEAN    1560
EQVCLYALGE SIESIRQKCP VDDADASSPP KTVPCLCRYA MTPERVTRLR MNHVTSIIVC    1620
SSFPLPKYKI EGVQKVKCSK VMLFDHNVPS RVSPREYRSS QESAQEASTI TSLTHSQFDL    1680
SVDGEILPVP SDLDADAPAL EPALDDGATH TLPSTTGNLA AVSDWVMSTV PVAPPRRRRG    1740
RNLTVTCDER EGNITPMASV RFFRAELCPV VQETAETRDT AMSLQAPPST ATEPNHPPIS    1800
FGASSETFPI TFGDFNEGEI ESLSSELLTF GDFLPGEVDD LTDSDWSTCS DTDDELLDRA    1860
GGYIFSSDTG PGHLQQKSVR QSVLPVNTLE EVHEEKCYPP KLDEAKEQLL LKKLQESASM    1920
ANRSRYQSRK VENMKAAIIQ RLKRGCRLYL MSETPKVPTY RTTYPAPVYS PPINVRLSNP    1980
ESAVAACNEF LARNYPTVSS YQITDEYDAY LDMVDGSESC LDRATFNPSK LRSYPKQHAY    2040
HAPSIRSAVP SPFQNTLQNV LAAATKRNCN VTQMRELPTL DSAVFNVECF KKFACNQEYW    2100
EEFAASPIRI TTENLATYVT KLKGPKAAAL FAKTHNLLPL QEVPMDRFTV DMKRDVKVTP    2160
GTKHTEERPK VQVIQAAEPL ATAYLCGIHR ELVRRLNAVL LPNVHTLFDM SAEDFDAIIA    2220
AHFKPGDTVL ETDIASFDKS QDDSLALTAL MLLEDLGVDH SLLDLIEAAF GEISSCHLPT    2280
GTRFKFGAMM KSGMFLTLFV NTLLNITIAS RVLEDRLTKS ACAAFIGDDN IIHGVVSDEL    2340
MAARCATWMN MEVKIIDAVV SLKAPYFCGG FILHDTVTGT ACRVADPLKR LFKLGKPLAA    2400
GDEQDEDRRR ALADEVIRWQ RTGLIDELEK AVYSRYEVQG ISVVVMSMAT FASSRSNFEK    2460
LRGPVITLYG GPK                                                       2473
```

FIGURE 47

```
MDPVYVDIDA DSAFLKALQR AYPMFEVEPR QVTPNDHANA RAFSHLAIKL IEQEIDPDST      60
ILDIGSAPAR RMMSDRKYHC VCPMRSAEDP ERLANYARKL ASAAGKVLDR NISGKIGDLQ     120
AVMAVPDTET PTFCLHTDVS CRQRADVAIY QDVYAVHAPT SLYHQAIKGV RVAYWVGFDT     180
TPFMYNAMAG AYPSYSTNWA DEQVLKAKNI GLCSTDLTEG RRGKLSIMRG KKLKPCDRVL     240
FSVGSTLYPE SRKLLKSWHL PSVFHLKGKL SFTCRCDTVV SCEGYVVKRI TMSPGLYGKT     300
TGYAVTHHAD GFLMCKTTDT VDGERVSFSV CTYVPATICD QMTGILATEV TPEDAQKLLV     360
GLNQRIVVNG RTQRNTNTMK NYLLPVVAQA FSKWAKECRK DMEDEKLLGV RERTLTCCCL     420
WAFKKQKTHT VYKRPDTQSI QKVQAEFDSF VVPSLWSSGL SIPLRTRIKW LLSKVPKTDL     480
IPYSGDAREA RDAEKEAEEE REAELTREAL PPLQAAQEDV QVEIDVEQLE DRAGAGIIET     540
PRGAIKVTAQ PTDHVVGEYL VLSPQTVLRS QKLSLIHALA EQVKTCTHNG RAGRYAVEAY     600
DGRVLVPSGY AISPEDFQSL SESATMVYNE REFVNRKLHH IAMHGPALNT DEESYELVRA     660
ERTEHEYVYD VDQRRCCKKE EAAGLVLVGD LTNPPYHEFA YEGLKIRPAC PYKIAVIGVF     720
GVPGSGKSAI IKNLVTRQDL VTSGKKENCQ EITTDVMRQR GLEISARTVD SLLLNGCNRP     780
VDVLYVDEAF ACHSGTLLAL IALVRPRQKV VLCGDPKQCG FFNMMQMKVN YNHNICTQVY     840
HKSISRRCTL PVTAIVSSLH YEGKMRTTNE YNKPIVVDTT GSTKPDPGDL VLTCFRGWVK     900
QLQIDYRGYE VMTAAASQGL TRKGVYAVRQ KVNENPLYAS TSEHVNVLLT RTEGKLVWKT     960
LSGDPWIKTL QNPPKGNFKA TIKEWEVEHA SIMAGICSHQ MTFDTFQNKA NVCWAKSLVP    1020
ILETAGIKLN DRQWSQIIQA FKEDKAYSPE VALNEICTRM YGVDLDSGLF SKPLVSVYYA    1080
DNHWDNRPGG KMFGFNPEAA SILERKYPFT KGKWNINKQI CVTTRRIEDF NPTTNIIPAN    1140
RRLPHSLVAE HRPVKGERME WLVNKINGHH VLLVSGYNLA LPTKRVTWVA PLGVRGADYT    1200
YNLELGLPAT LGRYDLVVIN IHTPFRIHHY QQCVDHAMKL QMLGGDSLRL LKPGGSLLIR    1260
AYGYADRTSE RVICVLGRKF RSSRALKPPC VTSNTEMFFL FSNFDNGRRN FTTHVMNNQL    1320
NAAFVGQVTR AGCAPSYRVK RMDIAKNDEE CVVNAANPRG LPGDGVCKAV YKKWPESFKN    1380
SATPVGTAKT VMCGTYPVIH AVGPNFSNYS ESEGDRELAA AYREVAKEVT RLGVNSVAIP    1440
LLSTGVYSGG KDRLTQSLNH LFTAMDSTDA DVVIYCRDKE WEKKISEAIQ MRTQVELLDE    1500
HISIDCDIVR VHPDSSLAGR KGYSTTEGAL YSYLEGTRFH QTAVDMAEIH TMWPKQTEAN    1560
EQVCLYALGE SIESIRQKCP VDDADASSPP KTVPCLCRYA MTPERVTRLR MNHVTSIIVC    1620
SSFPLPKYKI EGVQKVKCSK VMLFDHNVPS RVSPREYRSS QESAQEASTI TSLTHSQFDL    1680
SVDGEILPVP SDLDADAPAL EPALDDGATH TLPSTTGNLA AVSDWVMSTV PVAPPRRRRG    1740
RNLTVTCDER EGNITPMASV RFFRAELCPV VQETAETRDT AMSLQAPPST ATEPNHPPIS    1800
FGASSETFPI TFGDFNEGEI ESLSSELLTF GDFLPGEVDD LTDSDWSTCS DTDDELLDRA    1860
GGYIFSSDTG PGHLQQKSVR QSVLPVNTLE EVHEEKCYPP KLDEAKEQLL LKKLQESASM    1920
ANRSRYQSRK VENMKAAIIQ RLKRGCRLYL MSETPKVPTY RTTYPAPVYS PPINVRLSNP    1980
ESAVAACNEF LARNYPTVSS YQITDEYDAY LDMVDGSESC LDRATFNPSK LRSYPKQHAY    2040
HAPSIRSAVP SPFQNTLQNV LAAATKRNCN VTQMRELPTL DSAVFNVECF KKFACNQEYW    2100
EEFAASPIRI TTENLATYVT KLKGPKAAAL FAKTHNLLPL QEVPMDRFTV DMKRDVKVTP    2160
GTKHTEERPK VQVIQAAEPL ATAYLCGIHR ELVRRLNAVL LPNVHTLFDM SAEDFDAIIA    2220
AHFKPGDTVL ETDIASFDKS QDDSLALTAL MLLEDLGVDH SLLDLIEAAF GEISSCHLPT    2280
GTRFKFGAMM KSGMFLTLFV NTLLNITIAS RVLEDRLTKS ACAAFIGDDN IIHGVVSDEL    2340
MAARCATWMN MEVKIIDAVV SLKAPYFCGG FILHDTVTGT ACRVADPLKR LFKLGKPLAA    2400
GDEQDEDRRR ALADEVIRWQ RTGLIDELEK AVYSRYEVQG ISVVVMSMAT FASSRSNFEK    2460
LRGPVITLYG GPK                                                       2473
```

FIGURE 48

```
MDPVYVDIDA DSAFLKALQR AYPMFEVEPR QVTPNDHANA RAFSHLAIKL IEQEIDPDST      60
ILDIGSAPAR RMMSDRKYHC VCPMRSAEDP ERLANYARKL ASAAGKVLDR NISGKIGDLQ     120
AVMAVPDTET PTFCLHTDVS CRQRADVAIY QDVYAVHAPT SLYHQAIKGV RVAYWVGFDT     180
TPFMYNAMAG AYPSYSTNWA DEQVLKAKNI GLCSTDLTEG RRGKLSIMRG KKLKPCDRVL     240
FSVGSTLYPE SRKLLKSWHL PSVFHLKGKL SFTCRCDTVV SCEGYVVKRI TMSPGLYGKT     300
IGYAVTHHAD GFLMCKTTDT VDGERVSFSV CTYVPATICD QMTGILATEV TPEDAQKLLV     360
GLNQRIVVNG RTQRNTNTMK NYLLPVVAQA FSKWAKECRK DMEDEKLLGV RERTLTCCCL     420
WAFKKQKTHT VYKRPDTQSI QKVQAEFDSF VVPSLWSSGL SIPLRTRIKW LLSKVPKTDL     480
IPYSGDAREA RDAEKEAEEE REAELTREAL PPLQAAQEDV QVEIDVEQLE DRAGAGIIET     540
PRGAIKVTAQ PTDHVVGEYL VLSPQTVLRS QKLSLIHALA EQVKTCTHNG RAGRYAVEAY     600
DGRVLVPSGY AISPEDFQSL SESATMVYNE REFVNRKLHH IAMHGPALNT DEESYELVRA     660
ERTEHEYVYD VDQRRCCKKE EAAGLVLVGD LTNPPYHEFA YEGLKIRPAC PYKIAVIGVF     720
GVPGSGKSAI IKNLVTRQDL VTSGKKENCQ EITTDVMRQR GLEISARTVD SLLLNGCNRP     780
VDVLYVDEAF ACHSGTLLAL IALVRPRQKV VLCGDPKQCG FFNMMQMKVN YNHNICTQVY     840
HKSISRRCTL PVTAIVSSLH YEGKMRTTNE YNKPIVVDTT GSTKPDPGDL VLTCFRGWVK     900
QLQIDYRGYE VMTAAASQGL TRKGVYAVRQ KVNENPLYAS TSEHVNVLLT RTEGKLVWKT     960
LSGDPWIKTL QNPPKGNFKA TIKEWEVEHA SIMAGICSHQ MTFDTFQNKA NVCWAKSLVP    1020
ILETAGIKLN DRQWSQIIQA FKEDKAYSPE VALNEICTRM YGVDLDSGLF SKPLVSVYYA    1080
DNHWDNRPGG KMFGFNPEAA SILERKYPFT KGKWNINKQI CVTTRRIEDF NPTTNIIPAN    1140
RRLPHSLVAE HRPVKGERME WLVNKINGHH VLLVSGNNLA LPTKRVTWVA PLGVRGADYT    1200
YNLELGLPAT LGRYDLVVIN IHTPFRIHHY QQCVDHAMKL QMLGGDSLRL LKPGGSLLIR    1260
AYGYADRTSE RVICVLGRKF RSSRALKPPC VTSNTEMFFL FSNFDNGRRN FTTHVMNNQL    1320
NAAFVGQVTR AGCAPSYRVK RMDIAKNDEE CVVNAANPRG LPGDGVCKAV YKKWPESFKN    1380
SATPVGTAKT VMCGTYPVIH AVGPNFSNYS ESEGDRELAA AYREVAKEVT RLGVNSVAIP    1440
LLSTGVYSGG KDRLTQSLNH LFTAMDSTDA DVVIYCRDKE WEKKISEAIQ MRTQVELLDE    1500
HISIDCDIVR VHPDSSLAGR KGYSTTEGAL YSYLEGTRFH QTAVDMAEIH TMWPKQTEAN    1560
EQVCLYALGE SIESIRQKCP VDDADASSPP KTVPCLCRYA MTPERVTRLR MNHVTSIIVC    1620
SSFPLPKYKI EGVQKVKCSK VMLFDHNVPS RVSPREYRSS QESAQEASTI TSLTHSQFDL    1680
SVDGEILPVP SDLDADAPAL EPALDDGATH TLPSTTGNLA AVSDWVMSTV PVAPPRRRRG    1740
RNLTVTCDER EGNITPMASV RFFRAELCPV VQETAETRDT AMSLQAPPST ATPNHPPISF    1800
GASSETFPIT FGDFNEGEIE SLSSELLTFG DFLPGEVDDL TDSDWSTCSD TDDELLDRAG    1860
GYIFSSDTGP GHLQQKSVRQ SVLPVNTLEE VHEEKCYPPK LDEAKEQLLL KKLQESASMA    1920
NRSRYQSRKV ENMKAAIIQR LKRGCRLYLM SETPKVPTYR TTYPAPVYSP PINVRLSNPE    1980
SAVAACNEFL ARNYPTVSSY QITDEYDAYL DMVDGSESCL DRATFNPSKL RSYPKQHAYH    2040
APSIRSAVPS PFQNTLQNVL AAATKRNCNV TQMRELPTLD SAVFNVECFK KFACNQEYWE    2100
EFAASPIRIT TENLATYVTK LKGPKAAALF AKTHNLLPLQ EVPMDRFTVD MKRDVKVTPG    2160
TKHTEERPKV QVIQAAEPLA TAYLCGIHRE LVRRLNAVLL PNVHTLFDMS AEDFDAIIAA    2220
HFKPGDTVLE TDIASFDKSQ DDSLALTALM LLEDLGVDHS LLDLIEAAFG EISSCHLPTG    2280
TRFKFGAMMK SGMFLTLFVN TLLNITIASR VLEDRLTKSA CAAFIGDDNI IHGVVSDELM    2340
AARCATWMNM EVKIIDAVVS LKAPYFCGGF ILHDTVTGTA CRVADPLKRL FKLGKPLAAG    2400
DEQDEDRRRA LADEVIRWQR TGLIDELEKA VYSRYEVQGI SVVVMSMATF ASSRSNFEKL    2460
RGPVITLYGG PK                                                        2472
```

FIGURE 49

```
MFEVEPRQVT PNDHANARAF SHLAIKLIEQ EIDPDSTILD IGSAPARRMM SDRKYHCVCP      60
MRSAEDPERL ANYARKLASA AGKVLDRNIS GKIGDLQAVM AVPDTETPTF CLHTDVSCRQ     120
RADVAIYQDV YAVHAPTSLY HQAIKGVRVA YWVGFDTTPF MYNAMAGAYP SYSTNWADEQ     180
VLKAKNIGLC STDLTEGRRG KLSIMRGKKL KPCDRVLFSV GSTLYPESRK LLKSWHLPSV     240
FHLKGKLSFT CRCDTVVSCE GYVVKRITMS PGLYGKTTGY AVTHHADGFL MCKTTDTVDG     300
ERVSFSVCTY VPATICDQMT GILATEVTPE DAQKLLVGLN QRIVVNGRTQ RNTNTMKNYL     360
LPVVAQAFSK WAKECRKDME DEKLLGVRER TLTCCCLWAF KKQKTHTVYK RPDTQSIQKV     420
QAEFDSFVVP SLWSSGLSIP LRTRIKWLLS KVPKTDLIPY SGDAREARDA EKEAEEEREA     480
ELTREALPPL QAAQEDVQVE IDVEQLEDRA GAGIIETPRG AIKVTAQPTD HVVGEYLVLS     540
PQTVLRSQKL SLIHALAEQV KTCTHNGRAG RYAVEAYDGR VLVPSGYAIS PEDFQSLSES     600
ATMVYNEREF VNRKLHHIAM HGPALNTDEE SYELVRAERT EHEYVYDVDQ RRCCKKEEAA     660
GLVLVGDLTN PPYHEFAYEG LKIRPACPYK IAVIGVFGVP GSGKSAIIKN LVTRQDLVTS     720
GKKENCQEIT TDVMRQRGLE ISARTVDSLL LNGCNRPVDV LYVDEAFACH SGTLLALIAL     780
VRPRQKVVLC GDPKQCGFFN MMQMKVNYNH NICTQVYHKS ISRRCTLPVT AIVSSLHYEG     840
KMRTTNEYNK PIVVDTTGST KPDPGDLVLT CFRGWVKQLQ IDYRGYEVMT AAASQGLTRK     900
GVYAVRQKVN ENPLYASTSE HVNVLLTRTE GKLVWKTLSG DPWIKTLQNP PKGNFKATIK     960
EWEVEHASIM AGICSHQMTF DTFQNKANVC WAKSLVPILE TAGIKLNDRQ WSQIIQAFKE    1020
DKAYSPEVAL NEICTRMYGV DLDSGLFSKP LVSVYYADNH WDNRPGGKMF GFNPEAASIL    1080
ERKYPFTKGK WNINKQICVT TRRIEDFNPT TNIIPANRRL PHSLVAEHRP VKGERMEWLV    1140
NKINGHHVLL VSGYNLALPT KRVTWVAPLG VRGADYTYNL ELGLPATLGR YDLVVINIHT    1200
PFRIHHYQQC VDHAMKLQML GGDSLRLLKP GGSLLIRAYG YADRTSERVI CVLGRKFRSS    1260
RALKPPCVTS NTEMFFLFSN FDNGRRNFTT HVMNNQLNAA FVGQVTRAGC APSYRVKRMD    1320
IAKNDEECVV NAANPRGLPG DGVCKAVYKK WPESFKNSAT PVGTAKTVMC GTYPVIHAVG    1380
PNFSNYSESE GDRELAAAYR EVAKEVTRLG VNSVAIPLLS TGVYSGGKDR LTQSLNHLFT    1440
AMDSTDADVV IYCRDKEWEK KISEAIQMRT QVELLDEHIS IDCDIVRVHP DSSLAGRKGY    1500
STTEGALYSY LEGTRFHQTA VDMAEIHTMW PKQTEANEQV CLYALGESIE SIRQKCPVDD    1560
ADASSPPKTV PCLCRYAMTP ERVTRLRMNH VTSIIVCSSF PLPKYKIEGV QKVKCSKVML    1620
FDHNVPSRVS PREYRSSQES AQEASTITSL THSQFDLSVD GEILPVPSDL DADAPALEPA    1680
LDDGATHTLP STTGNLAAVS DWVMSTVPVA PPRRRRGRNL TVTCDEREGN ITPMASVRFF    1740
RAELCPVVQE TAETRDTAMS LQAPPSTATE PNHPPISFGA SSETFPITFG DFNEGEIESL    1800
SSELLTFGDF LPGEVDDLTD SDWSTCSDTD DELLDRAGGY IFSSDTGPGH LQQKSVRQSV    1860
LPVNTLEEVH EEKCYPPKLD EAKEQLLLKK LQESASMANR SRYQSRKVEN MKAAIIQRLK    1920
RGCRLYLMSE TPKVPTYRTT YPAPVYSPPI NVRLSNPESA VAACNEFLAR NYPTVSSYQI    1980
TDEYDAYLDM VDGSESCLDR ATFNPSKLRS YPKQHAYHAP SIRSAVPSPF QNTLQNVLAA    2040
ATKRNCNVTQ MRELPTLDSA VFNVECFKKF ACNQEYWEEF AASPIRITTE NLATYVTKLK    2100
GPKAAALFAK THNLLPLQEV PMDRFTVDMK RDVKVTPGTK HTEERPKVQV IQAAEPLATA    2160
YLCGIHRELV RRLNAVLLPN VHTLFDMSAE DFDAIIAAHF KPGDTVLETD IASFDKSQDD    2220
SLALTALMLL EDLGVDHSLL DLIEAAFGEI SSCHLPTGTR FKFGAMMKSG MFLTLFVNTL    2280
LNITIASRVL EDRLTKSACA AFIGDDNIIH GVVSDELMAA RCATWMNMEV KIIDAVVSLK    2340
APYFCGGFIL HDTVTGTACR VADPLKRLFK LGKPLAAGDE QDEDRRRALA DEVIRWQRTG    2400
LIDELEKAVY SRYEVQGISV VVMSMATFAS SRSNFEKLRG PVITLYGGPK               2450
```

FIGURE 50

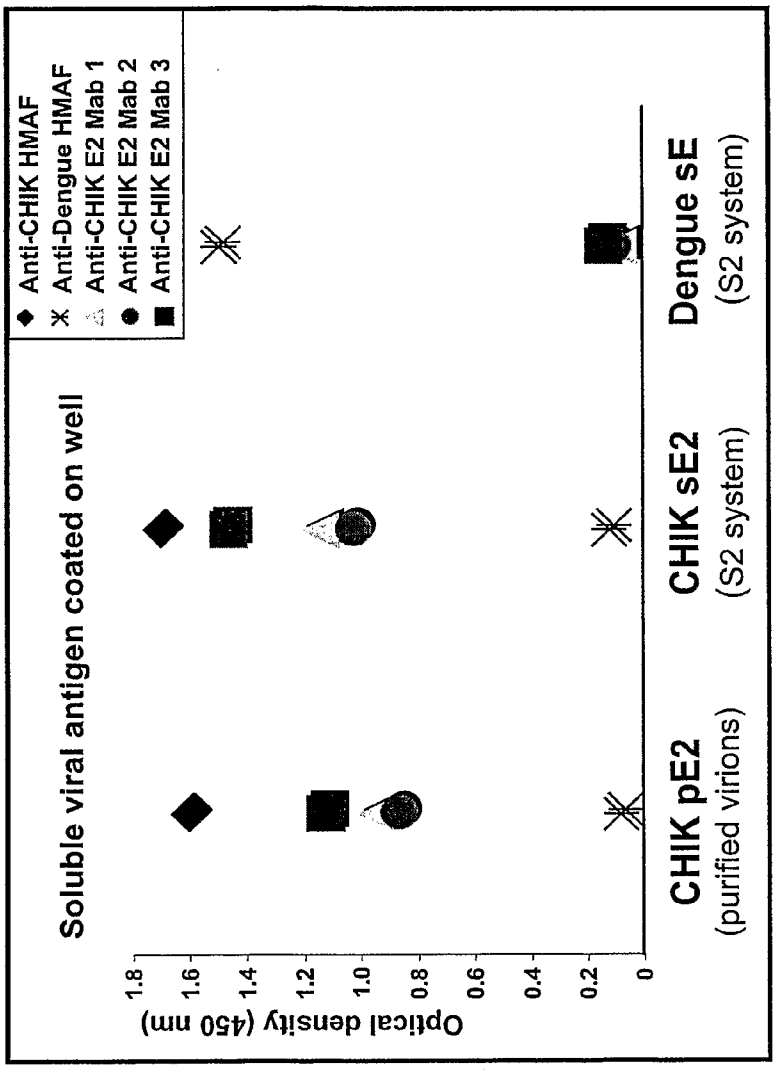

Anti-CHIK HMAF
Anti-Dengue HMAF
Anti-CHIK E2 Mab 1
Anti-CHIK E2 Mab 2
Anti-CHIK E2 Mab 3
Soluble viral antigen coated on well
Optical density (450 nm)
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
CHIK pE2
(purified virions)
CHIK sE2
(S2 system)
Dengue sE
(S2 system)

FIGURE 51

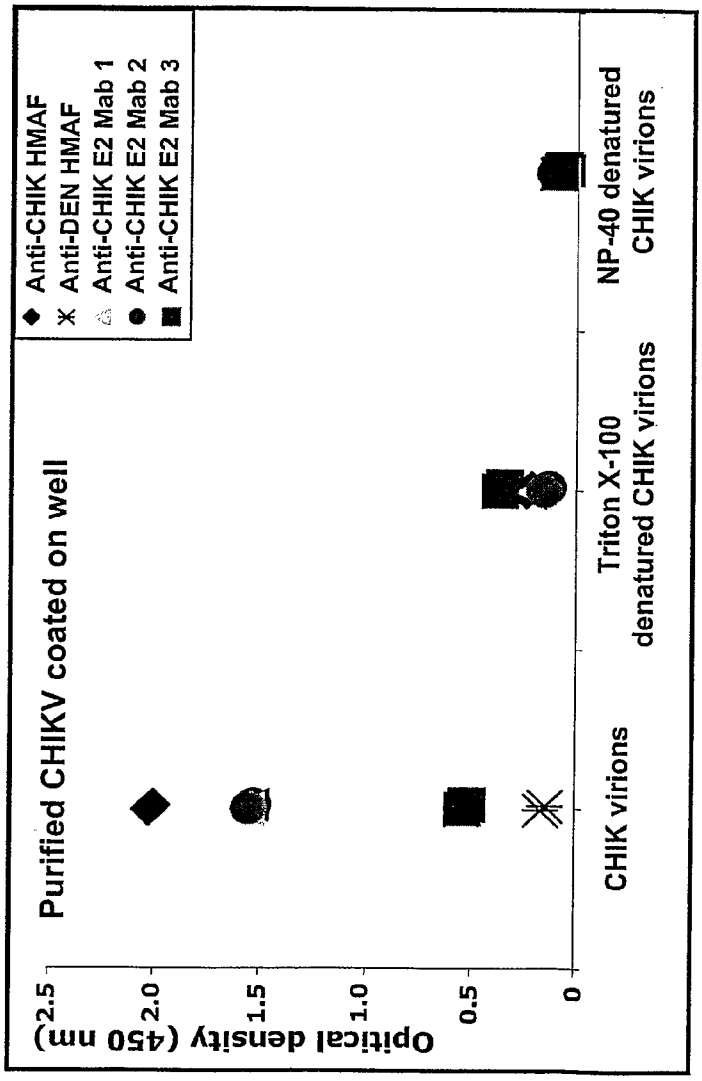

Anti-CHIK HMAF
Anti-DEN HMAF
Anti-CHIK E2 Mab 1
Anti-CHIK E2 Mab 2
Anti-CHIK E2 Mab 3
Purified CHIKV coated on well
Optical density (450 nm)
2.5
2.0
1.5
1.0
0.5
0
CHIK virions
Triton X-100 denatured CHIK virions
NP-40 denatured CHIK virions

FIGURE 52

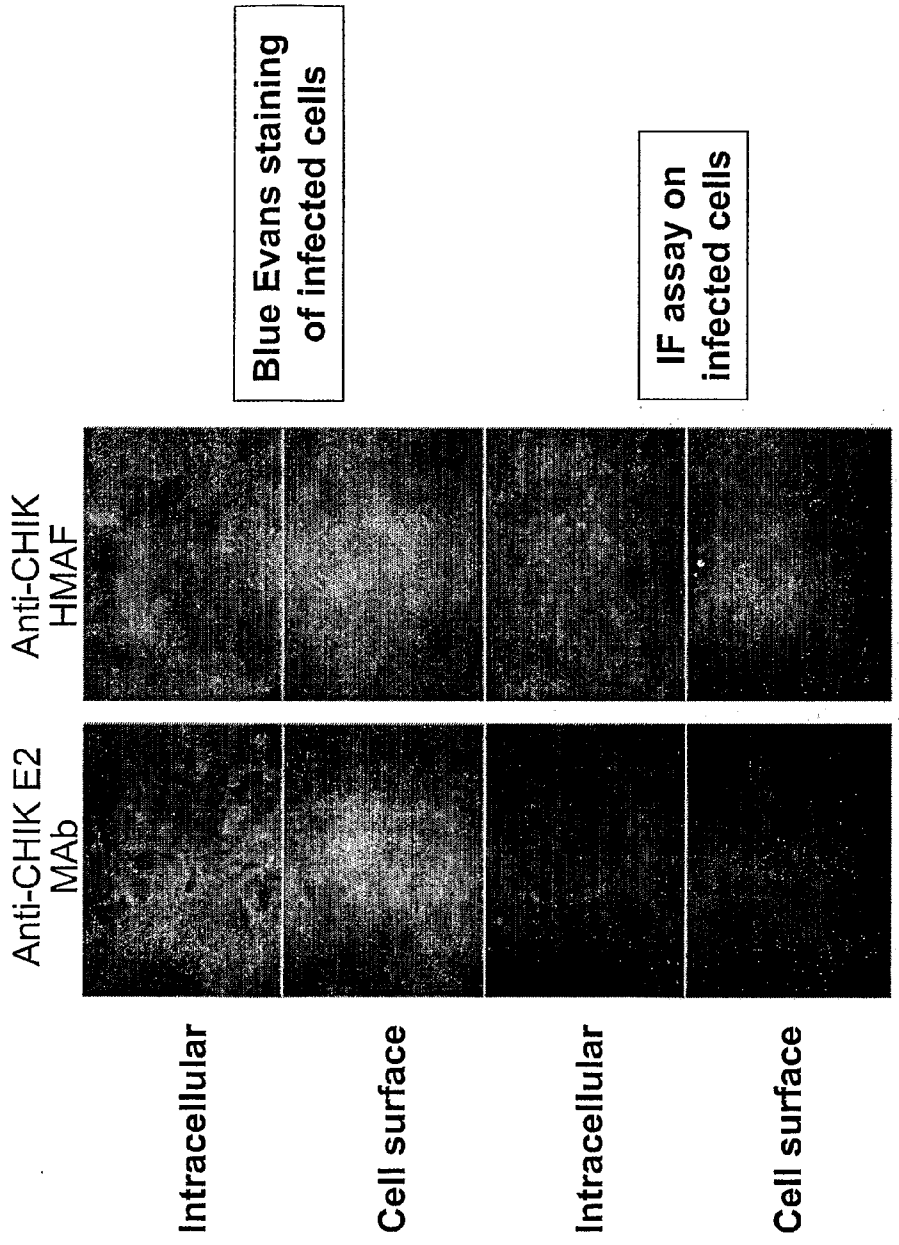

Blue Evans staining of infected cells
IF assay on infected cells
Anti-CHIK HMAF
Anti-CHIK E2 MAb
Intracellular
Cell surface
Intracellular
Cell surface

FIGURE 53

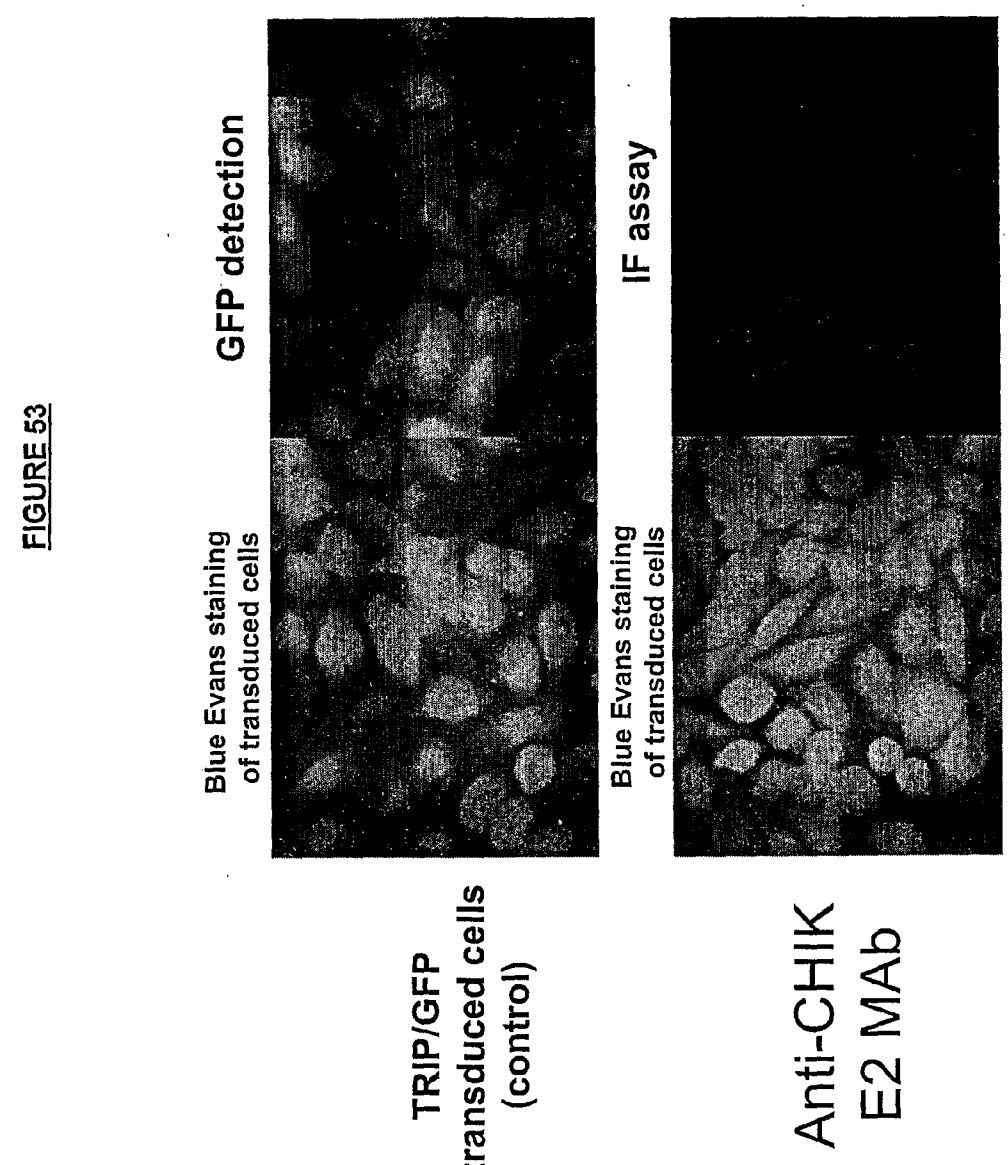

GFP detection
IF assay
Blue Evans staining of transduced cells
Blue Evans staining of transduced cells
TRIP/GFP -transduced cells (control)
Anti-CHIK E2 MAb

FIGURE 54

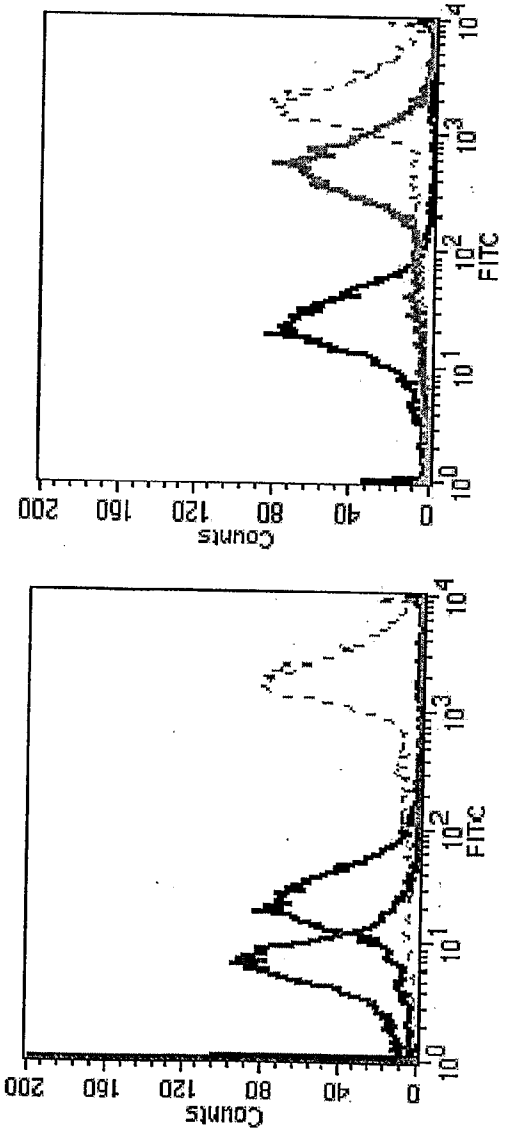
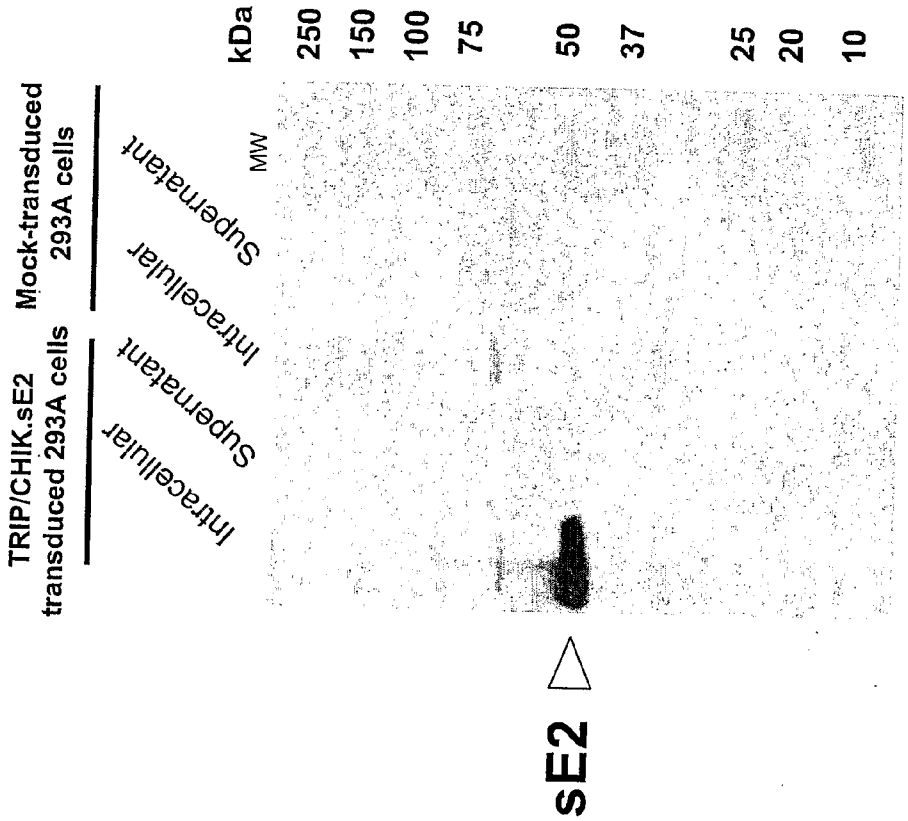

Key
Name
Anti-dengue HMAF
Anti-CHIK HMAF
Secondary antibody only
Anti-CHIK E2 Mab3
Counts
FITC

FIGURE 55


kDa
250
150
100
75
50
37
25
20
10
MW
Supernatant
Intracellular
Supernatant
Intracellular
Mock-transduced 293A cells
TRIP/CHIK.sE2 transduced 293A cells
sE2