## Title

Functional characterization of the HIV genome by genetic footprinting

## Author

Laurent, Louise Chang

## Publication Date

1998

# Functional Characterization of the HIV Genome by Genetic Footprinting

by

Louise Chang Laurent

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA SAN FRANCISCO

[redacted]

Date                                                            University Librarian

Degree Conferred: . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

To Marc

# Acknowledgements

Many thanks are due to the many members of the Brown lab, past and present. Foremost among these are fellow genetic footprinters Rachel Adams Crowley and Mari Olsen. Working with Rachel has made the struggle through graduate school a much more rewarding (and yes, even enjoyable) experience. I have benefited from her many abilities and personal strengths as we wrestled with the myriad difficulties, technical and ideological, involved in genetic footprinting. I will never be able to express how much I appreciate all of the help Mari has provided over the last few, very busy, years. I must thank Brian Scottoline and Liz Wilson for sharing their benches with me when I was orphaned by the Varmus lab. Thanks to Pat Brown, who has provided a stimulating environment in which to learn and taught me to true meaning of independence. Thanks to the other "virus people," including Richard.

Many thanks are due to the many mentors I have found in the scientific world. Thanks to Mrs. McPherson, my ninth-grade chemistry teacher. Thanks to Mr. Hozinsky, my tenth-grade biology teacher and the one who suggested that I roam the halls at the University of Chicago asking for someone to teach me about lab stuff. Thanks to Vikas Sukhatme, who took me into his lab and accepted the challenge of trying to teach a high school student molecular biology. Thanks to Jeremy Knowles and Betsy Komives, who were my mentors during my undergraduate years. Thanks to Mike Bishop and Harold Varmus for

convincing me to come to UCSF for my graduate training and playing large parts in that experience.

Many thanks are due to my friends. Here I tread lightly, as I am sure to leave someone out. I will restrict myself to thanking the members of my MSTP class and their spouses for making the past several years more than just school.

Many thanks are due to Jana and Sue, who have made life incredibly easy.

Many thanks are due to the many members of my family. My husband, Marc, has provided much support, moral and computeral. My daughter, Clara, has made the last eight months a joy and a challenge. This challenge has been greatly decreased by help received from our friend and guardian angel, Agnes Leturgie, and advice received from my elder sister and mother of four, Cindy. My family-in-law, and especially my mother-in-law, has been steadfastly supportive. And finally, and most importantly, infinite thanks to my parents.

# Abstract

## Functional Characterization of the HIV Genome by Genetic Footprinting

In this report, I present a detailed analysis of the functional characteristics of the 1000 nucleotides at the 5' end of the HIV RNA genome. The effects of one hundred and thirty-four independent insertions mutations were examined in a quantitative manner at three points in the viral replication cycle. I studied the abilities of mutants 1) to make stable viral RNA, 2) to assemble and release viral-RNA-containing viral particles, 3) to enter host cells, complete reverse transcription, enter the nuclei of host cells, and generate proviruses in the host genome by integration. In order to carry out a thorough investigation on a large number of mutations, a modification of the genetic footprinting technique was employed. Using this method, all of the mutants were constructed and analyzed en masse, greatly decreasing the labor typically involved in mutagenesis studies. The presence of several functional features previously assigned to the region of the HIV genome under investigation was confirmed, and evidence for a number of novel features was found. Among these new features were cis-acting sequences that appeared to contribute to formation of stable viral transcripts, viral RNA packaging, or an early step in viral replication. These sequences were distinct from previously identified sequences that have been shown to be important for these steps in the viral life cycle. An unanticipated trans-acting role for sequences near the N-terminus of matrix in the formation of stable viral RNA transcripts was also seen. Finally, in contrast to previous reports, the results of

this study suggested that mutations detrimental to viral replication in sequences encoding the matrix and capsid proteins principally interfered with assembly.

# Table of Contents

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

# List of Figures

# Chapter 1

## General Overview of Retrovirology

The first reports describing the disease that is now called AIDS (acquired immunodeficiency syndrome) were made in 1981 (Gottlieb et al. 1981; Masur et al. 1981; Siegal et al. 1981). The afflicted patients displayed an unusual series of symptoms and findings. They were all young, previously healthy men, a disproportionate number of whom were homosexual and/or substance abusers. They suffered from fulminant herpes simplex, candida, and cytomegalovirus infections and pneumocystis carinnii pneumonia. At the time, these types of infections were very rare and found only in severely immunocompromised persons, such as premature infants and transplant patients undergoing immunosuppressive therapy. In the subsequent months and years, many more patients with the same constellation of findings were described. The pathogenic agent responsible for this epidemic was isolated in 1984 (Popovic et al. 1984) and in very short order the genome of this pathogen, the human immunodeficiency virus (HIV), was cloned and sequenced (Shaw et al. 1984). In the intervening years, HIV has been the focus of much attention and progress in understanding HIV has been rapid on many fronts.

HIV is a retrovirus, a class of viruses previously called RNA tumor viruses. This older and somewhat inaccurate nomenclature refers to two characteristics of retroviruses: the genomes of these diploid viruses consist of two molecules of linear single-stranded RNA; and many (though not all) retroviruses are associated with neoplasms. The current appellation “retrovirus” is the result of the discovery of reverse transcriptase in 1970 (Baltimore 1970, Temin and Mizutani 1970). All retroviruses encode a reverse transcriptase, or RNA-directed DNA polymerase, that converts the single-stranded RNA viral genome found in virions into a double-stranded DNA form. Another virally-encoded enzyme, integrase, then integrates this linear DNA into the genome of a host cell, where it is called a provirus.

The rapid progress in understanding the biology of HIV was largely due to the work done on other retroviruses in the seventy or so years before AIDS was described. During the first decade of the twentieth century, two studies on tumors in chickens led to the discovery of the avian sarcoma/leukosis viruses (ASLV) (Ellerman and Bang 1908, Rous 1911). From the time of these initial reports until the 1960's, many more retroviruses were discovered and described in terms of their host range and the natural histories of the diseases they caused. Due to advances in biochemical, structural, cell culture, and molecular techniques, the emphasis in retrovirology in recent decades has been on cellular pathogenesis and a molecular description of retroviruses and retroviral replication. By the time HIV was isolated and shown to be the pathogenic agent in AIDS, the major protein components of retroviruses, their basic replication cycle, the general structure of the retroviral genome, and the nucleotide sequences of certain specific retroviruses had been described. Since the discovery of HIV, the field of retrovirology has expanded and progress has accelerated even more.

Mutants have been very useful tools in the study of retroviruses. With the advent of molecular biology, directed mutations have been made and screens of random mutations have been done to determine the functions of various parts of the retroviral genome. For HIV, the combination of knowing the nucleotide sequence of the genome and having structural information on most of the proteins makes it very interesting to have a detailed description of the phenotypes of mutants throughout the genome. Ideally, one would be able to study the effects of different types of mutations (point mutations, deletions, insertions, replacements) at every position in the genome on different points in the replication cycle (transcription, translation, packaging, budding, host-cell attachment, entry, uncoating, reverse transcription, nuclear entry, integration). Making individual mutants and studying their phenotypes one by one is now possible, but not experimentally feasible. Therefore, we have developed a method to study large numbers of mutants in parallel. This technique is restricted to two classes of mutations (insertions and replacements), but allows the collection of quantitative information on the effects of many mutations on several steps of the viral replication cycle in a highly parallel manner.

## Overview of Mutagenesis

Mutagenesis is a versatile and powerful tool in studying the function of nucleic acids. Mutagenesis can be performed either in vivo or in vitro, on a small piece of cloned DNA or on the intact genome of an organism, randomly or in a directed fashion. There are three types of mutagens in common use: chemicals (e.g. alkylating agents), radiation (e.g. X-rays or UV radiation), and enzymes (e.g. Taq polymerase in PCR mutagenesis or transposases in transposon-mediated mutagenesis). Standard methods of random mutagenesis involve subjecting the nucleic acid of interest to mutagenesis and either selecting/screening the resulting mutant population for a particular phenotype (e.g. resistance to an antibiotic) or isolating mutant clones and testing the characteristics of individual mutants one at a time (e.g. rate of replication). Directed mutagenesis involves constructing and testing mutants one at a time. These two basic strategies are useful in certain circumstances: the isolation of mutants where one has a good positive selection and the examination of a limited number of individual mutants. However, these techniques quickly become tedious if the goal is to quantitatively determine the behavior of large numbers of mutants. Recently, several methods aimed at assessing large numbers of mutants in parallel have been reported, including signature-tagged transposon mutagenesis (Hensel et al. 1995), genome-scale genetic footprinting (Smith et al. 1996), and high-resolution genetic footprinting (Singh et al. 1997).

The objective of this study was to obtain a detailed functional map of a portion of the HIV genome by examining between 100 and 200 insertional mutations distributed over a 1000 nucleotide region, which encompassed the 5'-LTR (Long Terminal Repeat), the p17 (matrix, MA) gene, and p24 (capsid, CA) gene of HIV. Each mutant was to be mapped at single-nucleotide resolution and quantitatively assessed for its affect on viral replication. Given these criteria, the high-resolution genetic footprinting technique was the method of choice.

## High-resolution Genetic Footprinting

High resolution genetic footprinting was developed as a method to make and gather quantitative information on large numbers of mutants en masse. The basic concept consists of constructing a library of insertion or replacement mutants, where the different mutants contain the same insertion or replacement sequence, differing only in the position of the mutation. The idea for using an integrase or transposase enzyme to make the mutations and analyzing the population of mutants by PCR originated from a paper by Pryciak and Varmus (Pryciak and Varmus 1992). These authors were actually studying the effect of DNA conformation on target site preferences of retroviral integrases. However, their work showed that large numbers of integration events could be tracked in parallel and mapped to single-nucleotide resolution.

Moloney Murine Leukemia Virus integrase can be used to integrate short double-stranded oligonucleotides in a concerted fashion into a circular double-stranded DNA target in vitro. In this concerted reaction, the terminal two nucleotides of the upper strands of two double-stranded oligonucleotides are clipped off, leaving two-nucleotide 5' overhangs. The newly exposed terminal 3' hydroxyl groups of these oligonucleotides are then used to attack 5' phosphates in the target DNA staggered by 4 base-pairs, producing a linear target DNA with an oligonucleotide covalently joined to each end. There are 4-nucleotide gaps in the target DNA and an extra 2-nucleotide 5' extension on the oligonucleotides at each end. Both of these features can be eliminated by doing a run-off reaction using a DNA polymerase such as Taq polymerase. MLV integrase is relatively insensitive to the sequence of the target DNA, resulting in integration events at many different sites.

In 1997, Singh, Crowley, and Brown demonstrated the utility of MLV integrase as a tool for genetic footprinting in high-resolution functional mapping of the SupF gene, which encodes an amber suppressor tRNA. The oligonucleotide used for integration contained three types of sequences: a viral end sequence that allowed MLV integrase to recognize the oligonucleotide as a substrate; a Bsg I restriction enzyme site; and a Not I restriction enzyme site. Insertion mutants were made by digesting the products of the concerted integration reaction with Not I, creating complementary cohesive ends, and recircularizing the target DNA by ligation. The resulting insertions included a 4-base-pair duplication in the target DNA and a central Not I site. Replacement mutants were generated by digesting with Bsg I, a type IIs restriction enzyme that cuts 16/14 nucleotides away from its recognition sequence, allowing cleavage within the target DNA sequence. The 12 base-pairs deleted from the target DNA sequence in this way were replaced by ligating in a 12-base-pair oligonucleotide containing an Nde I site. Both the insertion and replacement libraries were subjected to a selection that required the function of the SupF gene. The libraries before and after selection were analyzed and compared using PCR-based assays. To analyze the insertion library samples, PCR was performed using one oligonucleotide primer complementary to the sequence of the insert oligonucleotide and a second, $^{32}$P-labelled, primer complementary to a fixed position in the target DNA. Each mutant in the library gave a product of unique size that depended on the position of the insertion. Since the library consisted of mutants at many different positions, subjecting the products of the PCR reactions to electrophoresis through a denaturing polyacrylamide gel resulted in a ladder of bands. Bands that represented clones defective in SupF function were present in the pre-selection library and absent in the post-selection library, giving a functional footprint of the SupF gene. The oligonucleotide used for the replacement library was too short for efficient priming for PCR. Therefore, an alternative PCR strategy (which I will refer to as the "flanking PCR/restriction digestion" method) was designed to analyze the replacement library samples. Two fixed-position primers to target DNA sequences were used, one of which was labelled with $^{32}$P. The PCR products were digested with Nde I, which cleaved within the replacement sequence, yielding a unique-sized radioactively-labeled product for each mutant, the size of which again depended on the position of the replacement.

## Overview of the Genetic Footprinting Aspect of the Current system

Several modifications to the method reported by Singh, Crowley, and Brown (1997) were required to adapt it for the study of the HIV genome. Most significantly, the enzyme used for mutagenesis was changed from MLV integrase to MuA transposase. The basic strategy for introducing insertions into a target sequence remained the same, including a concerted integration reaction followed by gap-repair, restriction endonuclease digestion, and ligation reactions (figure 1). MLV integrase performs the concerted reaction inefficiently, requiring amplification of the integration products by PCR. Since the mixture of integration products is composed of circularly permuted linear pieces of DNA, troublesome PCR side-reactions tend to occur, with template DNAs priming off of one another. These reactions occur less frequently when smaller template DNAs are used, limiting the target DNA size to approximately 1000 base-pairs. This size limitation was unduly restrictive for the experiments on HIV, which involved mutagenizing stretches of the genome of up to 1.5 kilobases cloned into a 2.5 kilobase vector. MuA transposase is a much more efficient and robust enzyme, allowing the intermediate PCR amplification step to be eliminated. The most serious drawback to MuA transposase is that it is more finicky about the sequence of the target DNA. This property leads to an uneven representation of mutants, such that fewer mutants can be conveniently analyzed. An incidental difference between MLV integrase and MuA transposase is that MuA transposase produces a 5 base-pair rather than a 4 base-pair duplication. A second modification was the optimization of the analysis procedure. The insertions made in HIV contained only 10 unique base-pairs, too short for efficient priming. However, the insertions contained a Not I site, permitting the use of the flanking PCR/restriction enzyme digestion analysis method. The samples in the HIV experiment were more complex that those in the SupF studies, leading to higher background from incomplete PCR extension products. The level of these background products was greatly reduced by performing the PCR using one $^{32}$P-labelled target DNA primer and one biotinylated target DNA primer, treating the products of the PCR reactions with a single-stranded binding resin, adsorbing the PCR products to streptavidin-agarose beads, and digesting the products off of the beads with Not I (figure 2). Using this technique, bands visible on the denaturing polyacrylamide gel result from PCR products containing both a radioactive and a biotinylated primer, eliminating incomplete extension products.

The proviral HIV clone used in the work described here is approximately 9000 base-pairs long, and is carried in an approximately 2500 base-pair vector, making a total of 11500 base-pairs in the plasmid. During the mutagenesis procedure, it is necessary to separate the products of concerted integration by MuA transposase (linear) from the unintegrated target molecules (supercoiled circular) and products of single integration events (branched circular) by agarose gel electrophoresis. It is difficult to cleanly separate these species if the target molecule is more than 5000 base-pairs in length. Moreover, during the analysis step, only 200 to 300 base-pairs are examined at any given time. If 11500 base-pairs are mutagenized, the fraction of PCR products containing insertions in a average 300 base-pair segment would be 300/11500, or 2.6%. This value would result in an unacceptably low signal-to-noise ratio on the footprinting gel. Therefore, segments of the HIV genome ranging from 500 to 1600 base-pairs were subcloned for mutagenesis (corresponding to a plasmid size of 3000 to 4100 base-pairs). In order to ensure a good representation of mutant clones in the libraries, we wanted to achieve an average of at least 100 "hits" per base-pair. For a 4100 base-pair construct, therefore, we would aim for a library with at least 410,000 elements, a number which we found to be experimentally feasible to attain.

The mutagenized proviral segments were recloned into a plasmid containing the complete sequence of the provirus. Since a fraction of the "hits" were in vector sequences (the fraction being approximately proportional to the percentage of the entire plasmid composed of vector sequences), approximately 60% to 80% of the clones in the resulting libraries contained no insertions. This situation was to our advantage, since the wild-type clones did not interfere with testing mutations in cis-acting elements, and were actually desired to provide helper functions during the first round of infection for testing mutations in coding sequences. Fewer undesired side-products were obtained during cloning if the mutagenesis was done on proviral fragments carried in an ampicillin-selectable vector and the intact provirus was carried in a kanamycin-selectable vector. The principle potential troublemakers resulted from ligations between two insert-containing vector fragments (i.e. vector fragments that had been "hit" during mutagenesis), which could then homologously recombine using the insert sequences, generating very small plasmids which replicate very quickly and take over the culture.

A library of 15-nucleotide insertion mutants was made using MuA transposase in a replication-defective HIV background carrying the puromycin resistance gene in place of the env gene. The insertions contained a Not I restriction enzyme site, which was used in the analysis phase of the experiments. This library of replication-defective mutagenized proviruses was introduced into producer cells (details on the design of specific experiments are given in Chapter 4). Pseudotyping with VSV-G, a single round of viral production and infection was carried out. Nucleic acid samples were collected at various steps (figure 3), and footprinting these samples allowed us to examine the effect of different mutations on several steps in the replication cycle in parallel (figure 4). For example, samples of producer cell RNA ("cellular RNA") contained lower proportions of transcripts from proviruses containing mutations that interfere with transcription, mRNA stability, or polyadenylation than samples of producer cell genomic DNA. Similarly, RNAs with mutations that preclude efficient translation, dimerization and packaging of viral RNA, assembly of viral particles, or viral budding were underrepresented in pools of RNA in extracellular virions ("virion RNA") compared with pools of viral RNA in producer cells ("cellular RNA"). After infection of a fresh population of host cells with these virions, mutants defective in such processes as packaging of the tRNA primer, entry, uncoating, reverse transcription, nuclear entry, or integration were less well represented in pools of integrated viral DNA ("infected cell genomic DNA ") than in pools of virion RNA. This scheme permitted the assignment of defects in viral replication caused by individual mutants to phases in the viral life cycle without the necessity of testing each mutant alone.

One could isolate interesting mutant clones in one of two ways. First, if one identified specific clones by footprinting, one could PCR those clones out of the mutant library using primers that would prime only from clones with an insert at the desired location (see figure 5). Second, if one were interested in isolating clones that were enriched by a selection scheme, one could PCR a region out of a sample of post-selection nucleic acid and clone the PCR products en masse. To eliminate wild-type clones, one could digest the population of plasmids with Not I and purify the linearized plasmids (those that have a Not I-containing insert).

In the experiments described here, mutants were selected for their ability to perform various steps in the viral replication cycle. Selection strategies other than the one described here can be easily used. For example, to study viral resistance to therapeutic agents one could subject a libary of mutants in protease to a protease inhibitor and use footprinting to identify regions where insertions lead to resistant mutants.

# Chapter 2

## Introduction to High-resolution Genetic Footprinting of HIV

The retroviral life cycle is fairly well understood mechanistically and genetically. Mechanistically, the molecular events involved in virion production and infection are known in outline. Various processes, particularly transcription, assembly, reverse transcription, and integration, have been investigated and described in some detail (reviewed recently in Coffin et al. 1997). The genomes of several retroviruses have been subjected to extensive mutagenesis, both natural and experimental. As a result, functional regions of the viral genome, such as the long terminal repeats (LTRs) and sequences encoding the viral proteins, have been mapped. However, the mutations that have been made thus far are unevenly distributed across the genome and diverse (e.g. point mutations, insertions, and deletions of different sizes and sequences). Moreover, the effects of many of these mutations have not been studied in a uniform or comprehensive manner.

In the experiments described in this report, the goal was to create a high-resolution map of a one kilobase segment near the 5' end of the HIV RNA genome defining features essential for major steps in the viral replication cycle. This region of the HIV genome contains several previously identified functional elements (see figure 6), including several cis-acting elements and sequences encoding the matrix and capsid proteins. By studying a large number of mutants of uniform construction in a thorough and quantitative manner, we strove to gain detailed insight into known elements in the viral genome and to define novel features.

Many of the cis-acting sequences overlap with each other of with coding sequences. The multifunctional nature of certain sequences in the HIV genome can create difficulties in assigning unambiguous functions to these sequences. The TAR stem-loop structure is important in transcription of the viral genome (Berkhout et al. 1989; Selby et al. 1989; Roy et al. 1990a; Roy et al. 1990b; Feng and Holland 1988; Dingwall et al. 1989; Cordingly et al. 1990; Gait and Karn 1993) and overlaps with the sequences in R that are used during the first strand-transfer event in reverse transcription (Coffin and Haseltine 1977; Haseltine et al. 1977; Schwartz et al. 1977; Stoll et al. 1977; Coffin et al. 1978). R also contains a polyadenylation signal. At the 3' end of U5 resides the sequence encoding the 3' att site, a short (~15 base-pair) sequence required by integrase for efficient integration of the viral genome into host cell genomic DNA (Bushman and Craigie 1991; LaFemina et al. 1991; Leavitt et al; 1992, Sherman et al. 1992; van den Ent et al. 1994; Vicenzi et al. 1994). Adjacent to the att site is the primer binding site, an eighteen nucleotide sequence complementary to the eighteen terminal nucleotides of tRNA-Lys, which is used to prime the negative strand during reverse transcription. This sequence also plays a role in the second strand-transfer step of reverse transcription (Rhim et al. 1991). The region of the genome from the end of the LTR into the beginning of the matrix coding sequence contains an AP-1/AP-3 site, a DBF-1 site, and a SP-1 site (Verdin et al. 1990;Van Lint et al. 1991), a splice donor sequence used to produce the mRNA for the envelope protein, and sequences that contribute to dimerization and packaging of the viral single-stranded RNA genome (Lever et al. 1989; Luban and Goff 1994; McBride and Panganiban 1996; Laughrea et al. 1997a; Laughrea et al. 1997b; Clever and Parslow 1997).

The matrix and capsid proteins of retroviruses are translated as part of the gag polyprotein and subsequently cleaved from the polyprotein by a retrovirally-encoded protease. Matrix contains a N-terminal myristoyl group and a nearby basic region, both of which assist in targeting the unprocessed gag polyprotein to the host cell plasma membrane during assembly (Gottlinger et al. 1989; Bryant and Ratner 1990; Zhou et al. 1994). Matrix also interacts with the cytoplasmic tail of the viral envelope protein (Yu et al. 1992b; Facke et al. 1993). In some, but not all, experiments, HIV matrix has been demonstrated to assist in nuclear entry of the HIV pre-integration complex (Bukrinsky et al. 1993; Gallay et al. 1995a; Gallay et al. 1995b; von Schwedler et al. 1994; Fouchier et al. 1997; Freed et al. 1995). There are indications that the C terminus of matrix may play a role in uncoating (Yu et al. 1992a), and it has been suggested that matrix can bind to RNA (Bukrinskaya et al. 1992). HIV capsid is thought to be the major structural protein making up the viral core. Mutations in capsid have been shown to be defective in viral assembly or in an early step in viral replication, between entry and reverse transcription (Mammano et al. 1994; Wang and Barklis 1993; Reicin et al. 1995; Reicin et al. 1996; Dorfman et al. 1994a). Capsid interacts with a host protein, cyclophilin A, which is specifically incorporated into viral particles and seems to play a role in uncoating (Luban et al. 1993; Braaten et al. 1996; Franke et al. 1994; Thali et al. 1994).

High-resolution genetic footprinting has been used to map functionally important domains in the SupF gene (Singh et al. 1997). We have employed a modification of this method to define functional domains in a portion of the HIV genome. A library of insertion mutants was made in a region of the HIV genome using MuA transposase and selected en masse for the ability to undergo various phases of the viral life cycle. Each mutant contained a single insertion, which included a restriction endonuclease recognition sequence at a "random" position (in fact the MuA transposase demonstrates preferences for certain target sequences). An assay involving a PCR reaction and a restriction endonuclease digestion was then performed on nucleic acid samples of the library taken before and after each phase to asses the recovery of each mutant through that phase. This assay generated a product of unique length for each mutation; the length depended on the position of the insertion in the HIV sequence. Therefore, the nucleic acid samples analyzed, which were mixtures of mutants, produced mixtures of products of different lengths, which were resolved as bands on denaturing polyacrylamide gels. Mutants defective for a given phase of the viral life cycle were eliminated at that step, leading to a depletion of the corresponding bands. This scheme permitted the assignment of defects in viral replication caused by individual mutants to phases in the viral life cycle without the necessity of testing each mutant alone.

# Chapter 3

## Materials and Methods

### Plasmids

The HIV replication-defective proviral clone mutagenized in this report (HIV puro) was derived from pHIV-APΔenvΔVifΔVpr (Sutton et al. 1998) and subcloned into either Bluescript KS+ (Stratagene) or pBS -Kan (a Bluescript KS+-derived vector where the ampicillin-resistance gene was replaced by the kanamycin-resistance gene). pHIV-APΔenvΔVifΔVpr was constructed from HIV-AP, an HIV proviral clone containing the human placental alkaline phosphatase in place of nef (He and Landau 1995), by making a large deletion to eliminate most of env, vif, and vpr. To make HIV puro, the human placental alkaline phosphatase gene was replaced by the puromycin resistance gene driven by the SV40 promoter (Morgenstern and Land 1990). In addition, host DNA sequences flanking the proviral sequences were eliminated. PCR mutagenesis was used to eliminate the five Bsg I sites originally present in the plasmid (G → C at position 1222, C → G at position 2574, A → C at position 4856, A → T at position 5755, and A → C at position 5884. These changes did not detectably affect viral replication. Fragments of HIV puro were subcloned into Bluescript KS+, mutagenized (see below) in the context of these smaller plasmids, and subsequently cloned back into HIV puro to generate libraries of mutant proviruses.

### Mutagenesis

The mutagenesis procedure was a modification of the method described by Singh et al. 1997. MuA transposase was a generous gift from Kiyoshi Mizuuchi and Harri Savilahti. The double-stranded oligonucleotide (Not15) used for mutagenesis was made by annealing Not15A (5'-TGCGGCCGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC-3') and Not15B (5'-GTTTTCGCATTTATCGTGAAACGCTTTCGCGTTTTTCGTGCGCGGCCGCA-3') in 50 mM NaCl. The integration reaction was performed by incubating 25 pmol of Not15, 5 µg target plasmid, and 50 pmol MuA transposase (the volume of MuA transposase used was determined by a series of titration experiments) with 25 mM Tris pH 8.0, 100 µg/ml BSA, 15% glycerol (w/v), 144 mM NaCl, 0.1% Triton X-100 (v/v), 10 mM $MgCl_2$, and 15% DMSO (v/v) in a 0.5 ml reaction volume at 30 °C for 1 hour (Savalahti et al. 1995). Reaction products were phenol/chloroform extracted once, chloroform extracted once, precipitated in 0.3 M NaOAc pH 5.2 and 70% ethanol, washed with 70% ethanol, dried briefly under vacuum, and resuspended in 10 mM Tris.HCl/1 mM EDTA pH 8.0. Plasmids linearized by concerted integration events were separated from plasmids that had undergone single-ended integrations events or no integration events by agarose gel electrophoresis. The products of concerted integration events were purified (Qiaquick gel extraction kit) and the 5-nucleotide gaps resulting from the integration events were repaired by Taq DNA polymerase-mediated nick translation (incubation in 1x Taq DNA polymerase buffer (Perkin Elmer), 2.5 mM

$MgCl_2$, 2.5 mM dATP, 2.5 mM dCTP, 2.5 mM dGTP, 2.5 mM dTTP, and 2 units Taq DNA polymerase at 72 °C for 10 minutes in a 100 µl reaction volume). The products of these nick translation reactions were purified (Qiaquick PCR purification kit), then digested with Not I (New England Biolabs). Recircularization of the linear plasmids by ligation of the cohesive ends resulted in 15 base-pair insertions.

**Cell culture**

293, 293T, and HOS cells were grown in Dulbecco's Modified Eagle's Medium containing 4.5 g/l glucose and 10% Defined Fetal Calf Serum (Hyclone). 293T cells were used for all transient transfection experiments. 293 cells were used for all stable transfection experiments and infections by virions produced by transient transfection. HOS cells were used for infections by virions produced from infected or stably transfected cells. Cells were grown at 37 °C in 5% $CO_2$ in a water-jacketed incubator. Puromycin selection was performed using 2.5 ug/ml puromycin (Sigma) for 293 cells and 5 µg/ml puromycin for HOS cells.

**Transfections**

Transient and stable transfections using the Lipofectamine Plus kit (Gibco/BRL) were performed according to the recommended protocol. 30 µg total plasmid DNA, 60 µl Plus reagent, and 40 µl Lipofectamine were used for each 15 cm tissue culture dish. For stable transfections, puromycin selection was initiated 48 hours post-transfection. For transient transfections, the media was changed 48 hours post-transfection and virus was harvested 72 hours post-transfection.

### Infection

Viral stocks were diluted to the desired concentraton in media containing 4 μg/ml polybrene (Sigma) and used to infect cells for 2 hours at 37 °C. Puromycin selection was initiated 48 hours or post-infection.

### Nucleic acid preparation

Plasmid DNA: Plasmid DNA was purified using the Qiagen plasmid DNA kit and subsequently banded in a cesium chloride gradient (Sambrook et al. 1989).

Genomic DNA: The Qiagen Blood and Cell Culture Genomic DNA kit was used to prepare genomic DNA from tissue culture samples.

Total cellular RNA: Total cellular RNA was prepared using the Qiagen RNeasy total RNA kit.

Viral RNA: Viral RNA was prepared by pelleting virions by ultracentrifugation (28,000 rpm for 2 hours at 4 °C in a Beckman SW 28 rotor), pouring off the supernatant, resuspending the viral pellet in the residual media, and using the Qiagen Oligotex direct mRNA kit.

### Sequencing reactions

Sequencing reactions were performed using the Sequenase sequencing kit from USB.

### Reverse transcription

Reverse transcription of cellular RNA and virion RNA samples was performed using 100 ng template RNA with the HIV-specific oligonucleotides HIV521 (5'-GGGAGCTCTCTGGCTAACTAGGG -3') and HIV1573r (5'-CATCCTATTTGTTCCTGAAGGG -3') according to the manufacturer's instructions (Titan reverse transcription kit (Boehringer-Mannheim)).

### PCR

PCR was performed in 20 mM Tris.HCl pH 8.55, 150 ng/ml BSA, 16 mM $(NH_4)_2SO_4$, 3.5 mM $MgCl_2$, 625 uM each dNTP, 0.25 μM each primer, and 1 unit per 50 μl reaction Taq DNA polymerase (AmpliTaq from Perkin-Elmer). "Cold" PCR conditions consisted of 2 minutes at 94 °C followed by 30 cycles of 30 seconds at 94 °C, 30 seconds at 55 °C, and 2 minutes at 72 °C. "Hot" PCR conditions consisted of 2 minutes at 94 °C followed by 25 cycles of 30 seconds at 94 °C, 30 seconds at 55 °C, and 1 minute at 72 °C.

### Pretreatment of streptavidin-agarose beads

Streptavidin agarose beads (Sigma) were incubated in the presence of poly dI-dC (200 ug per ml streptavidin agarose slurry) in 1x binding buffer (12% glycerol (v/v), 12 mM Hepes pH 7.9, 4 mM Tris.HCl pH 8.0, 60 mM KCl, 1 mM EDTA, 1 mM DTT) for one hour at 25 °C. The beads were then washed four times in 1x binding buffer (1 ml buffer/ml slurry) and finally resuspended in 1x binding buffer to reconstitute the initial volume of slurry.

**Single-stranded Affinity Matrix (SSAM) treatment of PCR reactions**

8 µl of 8M Lithium chloride and 10 µl of SSAM (Clontech) were added to each 50 µl PCR reaction. The mixture was incubated for 10 minutes at room temperature with agitation every two minutes. The SSAM resin was then removed by passing the mixture through a 0.45 µm spin filter (Millipore). Alternatively, BNDC resin (Sigma) was suspended in 1M Lithium chloride (0.5 g resin in 2.5 ml 1M Lithium chloride) for 60 minutes at room temperature. 50 µl of this suspension was used per PCR reaction.

**Footprinting**

Initial amplification of nucleic acid samples was done according to the "cold" PCR protocol using HIV-specific primers HIV37 (5'-TGGAAGGGCTAATTCACTCCCAAAG -3'), HIV493 (5'-TCTCTCTGGTTAGACCAGATCTG -3'), HIV521(5'-GGGAGCTCTCTGGCTAACTAGGG -3') and HIV1573r (5'-CATCCTATTTGTTCCTGAAGGG -3'). 10 ng of plasmid samples, one-tenth of the products of reverse transcription reactions (equivalent to 10 ng input RNA), or 0.5 µg genomic DNA samples were used as templates. 10 ng of "cold" PCR products were used for "hot" PCR reactions. For "hot" PCR reactions, one HIV-specific primer was labeled with $^{32}$P (T4 polynucleotide kinase, New England Biolabs) while the other primer was biotinylated (Operon). High-specific-activity 32P-gamma-ATP (160 µCi/mmol, 23 pmol/µl, ICN) was used for radiolabelling at a stoichiometry of 1 pmol ATP/1 pmol oligonucleotide. "Hot" PCR products were treated with SSAM, purified (Qiaquick PCR purification kit), then adsorbed to 50 µl pretreated streptavidin-agarose beads (Sigma) in 1x binding buffer for one hour at 25 °C. The beads were then washed twice with 0.5 ml 1x binding buffer for 15 minutes at 25 °C, washed once with 0.5 ml 1x restriction enzyme buffer 3 (New England Biolabs), and incubated in 50 µl 1x restriction enzyme buffer 3 (New England Biolabs) containing 20 units Not I restriction enzyme (New England Biolabs) for 1 hour at 37 °C. The supernatant from this digestion step was separated from the beads by centrifugation through a Micro Bio-spin column (Bio-rad) at 3,000 rpm for 1 minute at room temperature in a tabletop microfuge. The supernatant was then precipitated in 0.3 M NaOAc pH 5.2 and 70% ethanol in the presence of 5 µg linear acrylamide, washed with 70% ethanol, dried briefly under vacuum, and resuspended in 3 µl 10 mM Tris.HCl/1 mM EDTA pH 8.0 + 3 µl 2x formamide loading dye (95% deionized formamide/25 mM EDTA pH 8.0/0.25% bromophenol blue/0.25% xylene cyanol). Samples were heated at 95 °C for 2 minutes, placed immediately onto ice, and analyzed by electrophoresis through 6% acylamide (19:1 acrylamide:bis-acrylamide)/1x TBE/7 M urea sequencing gels (2 µl sample per lane). Gels were dried for 1.5 hours at 80 °C under vacuum and exposed to Biomax MR film (Kodak).

### Quantitation

Autoradiographs were scanned using a flatbed scanner (Hewlett-Packard) at 300 dpi resolution, with brightness and contrast set at 125 (50%). Scanned images were read into a Matlab-based application (see Appendices D and E) by which individual bands were selected and quantitated for peak intensity values. Data from different footprinting reactions and different gels were normalized by fitting profiles of the relative intensities of bands within each run using an algorithm that minimizes the sum of the coefficients of variance for the mutants weighted for the number of measurements for each mutant (figure 7). The normalized data were then averaged. Data from triplicate experiments were normalized using the same algorithm and averaged (see Appendices D and E).

# Chapter 4

## Results

### Creation of a library of insertion mutants

The objective was to make a large number of mutations of the same type at diverse positions in a one kilobase stretch of the HIV genome and to assess the performance of each mutant at several points in the viral replication cycle. A library of 15-base-pair insertion mutants was constructed by in vitro transposition in a replication-defective HIV provirus containing the puromycin acetyltransferase gene driven by an internal promoter in place of the env gene. The mutations were made specifically in the segment of the HIV genome (positions 37-1550) including the 5'-LTR, the 5' untranslated region, the complete matrix gene and the 5' half of the capsid gene. Mutants are numbered according to the nucleotide position immediately 5' to the insertion.

The MuA enzyme was used to perform an in vitro transposition reaction, introducing a pair of double-stranded DNA oligonucleotides into a double-stranded circular target DNA molecule (figure 1). The oligonucleotides contained both sequences necessary for recognition by MuA and sequences recognized by the Not I restriction endonuclease. MuA inserts the oligonucleotides into the target DNAs in a staggered fashion, such that the products of the transposition reaction were gapped linear double-stranded DNA molecules, with an oligonucleotide located at either end. After filling in the gaps by nick translation, the reaction products were digested Not I, generating compatible cohesive ends, which were ligated. The final products were circular DNA molecules containing the inserted sequence, 5'-TGCGGCCGCA-3', flanked by five base-pair duplications of the target sequence. The insertions retained the NotI recognition sequence, which was used during the analysis procedure. Insertional mutants were generated using MuA transposase rather than MLV integrase, the enzyme used in the original footprinting experiments, since MuA transposase executes the necessary in vitro concerted integration reaction more robustly (Crowley et al., manuscript in preparation). Sixteen individual mutant clones, at positions 189, 238, 268, 358, 557, 622, 776, 926, 1012, 1045, 1067, 1175, 1264, 1267, 1277, and 1399, were isolated and sequenced. These clones were used as markers to determine the location of insertions during analysis.

Insertions were designed such that mutations in coding sequences would be in-frame insertions of five codons. The identity of the amino acids encoded by the insertions depended on both the reading frame and the sequences in the target DNA adjacent to the insertion site.

The positions of insertion mutants for which data were obtained are indicated in figure 6. Although the collection of mutants is extensive, the sequence space was not saturated, since MuA transposase does not make insertions at the same frequency at all sites. Moreover, since transcription starts at R in the 5'-LTR, the effects of mutations in U3 could not be assessed. Examination of nucleic acid samples before and after a single round of transcription by genetic footprinting confirmed this loss of mutants in U3 at transcription, indicating that nucleic acid samples were not contaminated with plasmid DNA from the initial transfections.

### Sampling populations of mutants at different steps in the viral replication cycle

To study the effects of insertions on cis-acting elements (e.g. transcriptional modulators, the packaging sequence, and the viral att site), the library was either transiently or stably transfected into producer cells. A plasmid encoding VSV-G was transiently transfected into the producer cells to pseudotype the env-defective virions. A single round of infection was then performed. Nucleic acid samples were collected at various steps during this experiment (see figure 3). Depletion of mutants at different steps in the viral replication cycle was followed by analyzing these nucleic acid samples by genetic footprinting.

A similar strategy was utilized to determine the effects of insertions in trans-acting sequences. Since more than one piece of DNA often enters a given cell during transfection, complementation can occur in a mixed population between trans-acting elements in a transfection experiment (see figure 8). In order to study the functions of trans-acting factors in the absence of complementation, a first round of transient transfection was conducted, cotransfecting the mutant library with a VSV-G expression construct. The goal was to produce a VSV-G pseudotyped, phenotypically mixed population in which mutants with defective trans-acting functions were rescued by complementation. Since approximately half of the clones in the library were wild-type (i.e. did not contain an insertion), this complementation was easy to achieve. These virions were then used to infect fresh host cells at a low multiplicity of infection (1 infectious unit for every 20 cells) such that each cell would receive only one viral genome. According to a Poisson distribution, 95.12% of the cells would receive 0 virions, 4.76% of the cells would receive 1 virion, and 0.12% of the cells would receive more than 1 virion. Hence, of the cells that received at least one virion, approximately 2.5% received more than one virion. The infected cells were selected using puromycin, and this pool of cells was used as the starting population of producer cells for a single round of infection. Nucleic acids were purified at various steps during this experiment and analyzed by genetic footprinting, allowing us to study the effects of mutations on the functions of trans-acting sequences.

From results obtained in these studies, it is now clear that complementation of trans-acting factors occurred very efficiently during the first round of infection in our transient transfection experiments but not to any appreciable degree in our stable transfection experiments. The number of proviruses per cell has not been directly measured. However, if the number of proviruses per cell is T for our transient transfection experiments and S for our stable transfection experiments, our results suggest that T is greater than S. In the simple case where the wild-type version of a gene is dominant and a mutant version is recessive, we would expect T to be greater than or equal to two and S to be equal to one. However, the viral proteins studied in our experiments probably function as oligomers, such that mutants might display dominant negative phenotypes. Thus, in our experiments, S might be larger than one, with T significantly larger than S. In fact, as mentioned above, multiple pieces of DNA can enter a single cell during transfection, leading us to expect S to be larger than one.

**Description of footprinting analysis procedure** (figure 2)

Nucleic acid samples collected at various points in the viral life cycle were subjected to an initial round of amplification by either PCR (for DNA samples) or RT-PCR (for RNA samples). PCR was then performed on these pre-amplified samples using one $^{32}$P-labelled DNA primer and one biotinylated DNA primer. The primers were complementary to HIV sequences and flanked the region to be analyzed. The products of this second PCR reaction were first treated with a single-stranded binding resin to remove incomplete extension products and then bound to streptavidin-agarose beads. The radioactively labeled portions of the PCR products containing Not I sites were digested off the beads with Not I, concentrated, and subjected to electrophoresis on denaturing polyacrylamide-urea gels. A typical gel is shown in figure 9.

**Cis-acting versus trans-acting elements**

Mutations in cis-acting and trans-acting features can often be distinguished by differential behavior in complemented versus uncomplemented infection cycles. One would expect mutations in cis-acting sequences to show their phenotypes in the presence or absence of complementation, while mutations in trans-acting sequences should be apparent only when uncomplemented. Trans-acting sequences are typically considered to be coding sequences. However, due to the pseudodiploid nature of retroviruses and peculiarities in certain steps of viral life cycle (such as assembly and reverse transcription), there can conceivably be trans-acting sequences in the HIV genome that act at the nucleic acid level.

Data showing the behavior of individual mutants in single-cycle infections are given in figure 10. Figure 10A shows survival of mutants through one round of complemented infection (first round transient transfection), while figure10B shows the behavior of mutants through one round of uncomplemented infection (second round transient transfection). Mutations that affect replication in both complemented and uncomplemented infections to a significant degree (greater than 55% depletion during one round of infection) appear to be localized to the region 5' to position 847. Since most of this region appears to be composed of noncoding sequences (up to position 828, where the matrix coding sequence begins), it is not surprising that we found cis-acting elements in this area of the genome. Insertions in sequences between positions 847 and 1524 display no clearly discernible effects in complemented infections, while many of these insertions interfere with infection in the uncomplemented situations. Since coding sequences for matrix and capsid lie in this stretch of the genome, one might have expected to find trans-acting functions here.

### Mutants in non-coding sequences defective in transcript formation or stability

The six mutants in the TAR region (492-542) were severely compromised in their ability to replicate. The primary deficiency was in transcript production or stability (only qualitative data is given as the region surveyed is too close to the end of the viral RNA for accurate quantitation). Most likely, these mutants are defective for tat binding, which would result in a low efficiency of transcription. The phenotype appears in the presence of complementation, reconfirming the cis-acting nature of the affected noncoding sequences.

In the transient transfection experiment, insertions at positions 564, 573, and 583 had detrimental effects on the second, but not the first, round of infection (figure 10). For the second round of the transient transfection experiment, the effects of the insertions at all three positions were most pronounced during transcript formation (figure 11). However, in the stable transfection experiment, the mutations at positions 564 and 583 resulted in decreases in fitness in the phase of the life cycle occurring between collection of the cellular RNA and viral RNA samples (figure 12). The most probable explanation for these observations is that these mutations, which are in and around the polyadenylation consensus sequence (563-568), interfered with polyadenylation. This defect would not be observed in the first round of infection since only the 5'-LTR was mutagenized and it was not until the first round of reverse transcription that mutations were transferred to the 3'-LTR, where the operative polyadenylation signal lies. In addition, the mutations at positions 564 and 583 might interrupt partially transcomplementable sequences that contribute to packaging of the viral RNA genome (see below for further discussion of packaging sequences). The dramatic depletion at a previous step (i.e. transcript formation) might be masking the same effect on viral assembly during the second round of the transient transfection experiment.

The other cis-acting mutations that appeared to affect transcript abundance (at positions 578, 727, 728, 730, 758, and 791) manifested moderately to severely decreased transcript levels in producer cells under all conditions tested. These mutations may affect the performance of cis-acting transcriptional enhancer elements.

**Mutations in cis-acting sequences that affect viral assembly**

A cis-acting RNA packaging signal has been previously mapped to the few hundred base pairs around the 5' splice donor site and the 5' end of gag. Here, we have observed that mutations in the interval between positions 739 and 846 were depleted between transcription and release of cell-free virus in all (complemented and uncomplemented) experiments (figure 12). This region encompasses the "kissing loop" dimerization and packaging signal, the 5' splice donor site, and two stem-loop structures which have been found to bind in vitro to gag and nucleocapsid proteins (Berkowitz and Goff 1994; Berkowitz et al. 1993; Clever et al. 1995; Sakaguchi et al. 1993).

The existence of a supplementary packaging signal is implied by a report by Vicenzi et al. 1994, where a deletion of the 5' one-third of U5 results in a 10-

fold decrease in RNA packaging. In our turn, we have found additional mutations in U5 (at positions 564, 583, 607, 621, and 640) that appear to be defective in viral RNA packaging (figure 12).

In general, the phenotypes of these packaging mutants were more severe in the absence of complementation (figure 10 and data not shown). If viral genomes with insertions at these positions are still able to form dimers, dimerization with wild-type viral genomes may partially rescue the packaging defect of these mutant genomes.

**Cis-acting mutants defective in late replication events**

Mutations at positions 607-654 and 758-791 resulted in a reduction in recovery during the early part of the viral life cycle, which includes viral entry, uncoating, reverse transcription, nuclear entry, and integration (figure 12). The mutations between positions 607-654 are located in U5, just 5' to the att site. Although no specific function for this region of U5 has been previously defined, its proximity to the att site raises the possibility that sequences in this region contribute to recognition of the viral genome by integrase. These mutations are also reasonably close to the primer binding site, and may interfere with initiation of reverse transcription (Leis et al. 1993). The second group of mutations, between positions 758-791, is in the "kissing loop" motif and the 5' splice donor sequence. A function for sequences in this area in early replication events has not been previously described.

The paucity of mutants displaying significant and specific defects in early steps of viral replication is probably due to the design of our experimental system. It is likely that elimination of mutants at steps in the viral life cycle occurring earlier in our series of experiments (e.g. transcription or assembly) prevents our recognition of additional defects in entry, reverse transcription, or integration. For example, two mutations (at positions 683 and 684) located in the primer binding site were severely depleted in the virion RNA sample (transient transfection experiment, data not shown), such that it was not possible to distinguish further reductions in the infected cell genomic DNA sample. Due to the sequence preferences of MuA transposase and the introduction of five base-pair duplications during the mutagenesis procedure, our pool of insertion mutants did not include any detectable mutations that destroyed the att site in U5, another feature in this segment of the genome known to be essential for integration.

### Mutations in matrix

Sequences at the 5' end of the matrix coding sequence (positions 827-838) appeared to contribute to viral RNA packaging in cis (see above). A few mutations near the 5' end of the matrix gene (positions 876-929) appeared to result in defects in the production of stable transcripts (figure 13). These trans-acting mutations, which could be rescued by complementation, were located in the sequences that encode the C-terminal end of helix 1, a loop between helix 1 and helix 2, and the N-terminal half of helix 2. This region contains many basic residues, and is at the edge of the globular domain of matrix that faces away from the trimer interfaces. The phenotype of these mutants suggests that matrix might have a role in enhancing transcription or stabilizing the viral RNA genome in the producer cell prior to budding. Supporting this possibility, it has been proposed that matrix has RNA-binding activity (Bukrinskaya et al.1992).

Most of the mutants with insertions from positions 937-1131 demonstrated primary losses in fitness in the portion of the life cycle from translation through assembly to budding (figures 13 and 15). These mutations could be rescued in trans and were in the portion of the matrix gene encoding the core of the globular domain of matrix. Mutations in this region are likely to interfere with the proper folding of the matrix protein and thus produce defects in viral assembly, as seen in our results.

As reported previously (Freed et al. 1994; Dorfman et al. 1994a), mutations in the C-terminal domain of matrix (positions 1141-1211 in this study), which consists of a long alpha-helical tail that extends away from the globular domain, were well-tolerated (figures13 and 15).

## Mutations in capsid

In accordance with other reports (Dorfman et al. 1994b; Mammano et al. 1994), we found that mutations in the sequence encoding the N-terminal half of the capsid protein were severely detrimental to viral replication (figure 10B). Capsid mutants with insertions between positions 1276 and 1464 were defective both at a step in viral production (assembly or release) and at an early step in replication (figures 14 and 15). Preliminary results indicate that the defect in early replication occurs before the completion of reverse transcription. This result is quite striking, particularly in comparison to the bulk of our matrix mutants, which appeared to be specifically defective at the assembly/budding step (figure 15).

Mutants with insertions in and immediately adjacent to the N-terminal β hairpin (1244-1264) and the cyclophilin A binding region (1479-1508) of capsid were able to form viral particles, but were defective in a step in early replication (figures 14 and 15). X-ray crystallographic studies (Wlodawer and Erickson 1993; Gitti et al. 1996) support the theory that the β hairpin structure forms only after proteolytic maturation of the viral particle. An extended, relatively disordered conformation during assembly may account for the fact that insertions in this region do not cause a drop in viral particle formation. The β hairpin and the cyclophilin A binding regions are the only regions in the N-terminal domain of capsid that protrude from a tightly packed helical core. These structural differences might explain the differential effects of insertions in these regions on assembly. It has been suggested that the disassembly of the viral core (uncoating) that occurs after viral entry and before the initiation of reverse transcription depends on an interaction between capsid and cyclophilin A (Braaten et al. 1996; Gamble et al. 1996). Therefore, one might expect insertions in the cyclophilin A binding region that interfere with this interaction to affect uncoating.

## Discussion

### Expected results and novel observations

In the course of these experiments, we have identified several features in the HIV genome, some of which have not been previously described. We mapped three types of cis-acting sequences: those that function in transcript formation/stability, those that are involved in viral RNA packaging, and those that are important for an early step in viral replication. Some of these sequences were found in areas previously mapped for these functions (e.g. TAR, the "kissing loop" motif) and others were found in novel locations. Several mutations near the N-terminus of matrix suggest an unforeseen trans-acting function for matrix in transcript formation or stabilization. In constrast to previous reports, we have observed that many mutations in the globular core of matrix have marked effects on assembly, and mutations in the helical core of the N-terminal domain of capsid cause defects in both assembly and an early step (perhaps disassembly) in viral replication (see below). Finally, mutations in the β hairpin and cyclophilin A binding regions of capsid primarily result in early replication defects. The behavior of the mutations in the cyclophilin A binding region are consistent with the postulated function of this region in uncoating.

How can the same mutation in capsid cause defects in both assembly and disassembly? The answer to this question may lie in the fact that assembly and disassembly are not simply reverse processes. Most obviously, assembly involves the aggregation of gag and gag-pol polyproteins whereas disassembly normally occurs after proteolysis of these polyproteins into several smaller entities. A mutation that decreases the efficiency of assembly may cause the viral proticles that do form to be aberrant in some way. This notion is supported by observations of abnormal viral core morphology in viruses with mutations in the N-terminal half of capsid (Dorfman et al. 1994b; Reicin et al. 1996). These particles may have problems that interfere with steps that are prerequisites to uncoating. For example, perhaps a decreased ratio of gag-pol to gag compromises proteolytic maturation. Alternatively, essential host factors such as cyclophilin A might be inefficiently incorporated.

### Sources of variability in the data

Approximately one kilobase of the HIV genome was analyzed using ten primer pairs. Each primer pair was used to examine an interval of 200 to 300 base-pairs. Each mutant was examined with at least two primer pairs. Moreover, each series of transfection/infection experiments was carried out in triplicate. It was therefore necessary to develop a normalization procedure (see Materials and Methods) so that data from separate gels and different replicates could be combined in determining the quantitative effect of each mutation. Data were normalized  based on previous findings that certain areas of the viral genome, such as the C-terminus of matrix, are consistently tolerant to small, in-frame insertions.

The normalized data was examined to determine whether the variability in the data arose primarily from variability in the transfection/infection experiments (which could result from sampling error) or variability in the genetic footprinting procedure. Data for thirty mutants from the stable transfection experiment cellular RNA sample were tabulated and the weighted average of the variances were calculated for the complete data set, data "within replicates," and data "within gels." The normalized intensity measurements for these thirty samples ranged between 7.1 and 111.6. The abundance of each mutant was measured four or five times per replicate. Data within a replicate were derived from separate genetic footprinting reactions using different primer pairs performed on the same nucleic acid sample and run on separate gels. Data within a gel were derived from separate genetic footprinting reactions using the same primer pairs performed on different nucleic acid samples and run on the same gel. The weighted average of the variances was 63.8 for the complete data set, 65.9 "within replicates", and 16.7 "within gels". Therefore, most of the variability appears to arise from differences between gels or primers rather than sampling error incurred during the selection procedure, variability between PCR reactions, or inconsistencies in other nucleic acid manipulations.

### Mutations that appear to confer an increase in replication-competence

Mutations at a few positions appear to result in proviruses with an enhanced ability to carry out certain step in viral replication. This finding is somewhat unexpected, as one might expect the wild-type virus to be optimized for replication. However, the system used for the experiments described here is significantly different from the environment in which wild-type HIV evolved. Viral production and infection was carried out in a tissue culture system, rather than in the context of a whole organism. In this context, the virus does not need to contend with the same complexity of virus-host interactions, such as evasion of the host immune defenses. The sequences encoding env and the accessory factors vif, vpr, vpu, and nef were removed from the proviral clone used, and VSV-G protein was used to pseudotype this defective proviral construct. The use of a pseudotyping system removes several constraints on the viral genome, including the preservation of a functional 5' splice donor sequence and retention of the env-interacting function of matrix. In summary, since the same constraints do not apply in the system used here and in the environment in which HIV evolved, mutations detrimental in one case may be beneficial in the other case.

**Incomplete depletion of mutant proviruses**

For mutations that severely compromise viral replication, the system presented here may overestimate the ability of these mutants to replicate. This error may stem from three sources. First, there is a certain amount of error in the analysis and quantitation procedures used. Second, the insertion sequence used in these experiments includes a 10 base-pair palindrome, which may form a nucleic acid hairpin structure. This type of mutation may be less disruptive to cis-acting sequences that depend on nucleic acid secondary structure than other types of mutations, such as deletions, substitutions, or non-palindromic insertions. Third, in the selection strategy, the uncomplemented infection cycles were carried out using either stably transfected cells or cells that had been infected at low m.o.i as producer cells. As discussed above, it is possible (and even likely) that some degree of complementation occurred in the stably transfected cells. As for the cells infected at low m.o.i., the measured m.o.i was 0.05. If the infection followed a Poisson distribution, approximately 2.5% of the cells that were infected by one virus were actually infected by more than one virus, allowing complementation to occur in those cells. Hence, for trans-acting factors, one would expect a background reading of approximately 2.5% of wild-type for recessive mutations.

**Observed discrepancies with previously published results**

The inconsistencies between our results and those found in other reports can be grouped into two classes. First, mutations at certain positions in the matrix gene resulted in severe defects in replication in our study, while it has been reported elsewhere that mutations at the same positions were tolerated (Freed et al. 1994). Second, we found that many mutations in the N-terminal half of capsid were defective in viral assembly. In contrast, others have reported that residues important for gag multimerization and viral assembly reside in the C-terminal domain of capsid (Jowett et al. 1992; Dorfman et al. 1994b; Von Poblotzki et al. 1993; Reicin et al. 1995), while viruses with mutations in the N-terminal domain of capsid were competent for viral assembly, although many formed viral particles with abnormal core morphologies (Dorfman et al. 1994b; Wang and Barklis 1993; Franke et al. 1994; Reicin et al. 1995, Reicin et al. 1996). The discrepancies between our results and those found in other reports may result from differences in experimental method or interpretation.

First, the precise locations and types of mutations differ between all the reports. Different point mutations at the same position in a given gene can lead to different phenotypes. The types of mutations employed vary widely between (and even within) reports, and include point mutations, small deletions, large deletions, and insertions in various combinations.

Second, the methods used to assess replication-competence differ between reports. Wang and Barklis (1993) performed single-round infectivity assays by measuring infection of a marker gene. Other groups followed exogenous RT activities or production of viral proteins in spreading infections over the course of several weeks (Freed et al. 1994 and Reicin et al. 1995; Dorfman et al. 1994a; Dorfman et al. 1994b). We looked at data from two types of experiments: one single-round infection without complementation and two single-round infections in series, the first of which was complemented and the second of which was not complemented. This approach stands in contrast to experiments with spreading infections, where it is difficult to know how many rounds of infection have occurred, which in turn makes it difficult to measure infectivity quantitatively.

Third, viral assembly has been measured in a variety of ways, including exogenous RT assays, RNase protection, western blotting for viral proteins, and electron micrography. These methods do not always assess whether the viral particles contain viral RNA; some are qualitative or yield highly variable results. We believe that none of these methods is as rigorous as the method employed in this report, where we assessed the relative representation of mutants in the viral RNA sample itself.

Fourth, in all of the other reports the viral particles studied were generated by transient transfection, while the viral RNA samples we footprinted were purified from virions produced from either stably transfected or cells infected at low m.o.i. In our experiments, the titer of virus produced by transient transfection was 10- to 100-fold higher than the titer of virus produced from stably transfected or infected cells. If this difference in titer reflected a difference in expression of the viral genome, the requirements for viral assembly and packaging of the viral RNA genome in our experiments were 10- to 100-fold more stringent than in the experiments described in the other reports. The mutants that have quantitative defects in assembly or packaging might appear to be competent for these functions by less stringent methods.

Finally, in the strategy presented here, a large number of mutants with insertions at diverse positions were followed en masse through two rounds of replication. This strategy permits a comprehensive examination of viral replication. Selection and analysis of the mutants in parallel provided built-in internal controls for variables such as sample recovery and efficiency of analysis procedures.

**Future directions**

Our examination of three nucleic acid samples per round of replication yielded a relatively crude breakdown of the HIV life cycle. Refinement of our picture of viral replication can be achieved by footprinting samples from more finely differentiated steps. For example, we could study the effects of mutations on nuclear export of viral RNA by comparing nuclear and cytoplasmic RNA samples from producer cells. Other interesting steps in the viral life cycle amenable to clarification by genetic footprinting are reverse transcription and nuclear entry. We could investigate these steps by collecting additional nucleic acid samples, such as intermediates in the reverse transcription reaction (minus-strand strong stop DNA and plus-strand strong stop DNA), full-length unintegrated viral DNA in the host cell cytoplasmic fraction, and full-length unintegrated viral DNA in host cell nuclear fraction. Of course, genetic footprinting can be performed on the rest of the HIV genome. In addition, we have developed methods to introduce and analyze a variety of mutations, including insertions of different lengths and sequences and substitutions of various types (Singh et al. 1997 and unpublished results).

**Generalizability of the genetic footprinting technique**

In the original report describing the genetic footprinting technique, this method was used to generate a high-resolution functional map of a small (200 base-pair) gene encoding an RNA molecule (Singh et al. 1997). A functional selection was carried out in a prokaryotic system, and the footprinted nucleic acid samples consisted of purified plasmids. Here we present modifications to genetic footprinting that permitted us to analyze a much larger (1000 base-pair) stretch of nucleic acid including both cis- and trans-acting sequences. The experiments presented here involved the isolation and analysis of complex nucleic acid samples, including cellular RNA, virion RNA, and genomic DNA samples from a selection scheme in eukaryotic cells. Thus, genetic footprinting can be used to map the functional features in any DNA sequence if an appropriate selection scheme exists. In such a scheme, the abundance of the sequence encoding a given mutant in a nucleic acid sample collected after selection varies directly with the ability of that mutant to survive the selection. In addition, we have developed methods to analyze genetic footprinting data in a quantitative manner. These tools not only reduce the labor involved in analysis of such data, but also allow a more objective assessment of the data.

# Chapter 5

## The Future of Genetic Footprinting

It is evident that the genetic footprinting method as it exists offers many advantages over traditional methods of mutagenesis and analysis of mutants. Genetic footprinting enables an investigator to perform the mutagenesis, functional selection, and analysis steps en masse, collecting quantitative data on hundreds of mutants at once. There are four major limitations to the present genetic footprinting technology.

First, the distribution of measurable mutants in a gene is largely limited by the sequence bias displayed by the enzyme utilized for mutagenesis. The current favorite enzyme, MuA transposase performs the desired concerted integration event robustly, but displays a sequence selectivity for integration that spans at least three orders of magnitude. The current analysis method covers two orders of magnitude and data can be obtained for only one out of six base-pair positions on average.

Second, the existing repertoire of enzymatic functions limits the design of mutations. The genesis of any mutant library must begin with the construction of a library of insertion mutants, where the palindromic insertions contain a five base-pair duplication in the target sequence and the recognition sequence for a restriction enzyme. This sequence must not occur anywhere else in the vector used for mutagenesis, and must be tolerated by MuA transposase. The range of mutants has been expanded by introducing a linker containing a type IIs restriction enzyme recognition sequence at each insertion. Type IIs restriction enzymes cleave some number of nucleotides away from their recognition sites, the number being specific to the enzyme. The linker can be designed such that digestion with the type IIs restriction enzyme either precisely excises the insertion or creates a deletion. Finally, a new linker of desired sequence is inserted into the gap. The most significant drawback to this approach is the lack of type IIs restriction enzymes that cut more than 16/14 nucleotides away from their recognition sequences.

Third, there are several restrictions imposed by the current analysis method. Two strategies for analysis have been employed, both of which rely on the polymerase chain reaction. The initial strategy used, the "direct PCR" approach, involved using one fixed primer complementary to a sequence in the target gene outside the region under inspection and one "mobile" primer complementary to the insertion sequence. The lengths of the products of this type of PCR reaction correspond to the positions of the inserts relative to the position of the fixed sequence. However, if one wishes to examine the effects of a short insertion or replacement, the direct PCR method proves to be unsatisfactory. A fifteen base-pair insertion contains a unique sequence of only ten base-pairs, too short for sufficiently specific priming. The solution to this problem has been to use the "flanking PCR/restriction digestion" technique. Here, a PCR reaction is performed using two fixed primers complementary to sequences in the target gene flanking the region of interest. The products of this reaction are digested with a restriction enzyme that recognizes a site in the insertion sequence. The obvious limitation to this method is that the insertion must be palindromic and contain a restriction site (if the sequence is not palidromic, each position of insertion will yield two products of different size, depending on the orientation of the insertion in the gene).

Finally, the gel-based detection method is cumbersome and introduces a significant amount of error into our results. In fact, using the current system, the analysis procedure appears to be responsible for much more variation than the selection scheme.

I believe there is a technically feasible alternate approach that eliminates the problems enumerated above. The investigator would be able to specify the positions and relative abundances of mutants, assuring more uniform coverage of the gene of interest. A much wider variety of mutations would be available, including insertions or replacements of as few as three base-pairs, with no limitations on the content of the introduced sequence. Even deletions can be examined, as long as one is willing to introduce a few unique base-pairs at the deletion site (see figure 16 for sketches of possible types of mutations). The analysis method would involve a flanking PCR step followed by hybridization of the PCR products to an array of oligonucleotides and scanning using a fluorescence detection system. A specific description of this approach follows.

First, one must make a library of mutants (see figure 17). This step involves the synthesis of two unique oligonucleotides for each mutant desired, in addition to two common oligonucleotides complementary to sequences flanking the region under mutagenesis. For example, in order to replace all the residues in a 300 amino acid protein with alanine, one would need 600 + 2 oligonucleotides. These mutagenic oligonucleotides are also used in the analysis process. The two unique oligonucleotides for a given mutant are complementary to each other, and contain two types of sequences. At the edges, the oligonucleotides are complementary to target gene sequences on either side of the site of mutagenesis. The middle of each oligonucleotide contains the insertion or replacement sequence. The lengths of the "edge" sequences are adjusted such that the oligonucleotides for all of the mutants have approximately the same melting temperature. Now all the mutagenic oligonucleotides complementary to the top strand of the target gene are mixed together in one pot and all the mutagenic oligonucleotides complementary to the bottom strand are mixed together in another pot, adjusting the ratios of the individual oligonucleotides according to the desired proportions of each mutant in our starting library. For instance, in order to start with twice as many mutants at position A as at position B, one would add twice as many oligonucleotides for position A as for position B. Then, two PCR reactions are performed. The template for both reactions is the gene to be mutagenized. One reaction contains the fixed oligonucleotide complementary to the bottom strand of the template and the mixture of mutagenic oligonucleotides complementary to the top strand of the template, and the other reaction contains the other set of oligonucleotide primers. In order to minimize the introduction of unwanted mutations due to misincorporation by the enzyme used for PCR, a proofreading polymerase is used and the number of cycles of PCR is minimized. Suppose that one starts with approximately 12.5 pmol of each primer and 0.015 pmol of template. After five cycles of PCR with an annealing temperature corresponding to the predicted melting temperature for the "edge" sequences of the mutagenic oligonucleotides, about one pmol of mutants should be present. Then, ten cycles of PCR with an annealing temperature corresponding to the predicted melting temperature for the complete mutagenic oligonucleotides are performed. After a purification step to eliminate any unincorporated primers, the products of the two PCR reactions are mixed together. Another PCR reaction including only the fixed, flanking oligonucleotides is performed, using an annealing temperature corresponding to the predicted melting temperature for the complete mutagenic oligonucleotides. The products of the initial round of PCR will prime off of each other if they overlap precisely, as they will when they correspond to the same mutant. Then, the flanking primers will amplify the population of mutants, which can be cloned into an appropriate vector for selection.

After a functional selection is performed, the relevant nucleic acid sample is purified. Both the original library of mutants and the selected nucleic acid sample are subjected separately to PCR using the flanking oligonucleotides. During this amplification step, fluorescent labels are incorporated, one color for the pre-selection sample and another color for the post-selection sample. The products of these two PCR reactions are then mixed and used as a probe to hybridize to an oligonucleotide array. The elements of this array are the original mutagenic oligonucleotides (one can of course use the fixed oligonucleotides as positive controls and normalization standards). PCR products containing mutations hybridize to the corresponding mutagenic oligonucleotides. It has been shown that existing array hybridization technology allows discrimination of one mismatch in an oligonucleotide octomer. This level of specificity should be adequate for the purposes of this type of experiment. The arrays are then scanned and quantitated. The ratios of the two colors at each spot on the array reflect the ability of the corresponding mutant to survive the selection.

## References

Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature 226, 1209-11.

Berkhout, B., Silverman, R. H., and Jeang, K. T. (1989). Tat trans-activates the human immunodeficiency virus through a nascent RNA target. Cell 59, 273-82.

Berkowitz, R. D., Luban, J., and Goff, S. P. (1993). Specific binding of human immunodeficiency virus type 1 gag polyprotein and nucleocapsid protein to viral RNAs detected by RNA mobility shift assays. Journal of Virology 67, 7190-200.

Berkowitz, R. D., and Goff, S. P. (1994). Analysis of binding elements in the human immunodeficiency virus type 1 genomic RNA and nucleocapsid protein. Virology 202, 233-46.

Braaten, D., Franke, E. K., and Luban, J. (1996). Cyclophilin A is required for an early step in the life cycle of human immunodeficiency virus type 1 before the initiation of reverse transcription. Journal of Virology 70, 3551-60.

Bryant, M., and Ratner, L. (1990). Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. Proceedings of the National Academy of Sciences of the United States of America 87, 523-7.

Bukrinskaya, A. G., Vorkunova, G. K., and Tentsov, Y. (1992). HIV-1 matrix protein p17 resides in cell nuclei in association with genomic RNA. Aids Research and Human Retroviruses 8, 1795-801.

Bukrinsky, M. I., Haggerty, S., Dempsey, M. P., Sharova, N., Adzhubel, A., Spitz, L., Lewis, P., Goldfarb, D., Emerman, M., and Stevenson, M. (1993). A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells [see comments]. Nature 365, 666-9.

Bushman, F. D., and Craigie, R. (1991). Activities of human immunodeficiency virus (HIV) integration protein in vitro: specific cleavage and integration of HIV DNA. Proceedings of the National Academy of Sciences of the United States of America 88, 1339-43.

Clever, J., Sassetti, C., and Parslow, T. G. (1995). RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. Journal of Virology 69, 2101-9.

Clever, J. L., and Parslow, T. G. (1997). Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. Journal of Virology 71, 3407-14.

Coffin, J. M., and Haseltine, W. A. (1977). Terminal redundancy and the origin of replication of Rous sarcoma virus RNA. Proceedings of the National Academy of Sciences of the United States of America 74, 1908-12.

Coffin, J. M., Hageman, T. C., Maxam, A. M., and Haseltine, W. A. (1978). Structure of the genome of Moloney murine leukemia virus: a terminally redundant sequence. Cell 13, 761-73.

Coffin, J. M., Hughes, S. H., and Varmus, H. (1997). Retroviruses (Plainview, N.Y.: Cold Spring Harbor Laboratory Press).

Cordingley, M. G., LaFemina, R. L., Callahan, P. L., Condra, J. H., Sardana, V. V., Graham, D. J., Nguyen, T. M., LeGrow, K., Gotlib, L., Schlabach, A. J., and et al. (1990). Sequence-specific interaction of Tat protein and Tat peptides with the transactivation-responsive sequence element of human immunodeficiency virus type 1 in vitro. Proceedings of the National Academy of Sciences of the United States of America 87, 8985-9.

Dingwall, C., Ernberg, I., Gait, M. J., Green, S. M., Heaphy, S., Karn, J., Lowe, A. D., Singh, M., Skinner, M. A., and Valerio, R. (1989). Human immunodeficiency virus 1 tat protein binds trans-activation-responsive region (TAR) RNA in vitro. Proceedings of the National Academy of Sciences of the United States of America 86, 6925-9.

Dorfman, T., Mammano, F., Haseltine, W. A., and Göttlinger, H. G. (1994a). Role of the matrix protein in the virion association of the human immunodeficiency virus type 1 envelope glycoprotein. Journal of Virology 68, 1689-96.

Dorfman, T., Bukovsky, A., Ohagen, A., Höglund, S., and Göttlinger, H. G. (1994b). Functional domains of the capsid protein of human immunodeficiency virus type 1. Journal of Virology 68, 8180-7.

Ellerman, V., and Bang, O. (1908). Experimentelle Leukamie bei Huhnern. Zentralbl. Bakteriol. Parasitenkd. Infectionskr. Hyg. Abt. Orig. 46, 595-609.

Fäcke, M., Janetzko, A., Shoeman, R. L., and Kräusslich, H. G. (1993). A large deletion in the matrix domain of the human immunodeficiency virus gag gene redirects virus particle assembly from the plasma membrane to the endoplasmic reticulum. Journal of Virology 67, 4972-80.

Feng, S., and Holland, E. C. (1988). HIV-1 tat trans-activation requires the loop sequence within tar. Nature 334, 165-7.

Fouchier, R. A., Meyer, B. E., Simon, J. H., Fischer, U., and Malim, M. H. (1997). HIV-1 infection of non-dividing cells: evidence that the amino-terminal basic region of the viral matrix protein is important for Gag processing but not for post-entry nuclear import. Embo Journal 16, 4531-9.

Franke, E. K., Yuan, H. E., and Luban, J. (1994). Specific incorporation of cyclophilin A into HIV-1 virions [see comments]. Nature 372, 359-62.

Freed, E. O., Orenstein, J. M., Buckler-White, A. J., and Martin, M. A. (1994). Single amino acid changes in the human immunodeficiency virus type 1 matrix protein block virus particle production. Journal of Virology 68, 5311-20.

Freed, E. O., Englund, G., and Martin, M. A. (1995). Role of the basic domain of human immunodeficiency virus type 1 matrix in macrophage infection. Journal of Virology 69, 3949-54.

Gait, M. J., and Karn, J. (1993). RNA recognition by the human immunodeficiency virus Tat and Rev proteins. Trends in Biochemical Sciences 18, 255-9.

Gallay, P., Swingler, S., Song, J., Bushman, F., and Trono, D. (1995a). HIV nuclear import is governed by the phosphotyrosine-mediated binding of matrix to the core domain of integrase. Cell 83, 569-76.

Gallay, P., Swingler, S., Aiken, C., and Trono, D. (1995b). HIV-1 infection of nondividing cells: C-terminal tyrosine phosphorylation of the viral matrix protein is a key regulator. Cell 80, 379-88.

Gamble, T. R., Vajdos, F. F., Yoo, S., Worthylake, D. K., Houseweart, M., Sundquist, W. I., and Hill, C. P. (1996). Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. Cell 87, 1285-94.

Gitti, R. K., Lee, B. M., Walker, J., Summers, M. F., Yoo, S., and Sundquist, W. I. (1996). Structure of the amino-terminal core domain of the HIV-1 capsid protein. Science 273, 231-5.

Gottlieb, M. S., Schroff, R., Schanker, H. M., Weisman, J. D., Fan, P. T., Wolf, R. A., and Saxon, A. (1981). Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. New England Journal of Medicine 305, 1425-31.

Göttlinger, H. G., Sodroski, J. G., and Haseltine, W. A. (1989). Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. Proceedings of the National Academy of Sciences of the United States of America 86, 5781-5.

Haseltine, W. A., Panet, A., Smoler, D., Baltimore, D., Peters, G., Harada, F., and Dahlberg, J. E. (1977). Interaction of tryptophan tRNA and avian myeloblastosis virus reverse transcriptase: further characterization of the binding reaction. Biochemistry 16, 3625-32.

He, J., and Landau, N. R. (1995). Use of a novel human immunodeficiency virus type 1 reporter virus expressing human placental alkaline phosphatase to detect an alternative viral receptor. Journal of Virology 69, 4587-92.

Hensel, M., Shea, J. E., Gleeson, C., Jones, M. D., Dalton, E., and Holden, D. W. (1995). Simultaneous identification of bacterial virulence genes by negative selection. Science 269, 400-3.

Jowett, J. B., Hockley, D. J., Nermut, M. V., and Jones, I. M. (1992). Distinct signals in human immunodeficiency virus type 1 Pr55 necessary for RNA binding and particle formation [published erratum appears in J Gen Virol 1993 May;74(Pt 5):943]. Journal of General Virology 73, 3079-86.

LaFemina, R. L., Callahan, P. L., and Cordingley, M. G. (1991). Substrate specificity of recombinant human immunodeficiency virus integrase protein. Journal of Virology 65, 5624-30.

Laughrea, M., and Jetté, L. (1997a). HIV-1 genome dimerization: kissing-loop hairpin dictates whether nucleotides downstream of the 5' splice junction contribute to loose and tight dimerization of human immunodeficiency virus RNA. Biochemistry 36, 9501-8.

Laughrea, M., Jetté, L., Mak, J., Kleiman, L., Liang, C., and Wainberg, M. A. (1997b). Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and dimerization. Journal of Virology 71, 3397-406.

Leavitt, A. D., Rose, R. B., and Varmus, H. E. (1992). Both substrate and target oligonucleotide sequences affect in vitro integration mediated by human immunodeficiency virus type 1 integrase protein produced in Saccharomyces cerevisiae. Journal of Virology 66, 2359-68.

Leis, J., Aiyar, A., and Cobrinik, D. (1993). Regulation of initiation of reverse transcription in retroviruses. In Reverse Transcriptase, A. M. Skalka and S. P. Goff, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 33-48.

Lever, A., Gottlinger, H., Haseltine, W., and Sodroski, J. (1989). Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. Journal of Virology 63, 4085-7.

Luban, J., Bossolt, K. L., Franke, E. K., Kalpana, G. V., and Goff, S. P. (1993). Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. Cell 73, 1067-78.

Luban, J., and Goff, S. P. (1994). Mutational analysis of cis-acting packaging signals in human immunodeficiency virus type 1 RNA. Journal of Virology 68, 3784-93.

Mammano, F., Ohagen, A., Höglund, S., and Göttlinger, H. G. (1994). Role of the major homology region of human immunodeficiency virus type 1 in virion morphogenesis. Journal of Virology 68, 4927-36.

Masur, H., Michelis, M. A., Greene, J. B., Onorato, I., Stouwe, R. A., Holzman, R. S., Wormser, G., Brettman, L., Lange, M., Murray, H. W., and Cunningham-Rundles, S. (1981). An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. New England Journal of Medicine 305, 1431-8.

McBride, M. S., and Panganiban, A. T. (1996). The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures [published erratum appears in J Virol 1997 Jan;71(1):858]. Journal of Virology 70, 2963-73.

Morgenstern, J. P., and Land, H. (1990). Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. Nucleic Acids Research 18, 3587-96.

Popovic, M., Sarngadharan, M. G., Read, E., and Gallo, R. C. (1984). Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science 224, 497-500.

Pryciak, P. M., and Varmus, H. E. (1992). Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. Cell 69, 769-80.

Reicin, A. S., Paik, S., Berkowitz, R. D., Luban, J., Lowy, I., and Goff, S. P. (1995). Linker insertion mutations in the human immunodeficiency virus type 1 gag gene: effects on virion particle assembly, release, and infectivity. Journal of Virology 69, 642-50.

Reicin, A. S., Ohagen, A., Yin, L., Hoglund, S., and Goff, S. P. (1996). The role of Gag in human immunodeficiency virus type 1 virion morphogenesis and early steps of the viral life cycle. Journal of Virology 70, 8645-52.

Rhim, H., Park, J., and Morrow, C. D. (1991). Deletions in the tRNA(Lys) primer-binding site of human immunodeficiency virus type 1 identify essential regions for reverse transcription. Journal of Virology 65, 4555-64.

Rous, P. (1911). A sarcoma of the fowl transmissible by an agent separable from the tumor cells. J. Exp. Med. 13, 397-411.

Roy, S., Delling, U., Chen, C. H., Rosen, C. A., and Sonenberg, N. (1990). A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. Genes and Development 4, 1365-73.

Roy, S., Parkin, N. T., Rosen, C., Itovitch, J., and Sonenberg, N. (1990). Structural requirements for trans activation of human immunodeficiency virus type 1 long terminal repeat-directed gene expression by tat: importance of base pairing, loop sequence, and bulges in the tat-responsive sequence. Journal of Virology 64, 1402-6.

Sakaguchi, K., Zambrano, N., Baldwin, E. T., Shapiro, B. A., Erickson, J. W., Omichinski, J. G., Clore, G. M., Gronenborn, A. M., and Appella, E. (1993). Identification of a binding site for the human immunodeficiency virus type 1 nucleocapsid protein. Proceedings of the National Academy of Sciences of the United States of America 90, 5219-23.

Sambrook, J., Maniatis, T., and Fritsch, E. F. (1989). Molecular cloning : a laboratory manual, 2nd Edition (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory).

Savilahti, H., Rice, P. A., and Mizuuchi, K. (1995). The phage Mu transpososome core: DNA requirements for assembly and function. Embo Journal 14, 4893-903.

Schwartz, D. E., Zamecnik, P. C., and Weith, H. L. (1977). Rous sarcoma virus genome is terminally redundant: the 3' sequence. Proceedings of the National Academy of Sciences of the United States of America 74, 994-8.

Selby, M. J., Bain, E. S., Luciw, P. A., and Peterlin, B. M. (1989). Structure, sequence, and position of the stem-loop in tar determine transcriptional elongation by tat through the HIV-1 long terminal repeat. Genes and Development 3, 547-58.

Shaw, G. M., Hahn, B. H., Arya, S. K., Groopman, J. E., Gallo, R. C., and Wong-Staal, F. (1984). Molecular characterization of human T-cell leukemia (lymphotropic) virus type III in the acquired immune deficiency syndrome. Science 226, 1165-71.

Sherman, P. A., Dickson, M. L., and Fyfe, J. A. (1992). Human immunodeficiency virus type 1 integration protein: DNA sequence requirements for cleaving and joining reactions. Journal of Virology 66, 3593-601.

Siegal, F. P., Lopez, C., Hammer, G. S., Brown, A. E., Kornfeld, S. J., Gold, J., Hassett, J., Hirschman, S. Z., Cunningham-Rundles, C., Adelsberg, B. R., and et al. (1981). Severe acquired immunodeficiency in male homosexuals, manifested by chronic perianal ulcerative herpes simplex lesions. New England Journal of Medicine 305, 1439-44.

Singh, I. R., Crowley, R. A., and Brown, P. O. (1997). High-resolution functional mapping of a cloned gene by genetic footprinting. Proceedings of the National Academy of Sciences of the United States of America 94, 1304-9.

Smith, V., Chou, K. N., Lashkari, D., Botstein, D., and Brown, P. O. (1996). Functional analysis of the genes of yeast chromosome V by genetic footprinting. Science 274, 2069-74.

Stoll, E., Billeter, M. A., Palmenberg, A., and Weissmann, C. (1977). Avian myeloblastosis virus RNA is terminally redundant: implications for the mechanism of retrovirus replication. Cell 12, 57-72.

Sutton, R. E., Wu, H. T., Rigg, R., Böhnlein, E., and Brown, P. O. (1998). Human immunodeficiency virus type 1 vectors efficiently transduce human hematopoietic stem cells. Journal of Virology 72, 5781-8.

Temin, H. M., and Mizutani, S. (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature 226, 1211-3.

Thali, M., Bukovsky, A., Kondo, E., Rosenwirth, B., Walsh, C. T., Sodroski, J., and Göttlinger, H. G. (1994). Functional association of cyclophilin A with HIV-1 virions [see comments]. Nature 372, 363-5.

van den Ent, F. M., Vink, C., and Plasterk, R. H. (1994). DNA substrate requirements for different activities of the human immunodeficiency virus type 1 integrase protein. Journal of Virology 68, 7825-32.

Van Lint, C., Burny, A., and Verdin, E. (1991). The intragenic enhancer of human immunodeficiency virus type 1 contains functional AP-1 binding sites. Journal of Virology 65, 7066-72.

Verdin, E., Becker, N., Bex, F., Droogmans, L., and Burny, A. (1990). Identification and characterization of an enhancer in the coding region of the genome of human immunodeficiency virus type 1. Proceedings of the National Academy of Sciences of the United States of America 87, 4874-8.

Vicenzi, E., Dimitrov, D. S., Engelman, A., Migone, T. S., Purcell, D. F., Leonard, J., Englund, G., and Martin, M. A. (1994). An integration-defective U5 deletion mutant of human immunodeficiency virus type 1 reverts by eliminating additional long terminal repeat sequences. Journal of Virology 68, 7879-90.

von Poblotzki, A., Wagner, R., Niedrig, M., Wanner, G., Wolf, H., and Modrow, S. (1993). Identification of a region in the Pr55gag-polyprotein essential for HIV-1 particle formation. Virology 193, 981-5.

von Schwedler, U., Kornbluth, R. S., and Trono, D. (1994). The nuclear localization signal of the matrix protein of human immunodeficiency virus type 1 allows the establishment of infection in macrophages and quiescent T lymphocytes. Proceedings of the National Academy of Sciences of the United States of America 91, 6992-6.

Wang, C. T., and Barklis, E. (1993). Assembly, processing, and infectivity of human immunodeficiency virus type 1 gag mutants. Journal of Virology 67, 4264-73.

Wlodawer, A., and Erickson, J. W. (1993). Structure-based inhibitors of HIV-1 protease. Annual Review of Biochemistry 62, 543-85.

Yu, X., Yu, Q. C., Lee, T. H., and Essex, M. (1992a). The C terminus of human immunodeficiency virus type 1 matrix protein is involved in early steps of the virus life cycle. Journal of Virology 66, 5667-70.

Yu, X., Yuan, X., Matsuda, Z., Lee, T. H., and Essex, M. (1992b). The matrix protein of human immunodeficiency virus type 1 is required for incorporation of viral envelope protein into mature virions. Journal of Virology 66, 4966-71.

Zhou, W., Parent, L. J., Wills, J. W., and Resh, M. D. (1994). Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. Journal of Virology 68, 2556-69.

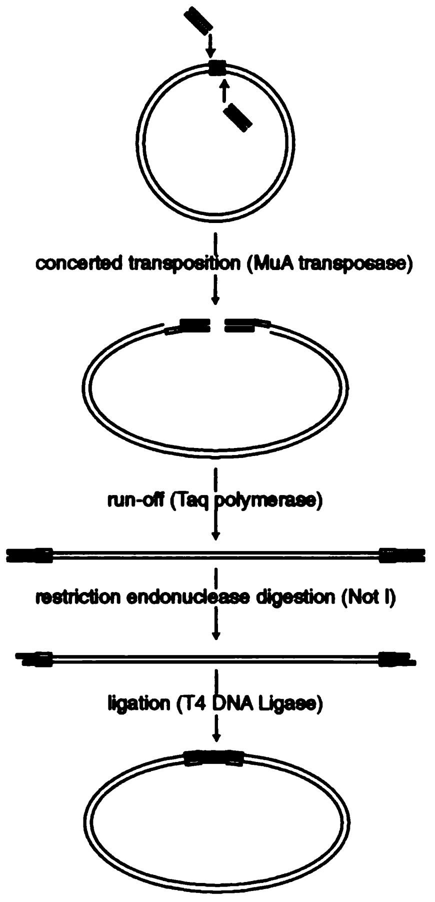

concerted transposition (MuA transposase)
run-off (Taq polymerase)
restriction endonuclease digestion (Not I)
ligation (T4 DNA Ligase)

Figure 1. Mutagenesis scheme using MuA tranposase. Oligonucleotides used for mutagenesis are bold lines (▬), the target DNA plasmid is drawn as thin lines, and sequences in the target DNA duplicated during mutagenesis are empty lines (▭).

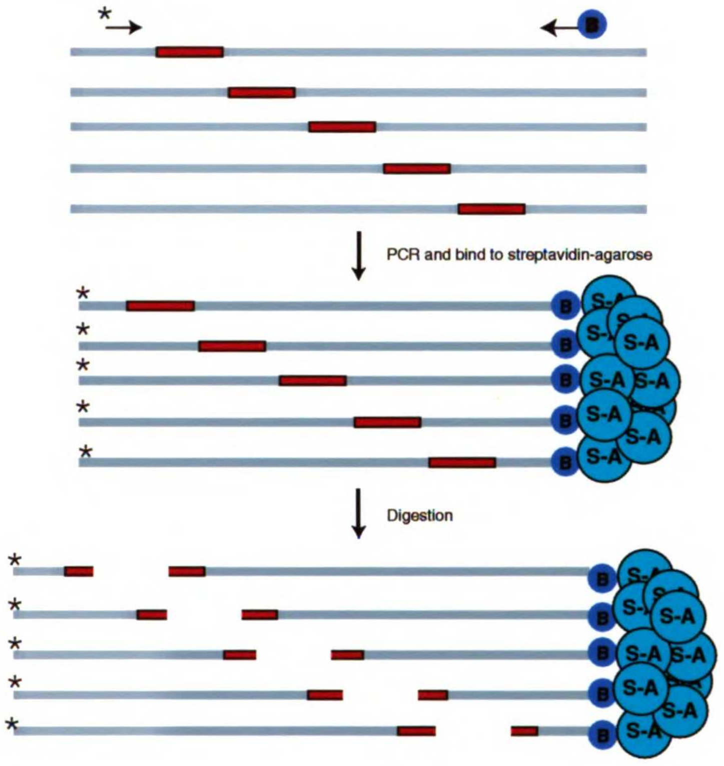


Figure 2. Genetic footprinting scheme using flanking PCR and restriction digestion. A collection of insertion mutants is subjected to PCR using one radioactively labelled primer (*→) and one biotinylated primer (←B). The PCR products are bound to streptavidin-agarose resin (S-A) and digested with a restriction enzyme that recognizes a site in the insertion sequence. The radioactively labelled ends of the PCR products are released.

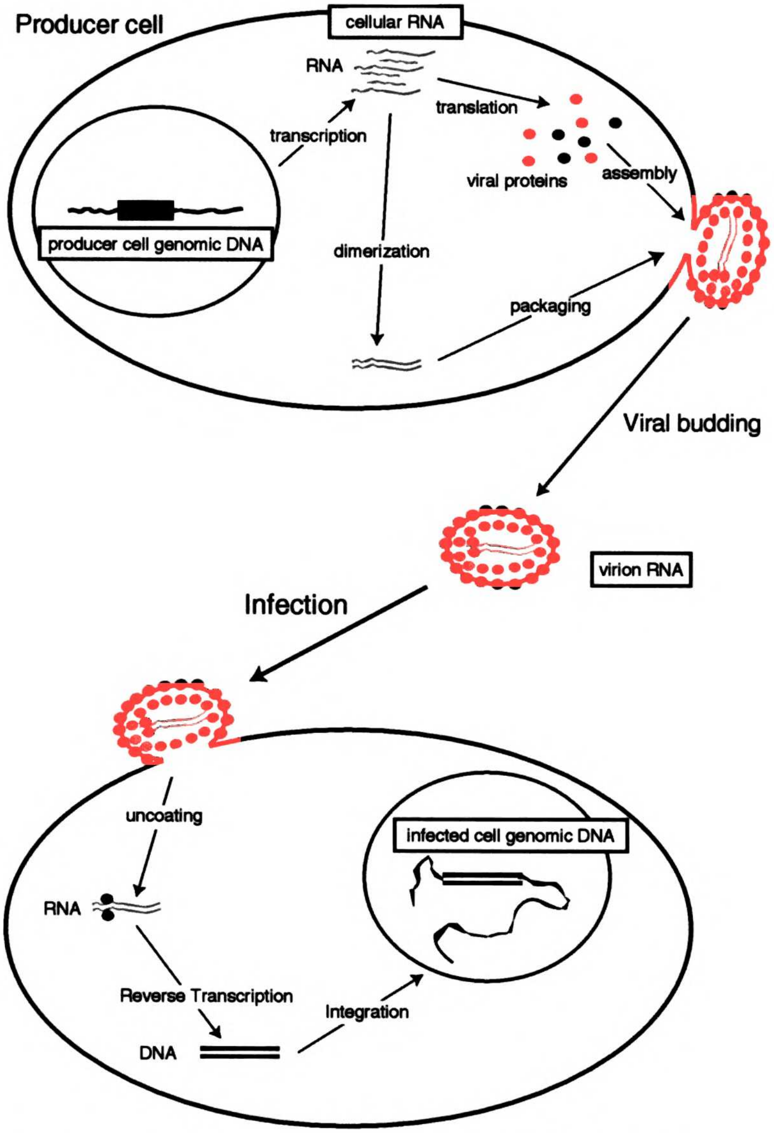


Figure 3. The retroviral life cycle. Nucleic acid samples analyzed in this study are boxed.

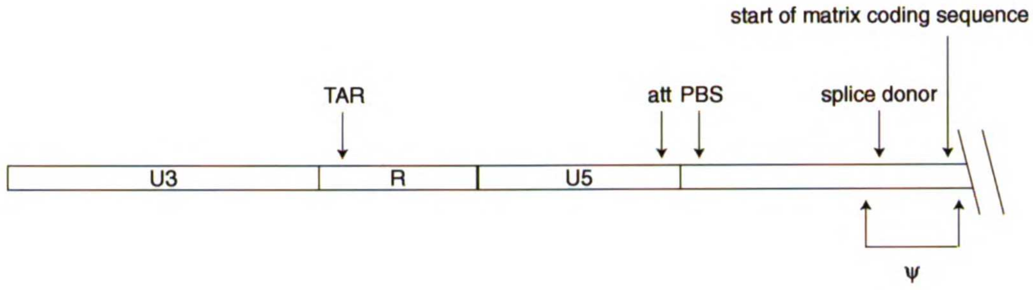


Figure 4A. .Schematic representation of the 5' end of the HIV genome, which includes TAR, the U5 att site, the primer binding site (PBS), the splice donor for the env message, the start of the matrix coding sequence, and the packaging sequence (ψ).

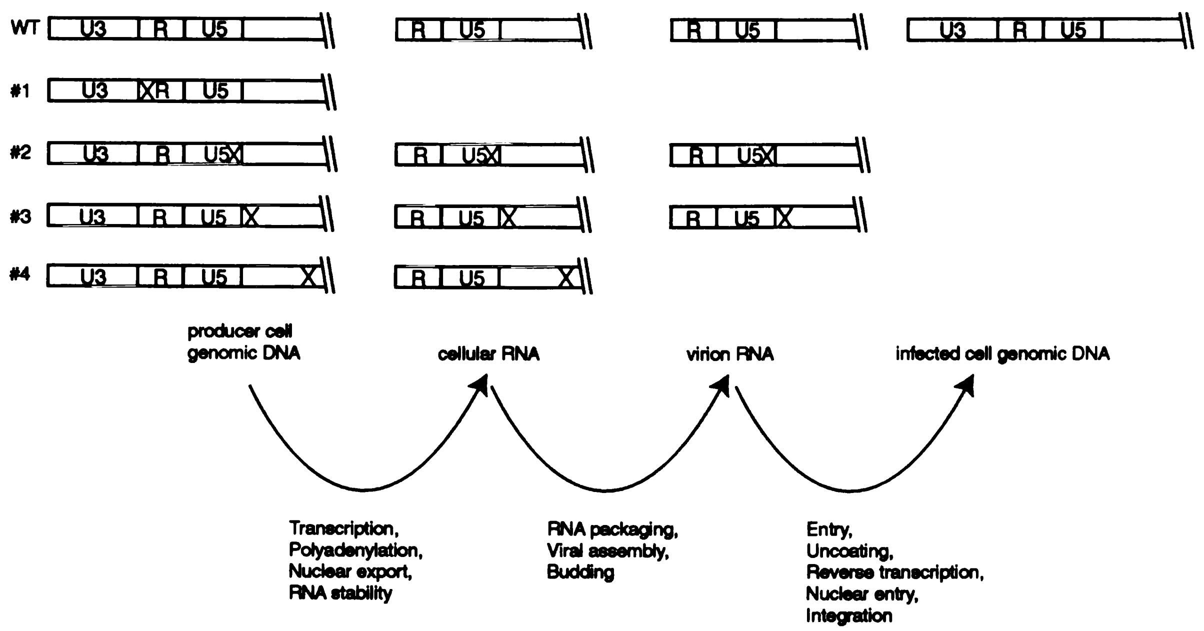


Figure 4B. Selection of mutants defective in TAR (#1), att site(#2), primer binding site (#3), and packaging sequence (#4) functions at different steps during the viral life cycle and the nucleic acid samples where one would observe these selections. The illustrated mutations and the phenotypes depicted for them are based on results from previous reports. "WT" indicates a wild-type, replication-competent viral genome.

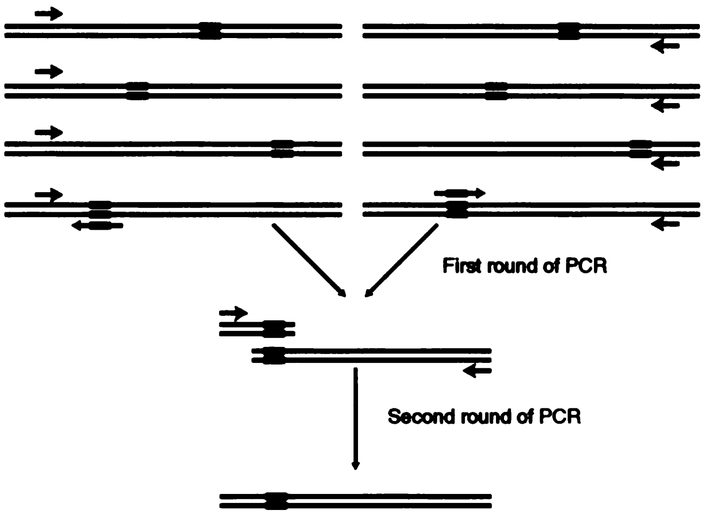


Figure 5. Isolating a specific mutant by PCR. Sequences introduced during the mutagenesis procedure are in bold lines (■). Primers are indicated by arrows. Note that primers that contains both wild-type and mutant sequences (←▬–) will selectively prime off of template DNAs that have a mutation at the selected site.

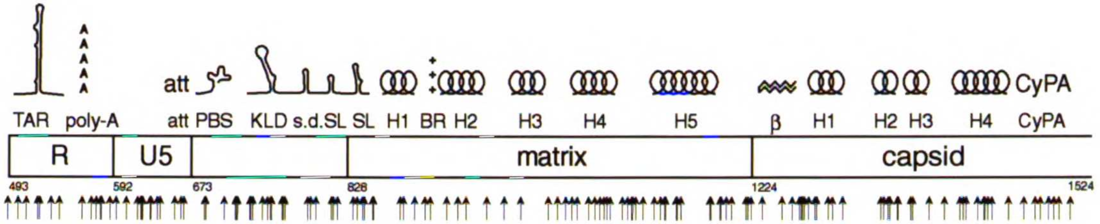
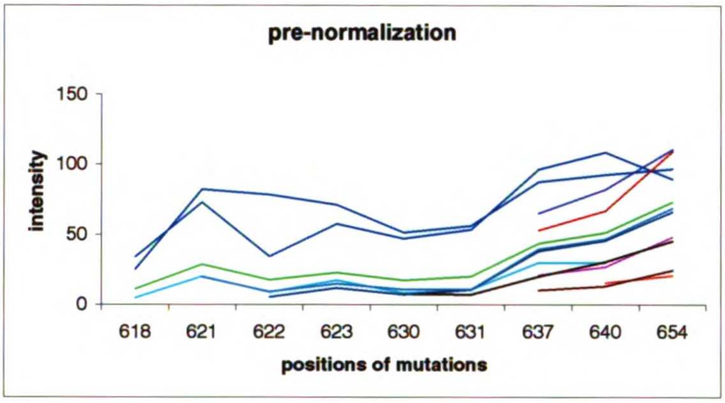


Figure 6. Diagram of mutations evaluated in this study mapped onto the HIV genome. Features described previously are TAR (TAR), the polyadenylation signal (poly-A), the att site (att), the primer binding site (PBS), the kissing loop domain (KLD), the splice donor (s.d.)for the env message, two gag-binding stem-loop structures (SL), helices 1-5 (H1-H5) of matrix, the basic region of matrix (BR), the beta hairpin of capsid (β), helices 1-4 (H1-H4) of capsid, and the cyclophilin A binding region of capsid (CyPA). Numbers indicate nucleotide positions at the borders of major regions in the HIV genome. Arrows indicate positions of insertional mutations evaluated in this study.



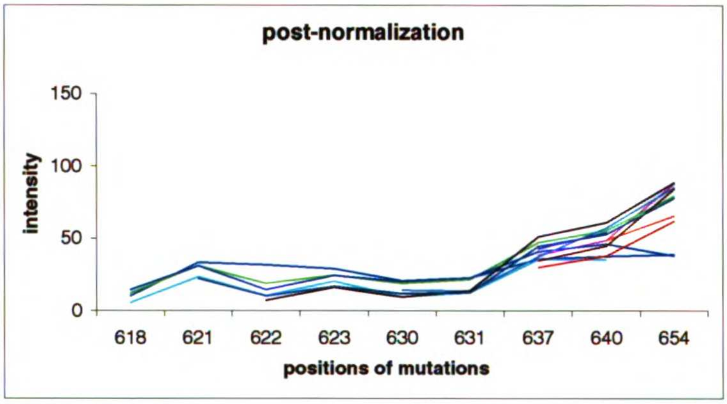


Figure 7. Example of quantitative data before and after normalization. Data shown is from genetic footprinting experiments on the library of mutagenized proviruses. Nucleotide positions of mutations for which data were obtained are given on the X-axis. Intensity of bands measured from autoradiograms is given in arbitrary units on the Y-axis. The different colored traces represent data measured from different gels.

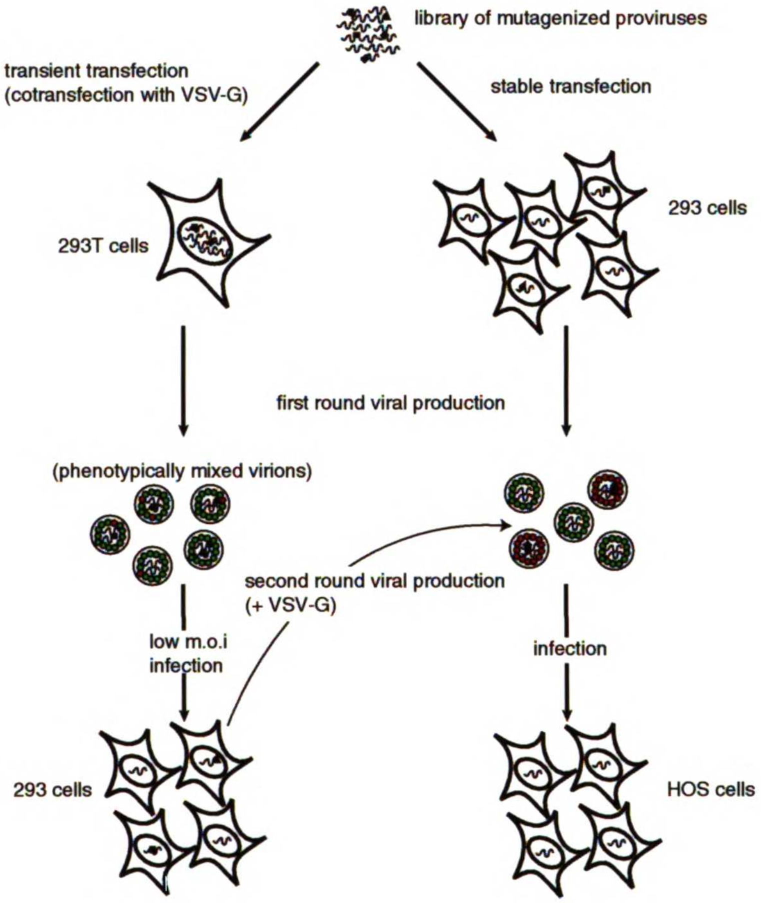

library of mutagenized proviruses
transient transfection
(cotransfection with VSV-G)
stable transfection
293T cells
293 cells
first round viral production
(phenotypically mixed virions)
second round viral production
(+ VSV-G)
low m.o.i
infection
infection
293 cells
HOS cells

Figure 8. Diagram of transfection and infection experiments. The library of mutagenized proviruses was either transiently or stably transfected into cells to produce populations of mutant virions. In our experiments, the viral genomes of mutants defective in trans-acting factors were efficiently rescued during the transient transfection by phenotypic mixing, but were not detectably rescued during the stable transfection. The virus produced in the transient transfection experiment was used to infect fresh, uninfected cells at a low multiplicity, resulting in a population of producer cells that contained a single provirus per cell. Wild-type viralgenome(∿), replication-defective viral genome with mutation in trans-acting factor (∿),wild-type viral protein (•), mutant viral protein (•).

| % mutant library sample | | | | | |
|---|---|---|---|---|---|
| mutant library | cellular RNA | virion RNA | genomic DNA | | Nucleic Acid |
| 100 | 80 | 13 | 6 | **1120** | AAGACACCAA TGCGGCCGCA ACCAAGGAAG |
| 100 | 158 | 23 | 17 | **1126** | CCAAGGAAGC TGCGGCCGCA GAAGCCTTAG |
| 100 | 91 | 60 | 62 | **1141** | ATAAGATAGA TGCGGCCGCA ATAGAGGAAG |
| 100 | 60 | 42 | 58 | **1142** | TAAGATAGAG TGCGGCCGCA TAGAGGAAGA |
| 100 | 93 | 80 | 76 | **1146** | ATAGAGGAAG TGCGGCCGCA GGAAGAGCAA |
| 100 | 149 | 136 | 137 | **1174** | AAAAGGCACA TGCGGCCGCA GCACAGCAAG |
| 100 | 105 | 45 | 65 | **1175** | AAAGGCACAG TGCGGCCGCA CACAGCAAGC |
| 100 | 109 | 188 | 121 | **1184** | GCAAGCAGCA TGCGGCCGCA CAGCAGCTGA |
| 100 | 105 | 446 | 207 | **1185** | CAAGCAGCAG TGCGGCCGCA AGCAGCTGAC |
| 100 | 206 | 194 | 249 | **1190** | CAGCAGCTGA TGCGGCCGCA GCTGACACAG |
| 100 | 135 | 89 | 123 | **1195** | CTGACACAGG TGCGGCCGCA ACAGGAAACA |
| 100 | 78 | 64 | 63 | **1209** | AACAGCCAGG TGCGGCCGCA CCAGGTCAGC |
| 100 | 85 | 108 | 136 | **1211** | CAGCCAGGTC TGCGGCCGCA AGGTCAGCCA |

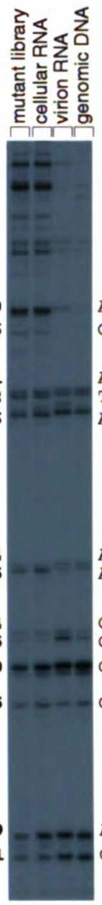


Figure 9. Genetic footprinting of library of mutagenized proviruses and nucleic acid samples from the transient transfection experiment, second round (the uncomplemented round) of viral production and infection (cellular RNA, viral RNA, and infected cell genomic DNA). Numbers directly to the left of the gel indicate exact positions of insertions. The first nucleotide of the HIV provirus is at position 37. Quantitative data averaged from normalized measurements are also given to the left of the gel. The nucleic acid sequences of the mutants are written to right of the gel. The sequences derived from the insertion oligonucleotide are bpxed, while the target sequence duplications are underlined.

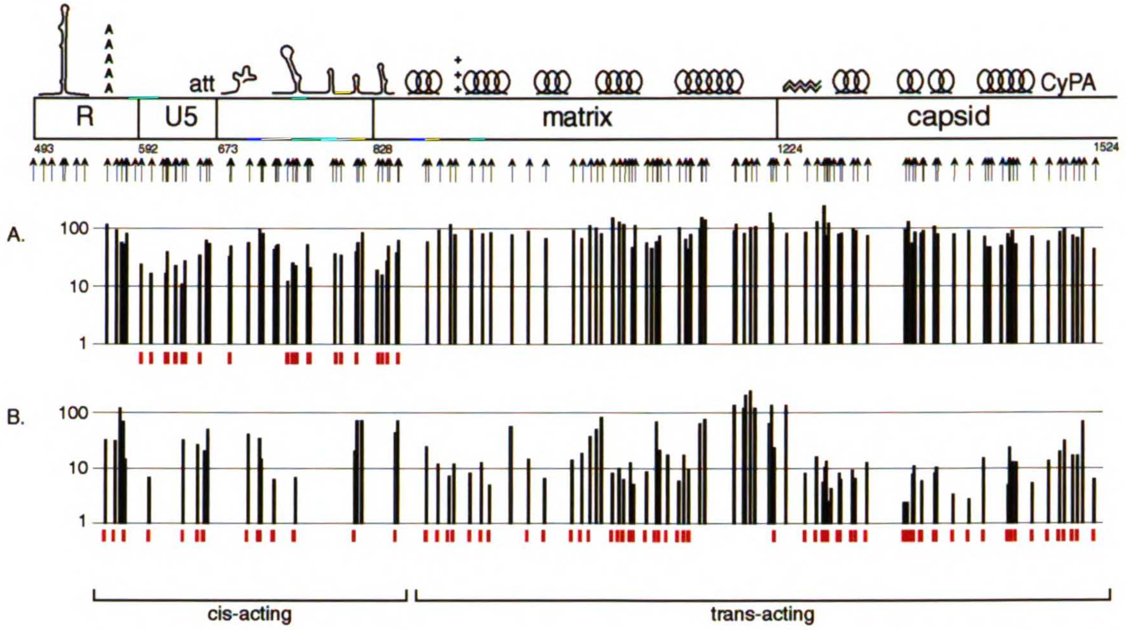


Figure 10. Percent recovery of mutants through single-cycle infections. Data are not shown for mutants where the coefficient of variation between triplicate experiments was greater than 0.5, except in cases where the observed phenotypes were confirmed by re-analysis. Graphs are plotted on a log scale. Red bars indicate mutants that display significant depletions (<45% recovery). Mutations that compromise replication both in the presence and the absence of complementation are considered to be located in cis-acting sequences, while those that affect only uncomplemented replication cycles are considered to be located in trans-acting sequences. A. Percent recovery of mutants after a single round of infectionin the presence of complementation. Data are from the transient transfection experiment, first round of infection. B. Percent recovery of mutants after a single round of infection in the absence of complementation. Data are from the transient transfection experiment, second round of infection. Data are not given for mutants whose abundance was very low after the first round of infection.

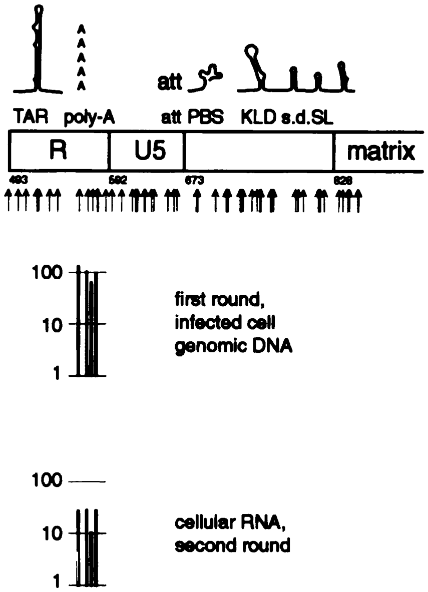


Figure 11. Behavior of selected mutants in cis-acting sequences in the transient transfection experiment. These mutants are replication competent in the first round of infection, but defective for transcript formation in the second round, possibly indicating that the cis-acting element interrupted by the insertions is active in the 3' LTR.

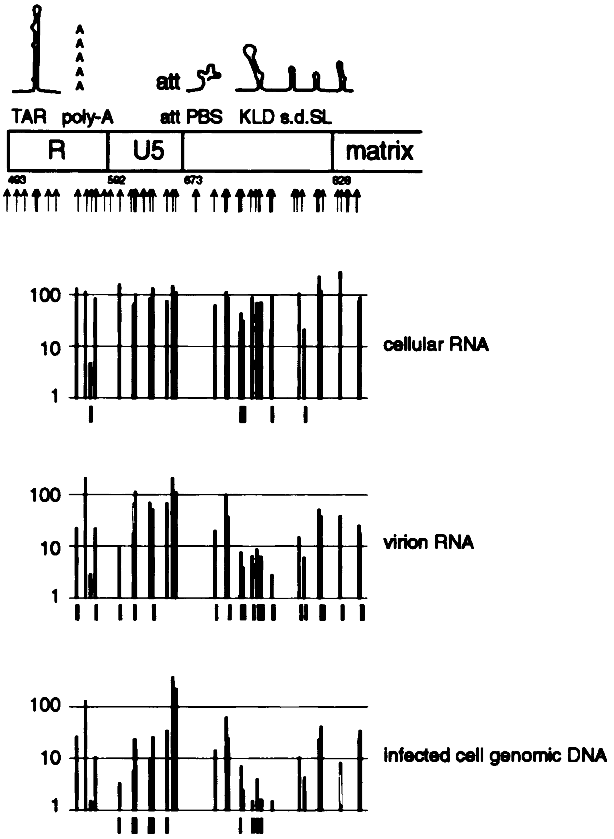


Figure 12. Percent recovery of mutations in cis-acting sequences at several steps of a single-cycle uncomplemented infection. Samples were collected from the stable transfection experiment, second round of viral production and infection. Percent recovery was calculated by dividing the abundance of a mutant in a given nucleic acid sample by the abundance of that mutant in the genomic DNA sample from the stable transfection. Data are not shown for points where the abundance of a particular mutant was very low in the preceding nucleic acid sample, or the coefficient of variation between triplicate experiments was greater than 0.5. Graphs are plotted on a log scale. Red bars indicate mutants that display significant depletions (<50% of preceding nucleic acid sample). Mutants which were depleted in the cellular RNA sample are considered to be defective in transcript formation, mutants which were depleted in the virion RNA sample are considered to be defective in packaging, and mutants which were depleted in the infected cell genomic DNA sample are considered to be defective in an early step in viral replication.

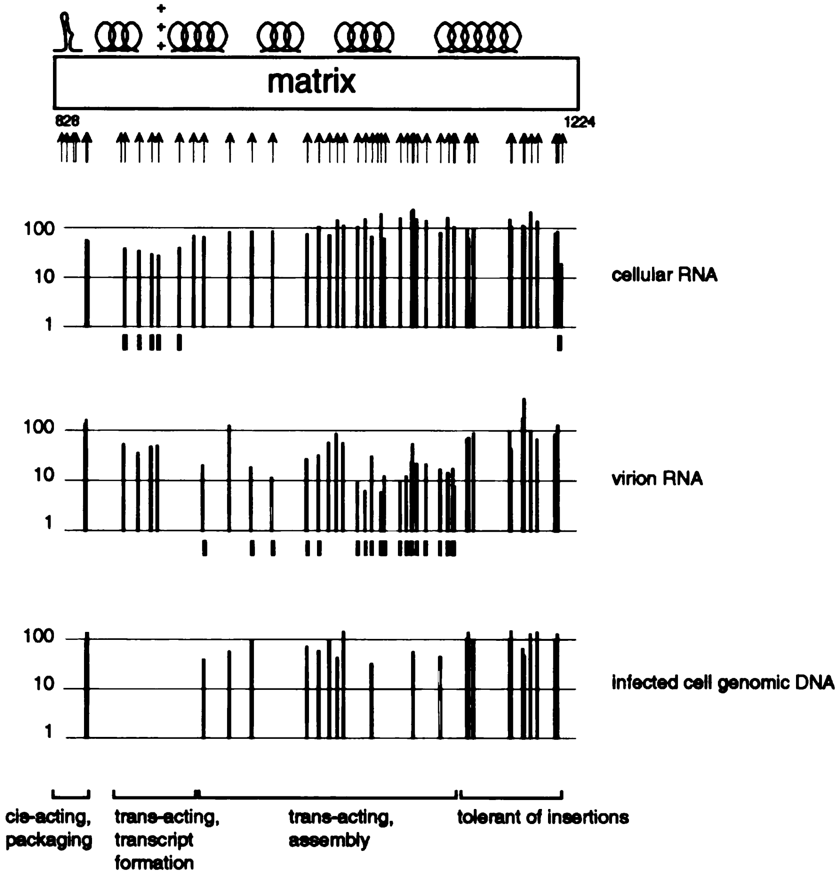


Figure 13. Percent recovery of mutations in matrix at several steps of a single-cycle uncomplemented infection. Samples were collected from the transient transfection experiment, second round of viral production and infection. Percent recovery was calculated by dividing the abundance of a mutant in a given nucleic acid sample by the abundance of that mutant in the infected cell genomic DNA sample from the first round of infection. Data are not shown for points where the abundance of a particular mutant was very low in the preceding nucleic acid sample, or the coefficient of variation between triplicate experiments was greater than 0.5. Graphs are plotted on a log scale. Red bars indicate mutants that display significant depletions (<45% of preceding nucleic acid sample). Mutants which were depleted in the cellular RNA sample are considered to be defective in transcript formation, mutants which were depleted in the virion RNA sample are considered to be defective in assembly, and mutants which were depleted in the infected cell genomic DNA sample are considered to be defective in an early step in viral replication.

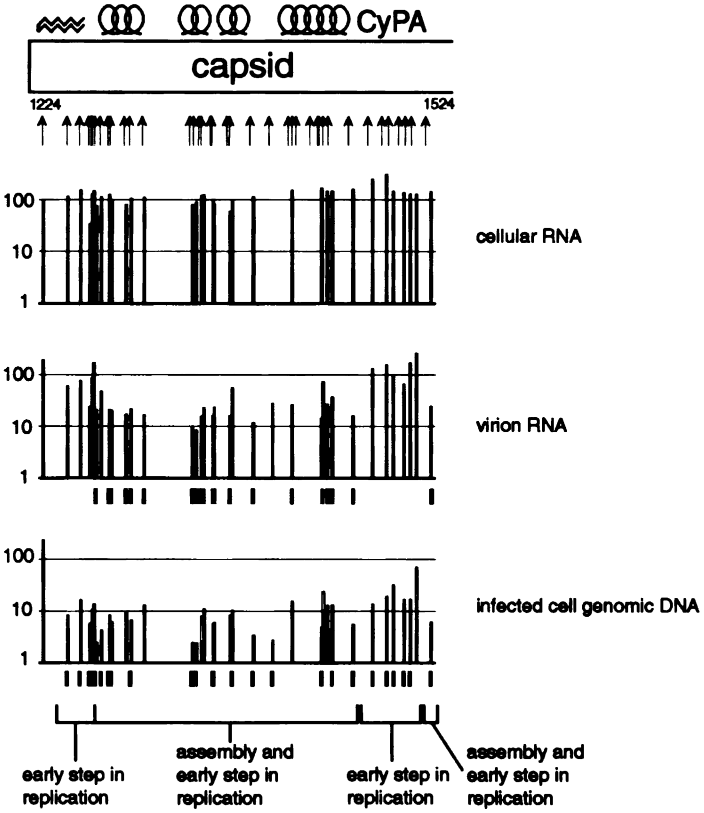


Figure 14. Percent recovery of mutations in capsid at several steps of a single-cycle uncomplemented infection. Samples were collected from the transient transfection experiment, second round of viral production and infection. Percent recovery was calculated by dividing the abundance of a mutant in a given nucleic acid sample by the abundance of that mutant in the infected cell genomic DNA sample from the first round of infection. Data are not shown for points where the abundance of a particular mutant was very low in the preceding nucleic acid sample, or the coefficient of variation between triplicate experiments was greater than 0.5. Graphs are plotted on a log scale. Red bars indicate mutants that display significant depletions (<45% of preceding nucleic acid sample). Mutants which were depleted in the cellular RNA sample are considered to be defective in transcript formation, mutants which were depleted in the virion RNA sample are considered to be defective in assembly, and mutants which were depleted in the infected cell genomic DNA sample are considered to be defective in an early step in viral replication. All mutations in capsid that affected viral replication showed their effects in trans.

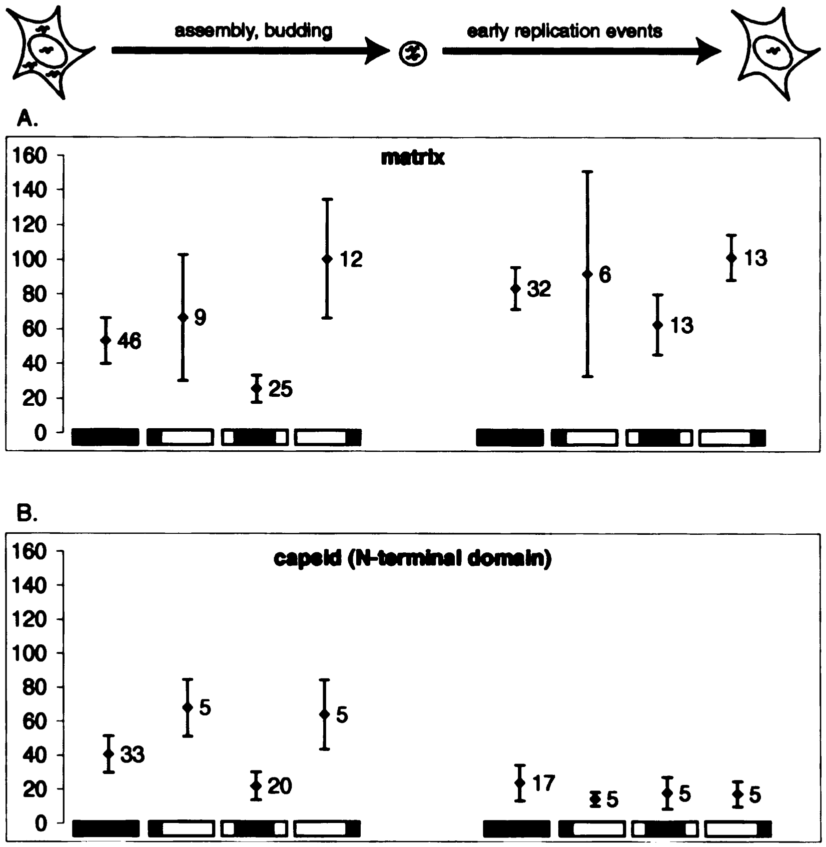


Figure 15. Percent recovery of matrix and capsid mutants through the viral assembly process and the early steps of the viral life cycle. Data are derived from single-round uncomplemented viral production and infection cycles. A schematic of the phases of the life cycle tested is drawn above the graphs. Numbers to the right of each point indicate the number of mutants from which data were averaged. Below each data point is a schematic of the region of matrix or capsid evaluated. Error bars indicate 95% confidence intervals. A. Data for the matrix protein. Data are shown for insertions between amino acid positions 1-132 (complete matrix protein), 1-35 (N-terminal region), 36-102 (central region), and 104-130 (C-terminal region). B. Data for the N-terminal half of the capsid protein. Data are shown for insertions bewteen amino acid positions 1-101 (N-terminal half), 1-15 (N-terminal beta hairpin), 17-81 (central helical region), and 85-96 (Cyclophilin A binding region).

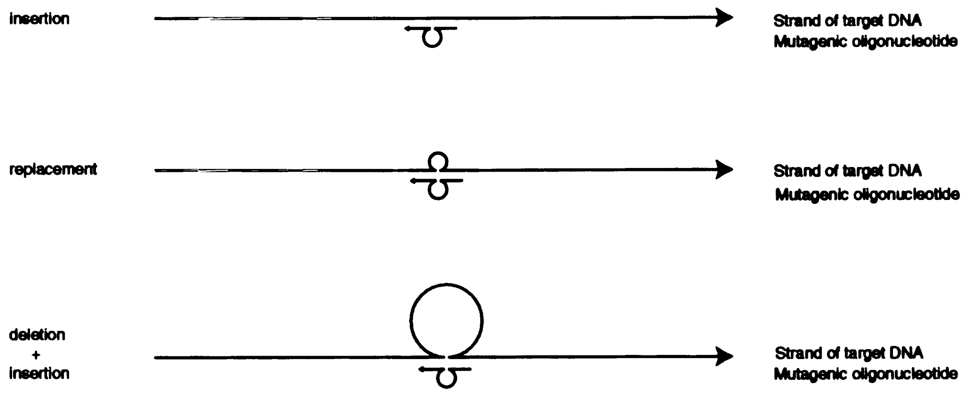


Figure 16. Design of oligonucleotides used to make different types of mutations by PCR.

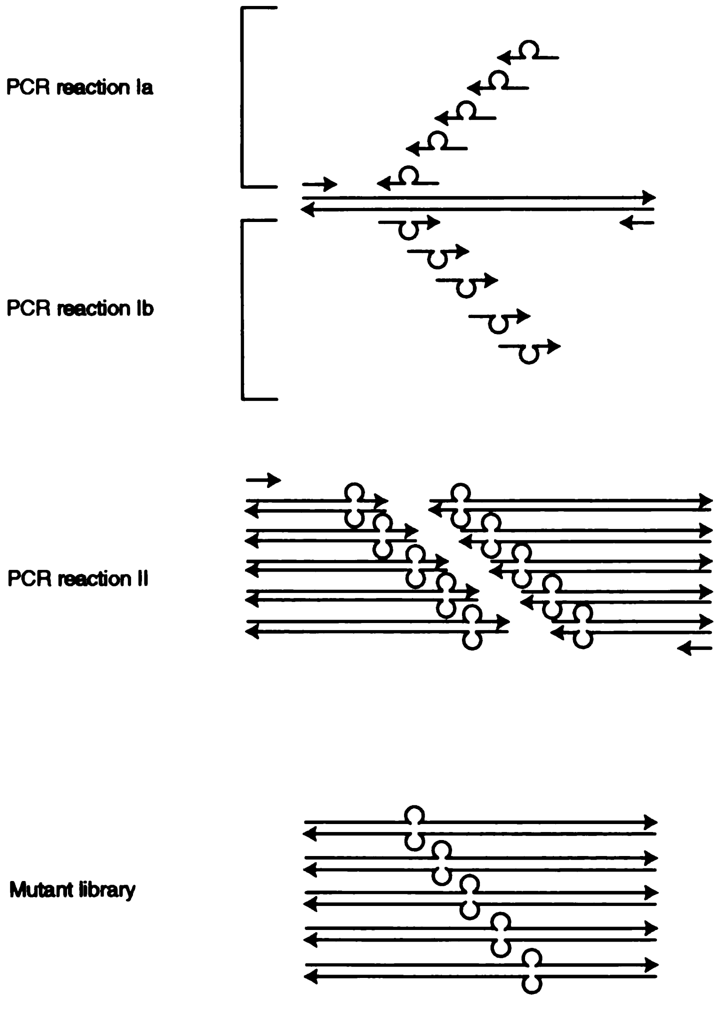


Figure 17. Strategy for achieving saturating mutagenesis of a stretch of DNA using PCR.

# Appendix A.
# Complete nucleotide sequence of HIV puro (the plasmid used for mutagenesis).

```
   1 ACATGTAGCC CCAGTTCTAC TTACACCAAG AAAGGCTGGA AGGGCTAATT CACTCCCAAA
  61 GAAGACAAGA TATCCTTGAT CTGTGGATCT ACCACACACA AGGCTACTTC CCTGATTGGC
 121 AGAACTACAC ACCAGGGCCA GGGGTCAGAT ATCCACTGAC CTTTGGATGG TGCTACAAGC
 181 TAGTACCAGT TGAGCCAGAT AAGGTAGAAG AGGCCAATAA AGGAGAGAAC ACCAGCTTGT
 241 TACACCCTGT GAGCCTGCAT GGAATGGATG ACCCTGAGAG AGAAGTGTTA GAGTGGAGGT
 301 TTGACAGCCG CCTAGCATTT CATCACGTGG CCCGAGAGCT GCATCCGGAG TACTTCAAGA
 361 ACTGCTGACA TCGAGCTTGC TACAAGGGAC TTTCCGCTGG GGACTTTCCA GGGAGGCGTG
 421 GCCTGGGCGG GACTGGGGAG TGGCGAGCCC TCAGATGCTG CATATAAGCA GCTGCTTTTT
 481 GCCTGTACTG GGTCTCTCTG GTTAGACCAG ATCTGAGCCT GGGAGCTCTC TGGCTAACTA
 541 GGGAACCCAC TGCTTAAGCC TCAATAAAGC TTGCCTTGAG GGAGTGCTTC AAGTAGTGTG
 601 TGCCCGTCTG TTGTGACTCT GGTAACTAGA GATCCCTCAG ACCCTTTTAG TCAGTGTGGA
 661 AAATCTCTAG CAGTGGCGCC CGAACAGGGA CTTGAAAGCG AAAGTAAAGC CAGAGGAGAT
 721 CTCTCGACGC AGGACTCGGC TTGCTGAAGC GCGCACGGCA AGAGGCGAGG GGCGGCGACT
 781 GGTGAGTACG CCAAAAATTT TGACTAGCGG AGGCTAGAAG GAGAGAGATG GGTGCGAGAG
 841 CGTCGGTATT AAGCGGGGGA GAATTAGATA AATGGGAAAA AATTCGGTTA AGGCCAGGGG
 901 GAAAGAAACA ATATAAACTA AAACATATAG TATGGGCAAG CAGGGAGCTA GAACGATTCG
 961 CAGTTAATCC TGGCCTTTTA GAGACATCAG AAGGCTGTAG ACAAATACTG GGACAGCTAC
1021 AACCATCCCT TCAGACAGGA TCAGAAGAAC TTAGATCATT ATATAATACA ATAGCAGTCC
1081 TCTATTGTGT GCATCAAAGG ATAGATGTAA AAGACACCAA GGAAGCCTTA GATAAGATAG
1141 AGGAAGAGCA AAACAAAAGT AAGAAAAAGG CACAGCAAGC AGCAGCTGAC ACAGGAAACA
1201 ACAGCCAGGT CAGCCAAAAT TACCCTATAG TCCAGAACCT CCAGGGGCAA ATGGTACATC
1261 AGGCCATATC ACCTAGAACT TTAAATGCAT GGGTAAAAGT AGTAGAAGAG AAGGCTTTCA
1321 GCCCAGAAGT AATACCCATG TTTTCAGCAT TATCAGAAGG AGCCACCCCA CAAGATTTAA
1381 ATACCATGCT AAACACAGTG GGGGGACATC AAGCAGCCAT GCAAATGTTA AAAGAGACCA
1441 TCAATGAGGA AGCTGCAGAA TGGGATAGAT TGCATCCAGT GCATGCAGGG CCTATTGCAC
1501 CAGGCCAGAT GAGAGAACCA AGGGGAAGTG ACATAGCAGG AACTACTAGT ACCCTTCAGG
1561 AACAAATAGG ATGGATGACA CATAATCCAC CTATCCCAGT AGGAGAAATC TATAAAAGAT
1621 GGATAATCCT GGGATTAAAT AAAATAGTAA GAATGTATAG CCCTACCAGC ATTCTGGACA
1681 TAAGACAAGG ACCAAAGGAA CCCTTTAGAG ACTATGTAGA CCGATTCTAT AAAACTCTAA
1741 GAGCCGAGCA AGCTTCACAA GAGGTAAAAA ATTGGATGAC AGAAACCTTG TTGGTCCAAA
1801 ATGCGAACCC AGATTGTAAG ACTATTTTAA AAGCACTGGG ACCAGGAGCG ACACTAGAAG
1861 AAATGATGAC AGCATGTCAG GGAGTGGGGG GACCCGGCCA TAAAGCAAGA GTTTTGGCTG
1921 AAGCAATGAG CCAAGTAACA AATCCAGCTA CCATAATGAT ACAGAAAGGC AATTTTAGGA
1981 ACCAAAGAAA GACTGTTAAG TGTTTCAATT GTGGCAAAGA AGGGCACATA GCCAAAAATT
2041 GCAGGGCCCC TAGGAAAAAG GGCTGTTGGA AATGTGGAAA GGAAGGACAC CAAATGAAAG
2101 ATTGTACTGA GAGACAGGCT AATTTTTTAG GGAAGATCTG GCCTTCCCAC AAGGGAAGGC
2161 CAGGGAATTT TCTTCAGAGC AGACCAGAGC CAACAGCCCC ACCAGAAGAG AGCTTCAGGT
2221 TTGGGGAAGA GACAACAACT CCCTCTCAGA GGCAGGAGCC GATAGACAAG GAACTGTATC
2281 CTTTAGCTTC CCTCAGATCA CTCTTTGGCA GCGACCCCTC GTCACAATAA AGATAGGGGG
2341 GCAATTAAAG GAAGCTCTAT TAGATACAGG AGCAGATGAT ACAGTATTAG AAGAAATGAA
2401 TTTGCCAGGA AGATGGAAAC CAAAAATGAT AGGGGGAATT GGAGGTTTTA TCAAAGTAAG
2461 ACAGTATGAT CAGATACTCA TAGAAATCTG CGGACATAAA GCTATAGGTA CAGTATTAGT
2521 AGGACCTACA CCTGTCAACA TAATTGGAAG AAATCTGTTG ACTCAGATTG GGTGCACTTT
2581 AAATTTTCCC ATTAGTCCTA TTGAGACTGT ACCAGTAAAA TTAAAGCCAG GAATGGATGG
2641 CCCAAAAGTT AAACAATGGC CATTGACAGA AGAAAAAATA AAAGCATTAG TAGAAATTTG
2701 TACAGAAATG GAAAAGGAAG GAAAAATTTC AAAAATTGGG CCTGAAAATC CATACAATAC
2761 TCCAGTATTT GCCATAAAGA AAAAAGACAG TACTAAATGG AGAAAATTAG TAGATTTCAG
2821 AGAACTTAAT AAGAGAACTC AAGATTTCTG GGAAGTTCAA TTAGGAATAC CACATCCTGC
2881 AGGGTTAAAA CAGAAAAAAT CAGTAACAGT ACTGGATGTG GGCGATGCAT ATTTTTCAGT
2941 TCCCTTAGAT AAAGACTTCA GGAAGTATAC TGCATTTACC ATACCTAGTA TAAACAATGA
```

```
3001 GACACCAGGG ATTAGATATC AGTACAATGT GCTTCCACAG GGATGGAAAG GATCACCAGC
3061 AATATTCCAG TGTAGCATGA CAAAAATCTT AGAGCCTTTT AGAAAACAAA ATCCAGACGT
3121 AGTCATCTAT CAATACATGG ATGATTTGTA TGTAGGATCT GACTTAGAAA TAGGGCAGCA
3181 TAGAACAAAA ATAGAGGAAC TGAGACAACA TCTGTTGAGG TGGGGATTTA CCACACCAGA
3241 CAAAAAACAT CAGAAAGAAC CTCCATTCCT TTGGATGGGT TATGAACTCC ATCCTGATAA
3301 ATGGACAGTA CAGCCTATAG TGCTGCCAGA AAAGGACAGC TGGACTGTCA ATGACATACA
3361 GAAATTAGTG GGAAAATTGA ATTGGGCAAG TCAGATTTAT GCAGGGATTA AAGTAAGGCA
3421 ATTATGTAAA CTTCTTAGGG GAACCAAAGC ACTAACAGAA GTAGTACCAC TAACAGAAGA
3481 AGCAGAGCTA GAACTGGCAG AAAACAGGGA GATTCTAAAA GAACCGGTAC ATGGAGTGTA
3541 TTATGACCCA TCAAAAGACT TAATAGCAGA AATACAGAAG CAGGGGCAAG GCCAATGGAC
3601 ATATCAAATT TATCAAGAGC CATTTAGAAA TCTGAAAACA GGAAAGTATG CAAGAATGAA
3661 GGGTGCCCAC ACTAATGATG TGAAACAATT AACAGAGGCA GTACAAAAAA TAGCCACAGA
3721 AAGCATAGTA ATATGGGGAA AGACTCCTAA ATTTAAATTA CCCATACAAA AGGAAACATG
3781 GGAAGCATGG TGGACAGAGT ATTGGCAAGC CACCTGGATT CCTGAGTGGG AGTTTGTCAA
3841 TACCCCTCCC TTAGTGAAGT TATGGTACCA GTTAGAGAAA GAACCCATAA TAGGAGCAGA
3901 AACTTTCTAT GTAGATGGGG CAGCCAATAG GGAAACTAAA TTAGGAAAAG CAGGATATGT
3961 AACTGACAGA GGAAGACAAA AAGTTGTCCC CCTAACGGAC ACAACAAATC AGAAGACTGA
4021 GTTACAAGCA ATTCATCTAG CTTTGCAGGA TTCGGGATTA GAAGTAAACA TAGTGACAGA
4081 CTCACAATAT GCATTGGGAA TCATTCAAGC ACAACCAGAT AAGAGTGAAT CAGAGTTAGT
4141 CAGTCAAATA ATAGAGCAGT TAATAAAAAA GGAAAAAGTC TACCTGGCAT GGGTACCAGC
4201 ACACAAAGGA ATTGGAGGAA ATGAACAAGT AGATAAATTG GTCAGTGCTG GAATCAGGAA
4261 AGTACTATTT TTAGATGGAA TAGATAAGGC CCAAGAAGAA CATGAGAAAT ATCACAGTAA
4321 TTGGAGAGCA ATGGCTAGTG ATTTTAACCT ACCACCTGTA GTAGCAAAAG AAATAGTAGC
4381 CAGCTGTGAT AAATGTCAGC TAAAAGGGGA AGCCATGCAT GGACAAGTAG ACTGTAGCCC
4441 AGGAATATGG CAGCTAGATT GTACACATTT AGAAGGAAAA GTTATCTTGG TGGCAGTTCA
4501 TGTAGCCAGT GGATATATAG AAGCAGAAGT AATTCCAGCA GAGACAGGGC AAGAAACAGC
4561 ATACTTCCTC TTAAAATTAG CAGGAAGATG GCCAGTAAAA ACAGTACATA CAGACAATGG
4621 CAGCAATTTC ACCAGTACTA CAGTTAAGGC CGCCTGTTGG TGGGCGGGAA TCAAGCAGGA
4681 ATTTGGCATT CCCTACAATC CCCAAAGTCA AGGAGTAATA GAATCTATGA ATAAAGAATT
4741 AAAGAAAATT ATAGGACAGG TAAGAGATCA GGCTGAACAT CTTAAGACAG CAGTACAAAT
4801 GGCAGTATTC ATCCACAATT TTAAAAGAAA AGGGGGGATT GGGGGGTACA GTGCCGGGGA
4861 AAGAATAGTA GACATAATAG CAACAGACAT ACAAACTAAA GAATTACAAA AACAAATTAC
4921 AAAAATTCAA AATTTTCGGG TTTATTACAG GGACAGCAGA GATCCAGTTT GGAAAGGACC
4981 AGCAAAGCTC CTCTGGAAAG GTGAAGGGGC AGTAGTAATA CAAGATAATA GTGACATAAA
5041 AGTAGTGCCA AGAAGAAAAG CAAAGATCAT CAGGGATTAT GGAAAACAGA TGGCAGGTGA
5101 TGATTGTGTG GCAAGTAGAC AGGATGAGGA TTAACACATG GAAAAGATTA GTAAAACACC
5161 ATATGGGAGT GGAAGCCATA ATAAGAATTC TGCAACAACT GCCGTTTATC CATTTCAGAA
5221 TTGGGTGTCG ACATAGCAGA ATAGGCGTTA CTCGACAGAG GAGAGCAAGA AATGGAGCCA
5281 GTAGATCCTA GACTAGAGCC CTGGAAGCAT CCAGGAAGTC AGCCTAAAAC TGCTTGTACC
5341 AATTGCTATT GTAAAAAGTG TTGCTTTCAT TGCCAAGTTT GTTTCATGAC AAAAGCCTTA
5401 GGCATCTCCT ATGGCAGGAA GAAGCGGAGA CAGCGACGAA GAGCTCATCA GAACAGTCAG
5461 ACTCATCAAG CTTCTCTATC AAAGCAGTAA GTAGTACATG GGCGCGCCCA TGTGGCAGGA
5521 AGTAGGAAAA GCAATGTATG CCCCTCCCAT CAGTGGACAA ATTAGATGTT CATCAAATAT
5581 TACTGGGCTG CTATTAACAA GAGATGGTGG TAATAACAAC AATGGGTCCG AGATCTTCAG
5641 ACCTGGAGGA GGCGATATGA GGGACAATTG GAGAAGTGAA TTATATAAAT ATAAAGTAGT
5701 AAAAATTGAA CCATTAGGAG TAGCACCCAC CAAGGCAAAG AGAAGAGTGG TGCTGAGAGA
5761 AAAAAGAGCA GTGGGAATAG GAGCTTTGTT CCTTGGGTTC TTGGGAGCAG CAGGAAGCAC
5821 TATGGGCGCA GCGTCAATGA CGCTGACGGT ACAGGCCAGA CAATTATTGT CTGATATAGT
5881 GCCGCAGCAG AACAATTTGC TGAGGGCTAT TGAGGCGCAA CAGCATCTGT TGCAACTCAC
5941 AGTCTGGGGC ATCAAACAGC TCCAGGCAAG AATCCTGGCT GTGGAAAGAT ACCTAAAGGA
6001 TCAACAGCTC CTGGGGATTT GGGGTTGCTC TGGAAAACTC ATTTGCACCA CTGCTGTGCC
6061 TTGGAATGCT AGTTGGAGTA ATAAATCTCT GGAACAGATT TGGAATAACA TGACCTGGAT
6121 GGAGTGGGAC AGAGAAATTA ACAATTACAC AAGCTTAATA CACTCCTTAA TTGAAGAATC
6181 GCAAAACCAG CAAGAAAAGA ATGAACAAGA ATTATTGGAA TTAGATAAAT GGGCAAGTTT
6241 GTGGAATTGG TTTAACATAA CAAATTGGCT GTGGTATATA AAATTATTCA TAATGATAGT
6301 AGGAGGCTTG GTAGGTTTAA GAATAGTTTT TGCTGTACTT TCTATAGTGA ATAGAGTTAG
6361 GCAGGGATAT TCACCATTAT CGTTTCAGAC CCACCTCCCA ATCCCGAGGG GACCCGACAG
```

```
6421 GCCCGAAGGA ATAGAAGAAG AAGGTGGAGA GAGAGGCAGA GACAGATCCA TTCGATTAGT
6481 GAACGGATCC TTGGCACTTA TCTGGGACGA TCTGCGGAGC CTGTGCCTCT TCAGCTACCA
6541 CCGCTTGAGA GACTTACTCT TGATTGTAAC GAGGATTGTG GAACTTCTGG GACGCAGGGG
6601 GTGGGAAGCC CTCAAATATT GGTGGAATCT CCTACAGTAT TGGAGTCAGG AACTAAAGAA
6661 TAGTGCTGTT AGCTTGCTCA ATGCCACAGC CATAGCAGTA GCTGAGGGGA CAGATAGGGT
6721 TATAGAAGTA GTACAAGGAG CTTGTAGAGC TATTCGCCAC ATACCTAGAA GAATAAGACA
6781 GGGCTTGGAA AGGATTTTGC TATAAGATGG GTGGCAAGTG GTCAAAAAGT AGTGTGATTG
6841 GATGGCTTAC TGTAAGGGAA AGAATGAGAC GAGCTGAGCC AGCAGCAGAT GGGGTGGGAG
6901 CAGCATGCGG CCCTCTAGAC GACCCTGTGG AATGTGTGTC AGTTAGGGTG TGGAAAGTCC
6961 CCAGGCTCCC CAGCAGGCAG AAGTATGCAA AGCATGCATC TCAATTAGTC AGCAACCAGG
7021 TGTGGAAAGT CCCCAGGCTC CCCAGCAGGC AGAAGTATGC AAAGCATGCA TCTCAATTAG
7081 TCAGCAACCA TAGTCCCGCC CCTAACTCCG CCCATCCCGC CCCTAACTCC GCCCAGTTCC
7141 GCCCATTCTC CGCCCCATGG CTGACTAATT TTTTTTATTT ATGCAGAGGC CGAGGCCGCC
7201 TCGGCCTCTG AGCTATTCCA GAAGTAGTGA GGAGGCTTTT TTGGAGGCCT AGGCTTTTGC
7261 AAAAAGCTCT TGACATGATA GAAGCACTCT ACTATATTCT CAATAGGTAG CTTACCATGA
7321 CCGAGTACAA GCCCACGGTG CGCCTCGCCA CCCGCGACGA CGTCCCCCGG GCCGTACGCA
7381 CCCTCGCCGC CGCGTTCGCC GACTACCCCG CCACGCGCCA CACCGTCGAC CCGGACCGCC
7441 ACATCGAGCG GGTCACCGAG CTGCAAGAAC TCTTCCTCAC GCGCGTCGGG CTCGACATCG
7501 GCAAGGTGTG GGTCGCGGAC GACGGCGCCG CGGTGGCGGT CTGGACCACG CCGGAGAGCG
7561 TCGAAGCGGG GGCGGTGTTC GCCGAGATCG GCCCGCGCAT GGCCGAGTTG AGCGGTTCCC
7621 GGCTGGCCGC GCAGCAACAG ATGGAAGGCC TCCTGGCGCC GCACCGGCCC AAGGAGCCCG
7681 CGTGGTTCCT GGCCACCGTC GGCGTCTCGC CCGACCACCA GGGCAAGGGT CTGGGCAGCG
7741 CCGTCGTGCT CCCCGGAGTG GAGGCGGCCG AGCGCGCCGG GGTGCCCGCC TTCCTGGAGA
7801 CCTCCGCGCC CCGCAACCTC CCCTTCTACG AGCGGCTCGG CTTCACCGTC ACCGCCGACG
7861 TCGAGGTGCC CGAAGGACCG CGCACCTGGT GCATGACCCG CAAGCCCGGT GCCTGACGCC
7921 CGCCCCACGA CCCGCAGCGC CCGACCGAAA GGAGCGCACG ACCCATCGCT CGAGACCTAG
7981 AAAAACATGG AGCAATCACA AGTAGCAATA CAGCAGCTAA CAATGCTGCT TGTGCCTGGC
8041 TAGAAGCACA AGAGGAGGAA GAGGTGGGTT TTCCAGTCAC ACCTCAGGTA CCTTTAAGAC
8101 CAATGACTTA CAAGGCAGCT GTAGATCTTA GCCACTTTTT AAAAGAAAAG GGGGGACTGG
8161 AAGGGCTAAT TCACTCCCAA AGAAGACAAG ATATCCTTGA TCTGTGGATC TACCACACAC
8221 AAGGCTACTT CCCTGATTGG CAGAACTACA CACCAGGGCC AGGGGTCAGA TATCCACTGA
8281 CCTTTGGATG GTGCTACAAG CTAGTACCAG TTGAGCCAGA TAAGGTAGAA GAGGCCAATA
8341 AAGGAGAGAA CACCAGCTTG TTACACCCTG TGAGCCTGCA TGGAATGGAT GACCCTGAGA
8401 GAGAAGTGTT AGAGTGGAGG TTTGACAGCC GCCTAGCATT TCATCACGTG GCCCGAGAGC
8461 TGCATCCGGA GTACTTCAAG AACTGCTGAC ATCGAGCTTG CTACAAGGGA CTTTCCGCTG
8521 GGGACTTTCC AGGGAGGCGT GGCCTGGGCG GGACTGGGGA GTGGCGAGCC CTCAGATGCT
8581 GCATATAAGC AGCTGCTTTT TGCCTGTACT GGGTCTCTCT GGTTAGACCA GATCTGAGCC
8641 TGGGAGCTCT CTGGCTAACT AGGGAACCCA CTGCTTAAGC CTCAATAAAG CTTGCCTTGA
8701 GGGAGTGCTT CAAGTAGTGT GTGCCCGTCT GTTGTGACTC TGGTAACTAG AGATCCCTCA
8761 GACCCTTTTA GTCAGTGTGG AAAATCTCTA GCACCCAGGA GGTAGAGGTT GCAGTGAGCC
8821 AAGATCGCGC CACTGCATTC CAGCCTGGGC AAGAAAACAA GACTGTTTAA AATAATAATA
8881 ATAAGTTAAG GGTATTAAAT ATATTTATAC ATGGAGGTCA TAAAAATATA TATATTTGGG
8941 CTGGGCGCAG TGGCTCACAC ATGCGCCCGG CCCTTTGGGA GGCCGAGGCA GGTGGATCAC
9001 CTGAGTTTGG GAGTTCCAGA CCAGCCTGAC CAACATGGAG AAACCCCTTC TCTGTGTATT
9061 TTTAGTAGAT TTTATTTTAT GTGTATTTTA TTCACAGGTA TTTCTGGAAA ACTGAAACTG
9121 TTTTTCTTCT ACTCTGATAC CACAAGAATC ATCAGCACAG AGGAAGACTT CTGTGATCAA
9181 ATGTGGTGGG AGAGGGAGGT TTTCACCAGC ACATGAGCAG TCAGTTCTGC CGCAGACTCG
9241 GCGGGTGTCC TTCGGTTCAG TTCCAACACC GCCTGCCTGG AGAGAGGTCA GACCACAGGG
9301 TGAGGGCTCA GTCCCCAAGA CATAAACACC CAAGACATAA ACACCCAACA GGTCCACCCC
9361 GCCTGCTGCC CAGGCAGAGC CGATTCACCA AGACGGGAAT TAGGATAGAG AAAGAGTAAG
9421 TCACACAGAG CCGGCTTTCC CCGTCAAGCT CTAAATCGGG GGCTCCCTTT AGGGTTCCGA
9481 TTTAGTGCTT TACGGCACCT CGACCCCAAA AAACTTGATT AGGGTGATGG TTCACGTAGT
9541 GGGCCATCGC CCTGATAGAC GGTTTTTCGC CCTTTGACGT TGGAGTCCAC GTTCTTTAAT
9601 AGTGGACTCT TGTTCCAAAC TGGAACAACA CTCAACCCTA TCTCGGTCTA TTCTTTTGAT
9661 TTATAAGGGA TTTTGCCGAT TTCGGCCTAT TGGTTAAAAA ATGAGCTGAT TTAACAAAAA
9721 TTTAACGCGA ATTTTAACAA AATATTAACG TTTACAATTT CAGGTGGCAC TTTTCGGGGA
9781 AATGTGCGCG GAACCCCTAT TTGTTTATTT TTCTAAATAC ATTCAAATAT GTATCCGCTC
```

```
 9841 ATGAGACAAT AACCCTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG TATGAGTATT
 9901 CAACATTTCC GTGTCGCCCT TATTCCCTTT TTTGCGGCAT TTTGCCTTCC TGTTTTTGCT
 9961 CACCCAGAAA CGCTGGTGAA AGTAAAAGAT GCTGAAGATC AGTTGGGTGC ACGAGTGGGT
10021 TACATCGAAC TGGATCTCAA CAGCGGTAAG ATCCTTGAGA GTTTTCGCCC CGAAGAACGT
10081 TTTCCAATGA TGAGCACTTT TAAAGTTCTG CTATGTGGCG CGGTATTATC CCGTATTGAC
10141 GCCGGGCAAG AGCAACTCGG TCGCCGCATA CACTATTCTC AGAATGACTT GGTTGAGTAC
10201 TCACCAGTCA CAGAAAAGCA TCTTACGGAT GGCATGACAG TAAGAGAATT ATGCAGTGCT
10261 GCCATAAGCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT CGGAGGACCG
10321 AAGGAGCTAA CCGCTTTTTT TCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG
10381 GAACCGGAGC TGAATGAAGC CATACCAAAC GACGAGCGTG ACACCACGAT GCCTGTAGCA
10441 ATGGCAACAA CGTTGCGCAA ACTATTAACT GGCGAACTAC TTACTCTAGC TTCCCGGCAA
10501 CAATTAATAG ACTGGATGGA GGCGGATAAA GTTGCAGGAC CACTTCTGCG CTCGGCCCTT
10561 CCGGCTGGCT GGTTTATTGC TGATAAATCT GGAGCCGGTG AGCGTGGGTC TCGCGGTATC
10621 ATTGCAGCAC TGGGGCCAGA TGGTAAGCCC TCCCGTATCG TAGTTATCTA CACGACGGGC
10681 AGTCAGGCAA CTATGGATGA ACGAAATAGA CAGATCGCTG AGATAGGTGC CTCACTGATT
10741 AAGCATTGGT AACTGTCAGA CCAAGTTTAC TCATATATAC TTTAGATTGA TTTAAAACTT
10801 CATTTTTAAT TTAAAAGGAT CTAGGTGAAG ATCCTTTTTG ATAATCTCAT GACCAAAATC
10861 CCTTAACGTG AGTTTTCGTT CCACTGAGCG TCAGACCCCG TAGAAAAGAT CAAAGGATCT
10921 TCTTGAGATC CTTTTTTTCT GCGCGTAATC TGCTGCTTGC AAACAAAAAA ACCACCGCTA
10981 CCAGCGGTGG TTTGTTTGCC GGATCAAGAG CTACCAACTC TTTTTCCGAA GGTAACTGGC
11041 TTCAGCAGAG CGCAGATACC AAATACTGTC CTTCTAGTGT AGCCGTAGTT AGGCCACCAC
11101 TTCAAGAACT CTGTAGCACC GCCTACATAC CTCGCTCTGC TAATCCTGTT ACCAGTGGCT
11161 GCTGCCAGTG GCGATAAGTC GTGTCTTACC GGGTTGGACT CAAGACGATA GTTACCGGAT
11221 AAGGCGCAGC GGTCGGGCTG AACGGGGGGT TCGTGCACAC AGCCCAGCTT GGAGCGAACG
11281 ACCTACACCG AACTGAGATA CCTACAGCGT GAGCATTGAG AAAGCGCCAC GCTTCCCGAA
11341 GGGAGAAAGG CGGACAGGTA TCCGGTAAGC GGCAGGGTCG GAACAGGAGA GCGCACGAGG
11401 GAGCTTCCAG GGGGAACGC CTGGTATCTT TATAGTCCTG TCGGGTTTCG CCACCTCTGA
11461 CTTGAGCGTC GATTTTTGTG ATGCTCGTCA GGGGGGCCGA GCCTATGGAA AAACGCCAGC
11521 AACGCGGCCT TTTTACGGTT CCTGGCCTTT TGCTGGCCTT TTGCTCACAT GT
```

## Appendix B. Local nucleotide and peptide sequences of mutants

| | | left flanking nucleotide sequence | insertion | right flanking nucleotide sequence | peptide sequence |
|---|---|---|---|---|---|
| 1 | A | A | TGCGGCCGCA | ACATGTAGCCCCAGTT | AAATCSPS |
| 2 | C | AC | TGCGGCCGCA | ACATGTAGCCCCAGTTC | CGRNM*PQF |
| 3 | A | ACA | TGCGGCCGCA | ACATGTAGCCCCAGTTCT | MRPQHVAPV |
| 4 | T | ACAT | TGCGGCCGCA | ACATGTAGCCCCAGTTCTA | IAAATCSPSS |
| 5 | G | ACATG | TGCGGCCGCA | ACATGTAGCCCCAGTTCTAC | MCGRNM*PQFY |
| 6 | T | ACATGT | TGCGGCCGCA | CATGTAGCCCCAGTTCTACT | MLRPHM*PQFY |
| 7 | A | ACATGTA | TGCGGCCGCA | ATGTAGCCCCAGTTCTACTT | MYAAAM*PQFY |
| 8 | G | ACATGTAG | TGCGGCCGCA | TGTAGCCCCAGTTCTACTTA | M*CGRM*PQFYL |
| 9 | C | ACATGTAGC | TGCGGCCGCA | GTAGCCCCAGTTCTACTTAC | M*LRPQ*PQFYL |
| 10 | C | ACATGTAGCC | TGCGGCCGCA | TAGCCCCAGTTCTACTTACA | M*PAAA*PQFYL |
| 11 | C | ACATGTAGCCC | TGCGGCCGCA | AGCCCCAGTTCTACTTACAC | M*PCGRKPQFYLH |
| 12 | C | ACATGTAGCCCC | TGCGGCCGCA | GCCCCAGTTCTACTTACACC | M*PLRPQPQFYLH |
| 13 | A | ACATGTAGCCCCA | TGCGGCCGCA | CCCCAGTTCTACTTACACCA | M*PHAAAPQFYLH |
| 14 | G | ACATGTAGCCCCAG | TGCGGCCGCA | CCCAGTTCTACTTACACCAA | M*PQCGRTQFYLHQ |
| 15 | T | ACATGTAGCCCCAGT | TGCGGCCGCA | CCAGTTCTACTTACACCAAG | M*PQLRPHQFYLHQ |
| 16 | T | ACATGTAGCCCCAGTT | TGCGGCCGCA | CAGTTCTACTTACACCAAGA | M*PQFAAAQFYLHQ |
| 17 | C | ACATGTAGCCCCAGTTC | TGCGGCCGCA | AGTTCTACTTACACCAAGAA | M*PQFCGRKFYLHQE |
| 18 | T | ACATGTAGCCCCAGTTCT | TGCGGCCGCA | GTTCTACTTACACCAAGAAA | M*PQFLRPQFYLHQE |
| 19 | A | ACATGTAGCCCCAGTTCTA | TGCGGCCGCA | TTCTACTTACACCAAGAAAG | M*PQFYAAAFYLHQE |
| 20 | C | ACATGTAGCCCCAGTTCTAC | TGCGGCCGCA | TCTACTTACACCAAGAAAGG | M*PQFYCGRIYLHQER |
| 21 | T | CATGTAGCCCCAGTTCTACT | TGCGGCCGCA | CTACTTACACCAAGAAAGGC | M*PQFYLRPHYLHQER |
| 22 | T | ATGTAGCCCCAGTTCTACTT | TGCGGCCGCA | TACTTACACCAAGAAAGGCT | M*PQFYFAAAYLHQER |
| 23 | A | TGTAGCCCCAGTTCTACTTA | TGCGGCCGCA | ACTTACACCAAGAAAGGCTG | *PQFYLCGRNLHQERL |
| 24 | C | GTAGCCCCAGTTCTACTTAC | TGCGGCCGCA | CTTACACCAAGAAAGGCTGG | *PQFYLLRPHLHQERL |
| 25 | A | TAGCCCCAGTTCTACTTACA | TGCGGCCGCA | TTACACCAAGAAAGGCTGGA | *PQFYLHAAALHQERL |
| 26 | C | AGCCCCAGTTCTACTTACAC | TGCGGCCGCA | TACACCAAGAAAGGCTGGAA | PQFYLHCGRIHQERLE |
| 27 | C | GCCCCAGTTCTACTTACACC | TGCGGCCGCA | ACACCAAGAAAGGCTGGAAG | PQFYLHLRPQHQERLE |
| 28 | A | CCCCAGTTCTACTTACACCA | TGCGGCCGCA | CACCAAGAAAGGCTGGAAGG | PQFYLHHAAAHQERLE |
| 29 | A | CCCAGTTCTACTTACACCAA | TGCGGCCGCA | ACCAAGAAAGGCTGGAAGGG | QFYLHQCGRNQERLEG |
| 30 | G | CCAGTTCTACTTACACCAAG | TGCGGCCGCA | CCAAGAAAGGCTGGAAGGGC | QFYLHQVRPHQERLEG |
| 31 | A | CAGTTCTACTTACACCAAGA | TGCGGCCGCA | CAAGAAAGGCTGGAAGGGCT | QFYLHQDAAAQERLEG |
| 32 | A | AGTTCTACTTACACCAAGAA | TGCGGCCGCA | AAGAAAGGCTGGAAGGGCTA | FYLHQECGRKERLEGL |
| 33 | A | GTTCTACTTACACCAAGAAA | TGCGGCCGCA | AGAAAGGCTGGAAGGGCTAA | FYLHQEMRPQERLEGL |
| 34 | G | TTCTACTTACACCAAGAAAG | TGCGGCCGCA | GAAAGGCTGGAAGGGCTAAT | FYLHQESAAAERLEGL |
| 35 | G | TCTACTTACACCAAGAAAGG | TGCGGCCGCA | AAAGGCTGGAAGGGCTAATT | YLHQERCGRKRLEGLI |
| 36 | C | CTACTTACACCAAGAAAGGC | TGCGGCCGCA | AAGGCTGGAAGGGCTAATTC | YLHQERLRPQRLEGLI |
| 37 | T | TACTTACACCAAGAAAGGCT | TGCGGCCGCA | AGGCTGGAAGGGCTAATTCA | YLHQERLAAARLEGLI |
| 38 | G | ACTTACACCAAGAAAGGCTG | TGCGGCCGCA | GGCTGGAAGGGCTAATTCAC | LHQERLCGRRLEGLIH |
| 39 | G | CTTACACCAAGAAAGGCTGG | TGCGGCCGCA | GCTGGAAGGGCTAATTCACT | LHQERLVRPQLEGLIH |
| 40 | A | TTACACCAAGAAAGGCTGGA | TGCGGCCGCA | CTGGAAGGGCTAATTCACTC | LHQERLDAAALEGLIH |
| 41 | A | TACACCAAGAAAGGCTGGAA | TGCGGCCGCA | TGGAAGGGCTAATTCACTCC | HQERLECGRMEGLIHS |
| 42 | G | ACACCAAGAAAGGCTGGAAG | TGCGGCCGCA | GGAAGGGCTAATTCACTCCC | HQERLEVRPQEGLIHS |
| 43 | G | CACCAAGAAAGGCTGGAAGG | TGCGGCCGCA | GAAGGGCTAATTCACTCCCA | HQERLEGAAAEGLIHS |
| 44 | G | ACCAAGAAAGGCTGGAAGGG | TGCGGCCGCA | AAGGGCTAATTCACTCCCAA | QERLEGCGRKGLIHSQ |
| 45 | C | CCAAGAAAGGCTGGAAGGGC | TGCGGCCGCA | AGGGCTAATTCACTCCCAAA | QERLEGLRPQGLIHSQ |
| 46 | T | CAAGAAAGGCTGGAAGGGCT | TGCGGCCGCA | GGGCTAATTCACTCCCAAAG | QERLEGLAAAGLIHSQ |
| 47 | A | AAGAAAGGCTGGAAGGGCTA | TGCGGCCGCA | GGCTAATTCACTCCCAAAGA | ERLEGLCGRRLIHSQR |
| 48 | A | AGAAAGGCTGGAAGGGCTAA | TGCGGCCGCA | GCTAATTCACTCCCAAAGAA | ERLEGLMRPQLIHSQR |
| 49 | T | GAAAGGCTGGAAGGGCTAAT | TGCGGCCGCA | CTAATTCACTCCCAAAGAAG | ERLEGLIAAALIHSQR |
| 50 | T | AAAGGCTGGAAGGGCTAATT | TGCGGCCGCA | TAATTCACTCCCAAAGAAGA | RLEGLICGRIIHSQRR |
| 51 | C | AAGGCTGGAAGGGCTAATTC | TGCGGCCGCA | AATTCACTCCCAAAGAAGAC | RLEGLILRPQIHSQRR |
| 52 | A | AGGCTGGAAGGGCTAATTCA | TGCGGCCGCA | ATTCACTCCCAAAGAAGACA | RLEGLIHAAAIHSQRR |
| 53 | C | GGCTGGAAGGGCTAATTCAC | TGCGGCCGCA | TTCACTCCCAAAGAAGACAA | LEGLIHCGRIHSQRRQ |
| 54 | T | GCTGGAAGGGCTAATTCACT | TGCGGCCGCA | TCACTCCCAAAGAAGACAAG | LEGLIHLRPHHSQRRQ |
| 55 | C | CTGGAAGGGCTAATTCACTC | TGCGGCCGCA | CACTCCCAAAGAAGACAAGA | LEGLIHSAAAHSQRRQ |
| 56 | C | TGGAAGGGCTAATTCACTCC | TGCGGCCGCA | ACTCCCAAAGAAGACAAGAT | EGLIHSCGRNSQRRQD |
| 57 | C | GGAAGGGCTAATTCACTCCC | TGCGGCCGCA | CTCCCAAAGAAGACAAGATA | EGLIHSLRPHSQRRQD |
| 58 | A | GAAGGGCTAATTCACTCCCA | TGCGGCCGCA | TCCCAAAGAAGACAAGATAT | EGLIHSHAAASQRRQD |
| 59 | A | AAGGGCTAATTCACTCCCAA | TGCGGCCGCA | CCCAAAGAAGACAAGATATC | GLIHSQCGRTQRRQDI |
| 60 | A | AGGGCTAATTCACTCCCAAA | TGCGGCCGCA | CCAAAGAAGACAAGATATCC | GLIHSQMRPHQRRQDI |
| 61 | G | GGGCTAATTCACTCCCAAAG | TGCGGCCGCA | CAAAGAAGACAAGATATCCT | GLIHSQSAAAQRRQDI |
| 62 | A | GGCTAATTCACTCCCAAAGA | TGCGGCCGCA | AAAGAAGACAAGATATCCTT | LIHSQRCGRKRRQDIL |
| 63 | A | GCTAATTCACTCCCAAAGAA | TGCGGCCGCA | AAGAAGACAAGATATCCTTG | LIHSQRMRPQRRQDIL |
| 64 | G | CTAATTCACTCCCAAAGAAG | TGCGGCCGCA | AGAAGACAAGATATCCTTGA | LIHSQRSAAARRQDIL |
| 65 | A | TAATTCACTCCCAAAGAAGA | TGCGGCCGCA | GAAGACAAGATATCCTTGAT | IHSQRRCGRRRQDILD |
| 66 | C | AATTCACTCCCAAAGAAGAC | TGCGGCCGCA | AAGACAAGATATCCTTGATC | IHSQRRLRPQRQDILD |
| 67 | A | ATTCACTCCCAAAGAAGACA | TGCGGCCGCA | AGACAAGATATCCTTGATCT | IHSQRRHAAARQDILD |
| 68 | A | TTCACTCCCAAAGAAGACAA | TGCGGCCGCA | GACAAGATATCCTTGATCTG | HSQRRQCGRRQDILDL |
| 69 | G | TCACTCCCAAAGAAGACAAG | TGCGGCCGCA | ACAAGATATCCTTGATCTGT | HSQRRQVRPQQDILDL |
| 70 | A | CACTCCCAAAGAAGACAAGA | TGCGGCCGCA | CAAGATATCCTTGATCTGTG | HSQRRQDAAAQDILDL |
| 71 | T | ACTCCCAAAGAAGACAAGAT | TGCGGCCGCA | AAGATATCCTTGATCTGTGG | SQRRQDCGRKDILDLW |
| 72 | A | CTCCCAAAGAAGACAAGATA | TGCGGCCGCA | AGATATCCTTGATCTGTGGA | SQRRQDMRPQDILDLW |
| 73 | T | TCCCAAAGAAGACAAGATAT | TGCGGCCGCA | GATATCCTTGATCTGTGGAT | SQRRQDIAAADILDLW |
| 74 | C | CCCAAAGAAGACAAGATATC | TGCGGCCGCA | ATATCCTTGATCTGTGGATC | QRRQDICGRNILDLWI |
| 75 | C | CCAAAGAAGACAAGATATCC | TGCGGCCGCA | TATCCTTGATCTGTGGATCT | QRRQDILRPHILDLWI |
| 76 | T | CAAAGAAGACAAGATATCCT | TGCGGCCGCA | ATCCTTGATCTGTGGATCTA | QRRQDILAAAILDLWI |
| 77 | T | AAAGAAGACAAGATATCCTT | TGCGGCCGCA | TCCTTGATCTGTGGATCTAC | RRQDILCGRILDLWIY |
| 78 | G | AAGAAGACAAGATATCCTTG | TGCGGCCGCA | CCTTGATCTGTGGATCTACC | RRQDILVRPHLDLWIY |
| 79 | A | AGAAGACAAGATATCCTTGA | TGCGGCCGCA | CTTGATCTGTGGATCTACCA | RRQDILDAAALDLWIY |
| 80 | T | GAAGACAAGATATCCTTGAT | TGCGGCCGCA | TTGATCTGTGGATCTACCAC | RQDILDCGRIDLWIYH |

| | | | | | |
|---|---|---|---|---|---|
| 81 | C | AAGACAAGATATCCTTGATC | TGCGGCCGCA | TGATCTGTGGATCTACCACA | RQDILDLRPHDLWIYH |
| 82 | T | AGACAAGATATCCTTGATCT | TGCGGCCGCA | GATCTGTGGATCTACCACAC | RQDILDLAAADLWIYH |
| 83 | G | GACAAGATATCCTTGATCTG | TGCGGCCGCA | ATCTGTGGATCTACCACACA | QDILDLCGRNLWIYHT |
| 84 | T | ACAAGATATCCTTGATCTGT | TGCGGCCGCA | TCTGTGGATCTACCACACAC | QDILDLLRPHLWIYHT |
| 85 | G | CAAGATATCCTTGATCTGTG | TGCGGCCGCA | CTGTGGATCTACCACACACA | QDILDLCAAALWIYHT |
| 86 | G | AAGATATCCTTGATCTGTGG | TGCGGCCGCA | TGTGGATCTACCACACACAA | DILDLWCGRMWIYHTQ |
| 87 | A | AGATATCCTTGATCTGTGGA | TGCGGCCGCA | GTGGATCTACCACACACAAG | DILDLWMRPQWIYHTQ |
| 88 | T | GATATCCTTGATCTGTGGAT | TGCGGCCGCA | TGGATCTACCACACACAAGG | DILDLWIAAAWIYHTQ |
| 89 | C | ATATCCTTGATCTGTGGATC | TGCGGCCGCA | GGATCTACCACACACAAGGC | ILDLWICGRRIYHTQG |
| 90 | T | TATCCTTGATCTGTGGATCT | TGCGGCCGCA | GATCTACCACACACAAGGCT | ILDLWILRPQIYHTQG |
| 91 | A | ATCCTTGATCTGTGGATCTA | TGCGGCCGCA | ATCTACCACACACAAGGCTA | ILDLWIYAAAIYHTQG |
| 92 | C | TCCTTGATCTGTGGATCTAC | TGCGGCCGCA | TCTACCACACACAAGGCTAC | LDLWIYCGRIYHTQGY |
| 93 | C | CCTTGATCTGTGGATCTACC | TGCGGCCGCA | CTACCACACACAAGGCTACT | LDLWIYLRPHYHTQGY |
| 94 | A | CTTGATCTGTGGATCTACCA | TGCGGCCGCA | TACCACACACAAGGCTACTT | LDLWIYHAAAYHTQGY |
| 95 | C | TTGATCTGTGGATCTACCAC | TGCGGCCGCA | ACCACACACAAGGCTACTTC | DLWIYHCGRNHTQGYF |
| 96 | A | TGATCTGTGGATCTACCACA | TGCGGCCGCA | CCACACACAAGGCTACTTCC | DLWIYHMRPHHTQGYF |
| 97 | C | GATCTGTGGATCTACCACAC | TGCGGCCGCA | CACACACAAGGCTACTTCCC | DLWIYHTAAAHTQGYF |
| 98 | A | ATCTGTGGATCTACCACACA | TGCGGCCGCA | ACACACAAGGCTACTTCCCT | LWIYHTCGRNTQGYFP |
| 99 | C | TCTGTGGATCTACCACACAC | TGCGGCCGCA | CACACAAGGCTACTTCCCTG | LWIYHTLRPHTQGYFP |
| 100 | A | CTGTGGATCTACCACACACA | TGCGGCCGCA | ACACAAGGCTACTTCCCTGA | LWIYHTHAAATQGYFP |
| 101 | A | TGTGGATCTACCACACACAA | TGCGGCCGCA | CACAAGGCTACTTCCCTGAT | WIYHTQCGRTQGYFPD |
| 102 | G | GTGGATCTACCACACACAAG | TGCGGCCGCA | ACAAGGCTACTTCCCTGATT | WIYHTQVRPQQGYFPD |
| 103 | G | TGGATCTACCACACACAAGG | TGCGGCCGCA | CAAGGCTACTTCCCTGATTG | WIYHTQGAAAQGYFPD |
| 104 | C | GGATCTACCACACACAAGGC | TGCGGCCGCA | AAGGCTACTTCCCTGATTGG | IYHTQGCGRKGYFPDW |
| 105 | T | GATCTACCACACACAAGGCT | TGCGGCCGCA | AGGCTACTTCCCTGATTGGC | IYHTQGLRPQGYFPDW |
| 106 | A | ATCTACCACACACAAGGCTA | TGCGGCCGCA | GGCTACTTCCCTGATTGGCA | IYHTQGYAAAGYFPDW |
| 107 | C | TCTACCACACACAAGGCTAC | TGCGGCCGCA | GCTACTTCCCTGATTGGCAG | YHTQGYCGRSYFPDWQ |
| 108 | T | CTACCACACACAAGGCTACT | TGCGGCCGCA | CTACTTCCCTGATTGGCAGA | YHTQGYLRPHYFPDWQ |
| 109 | T | TACCACACACAAGGCTACTT | TGCGGCCGCA | TACTTCCCTGATTGGCAGAA | YHTQGYFAAAYFPDWQ |
| 110 | C | ACCACACACAAGGCTACTTC | TGCGGCCGCA | ACTTCCCTGATTGGCAGAAC | HTQGYFCGRNFPDWQN |
| 111 | C | CCACACACAAGGCTACTTCC | TGCGGCCGCA | CTTCCCTGATTGGCAGAACT | HTQGYFLRPHFPDWQN |
| 112 | C | CACACACAAGGCTACTTCCC | TGCGGCCGCA | TTCCCTGATTGGCAGAACTA | HTQGYFPAAAFPDWQN |
| 113 | T | ACACACAAGGCTACTTCCCT | TGCGGCCGCA | TCCCTGATTGGCAGAACTAC | TQGYFPCGRIPDWQNY |
| 114 | G | CACACAAGGCTACTTCCCTG | TGCGGCCGCA | CCCTGATTGGCAGAACTACA | TQGYFPVRPHPDWQNY |
| 115 | A | ACACAAGGCTACTTCCCTGA | TGCGGCCGCA | CCTGATTGGCAGAACTACAC | TQGYFPDAAAPDWQNY |
| 116 | T | CACAAGGCTACTTCCCTGAT | TGCGGCCGCA | CTGATTGGCAGAACTACACA | QGYFPDCGRTDWQNYT |
| 117 | T | ACAAGGCTACTTCCCTGATT | TGCGGCCGCA | TGATTGGCAGAACTACACAC | QGYFPDLRPHDWQNYT |
| 118 | G | CAAGGCTACTTCCCTGATTG | TGCGGCCGCA | GATTGGCAGAACTACACACC | QGYFPDCAAADWQNYT |
| 119 | G | AAGGCTACTTCCCTGATTGG | TGCGGCCGCA | ATTGGCAGAACTACACACCA | GYFPDWCGRNWQNYTP |
| 120 | C | AGGCTACTTCCCTGATTGGC | TGCGGCCGCA | TTGGCAGAACTACACACCAG | GYFPDWLRPHWQNYTP |
| 121 | A | GGCTACTTCCCTGATTGGCA | TGCGGCCGCA | TGGCAGAACTACACACCAGG | GYFPDWHAAAWQNYTP |
| 122 | G | GCTACTTCCCTGATTGGCAG | TGCGGCCGCA | GGCAGAACTACACACCAGGG | YFPDWQCGRRQNYTPG |
| 123 | A | CTACTTCCCTGATTGGCAGA | TGCGGCCGCA | GCAGAACTACACACCAGGGC | YFPDWQMRPQQNYTPG |
| 124 | A | TACTTCCCTGATTGGCAGAA | TGCGGCCGCA | CAGAACTACACACCAGGGCC | YFPDWQNAAAQNYTPG |
| 125 | C | ACTTCCCTGATTGGCAGAAC | TGCGGCCGCA | AGAACTACACACCAGGGCCA | FPDWQNCGRKNYTPGP |
| 126 | T | CTTCCCTGATTGGCAGAACT | TGCGGCCGCA | GAACTACACACCAGGGCCAG | FPDWQNLRPQNYTPGP |
| 127 | A | TTCCCTGATTGGCAGAACTA | TGCGGCCGCA | AACTACACACCAGGGCCAGG | FPDWQNYAAANYTPGP |
| 128 | C | TCCCTGATTGGCAGAACTAC | TGCGGCCGCA | ACTACACACCAGGGCCAGGG | PDWQNYCGRNYTPGPG |
| 129 | A | CCCTGATTGGCAGAACTACA | TGCGGCCGCA | CTACACACCAGGGCCAGGGG | PDWQNYMRPHYTPGPG |
| 130 | C | CCTGATTGGCAGAACTACAC | TGCGGCCGCA | TACACACCAGGGCCAGGGGT | PDWQNYTAAAYTPGPG |
| 131 | A | CTGATTGGCAGAACTACACA | TGCGGCCGCA | ACACACCAGGGCCAGGGGTC | DWQNYTCGRNTPGPGV |
| 132 | C | TGATTGGCAGAACTACACAC | TGCGGCCGCA | CACACCAGGGCCAGGGGTCA | DWQNYTLRPHTPGPGV |
| 133 | C | GATTGGCAGAACTACACACC | TGCGGCCGCA | ACACCAGGGCCAGGGGTCAG | DWQNYTPAAATPGPGV |
| 134 | A | ATTGGCAGAACTACACACCA | TGCGGCCGCA | CACCAGGGCCAGGGGTCAGA | WQNYTPCGRTPGPGVR |
| 135 | G | TTGGCAGAACTACACACCAG | TGCGGCCGCA | ACCAGGGCCAGGGGTCAGAT | WQNYTPVRPQPGPGVR |
| 136 | G | TGGCAGAACTACACACCAGG | TGCGGCCGCA | CCAGGGCCAGGGGTCAGATA | WQNYTPGAAAPGPGVR |
| 137 | G | GGCAGAACTACACACCAGGG | TGCGGCCGCA | CAGGGCCAGGGGTCAGATAT | QNYTPGCGRTGPGVRY |
| 138 | C | GCAGAACTACACACCAGGGC | TGCGGCCGCA | AGGGCCAGGGGTCAGATATC | QNYTPGLRPQGPGVRY |
| 139 | C | CAGAACTACACACCAGGGCC | TGCGGCCGCA | GGGCCAGGGGTCAGATATCC | QNYTPGPAAAGPGVRY |
| 140 | A | AGAACTACACACCAGGGCCA | TGCGGCCGCA | GGCCAGGGGTCAGATATCCA | NYTPGPCGRRPGVRYP |
| 141 | G | GAACTACACACCAGGGCCAG | TGCGGCCGCA | GCCAGGGGTCAGATATCCAC | NYTPGPVRPQPGVRYP |
| 142 | G | AACTACACACCAGGGCCAGG | TGCGGCCGCA | CCAGGGGTCAGATATCCACT | NYTPGPGAAAPGVRYP |
| 143 | G | ACTACACACCAGGGCCAGGG | TGCGGCCGCA | CAGGGGTCAGATATCCACTG | YTPGPGCGRTGVRYPL |
| 144 | G | CTACACACCAGGGCCAGGGG | TGCGGCCGCA | AGGGGTCAGATATCCACTGA | YTPGPGVRPQGVRYPL |
| 145 | T | TACACACCAGGGCCAGGGGT | TGCGGCCGCA | GGGGTCAGATATCCACTGAC | YTPGPGVAAAGVRYPL |
| 146 | C | ACACACCAGGGCCAGGGGTC | TGCGGCCGCA | GGGTCAGATATCCACTGACC | TPGPGVCGRRVRYPLT |
| 147 | A | CACACCAGGGCCAGGGGTCA | TGCGGCCGCA | GGTCAGATATCCACTGACCT | TPGPGVMRPQVRYPLT |
| 148 | G | ACACCAGGGCCAGGGGTCAG | TGCGGCCGCA | GTCAGATATCCACTGACCTT | TPGPGVSAAAVRYPLT |
| 149 | A | CACCAGGGCCAGGGGTCAGA | TGCGGCCGCA | TCAGATATCCACTGACCTTT | PGPGVRCGRIRYPLTF |
| 150 | T | ACCAGGGCCAGGGGTCAGAT | TGCGGCCGCA | CAGATATCCACTGACCTTTG | PGPGVRLRPHRYPLTF |
| 151 | A | CCAGGGCCAGGGGTCAGATA | TGCGGCCGCA | AGATATCCACTGACCTTTGG | PGPGVRYAAARYPLTF |
| 152 | T | CAGGGCCAGGGGTCAGATAT | TGCGGCCGCA | GATATCCACTGACCTTTGGA | GPGVRYCGRRYPLTFG |
| 153 | C | AGGGCCAGGGGTCAGATATC | TGCGGCCGCA | ATATCCACTGACCTTTGGAT | GPGVRYLRPQYPLTFG |
| 154 | C | GGGCCAGGGGTCAGATATCC | TGCGGCCGCA | TATCCACTGACCTTTGGATG | GPGVRYPAAAYPLTFG |
| 155 | A | GGCCAGGGGTCAGATATCCA | TGCGGCCGCA | ATCCACTGACCTTTGGATGG | PGVRYPCGRNPLTFGW |
| 156 | C | GCCAGGGGTCAGATATCCAC | TGCGGCCGCA | TCCACTGACCTTTGGATGGT | PGVRYPLRPHPLTFGW |
| 157 | T | CCAGGGGTCAGATATCCACT | TGCGGCCGCA | CCACTGACCTTTGGATGGTG | PGVRYPLAAAPLTFGW |
| 158 | G | CAGGGGTCAGATATCCACTG | TGCGGCCGCA | CACTGACCTTTGGATGGTGC | GVRYPLCGRTLTFGWC |
| 159 | A | AGGGGTCAGATATCCACTGA | TGCGGCCGCA | ACTGACCTTTGGATGGTGCT | GVRYPLMRPQLTFGWC |
| 160 | C | GGGGTCAGATATCCACTGAC | TGCGGCCGCA | CTGACCTTTGGATGGTGCTA | GVRYPLTAAALTFGWC |
| 161 | C | GGGTCAGATATCCACTGACC | TGCGGCCGCA | TGACCTTTGGATGGTGCTAC | VRYPLTCGRMTFGWCY |
| 162 | T | GGTCAGATATCCACTGACCT | TGCGGCCGCA | GACCTTTGGATGGTGCTACA | VRYPLTLRPQTFGWCY |
| 163 | T | GTCAGATATCCACTGACCTT | TGCGGCCGCA | ACCTTTGGATGGTGCTACAA | VRYPLTFAAATFGWCY |
| 164 | T | TCAGATATCCACTGACCTTT | TGCGGCCGCA | CCTTTGGATGGTGCTACAAG | RYPLTFCGRTFGWCYK |
| 165 | G | CAGATATCCACTGACCTTTG | TGCGGCCGCA | CTTTGGATGGTGCTACAAGC | RYPLTFVRPHFGWCYK |
| 166 | G | AGATATCCACTGACCTTTGG | TGCGGCCGCA | TTTGGATGGTGCTACAAGCT | RYPLTFGAAAFGWCYK |

| | | | | | |
|---|---|---|---|---|---|
| 167 | A | GATATCCACTGACCTTTGGA | TGCGGCCGCA | TTGGATGGTGCTACAAGCTA | YPLTFGCGRIGWCYKL |
| 168 | T | ATATCCACTGACCTTTGGAT | TGCGGCCGCA | TGGATGGTGCTACAAGCTAG | YPLTFGLRPHGWCYKL |
| 169 | G | TATCCACTGACCTTTGGATG | TGCGGCCGCA | GGATGGTGCTACAAGCTAGT | YPLTFGCAAAGWCYKL |
| 170 | G | ATCCACTGACCTTTGGATGG | TGCGGCCGCA | GATGGTGCTACAAGCTAGTA | PLTFGWCGRRWCYKLV |
| 171 | T | TCCACTGACCTTTGGATGGT | TGCGGCCGCA | ATGGTGCTACAAGCTAGTAC | PLTFGWLRPQWCYKLV |
| 172 | G | CCACTGACCTTTGGATGGTG | TGCGGCCGCA | TGGTGCTACAAGCTAGTACC | PLTFGWCAAAWCYKLV |
| 173 | C | CACTGACCTTTGGATGGTGC | TGCGGCCGCA | GGTGCTACAAGCTAGTACCA | LTFGWCCGRRCYKLVP |
| 174 | T | ACTGACCTTTGGATGGTGCT | TGCGGCCGCA | GTGCTACAAGCTAGTACCAG | LTFGWCLRPQCYKLVP |
| 175 | A | CTGACCTTTGGATGGTGCTA | TGCGGCCGCA | TGCTACAAGCTAGTACCAGT | LTFGWCYAAACYKLVP |
| 176 | C | TGACCTTTGGATGGTGCTAC | TGCGGCCGCA | GCTACAAGCTAGTACCAGTT | TFGWCYCGRSYKLVPV |
| 177 | A | GACCTTTGGATGGTGCTACA | TGCGGCCGCA | CTACAAGCTAGTACCAGTTG | TFGWCYMRPHYKLVPV |
| 178 | A | ACCTTTGGATGGTGCTACAA | TGCGGCCGCA | TACAAGCTAGTACCAGTTGA | TFGWCYNAAAYKLVPV |
| 179 | G | CCTTTGGATGGTGCTACAAG | TGCGGCCGCA | ACAAGCTAGTACCAGTTGAG | FGWCYKCGRNKLVPVE |
| 180 | C | CTTTGGATGGTGCTACAAGC | TGCGGCCGCA | CAAGCTAGTACCAGTTGAGC | FGWCYKLRPHKLVPVE |
| 181 | T | TTTGGATGGTGCTACAAGCT | TGCGGCCGCA | AAGCTAGTACCAGTTGAGCC | FGWCYKLAAAKLVPVE |
| 182 | A | TTGGATGGTGCTACAAGCTA | TGCGGCCGCA | AGCTAGTACCAGTTGAGCCA | GWCYKLCGRKLVPVEP |
| 183 | G | TGGATGGTGCTACAAGCTAG | TGCGGCCGCA | GCTAGTACCAGTTGAGCCAG | GWCYKLVRPQLVPVEP |
| 184 | T | GGATGGTGCTACAAGCTAGT | TGCGGCCGCA | CTAGTACCAGTTGAGCCAGA | GWCYKLVAAALVPVEP |
| 185 | A | GATGGTGCTACAAGCTAGTA | TGCGGCCGCA | TAGTACCAGTTGAGCCAGAT | WCYKLVCGRIVPVEPD |
| 186 | C | ATGGTGCTACAAGCTAGTAC | TGCGGCCGCA | AGTACCAGTTGAGCCAGATA | WCYKLVLRPQVPVEPD |
| 187 | C | TGGTGCTACAAGCTAGTACC | TGCGGCCGCA | GTACCAGTTGAGCCAGATAA | WCYKLVPAAAVPVEPD |
| 188 | A | GGTGCTACAAGCTAGTACCA | TGCGGCCGCA | TACCAGTTGAGCCAGATAAG | CYKLVPCGRIPVEPDK |
| 189 | G | GTGCTACAAGCTAGTACCAG | TGCGGCCGCA | ACCAGTTGAGCCAGATAAGG | CYKLVPVRPQPVEPDK |
| 190 | T | TGCTACAAGCTAGTACCAGT | TGCGGCCGCA | CCAGTTGAGCCAGATAAGGT | CYKLVPVAAAPVEPDK |
| 191 | T | GCTACAAGCTAGTACCAGTT | TGCGGCCGCA | CAGTTGAGCCAGATAAGGTA | YKLVPVCGRTVEPDKV |
| 192 | G | CTACAAGCTAGTACCAGTTG | TGCGGCCGCA | AGTTGAGCCAGATAAGGTAG | YKLVPVVRPQVEPDKV |
| 193 | A | TACAAGCTAGTACCAGTTGA | TGCGGCCGCA | GTTGAGCCAGATAAGGTAGA | YKLVPVDAAAVEPDKV |
| 194 | G | ACAAGCTAGTACCAGTTGAG | TGCGGCCGCA | TTGAGCCAGATAAGGTAGAA | KLVPVECGRIEPDKVE |
| 195 | C | CAAGCTAGTACCAGTTGAGC | TGCGGCCGCA | TGAGCCAGATAAGGTAGAAG | KLVPVELRPHEPDKVE |
| 196 | C | AAGCTAGTACCAGTTGAGCC | TGCGGCCGCA | GAGCCAGATAAGGTAGAAGA | KLVPVEPAAAEPDKVE |
| 197 | A | AGCTAGTACCAGTTGAGCCA | TGCGGCCGCA | AGCCAGATAAGGTAGAAGAG | LVPVEPCGRKPDKVEE |
| 198 | G | GCTAGTACCAGTTGAGCCAG | TGCGGCCGCA | GCCAGATAAGGTAGAAGAGG | LVPVEPVRPQPDKVEE |
| 199 | A | CTAGTACCAGTTGAGCCAGA | TGCGGCCGCA | CCAGATAAGGTAGAAGAGGC | LVPVEPDAAAPDKVEE |
| 200 | T | TAGTACCAGTTGAGCCAGAT | TGCGGCCGCA | CAGATAAGGTAGAAGAGGCC | VPVEPDCGRTDKVEEA |
| 201 | A | AGTACCAGTTGAGCCAGATA | TGCGGCCGCA | AGATAAGGTAGAAGAGGCCA | VPVEPDMRPQDKVEEA |
| 202 | A | GTACCAGTTGAGCCAGATAA | TGCGGCCGCA | GATAAGGTAGAAGAGGCCAA | VPVEPDNAAADKVEEA |
| 203 | G | TACCAGTTGAGCCAGATAAG | TGCGGCCGCA | ATAAGGTAGAAGAGGCCAAT | PVEPDKCGRNKVEEAN |
| 204 | G | ACCAGTTGAGCCAGATAAGG | TGCGGCCGCA | TAAGGTAGAAGAGGCCAATA | PVEPDKVRPHKVEEAN |
| 205 | T | CCAGTTGAGCCAGATAAGGT | TGCGGCCGCA | AAGGTAGAAGAGGCCAATAA | PVEPDKVAAAKVEEAN |
| 206 | A | CAGTTGAGCCAGATAAGGTA | TGCGGCCGCA | AGGTAGAAGAGGCCAATAAA | VEPDKVCGRKVEEANK |
| 207 | G | AGTTGAGCCAGATAAGGTAG | TGCGGCCGCA | GGTAGAAGAGGCCAATAAAG | VEPDKVVRPQVEEANK |
| 208 | A | GTTGAGCCAGATAAGGTAGA | TGCGGCCGCA | GTAGAAGAGGCCAATAAAGG | VEPDKVDAAAVEEANK |
| 209 | A | TTGAGCCAGATAAGGTAGAA | TGCGGCCGCA | TAGAAGAGGCCAATAAAGGA | EPDKVECGRIEEANKG |
| 210 | G | TGAGCCAGATAAGGTAGAAG | TGCGGCCGCA | AGAAGAGGCCAATAAAGGAG | EPDKVEVRPQEEANKG |
| 211 | A | GAGCCAGATAAGGTAGAAGA | TGCGGCCGCA | GAAGAGGCCAATAAAGGAGA | EPDKVEDAAAEEANKG |
| 212 | G | AGCCAGATAAGGTAGAAGAG | TGCGGCCGCA | AAGAGGCCAATAAAGGAGAG | PDKVEECGRKEANKGE |
| 213 | G | GCCAGATAAGGTAGAAGAGG | TGCGGCCGCA | AGAGGCCAATAAAGGAGAGA | PDKVEEVRPQEANKGE |
| 214 | C | CCAGATAAGGTAGAAGAGGC | TGCGGCCGCA | GAGGCCAATAAAGGAGAGAA | PDKVEEAAAAEANKGE |
| 215 | C | CAGATAAGGTAGAAGAGGCC | TGCGGCCGCA | AGGCCAATAAAGGAGAGAAC | DKVEEACGRKANKGEN |
| 216 | A | AGATAAGGTAGAAGAGGCCA | TGCGGCCGCA | GGCCAATAAAGGAGAGAACA | DKVEEAMRPQANKGEN |
| 217 | A | GATAAGGTAGAAGAGGCCAA | TGCGGCCGCA | GCCAATAAAGGAGAGAACAC | DKVEEANAAAANKGEN |
| 218 | T | ATAAGGTAGAAGAGGCCAAT | TGCGGCCGCA | CCAATAAAGGAGAGAACACC | KVEEANCGRTNKGENT |
| 219 | A | TAAGGTAGAAGAGGCCAATA | TGCGGCCGCA | CAATAAAGGAGAGAACACCA | KVEEANMRPHNKGENT |
| 220 | A | AAGGTAGAAGAGGCCAATAA | TGCGGCCGCA | AATAAAGGAGAGAACACCAG | KVEEANNAAANKGENT |
| 221 | A | AGGTAGAAGAGGCCAATAAA | TGCGGCCGCA | ATAAAGGAGAGAACACCAGC | VEEANKCGRNKGENTS |
| 222 | G | GGTAGAAGAGGCCAATAAAG | TGCGGCCGCA | TAAAGGAGAGAACACCAGCT | VEEANKVRPHKGENTS |
| 223 | G | GTAGAAGAGGCCAATAAAGG | TGCGGCCGCA | AAAGGAGAGAACACCAGCTT | VEEANKGAAAKGENTS |
| 224 | A | TAGAAGAGGCCAATAAAGGA | TGCGGCCGCA | AAGGAGAGAACACCAGCTTG | EEANKGCGRKGENTSL |
| 225 | G | AGAAGAGGCCAATAAAGGAG | TGCGGCCGCA | AGGAGAGAACACCAGCTTGT | EEANKGVRPQGENTSL |
| 226 | A | GAAGAGGCCAATAAAGGAGA | TGCGGCCGCA | GGAGAGAACACCAGCTTGTT | EEANKGDAAAGENTSL |
| 227 | G | AAGAGGCCAATAAAGGAGAG | TGCGGCCGCA | GAGAGAACACCAGCTTGTTA | EANKGECGRRENTSLL |
| 228 | A | AGAGGCCAATAAAGGAGAGA | TGCGGCCGCA | AGAGAACACCAGCTTGTTAC | EANKGEMRPQENTSLL |
| 229 | A | GAGGCCAATAAAGGAGAGAA | TGCGGCCGCA | GAGAACACCAGCTTGTTACA | EANKGENAAAENTSLL |
| 230 | C | AGGCCAATAAAGGAGAGAAC | TGCGGCCGCA | AGAACACCAGCTTGTTACAC | ANKGENCGRKNTSLLH |
| 231 | A | GGCCAATAAAGGAGAGAACA | TGCGGCCGCA | GAACACCAGCTTGTTACACC | ANKGENMRPQNTSLLH |
| 232 | C | GCCAATAAAGGAGAGAACAC | TGCGGCCGCA | AACACCAGCTTGTTACACCC | ANKGENTAAANTSLLH |
| 233 | C | CCAATAAAGGAGAGAACACC | TGCGGCCGCA | ACACCAGCTTGTTACACCCT | NKGENTCGRNTSLLHP |
| 234 | A | CAATAAAGGAGAGAACACCA | TGCGGCCGCA | CACCAGCTTGTTACACCCTG | NKGENTMRPHTSLLHP |
| 235 | G | AATAAAGGAGAGAACACCAG | TGCGGCCGCA | ACCAGCTTGTTACACCCTGT | NKGENTSAAATSLLHP |
| 236 | C | ATAAAGGAGAGAACACCAGC | TGCGGCCGCA | CCAGCTTGTTACACCCTGTG | KGENTSCGRTSLLHPV |
| 237 | T | TAAAGGAGAGAACACCAGCT | TGCGGCCGCA | CAGCTTGTTACACCCTGTGA | KGENTSLRPHSLLHPV |
| 238 | T | AAAGGAGAGAACACCAGCTT | TGCGGCCGCA | AGCTTGTTACACCCTGTGAG | KGENTSFAAASLLHPV |
| 239 | G | AAGGAGAGAACACCAGCTTG | TGCGGCCGCA | GCTTGTTACACCCTGTGAGC | GENTSLCGRSLLHPVS |
| 240 | T | AGGAGAGAACACCAGCTTGT | TGCGGCCGCA | CTTGTTACACCCTGTGAGCC | GENTSLLRPHLLHPVS |
| 241 | T | GGAGAGAACACCAGCTTGTT | TGCGGCCGCA | TTGTTACACCCTGTGAGCCT | GENTSLFAAALLHPVS |
| 242 | A | GAGAGAACACCAGCTTGTTA | TGCGGCCGCA | TGTTACACCCTGTGAGCCTG | ENTSLLCGRMLHPVSL |
| 243 | C | AGAGAACACCAGCTTGTTAC | TGCGGCCGCA | GTTACACCCTGTGAGCCTGC | ENTSLLLRPQLHPVSL |
| 244 | A | GAGAACACCAGCTTGTTACA | TGCGGCCGCA | TTACACCCTGTGAGCCTGCA | ENTSLLHAAALHPVSL |
| 245 | C | AGAACACCAGCTTGTTACAC | TGCGGCCGCA | TACACCCTGTGAGCCTGCAT | NTSLLHCGRIHPVSLH |
| 246 | C | GAACACCAGCTTGTTACACC | TGCGGCCGCA | ACACCCTGTGAGCCTGCATG | NTSLLHLRPQHPVSLH |
| 247 | C | AACACCAGCTTGTTACACCC | TGCGGCCGCA | CACCCTGTGAGCCTGCATGG | NTSLLHPAAAHPVSLH |
| 248 | T | ACACCAGCTTGTTACACCCT | TGCGGCCGCA | ACCCTGTGAGCCTGCATGGA | TSLLHPCGRNPVSLHG |
| 249 | G | CACCAGCTTGTTACACCCTG | TGCGGCCGCA | CCCTGTGAGCCTGCATGGAA | TSLLHPVRPHPVSLHG |
| 250 | T | ACCAGCTTGTTACACCCTGT | TGCGGCCGCA | CCTGTGAGCCTGCATGGAAT | TSLLHPVAAAPVSLHG |
| 251 | G | CCAGCTTGTTACACCCTGTG | TGCGGCCGCA | CTGTGAGCCTGCATGGAATG | SLLHPVCGRTVSLHGM |
| 252 | A | CAGCTTGTTACACCCTGTGA | TGCGGCCGCA | TGTGAGCCTGCATGGAATGG | SLLHPVMRPHVSLHGM |

| | | | | | |
|---|---|---|---|---|---|
| 253 | G | AGCTTGTTACACCCTGTGAG | TGCGGCCGCA | GTGAGCCTGCATGGAATGGA | SLLHPVSAAAVSLHGM |
| 254 | C | GCTTGTTACACCCTGTGAGC | TGCGGCCGCA | TGAGCCTGCATGGAATGGAT | LLHPVSCGRMSLHGMD |
| 255 | C | CTTGTTACACCCTGTGAGCC | TGCGGCCGCA | GAGCCTGCATGGAATGGATG | LLHPVSLRPQSLHGMD |
| 256 | T | TTGTTACACCCTGTGAGCCT | TGCGGCCGCA | AGCCTGCATGGAATGGATGA | LLHPVSLAAASLHGMD |
| 257 | G | TGTTACACCCTGTGAGCCTG | TGCGGCCGCA | GCCTGCATGGAATGGATGAC | LHPVSLCGRSLHGMDD |
| 258 | C | GTTACACCCTGTGAGCCTGC | TGCGGCCGCA | CCTGCATGGAATGGATGACC | LHPVSLLRPHLHGMDD |
| 259 | A | TTACACCCTGTGAGCCTGCA | TGCGGCCGCA | CTGCATGGAATGGATGACCC | LHPVSLHAAALHGMDD |
| 260 | T | TACACCCTGTGAGCCTGCAT | TGCGGCCGCA | TGCATGGAATGGATGACCCT | HPVSLHCGRMHGMDDP |
| 261 | G | ACACCCTGTGAGCCTGCATG | TGCGGCCGCA | GCATGGAATGGATGACCCTG | HPVSLHVRPQHGMDDP |
| 262 | G | CACCCTGTGAGCCTGCATGG | TGCGGCCGCA | CATGGAATGGATGACCCTGA | HPVSLHGAAAHGMDDP |
| 263 | A | ACCCTGTGAGCCTGCATGGA | TGCGGCCGCA | ATGGAATGGATGACCCTGAG | PVSLHGCGRNGMDDPE |
| 264 | A | CCCTGTGAGCCTGCATGGAA | TGCGGCCGCA | TGGAATGGATGACCCTGAGA | PVSLHGMRPHGMDDPE |
| 265 | T | CCTGTGAGCCTGCATGGAAT | TGCGGCCGCA | GGAATGGATGACCCTGAGAG | PVSLHGIAAAGMDDPE |
| 266 | G | CTGTGAGCCTGCATGGAATG | TGCGGCCGCA | GAATGGATGACCCTGAGAGA | VSLHGMCGRRMDDPER |
| 267 | G | TGTGAGCCTGCATGGAATGG | TGCGGCCGCA | AATGGATGACCCTGAGAGAG | VSLHGMVRPQMDDPER |
| 268 | A | GTGAGCCTGCATGGAATGGA | TGCGGCCGCA | ATGGATGACCCTGAGAGAGA | VSLHGMDAAAMDDPER |
| 269 | T | TGAGCCTGCATGGAATGGAT | TGCGGCCGCA | TGGATGACCCTGAGAGAGAA | SLHGMDCGRMDDPERE |
| 270 | G | GAGCCTGCATGGAATGGATG | TGCGGCCGCA | GGATGACCCTGAGAGAGAAG | SLHGMDVRPQDDPERE |
| 271 | A | AGCCTGCATGGAATGGATGA | TGCGGCCGCA | GATGACCCTGAGAGAGAAGT | SLHGMDDAAADDPERE |
| 272 | C | GCCTGCATGGAATGGATGAC | TGCGGCCGCA | ATGACCCTGAGAGAGAAGTG | LHGMDDCGRNDPEREV |
| 273 | C | CCTGCATGGAATGGATGACC | TGCGGCCGCA | TGACCCTGAGAGAGAAGTGT | LHGMDDLRPHDPEREV |
| 274 | C | CTGCATGGAATGGATGACCC | TGCGGCCGCA | GACCCTGAGAGAGAAGTGTT | LHGMDDPAAADPEREV |
| 275 | T | TGCATGGAATGGATGACCCT | TGCGGCCGCA | ACCCTGAGAGAGAAGTGTTA | HGMDDPCGRNPEREVL |
| 276 | G | GCATGGAATGGATGACCCTG | TGCGGCCGCA | CCCTGAGAGAGAAGTGTTAG | HGMDDPVRPHPEREVL |
| 277 | A | CATGGAATGGATGACCCTGA | TGCGGCCGCA | CCTGAGAGAGAAGTGTTAGA | HGMDDPDAAAPEREVL |
| 278 | G | ATGGAATGGATGACCCTGAG | TGCGGCCGCA | CTGAGAGAGAAGTGTTAGAG | GMDDPECGRTEREVLE |
| 279 | A | TGGAATGGATGACCCTGAGA | TGCGGCCGCA | TGAGAGAGAAGTGTTAGAGT | GMDDPEMRPHEREVLE |
| 280 | G | GGAATGGATGACCCTGAGAG | TGCGGCCGCA | GAGAGAGAAGTGTTAGAGTG | GMDDPESAAAEREVLE |
| 281 | A | GAATGGATGACCCTGAGAGA | TGCGGCCGCA | AGAGAGAAGTGTTAGAGTGG | MDDPERCGRKREVLEW |
| 282 | G | AATGGATGACCCTGAGAGAG | TGCGGCCGCA | GAGAGAAGTGTTAGAGTGGA | MDDPERVRPQREVLEW |
| 283 | A | ATGGATGACCCTGAGAGAGA | TGCGGCCGCA | AGAGAAGTGTTAGAGTGGAG | MDDPERDAAAREVLEW |
| 284 | A | TGGATGACCCTGAGAGAGAA | TGCGGCCGCA | GAGAAGTGTTAGAGTGGAGG | DDPERECGRREVLEWR |
| 285 | G | GGATGACCCTGAGAGAGAAG | TGCGGCCGCA | AGAAGTGTTAGAGTGGAGGT | DDPEREVRPQEVLEWR |
| 286 | T | GATGACCCTGAGAGAGAAGT | TGCGGCCGCA | GAAGTGTTAGAGTGGAGGTT | DDPEREVAAAEVLEWR |
| 287 | G | ATGACCCTGAGAGAGAAGTG | TGCGGCCGCA | AAGTGTTAGAGTGGAGGTTT | DPEREVCGRKVLEWRF |
| 288 | T | TGACCCTGAGAGAGAAGTGT | TGCGGCCGCA | AGTGTTAGAGTGGAGGTTTG | DPEREVLRPQVLEWRF |
| 289 | T | GACCCTGAGAGAGAAGTGTT | TGCGGCCGCA | GTGTTAGAGTGGAGGTTTGA | DPEREVFAAAVLEWRF |
| 290 | A | ACCCTGAGAGAGAAGTGTTA | TGCGGCCGCA | TGTTAGAGTGGAGGTTTGAC | PEREVLCGRMLEWRFD |
| 291 | G | CCCTGAGAGAGAAGTGTTAG | TGCGGCCGCA | GTTAGAGTGGAGGTTTGACA | PEREVLVRPQLEWRFD |
| 292 | A | CCTGAGAGAGAAGTGTTAGA | TGCGGCCGCA | TTAGAGTGGAGGTTTGACAG | PEREVLDAAALEWRFD |
| 293 | G | CTGAGAGAGAAGTGTTAGAG | TGCGGCCGCA | TAGAGTGGAGGTTTGACAGC | EREVLECGRIEWRFDS |
| 294 | T | TGAGAGAGAAGTGTTAGAGT | TGCGGCCGCA | AGAGTGGAGGTTTGACAGCC | EREVLELRPQEWRFDS |
| 295 | G | GAGAGAGAAGTGTTAGAGTG | TGCGGCCGCA | GAGTGGAGGTTTGACAGCCG | EREVLECAAAEWRFDS |
| 296 | G | AGAGAGAAGTGTTAGAGTGG | TGCGGCCGCA | AGTGGAGGTTTGACAGCCGC | REVLEWCGRKWRFDSR |
| 297 | A | GAGAGAAGTGTTAGAGTGGA | TGCGGCCGCA | GTGGAGGTTTGACAGCCGCC | REVLEWMRPQWRFDSR |
| 298 | G | AGAGAAGTGTTAGAGTGGAG | TGCGGCCGCA | TGGAGGTTTGACAGCCGCCT | REVLEWSAAAWRFDSR |
| 299 | G | GAGAAGTGTTAGAGTGGAGG | TGCGGCCGCA | GGAGGTTTGACAGCCGCCTA | EVLEWRCGRRRFDSRL |
| 300 | T | AGAAGTGTTAGAGTGGAGGT | TGCGGCCGCA | GAGGTTTGACAGCCGCCTAG | EVLEWRLRPQRFDSRL |
| 301 | T | GAAGTGTTAGAGTGGAGGTT | TGCGGCCGCA | AGGTTTGACAGCCGCCTAGC | EVLEWRFAAARFDSRL |
| 302 | T | AAGTGTTAGAGTGGAGGTTT | TGCGGCCGCA | GGTTTGACAGCCGCCTAGCA | VLEWRFCGRRFDSRLA |
| 303 | G | AGTGTTAGAGTGGAGGTTTG | TGCGGCCGCA | GTTTGACAGCCGCCTAGCAT | VLEWRFVRPQFDSRLA |
| 304 | A | GTGTTAGAGTGGAGGTTTGA | TGCGGCCGCA | TTTGACAGCCGCCTAGCATT | VLEWRFDAAAFDSRLA |
| 305 | C | TGTTAGAGTGGAGGTTTGAC | TGCGGCCGCA | TTGACAGCCGCCTAGCATTT | LEWRFDCGRIDSRLAF |
| 306 | A | GTTAGAGTGGAGGTTTGACA | TGCGGCCGCA | TGACAGCCGCCTAGCATTTC | LEWRFDMRPHDSRLAF |
| 307 | G | TTAGAGTGGAGGTTTGACAG | TGCGGCCGCA | GACAGCCGCCTAGCATTTCA | LEWRFDSAAADSRLAF |
| 308 | C | TAGAGTGGAGGTTTGACAGC | TGCGGCCGCA | ACAGCCGCCTAGCATTTCAT | EWRFDSCGRNSRLAFH |
| 309 | C | AGAGTGGAGGTTTGACAGCC | TGCGGCCGCA | CAGCCGCCTAGCATTTCATC | EWRFDSLRPHSRLAFH |
| 310 | G | GAGTGGAGGTTTGACAGCCG | TGCGGCCGCA | AGCCGCCTAGCATTTCATCA | EWRFDSRAAASRLAFH |
| 311 | C | AGTGGAGGTTTGACAGCCGC | TGCGGCCGCA | GCCGCCTAGCATTTCATCAC | WRFDSRCGRSRLAFHH |
| 312 | C | GTGGAGGTTTGACAGCCGCC | TGCGGCCGCA | CCGCCTAGCATTTCATCACG | WRFDSRLRPHRLAFHH |
| 313 | T | TGGAGGTTTGACAGCCGCCT | TGCGGCCGCA | CGCCTAGCATTTCATCACGT | WRFDSRLAAARLAFHH |
| 314 | A | GGAGGTTTGACAGCCGCCTA | TGCGGCCGCA | GCCTAGCATTTCATCACGTG | RFDSRLCGRSLAFHHV |
| 315 | G | GAGGTTTGACAGCCGCCTAG | TGCGGCCGCA | CCTAGCATTTCATCACGTGG | RFDSRLVRPHLAFHHV |
| 316 | C | AGGTTTGACAGCCGCCTAGC | TGCGGCCGCA | CTAGCATTTCATCACGTGGC | RFDSRLAAAALAFHHV |
| 317 | A | GGTTTGACAGCCGCCTAGCA | TGCGGCCGCA | TAGCATTTCATCACGTGGCC | FDSRLACGRIAFHHVA |
| 318 | T | GTTTGACAGCCGCCTAGCAT | TGCGGCCGCA | AGCATTTCATCACGTGGCCC | FDSRLALRPQAFHHVA |
| 319 | T | TTTGACAGCCGCCTAGCATT | TGCGGCCGCA | GCATTTCATCACGTGGCCCG | FDSRLAFAAAAFHHVA |
| 320 | T | TTGACAGCCGCCTAGCATTT | TGCGGCCGCA | CATTTCATCACGTGGCCCGA | DSRLAFCGRTFHHVAR |
| 321 | C | TGACAGCCGCCTAGCATTTC | TGCGGCCGCA | ATTTCATCACGTGGCCCGAG | DSRLAFLRPQFHHVAR |
| 322 | A | GACAGCCGCCTAGCATTTCA | TGCGGCCGCA | TTTCATCACGTGGCCCGAGA | DSRLAFHAAAFHHVAR |
| 323 | T | ACAGCCGCCTAGCATTTCAT | TGCGGCCGCA | TTCATCACGTGGCCCGAGAG | SRLAFHCGRIHHVARE |
| 324 | C | CAGCCGCCTAGCATTTCATC | TGCGGCCGCA | TCATCACGTGGCCCGAGAGC | SRLAFHLRPHHHVARE |
| 325 | A | AGCCGCCTAGCATTTCATCA | TGCGGCCGCA | CATCACGTGGCCCGAGAGCT | SRLAFHHAAAHHVARE |
| 326 | C | GCCGCCTAGCATTTCATCAC | TGCGGCCGCA | ATCACGTGGCCCGAGAGCTG | RLAFHHCGRNHVAREL |
| 327 | G | CCGCCTAGCATTTCATCACG | TGCGGCCGCA | TCACGTGGCCCGAGAGCTGC | RLAFHHVRPHHVAREL |
| 328 | T | CGCCTAGCATTTCATCACGT | TGCGGCCGCA | CACGTGGCCCGAGAGCTGCA | RLAFHHVAAAHVAREL |
| 329 | G | GCCTAGCATTTCATCACGTG | TGCGGCCGCA | ACGTGGCCCGAGAGCTGCAT | LAFHHVCGRNVARELH |
| 330 | G | CCTAGCATTTCATCACGTGG | TGCGGCCGCA | CGTGGCCCGAGAGCTGCATC | LAFHHVVRPHVARELH |
| 331 | C | CTAGCATTTCATCACGTGGC | TGCGGCCGCA | GTGGCCCGAGAGCTGCATCC | LAFHHVAAAAVARELH |
| 332 | C | TAGCATTTCATCACGTGGCC | TGCGGCCGCA | TGGCCCGAGAGCTGCATCCG | AFHHVACGRMARELHP |
| 333 | C | AGCATTTCATCACGTGGCCC | TGCGGCCGCA | GGCCCGAGAGCTGCATCCGG | AFHHVALRPQARELHP |
| 334 | G | GCATTTCATCACGTGGCCCG | TGCGGCCGCA | GCCCGAGAGCTGCATCCGGA | AFHHVARAAAARELHP |
| 335 | A | CATTTCATCACGTGGCCCGA | TGCGGCCGCA | CCCGAGAGCTGCATCCGGAG | FHHVARCGRTRELHPE |
| 336 | G | ATTTCATCACGTGGCCCGAG | TGCGGCCGCA | CCGAGAGCTGCATCCGGAGT | FHHVARVRPHRELHPE |
| 337 | A | TTTCATCACGTGGCCCGAGA | TGCGGCCGCA | CGAGAGCTGCATCCGGAGTA | FHHVARDAAARELHPE |
| 338 | G | TTCATCACGTGGCCCGAGAG | TGCGGCCGCA | GAGAGCTGCATCCGGAGTAC | HHVARECGRRELHPEY |

| | | | | | |
|---|---|---|---|---|---|
| 339 | C | TCATCACGTGGCCCGAGAGC | TGCGGCCGCA | AGAGCTGCATCCGGAGTACT | HHVARELRPQELHPEY |
| 340 | T | CATCACGTGGCCCGAGAGCT | TGCGGCCGCA | GAGCTGCATCCGGAGTACTT | HHVARELAAAELHPEY |
| 341 | G | ATCACGTGGCCCGAGAGCTG | TGCGGCCGCA | AGCTGCATCCGGAGTACTTC | HVARELCGRKLHPEYF |
| 342 | C | TCACGTGGCCCGAGAGCTGC | TGCGGCCGCA | GCTGCATCCGGAGTACTTCA | HVARELLRPQLHPEYF |
| 343 | A | CACGTGGCCCGAGAGCTGCA | TGCGGCCGCA | CTGCATCCGGAGTACTTCAA | HVARELHAAALHPEYF |
| 344 | T | ACGTGGCCCGAGAGCTGCAT | TGCGGCCGCA | TGCATCCGGAGTACTTCAAG | VARELHCGRMHPEYFK |
| 345 | C | CGTGGCCCGAGAGCTGCATC | TGCGGCCGCA | GCATCCGGAGTACTTCAAGA | VARELHLRPQHPEYFK |
| 346 | C | GTGGCCCGAGAGCTGCATCC | TGCGGCCGCA | CATCCGGAGTACTTCAAGAA | VARELHPAAAHPEYFK |
| 347 | G | TGGCCCGAGAGCTGCATCCG | TGCGGCCGCA | ATCCGGAGTACTTCAAGAAC | ARELHPCGRNPEYFKN |
| 348 | G | GGCCCGAGAGCTGCATCCGG | TGCGGCCGCA | TCCGGAGTACTTCAAGAACT | ARELHPVRPHPEYFKN |
| 349 | A | GCCCGAGAGCTGCATCCGGA | TGCGGCCGCA | CCGGAGTACTTCAAGAACTG | ARELHPDAAAPEYFKN |
| 350 | G | CCCGAGAGCTGCATCCGGAG | TGCGGCCGCA | CGGAGTACTTCAAGAACTGC | RELHPECGRTEYFKNC |
| 351 | T | CCGAGAGCTGCATCCGGAGT | TGCGGCCGCA | GGAGTACTTCAAGAACTGCT | RELHPELRPQEYFKNC |
| 352 | A | CGAGAGCTGCATCCGGAGTA | TGCGGCCGCA | GAGTACTTCAAGAACTGCTG | RELHPEYAAAEYFKNC |
| 353 | C | GAGAGCTGCATCCGGAGTAC | TGCGGCCGCA | AGTACTTCAAGAACTGCTGA | ELHPEYCGRKYFKNC* |
| 354 | T | AGAGCTGCATCCGGAGTACT | TGCGGCCGCA | GTACTTCAAGAACTGCTGAC | ELHPEYLRPQYFKNC* |
| 355 | T | GAGCTGCATCCGGAGTACTT | TGCGGCCGCA | TACTTCAAGAACTGCTGACA | ELHPEYFAAAYFKNC* |
| 356 | C | AGCTGCATCCGGAGTACTTC | TGCGGCCGCA | ACTTCAAGAACTGCTGACAT | LHPEYFCGRNFKNC*H |
| 357 | A | GCTGCATCCGGAGTACTTCA | TGCGGCCGCA | CTTCAAGAACTGCTGACATC | LHPEYFMRPHFKNC*H |
| 358 | A | CTGCATCCGGAGTACTTCAA | TGCGGCCGCA | TTCAAGAACTGCTGACATCG | LHPEYFNAAAFKNC*H |
| 359 | G | TGCATCCGGAGTACTTCAAG | TGCGGCCGCA | TCAAGAACTGCTGACATCGA | HPEYFKCGRIKNC*HR |
| 360 | A | GCATCCGGAGTACTTCAAGA | TGCGGCCGCA | CAAGAACTGCTGACATCGAG | HPEYFKMRPHKNC*HR |
| 361 | A | CATCCGGAGTACTTCAAGAA | TGCGGCCGCA | AAGAACTGCTGACATCGAGC | HPEYFKNAAAKNC*HR |
| 362 | C | ATCCGGAGTACTTCAAGAAC | TGCGGCCGCA | AGAACTGCTGACATCGAGCT | PEYFKNCGRKNC*HRA |
| 363 | T | TCCGGAGTACTTCAAGAACT | TGCGGCCGCA | GAACTGCTGACATCGAGCTT | PEYFKNLRPQNC*HRA |
| 364 | G | CCGGAGTACTTCAAGAACTG | TGCGGCCGCA | AACTGCTGACATCGAGCTTG | PEYFKNCAAANC*HRA |
| 365 | C | CGGAGTACTTCAAGAACTGC | TGCGGCCGCA | ACTGCTGACATCGAGCTTGC | EYFKNCCGRNC*HRAC |
| 366 | T | GGAGTACTTCAAGAACTGCT | TGCGGCCGCA | CTGCTGACATCGAGCTTGCT | EYFKNCLRPHC*HRAC |
| 367 | G | GAGTACTTCAAGAACTGCTG | TGCGGCCGCA | TGCTGACATCGAGCTTGCTA | EYFKNCCAAAC*HRAC |
| 368 | A | AGTACTTCAAGAACTGCTGA | TGCGGCCGCA | GCTGACATCGAGCTTGCTAC | YFKNC*CGRS*HRACY |
| 369 | C | GTACTTCAAGAACTGCTGAC | TGCGGCCGCA | CTGACATCGAGCTTGCTACA | YFKNC*LRPH*HRACY |
| 370 | A | TACTTCAAGAACTGCTGACA | TGCGGCCGCA | TGACATCGAGCTTGCTACAA | YFKNC*HAAA*HRACY |
| 371 | T | ACTTCAAGAACTGCTGACAT | TGCGGCCGCA | GACATCGAGCTTGCTACAAG | FKNC*HCGRRHRACYK |
| 372 | C | CTTCAAGAACTGCTGACATC | TGCGGCCGCA | ACATCGAGCTTGCTACAAGG | FKNC*HLRPQHRACYK |
| 373 | G | TTCAAGAACTGCTGACATCG | TGCGGCCGCA | CATCGAGCTTGCTACAAGGG | FKNC*HRAAAHRACYK |
| 374 | A | TCAAGAACTGCTGACATCGA | TGCGGCCGCA | ATCGAGCTTGCTACAAGGGA | KNC*HRCGRNRACYKG |
| 375 | G | CAAGAACTGCTGACATCGAG | TGCGGCCGCA | TCGAGCTTGCTACAAGGGAC | KNC*HRVRPHRACYKG |
| 376 | C | AAGAACTGCTGACATCGAGC | TGCGGCCGCA | CGAGCTTGCTACAAGGGACT | KNC*HRAAAARACYKG |
| 377 | T | AGAACTGCTGACATCGAGCT | TGCGGCCGCA | GAGCTTGCTACAAGGGACTT | NC*HRACGRRACYKGL |
| 378 | T | GAACTGCTGACATCGAGCTT | TGCGGCCGCA | AGCTTGCTACAAGGGACTTT | NC*HRALRPQACYKGL |
| 379 | G | AACTGCTGACATCGAGCTTG | TGCGGCCGCA | GCTTGCTACAAGGGACTTTC | NC*HRACAAAACYKGL |
| 380 | C | ACTGCTGACATCGAGCTTGC | TGCGGCCGCA | CTTGCTACAAGGGACTTTCC | C*HRACCGRTCYKGLS |
| 381 | T | CTGCTGACATCGAGCTTGCT | TGCGGCCGCA | TTGCTACAAGGGACTTTCCG | C*HRACLRPHCYKGLS |
| 382 | A | TGCTGACATCGAGCTTGCTA | TGCGGCCGCA | TGCTACAAGGGACTTTCCGC | C*HRACYAAACYKGLS |
| 383 | C | GCTGACATCGAGCTTGCTAC | TGCGGCCGCA | GCTACAAGGGACTTTCCGCT | *HRACYCGRSYKGLSA |
| 384 | A | CTGACATCGAGCTTGCTACA | TGCGGCCGCA | CTACAAGGGACTTTCCGCTG | *HRACYMRPHYKGLSA |
| 385 | A | TGACATCGAGCTTGCTACAA | TGCGGCCGCA | TACAAGGGACTTTCCGCTGG | *HRACYNAAAYKGLSA |
| 386 | G | GACATCGAGCTTGCTACAAG | TGCGGCCGCA | ACAAGGGACTTTCCGCTGGG | HRACYKCGRNKGLSAG |
| 387 | G | ACATCGAGCTTGCTACAAGG | TGCGGCCGCA | CAAGGGACTTTCCGCTGGGG | HRACYKVRPHKGLSAG |
| 388 | G | CATCGAGCTTGCTACAAGGG | TGCGGCCGCA | AAGGGACTTTCCGCTGGGGA | HRACYKGAAAKGLSAG |
| 389 | A | ATCGAGCTTGCTACAAGGGA | TGCGGCCGCA | AGGGACTTTCCGCTGGGGAC | RACYKGCGRKGLSAGD |
| 390 | C | TCGAGCTTGCTACAAGGGAC | TGCGGCCGCA | GGGACTTTCCGCTGGGGACT | RACYKGLRPQGLSAGD |
| 391 | T | CGAGCTTGCTACAAGGGACT | TGCGGCCGCA | GGACTTTCCGCTGGGGACTT | RACYKGLAAAGLSAGD |
| 392 | T | GAGCTTGCTACAAGGGACTT | TGCGGCCGCA | GACTTTCCGCTGGGGACTTT | ACYKGLCGRRLSAGDF |
| 393 | T | AGCTTGCTACAAGGGACTTT | TGCGGCCGCA | ACTTTCCGCTGGGGACTTTC | ACYKGLLRPQLSAGDF |
| 394 | C | GCTTGCTACAAGGGACTTTC | TGCGGCCGCA | CTTTCCGCTGGGGACTTTCC | ACYKGLSAAALSAGDF |
| 395 | C | CTTGCTACAAGGGACTTTCC | TGCGGCCGCA | TTTCCGCTGGGGACTTTCCA | CYKGLSCGRISAGDFP |
| 396 | G | TTGCTACAAGGGACTTTCCG | TGCGGCCGCA | TTCCGCTGGGGACTTTCCAG | CYKGLSVRPHSAGDFP |
| 397 | C | TGCTACAAGGGACTTTCCGC | TGCGGCCGCA | TCCGCTGGGGACTTTCCAGG | CYKGLSAAAASAGDFP |
| 398 | T | GCTACAAGGGACTTTCCGCT | TGCGGCCGCA | CCGCTGGGGACTTTCCAGGG | YKGLSACGRTAGDFPG |
| 399 | G | CTACAAGGGACTTTCCGCTG | TGCGGCCGCA | CGCTGGGGACTTTCCAGGGA | YKGLSAVRPHAGDFPG |
| 400 | G | TACAAGGGACTTTCCGCTGG | TGCGGCCGCA | GCTGGGGACTTTCCAGGGAG | YKGLSAGAAAAGDFPG |
| 401 | G | ACAAGGGACTTTCCGCTGGG | TGCGGCCGCA | CTGGGGACTTTCCAGGGAGG | KGLSAGCGRTGDFPGR |
| 402 | G | CAAGGGACTTTCCGCTGGGG | TGCGGCCGCA | TGGGGACTTTCCAGGGAGGC | KGLSAGVRPHGDFPGR |
| 403 | A | AAGGGACTTTCCGCTGGGGA | TGCGGCCGCA | GGGGACTTTCCAGGGAGGCG | KGLSAGDAAAGDFPGR |
| 404 | C | AGGGACTTTCCGCTGGGGAC | TGCGGCCGCA | GGGACTTTCCAGGGAGGCGT | GLSAGDCGRRDFPGRR |
| 405 | T | GGGACTTTCCGCTGGGGACT | TGCGGCCGCA | GGACTTTCCAGGGAGGCGTG | GLSAGDLRPQDFPGRR |
| 406 | T | GGACTTTCCGCTGGGGACTT | TGCGGCCGCA | GACTTTCCAGGGAGGCGTGG | GLSAGDFAAADFPGRR |
| 407 | T | GACTTTCCGCTGGGGACTTT | TGCGGCCGCA | ACTTTCCAGGGAGGCGTGGC | LSAGDFCGRNFPGRRG |
| 408 | C | ACTTTCCGCTGGGGACTTTC | TGCGGCCGCA | CTTTCCAGGGAGGCGTGGCC | LSAGDFLRPHFPGRRG |
| 409 | C | CTTTCCGCTGGGGACTTTCC | TGCGGCCGCA | TTTCCAGGGAGGCGTGGCCT | LSAGDFPAAAFPGRRG |
| 410 | A | TTTCCGCTGGGGACTTTCCA | TGCGGCCGCA | TTCCAGGGAGGCGTGGCCTG | SAGDFPCGRIPGRRGL |
| 411 | G | TTCCGCTGGGGACTTTCCAG | TGCGGCCGCA | TCCAGGGAGGCGTGGCCTGG | SAGDFPVRPHPGRRGL |
| 412 | G | TCCGCTGGGGACTTTCCAGG | TGCGGCCGCA | CCAGGGAGGCGTGGCCTGGG | SAGDFPGAAAPGRRGL |
| 413 | G | CCGCTGGGGACTTTCCAGGG | TGCGGCCGCA | CAGGGAGGCGTGGCCTGGGC | AGDFPGCGRTGRRGLG |
| 414 | A | CGCTGGGGACTTTCCAGGGA | TGCGGCCGCA | AGGGAGGCGTGGCCTGGGCG | AGDFPGMRPQGRRGLG |
| 415 | G | GCTGGGGACTTTCCAGGGAG | TGCGGCCGCA | GGGAGGCGTGGCCTGGGCGG | AGDFPGSAAAGRRGLG |
| 416 | G | CTGGGGACTTTCCAGGGAGG | TGCGGCCGCA | GGAGGCGTGGCCTGGGCGGG | GDFPGRCGRRRRGLGG |
| 417 | C | TGGGGACTTTCCAGGGAGGC | TGCGGCCGCA | GAGGCGTGGCCTGGGCGGGA | GDFPGRLRPQRRGLGG |
| 418 | G | GGGGACTTTCCAGGGAGGCG | TGCGGCCGCA | AGGCGTGGCCTGGGCGGGAC | GDFPGRRAAARRGLGG |
| 419 | T | GGGACTTTCCAGGGAGGCGT | TGCGGCCGCA | GGCGTGGCCTGGGCGGGACT | DFPGRRCGRRRGLGGT |
| 420 | G | GGACTTTCCAGGGAGGCGTG | TGCGGCCGCA | GCGTGGCCTGGGCGGGACTG | DFPGRRVRPQRGLGGT |
| 421 | G | GACTTTCCAGGGAGGCGTGG | TGCGGCCGCA | CGTGGCCTGGGCGGGACTGG | DFPGRRGAAARGLGGT |
| 422 | C | ACTTTCCAGGGAGGCGTGGC | TGCGGCCGCA | GTGGCCTGGGCGGGACTGGG | FPGRRGCGRSGLGGTG |
| 423 | C | CTTTCCAGGGAGGCGTGGCC | TGCGGCCGCA | TGGCCTGGGCGGGACTGGGG | FPGRRGLRPHGLGGTG |
| 424 | T | TTTCCAGGGAGGCGTGGCCT | TGCGGCCGCA | GGCCTGGGCGGGACTGGGGA | FPGRRGLAAAGLGGTG |

| | | | | | |
|---|---|---|---|---|---|
| 425 | G | TTCCAGGGAGGCGTGGCCTG | TGCGGCCGCA | GCCTGGGCGGGACTGGGGAG | PGRRGLCGRSLGGTGE |
| 426 | G | TCCAGGGAGGCGTGGCCTGG | TGCGGCCGCA | CCTGGGCGGGACTGGGGAGT | PGRRGLVRPHLGGTGE |
| 427 | G | CCAGGGAGGCGTGGCCTGGG | TGCGGCCGCA | CTGGGCGGGACTGGGGAGTG | PGRRGLGAAALGGTGE |
| 428 | C | CAGGGAGGCGTGGCCTGGGC | TGCGGCCGCA | TGGGCGGGACTGGGGAGTGG | GRRGLGCGRMGGTGEW |
| 429 | G | AGGGAGGCGTGGCCTGGGCG | TGCGGCCGCA | GGGCGGGACTGGGGAGTGGC | GRRGLGVRPQGGTGEW |
| 430 | G | GGGAGGCGTGGCCTGGGCGG | TGCGGCCGCA | GGCGGGACTGGGGAGTGGCG | GRRGLGGAAAGGTGEW |
| 431 | G | GGAGGCGTGGCCTGGGCGGG | TGCGGCCGCA | GCGGGACTGGGGAGTGGCGA | RRGLGGCGRSGTGEWR |
| 432 | A | GAGGCGTGGCCTGGGCGGGA | TGCGGCCGCA | CGGGACTGGGGAGTGGCGAG | RRGLGGMRPHGTGEWR |
| 433 | C | AGGCGTGGCCTGGGCGGGAC | TGCGGCCGCA | GGGACTGGGGAGTGGCGAGC | RRGLGGTAAAGTGEWR |
| 434 | T | GGCGTGGCCTGGGCGGGACT | TGCGGCCGCA | GGACTGGGGAGTGGCGAGCC | RGLGGTCGRRTGEWRA |
| 435 | G | GCGTGGCCTGGGCGGGACTG | TGCGGCCGCA | GACTGGGGAGTGGCGAGCCC | RGLGGTVRPQTGEWRA |
| 436 | G | CGTGGCCTGGGCGGGACTGG | TGCGGCCGCA | ACTGGGGAGTGGCGAGCCCT | RGLGGTGAAATGEWRA |
| 437 | G | GTGGCCTGGGCGGGACTGGG | TGCGGCCGCA | CTGGGGAGTGGCGAGCCCTC | GLGGTGCGRTGEWRAL |
| 438 | G | TGGCCTGGGCGGGACTGGGG | TGCGGCCGCA | TGGGGAGTGGCGAGCCCTCA | GLGGTGVRPHGEWRAL |
| 439 | A | GGCCTGGGCGGGACTGGGGA | TGCGGCCGCA | GGGGAGTGGCGAGCCCTCAG | GLGGTGDAAAGEWRAL |
| 440 | G | GCCTGGGCGGGACTGGGGAG | TGCGGCCGCA | GGGAGTGGCGAGCCCTCAGA | LGGTGECGRREWRALR |
| 441 | T | CCTGGGCGGGACTGGGGAGT | TGCGGCCGCA | GGAGTGGCGAGCCCTCAGAT | LGGTGELRPQEWRALR |
| 442 | G | CTGGGCGGGACTGGGGAGTG | TGCGGCCGCA | GAGTGGCGAGCCCTCAGATG | LGGTGECAAAEWRALR |
| 443 | G | TGGGCGGGACTGGGGAGTGG | TGCGGCCGCA | AGTGGCGAGCCCTCAGATGC | GGTGEWCGRKWRALRC |
| 444 | C | GGGCGGGACTGGGGAGTGGC | TGCGGCCGCA | GTGGCGAGCCCTCAGATGCT | GGTGEWLRPQWRALRC |
| 445 | G | GGCGGGACTGGGGAGTGGCG | TGCGGCCGCA | TGGCGAGCCCTCAGATGCTG | GGTGEWRAAAWRALRC |
| 446 | A | GCGGGACTGGGGAGTGGCGA | TGCGGCCGCA | GGCGAGCCCTCAGATGCTGC | GTGEWRCGRRRALRCC |
| 447 | G | CGGGACTGGGGAGTGGCGAG | TGCGGCCGCA | GCGAGCCCTCAGATGCTGCA | GTGEWRVRPQRALRCC |
| 448 | C | GGGACTGGGGAGTGGCGAGC | TGCGGCCGCA | CGAGCCCTCAGATGCTGCAT | GTGEWRAAAARALRCC |
| 449 | C | GGACTGGGGAGTGGCGAGCC | TGCGGCCGCA | GAGCCCTCAGATGCTGCATA | TGEWRACGRRALRCCI |
| 450 | C | GACTGGGGAGTGGCGAGCCC | TGCGGCCGCA | AGCCCTCAGATGCTGCATAT | TGEWRALRPQALRCCI |
| 451 | T | ACTGGGGAGTGGCGAGCCCT | TGCGGCCGCA | GCCCTCAGATGCTGCATATA | TGEWRALAAAALRCCI |
| 452 | C | CTGGGGAGTGGCGAGCCCTC | TGCGGCCGCA | CCCTCAGATGCTGCATATAA | GEWRALCGRTLRCCI* |
| 453 | A | TGGGGAGTGGCGAGCCCTCA | TGCGGCCGCA | CCTCAGATGCTGCATATAAG | GEWRALMRPHLRCCI* |
| 454 | G | GGGGAGTGGCGAGCCCTCAG | TGCGGCCGCA | CTCAGATGCTGCATATAAGC | GEWRALSAAALRCCI* |
| 455 | A | GGGAGTGGCGAGCCCTCAGA | TGCGGCCGCA | TCAGATGCTGCATATAAGCA | EWRALRCGRIRCCI*A |
| 456 | T | GGAGTGGCGAGCCCTCAGAT | TGCGGCCGCA | CAGATGCTGCATATAAGCAG | EWRALRLRPHRCCI*A |
| 457 | G | GAGTGGCGAGCCCTCAGATG | TGCGGCCGCA | AGATGCTGCATATAAGCAGC | EWRALRCAAARCCI*A |
| 458 | C | AGTGGCGAGCCCTCAGATGC | TGCGGCCGCA | GATGCTGCATATAAGCAGCT | WRALRCCGRRCCI*AA |
| 459 | T | GTGGCGAGCCCTCAGATGCT | TGCGGCCGCA | ATGCTGCATATAAGCAGCTG | WRALRCLRPQCCI*AA |
| 460 | G | TGGCGAGCCCTCAGATGCTG | TGCGGCCGCA | TGCTGCATATAAGCAGCTGC | WRALRCCAAACCI*AA |
| 461 | C | GGCGAGCCCTCAGATGCTGC | TGCGGCCGCA | GCTGCATATAAGCAGCTGCT | RALRCCCGRSCI*AAA |
| 462 | A | GCGAGCCCTCAGATGCTGCA | TGCGGCCGCA | CTGCATATAAGCAGCTGCTT | RALRCCMRPHCI*AAA |
| 463 | T | CGAGCCCTCAGATGCTGCAT | TGCGGCCGCA | TGCATATAAGCAGCTGCTTT | RALRCCIAAACI*AAA |
| 464 | A | GAGCCCTCAGATGCTGCATA | TGCGGCCGCA | GCATATAAGCAGCTGCTTTT | ALRCCICGRSI*AAAF |
| 465 | T | AGCCCTCAGATGCTGCATAT | TGCGGCCGCA | CATATAAGCAGCTGCTTTTT | ALRCCILRPHI*AAAF |
| 466 | A | GCCCTCAGATGCTGCATATA | TGCGGCCGCA | ATATAAGCAGCTGCTTTTTG | ALRCCIYAAAI*AAAF |
| 467 | A | CCCTCAGATGCTGCATATAA | TGCGGCCGCA | TATAAGCAGCTGCTTTTTGC | LRCCI*CGRI*AAAFC |
| 468 | G | CCTCAGATGCTGCATATAAG | TGCGGCCGCA | ATAAGCAGCTGCTTTTTGCC | LRCCI*VRPQ*AAAFC |
| 469 | C | CTCAGATGCTGCATATAAGC | TGCGGCCGCA | TAAGCAGCTGCTTTTTGCCT | LRCCI*AAAA*AAAFC |
| 470 | A | TCAGATGCTGCATATAAGCA | TGCGGCCGCA | AAGCAGCTGCTTTTTGCCTG | RCCI*ACGRKAAAFCL |
| 471 | G | CAGATGCTGCATATAAGCAG | TGCGGCCGCA | AGCAGCTGCTTTTTGCCTGT | RCCI*AVRPQAAAFCL |
| 472 | C | AGATGCTGCATATAAGCAGC | TGCGGCCGCA | GCAGCTGCTTTTTGCCTGTA | RCCI*AAAAAAAAFCL |
| 473 | T | GATGCTGCATATAAGCAGCT | TGCGGCCGCA | CAGCTGCTTTTTGCCTGTAC | CCI*AACGRTAAFCLY |
| 474 | G | ATGCTGCATATAAGCAGCTG | TGCGGCCGCA | AGCTGCTTTTTGCCTGTACT | CCI*AAVRPQAAFCLY |
| 475 | C | TGCTGCATATAAGCAGCTGC | TGCGGCCGCA | GCTGCTTTTTGCCTGTACTG | CCI*AAAAAAAAFCLY |
| 476 | T | GCTGCATATAAGCAGCTGCT | TGCGGCCGCA | CTGCTTTTTGCCTGTACTGG | CI*AAACGRTAFCLYW |
| 477 | T | CTGCATATAAGCAGCTGCTT | TGCGGCCGCA | TGCTTTTTGCCTGTACTGGG | CI*AAALRPHAFCLYW |
| 478 | T | TGCATATAAGCAGCTGCTTT | TGCGGCCGCA | GCTTTTTGCCTGTACTGGGT | CI*AAAFAAAAFCLYW |
| 479 | T | GCATATAAGCAGCTGCTTTT | TGCGGCCGCA | CTTTTTGCCTGTACTGGGTC | I*AAAFCGRTFCLYWV |
| 480 | T | CATATAAGCAGCTGCTTTTT | TGCGGCCGCA | TTTTTGCCTGTACTGGGTCT | I*AAAFLRPHFCLYWV |
| 481 | G | ATATAAGCAGCTGCTTTTTG | TGCGGCCGCA | TTTTGCCTGTACTGGGTCTC | I*AAAFCAAAFCLYWV |
| 482 | C | TATAAGCAGCTGCTTTTTGC | TGCGGCCGCA | TTTGCCTGTACTGGGTCTCT | *AAAFCCGRICLYWVS |
| 483 | C | ATAAGCAGCTGCTTTTTGCC | TGCGGCCGCA | TTGCCTGTACTGGGTCTCTC | *AAAFCLRPHCLYWVS |
| 484 | T | TAAGCAGCTGCTTTTTGCCT | TGCGGCCGCA | TGCCTGTACTGGGTCTCTCT | *AAAFCLAAACLYWVS |
| 485 | G | AAGCAGCTGCTTTTTGCCTG | TGCGGCCGCA | GCCTGTACTGGGTCTCTCTG | AAAFCLCGRSLYWVSL |
| 486 | T | AGCAGCTGCTTTTTGCCTGT | TGCGGCCGCA | CCTGTACTGGGTCTCTCTGG | AAAFCLLRPHLYWVSL |
| 487 | A | GCAGCTGCTTTTTGCCTGTA | TGCGGCCGCA | CTGTACTGGGTCTCTCTGGT | AAAFCLYAAALYWVSL |
| 488 | C | CAGCTGCTTTTTGCCTGTAC | TGCGGCCGCA | TGTACTGGGTCTCTCTGGTT | AAFCLYCGRMYWVSLV |
| 489 | T | AGCTGCTTTTTGCCTGTACT | TGCGGCCGCA | GTACTGGGTCTCTCTGGTTA | AAFCLYLRPQYWVSLV |
| 490 | G | GCTGCTTTTTGCCTGTACTG | TGCGGCCGCA | TACTGGGTCTCTCTGGTTAG | AAFCLYCAAAYWVSLV |
| 491 | G | CTGCTTTTTGCCTGTACTGG | TGCGGCCGCA | ACTGGGTCTCTCTGGTTAGA | AFCLYWCGRNWVSLVR |
| 492 | G | TGCTTTTTGCCTGTACTGGG | TGCGGCCGCA | CTGGGTCTCTCTGGTTAGAC | AFCLYWVRPHWVSLVR |
| 493 | T | GCTTTTTGCCTGTACTGGGT | TGCGGCCGCA | TGGGTCTCTCTGGTTAGACC | AFCLYWVAAAWVSLVR |
| 494 | C | CTTTTTGCCTGTACTGGGTC | TGCGGCCGCA | GGGTCTCTCTGGTTAGACCA | FCLYWVCGRRVSLVRP |
| 495 | T | TTTTTGCCTGTACTGGGTCT | TGCGGCCGCA | GGTCTCTCTGGTTAGACCAG | FCLYWVLRPQVSLVRP |
| 496 | C | TTTTGCCTGTACTGGGTCTC | TGCGGCCGCA | GTCTCTCTGGTTAGACCAGA | FCLYWVSAAAVSLVRP |
| 497 | T | TTTGCCTGTACTGGGTCTCT | TGCGGCCGCA | TCTCTCTGGTTAGACCAGAT | CLYWVSCGRISLVRPD |
| 498 | C | TTGCCTGTACTGGGTCTCTC | TGCGGCCGCA | CTCTCTGGTTAGACCAGATC | CLYWVSLRPHSLVRPD |
| 499 | T | TGCCTGTACTGGGTCTCTCT | TGCGGCCGCA | TCTCTGGTTAGACCAGATCT | CLYWVSLAAASLVRPD |
| 500 | G | GCCTGTACTGGGTCTCTCTG | TGCGGCCGCA | CTCTGGTTAGACCAGATCTG | LYWVSLCGRTLVRPDL |
| 501 | G | CCTGTACTGGGTCTCTCTGG | TGCGGCCGCA | TCTGGTTAGACCAGATCTGA | LYWVSLVRPHLVRPDL |
| 502 | T | CTGTACTGGGTCTCTCTGGT | TGCGGCCGCA | CTGGTTAGACCAGATCTGAG | LYWVSLVAAALVRPDL |
| 503 | T | TGTACTGGGTCTCTCTGGTT | TGCGGCCGCA | TGGTTAGACCAGATCTGAGC | YWVSLVCGRMVRPDLS |
| 504 | A | GTACTGGGTCTCTCTGGTTA | TGCGGCCGCA | GGTTAGACCAGATCTGAGCC | YWVSLVMRPQVRPDLS |
| 505 | G | TACTGGGTCTCTCTGGTTAG | TGCGGCCGCA | GTTAGACCAGATCTGAGCCT | YWVSLVSAAAVRPDLS |
| 506 | A | ACTGGGTCTCTCTGGTTAGA | TGCGGCCGCA | TTAGACCAGATCTGAGCCTG | WVSLVRCGRIRPDLSL |
| 507 | C | CTGGGTCTCTCTGGTTAGAC | TGCGGCCGCA | TAGACCAGATCTGAGCCTGG | WVSLVRLRPHRPDLSL |
| 508 | C | TGGGTCTCTCTGGTTAGACC | TGCGGCCGCA | AGACCAGATCTGAGCCTGGG | WVSLVRPAAARPDLSL |
| 509 | A | GGGTCTCTCTGGTTAGACCA | TGCGGCCGCA | GACCAGATCTGAGCCTGGGA | VSLVRPCGRRPDLSLG |
| 510 | G | GGTCTCTCTGGTTAGACCAG | TGCGGCCGCA | ACCAGATCTGAGCCTGGGAG | VSLVRPVRPQPDLSLG |

| | | | | | |
|---|---|---|---|---|---|
| 511 | A | GTCTCTCTGGTTAGACCAGA | TGCGGCCGCA | CCAGATCTGAGCCTGGGAGC | VSLVRPDAAAPDLSLG |
| 512 | T | TCTCTCTGGTTAGACCAGAT | TGCGGCCGCA | CAGATCTGAGCCTGGGAGCT | SLVRPDCGRTDLSLGA |
| 513 | C | CTCTCTGGTTAGACCAGATC | TGCGGCCGCA | AGATCTGAGCCTGGGAGCTC | SLVRPDLRPQDLSLGA |
| 514 | T | TCTCTGGTTAGACCAGATCT | TGCGGCCGCA | GATCTGAGCCTGGGAGCTCT | SLVRPDLAAADLSLGA |
| 515 | G | CTCTGGTTAGACCAGATCTG | TGCGGCCGCA | ATCTGAGCCTGGGAGCTCTC | LVRPDLCGRNLSLGAL |
| 516 | A | TCTGGTTAGACCAGATCTGA | TGCGGCCGCA | TCTGAGCCTGGGAGCTCTCT | LVRPDLMRPHLSLGAL |
| 517 | G | CTGGTTAGACCAGATCTGAG | TGCGGCCGCA | CTGAGCCTGGGAGCTCTCTG | LVRPDLSAAALSLGAL |
| 518 | C | TGGTTAGACCAGATCTGAGC | TGCGGCCGCA | TGAGCCTGGGAGCTCTCTGG | VRPDLSCGRMSLGALW |
| 519 | C | GGTTAGACCAGATCTGAGCC | TGCGGCCGCA | GAGCCTGGGAGCTCTCTGGC | VRPDLSLRPQSLGALW |
| 520 | T | GTTAGACCAGATCTGAGCCT | TGCGGCCGCA | AGCCTGGGAGCTCTCTGGCT | VRPDLSLAAASLGALW |
| 521 | G | TTAGACCAGATCTGAGCCTG | TGCGGCCGCA | GCCTGGGAGCTCTCTGGCTA | RPDLSLCGRSLGALWL |
| 522 | G | TAGACCAGATCTGAGCCTGG | TGCGGCCGCA | CCTGGGAGCTCTCTGGCTAA | RPDLSLVRPHLGALWL |
| 523 | G | AGACCAGATCTGAGCCTGGG | TGCGGCCGCA | CTGGGAGCTCTCTGGCTAAC | RPDLSLGAAALGALWL |
| 524 | A | GACCAGATCTGAGCCTGGGA | TGCGGCCGCA | TGGGAGCTCTCTGGCTAACT | PDLSLGCGRMGALWLT |
| 525 | G | ACCAGATCTGAGCCTGGGAG | TGCGGCCGCA | GGGAGCTCTCTGGCTAACTA | PDLSLGVRPQGALWLT |
| 526 | C | CCAGATCTGAGCCTGGGAGC | TGCGGCCGCA | GGAGCTCTCTGGCTAACTAG | PDLSLGAAAAGALWLT |
| 527 | T | CAGATCTGAGCCTGGGAGCT | TGCGGCCGCA | GAGCTCTCTGGCTAACTAGG | DLSLGACGRRALWLTR |
| 528 | C | AGATCTGAGCCTGGGAGCTC | TGCGGCCGCA | AGCTCTCTGGCTAACTAGGG | DLSLGALRPQALWLTR |
| 529 | T | GATCTGAGCCTGGGAGCTCT | TGCGGCCGCA | GCTCTCTGGCTAACTAGGGA | DLSLGALAAAALWLTR |
| 530 | C | ATCTGAGCCTGGGAGCTCTC | TGCGGCCGCA | CTCTCTGGCTAACTAGGGAA | LSLGALCGRTLWLTRE |
| 531 | T | TCTGAGCCTGGGAGCTCTCT | TGCGGCCGCA | TCTCTGGCTAACTAGGGAAC | LSLGALLRPHLWLTRE |
| 532 | G | CTGAGCCTGGGAGCTCTCTG | TGCGGCCGCA | CTCTGGCTAACTAGGGAACC | LSLGALCAAALWLTRE |
| 533 | G | TGAGCCTGGGAGCTCTCTGG | TGCGGCCGCA | TCTGGCTAACTAGGGAACCC | SLGALWCGRIWLTREP |
| 534 | C | GAGCCTGGGAGCTCTCTGGC | TGCGGCCGCA | CTGGCTAACTAGGGAACCCA | SLGALWLRPHWLTREP |
| 535 | T | AGCCTGGGAGCTCTCTGGCT | TGCGGCCGCA | TGGCTAACTAGGGAACCCAC | SLGALWLAAAWLTREP |
| 536 | A | GCCTGGGAGCTCTCTGGCTA | TGCGGCCGCA | GGCTAACTAGGGAACCCACT | LGALWLCGRRLTREPT |
| 537 | A | CCTGGGAGCTCTCTGGCTAA | TGCGGCCGCA | GCTAACTAGGGAACCCACTG | LGALWLMRPQLTREPT |
| 538 | C | CTGGGAGCTCTCTGGCTAAC | TGCGGCCGCA | CTAACTAGGGAACCCACTGC | LGALWLTAAALTREPT |
| 539 | T | TGGGAGCTCTCTGGCTAACT | TGCGGCCGCA | TAACTAGGGAACCCACTGCT | GALWLTCGRITREPTA |
| 540 | A | GGGAGCTCTCTGGCTAACTA | TGCGGCCGCA | AACTAGGGAACCCACTGCTT | GALWLTMRPQTREPTA |
| 541 | G | GGAGCTCTCTGGCTAACTAG | TGCGGCCGCA | ACTAGGGAACCCACTGCTTA | GALWLTSAAATREPTA |
| 542 | G | GAGCTCTCTGGCTAACTAGG | TGCGGCCGCA | CTAGGGAACCCACTGCTTAA | ALWLTRCGRTREPTA* |
| 543 | G | AGCTCTCTGGCTAACTAGGG | TGCGGCCGCA | TAGGGAACCCACTGCTTAAG | ALWLTRVRPHREPTA* |
| 544 | A | GCTCTCTGGCTAACTAGGGA | TGCGGCCGCA | AGGGAACCCACTGCTTAAGC | ALWLTRDAAAREPTA* |
| 545 | A | CTCTCTGGCTAACTAGGGAA | TGCGGCCGCA | GGGAACCCACTGCTTAAGCC | LWLTRECGRREPTA*A |
| 546 | C | TCTCTGGCTAACTAGGGAAC | TGCGGCCGCA | GGAACCCACTGCTTAAGCCT | LWLTRELRPQEPTA*A |
| 547 | C | CTCTGGCTAACTAGGGAACC | TGCGGCCGCA | GAACCCACTGCTTAAGCCTC | LWLTREPAAAEPTA*A |
| 548 | C | TCTGGCTAACTAGGGAACCC | TGCGGCCGCA | AACCCACTGCTTAAGCCTCA | WLTREPCGRKPTA*AS |
| 549 | A | CTGGCTAACTAGGGAACCCA | TGCGGCCGCA | ACCCACTGCTTAAGCCTCAA | WLTREPMRPQPTA*AS |
| 550 | C | TGGCTAACTAGGGAACCCAC | TGCGGCCGCA | CCCACTGCTTAAGCCTCAAT | WLTREPTAAAPTA*AS |
| 551 | T | GGCTAACTAGGGAACCCACT | TGCGGCCGCA | CCACTGCTTAAGCCTCAATA | LTREPTCGRTTA*ASI |
| 552 | G | GCTAACTAGGGAACCCACTG | TGCGGCCGCA | CACTGCTTAAGCCTCAATAA | LTREPTVRPHTA*ASI |
| 553 | C | CTAACTAGGGAACCCACTGC | TGCGGCCGCA | ACTGCTTAAGCCTCAATAAA | LTREPTAAAATA*ASI |
| 554 | T | TAACTAGGGAACCCACTGCT | TGCGGCCGCA | CTGCTTAAGCCTCAATAAAG | TREPTACGRTA*ASIK |
| 555 | T | AACTAGGGAACCCACTGCTT | TGCGGCCGCA | TGCTTAAGCCTCAATAAAGC | TREPTALRPHA*ASIK |
| 556 | A | ACTAGGGAACCCACTGCTTA | TGCGGCCGCA | GCTTAAGCCTCAATAAAGCT | TREPTAYAAAA*ASIK |
| 557 | A | CTAGGGAACCCACTGCTTAA | TGCGGCCGCA | CTTAAGCCTCAATAAAGCTT | REPTA*CGRT*ASIKL |
| 558 | G | TAGGGAACCCACTGCTTAAG | TGCGGCCGCA | TTAAGCCTCAATAAAGCTTG | REPTA*VRPH*ASIKL |
| 559 | C | AGGGAACCCACTGCTTAAGC | TGCGGCCGCA | TAAGCCTCAATAAAGCTTGC | REPTA*AAAA*ASIKL |
| 560 | C | GGGAACCCACTGCTTAAGCC | TGCGGCCGCA | AAGCCTCAATAAAGCTTGCC | EPTA*ACGRKASIKLA |
| 561 | T | GGAACCCACTGCTTAAGCCT | TGCGGCCGCA | AGCCTCAATAAAGCTTGCCT | EPTA*ALRPQASIKLA |
| 562 | C | GAACCCACTGCTTAAGCCTC | TGCGGCCGCA | GCCTCAATAAAGCTTGCCTT | EPTA*ASAAAASIKLA |
| 563 | A | AACCCACTGCTTAAGCCTCA | TGCGGCCGCA | CCTCAATAAAGCTTGCCTTG | PTA*ASCGRTSIKLAL |
| 564 | A | ACCCACTGCTTAAGCCTCAA | TGCGGCCGCA | CTCAATAAAGCTTGCCTTGA | PTA*ASMRPHSIKLAL |
| 565 | T | CCCACTGCTTAAGCCTCAAT | TGCGGCCGCA | TCAATAAAGCTTGCCTTGAG | PTA*ASIAAASIKLAL |
| 566 | A | CCACTGCTTAAGCCTCAATA | TGCGGCCGCA | CAATAAAGCTTGCCTTGAGG | TA*ASICGRTIKLALR |
| 567 | A | CACTGCTTAAGCCTCAATAA | TGCGGCCGCA | AATAAAGCTTGCCTTGAGGG | TA*ASIMRPQIKLALR |
| 568 | A | ACTGCTTAAGCCTCAATAAA | TGCGGCCGCA | ATAAAGCTTGCCTTGAGGGA | TA*ASINAAAIKLALR |
| 569 | G | CTGCTTAAGCCTCAATAAAG | TGCGGCCGCA | TAAAGCTTGCCTTGAGGGAG | A*ASIKCGRIKLALRE |
| 570 | C | TGCTTAAGCCTCAATAAAGC | TGCGGCCGCA | AAAGCTTGCCTTGAGGGAGT | A*ASIKLRPQKLALRE |
| 571 | T | GCTTAAGCCTCAATAAAGCT | TGCGGCCGCA | AAGCTTGCCTTGAGGGAGTG | A*ASIKLAAAKLALRE |
| 572 | T | CTTAAGCCTCAATAAAGCTT | TGCGGCCGCA | AGCTTGCCTTGAGGGAGTGC | *ASIKLCGRKLALREC |
| 573 | G | TTAAGCCTCAATAAAGCTTG | TGCGGCCGCA | GCTTGCCTTGAGGGAGTGCT | *ASIKLVRPQLALREC |
| 574 | C | TAAGCCTCAATAAAGCTTGC | TGCGGCCGCA | CTTGCCTTGAGGGAGTGCTT | *ASIKLAAAALALREC |
| 575 | C | AAGCCTCAATAAAGCTTGCC | TGCGGCCGCA | TTGCCTTGAGGGAGTGCTTC | ASIKLACGRIALRECF |
| 576 | T | AGCCTCAATAAAGCTTGCCT | TGCGGCCGCA | TGCCTTGAGGGAGTGCTTCA | ASIKLALRPHALRECF |
| 577 | T | GCCTCAATAAAGCTTGCCTT | TGCGGCCGCA | GCCTTGAGGGAGTGCTTCAA | ASIKLAFAAAALRECF |
| 578 | G | CCTCAATAAAGCTTGCCTTG | TGCGGCCGCA | CCTTGAGGGAGTGCTTCAAG | SIKLALCGRTLRECFK |
| 579 | A | CTCAATAAAGCTTGCCTTGA | TGCGGCCGCA | CTTGAGGGAGTGCTTCAAGT | SIKLALMRPHLRECFK |
| 580 | G | TCAATAAAGCTTGCCTTGAG | TGCGGCCGCA | TTGAGGGAGTGCTTCAAGTA | SIKLALSAAALRECFK |
| 581 | G | CAATAAAGCTTGCCTTGAGG | TGCGGCCGCA | TGAGGGAGTGCTTCAAGTAG | IKLALRCGRMRECFK* |
| 582 | G | AATAAAGCTTGCCTTGAGGG | TGCGGCCGCA | GAGGGAGTGCTTCAAGTAGT | IKLALRVRPQRECFK* |
| 583 | A | ATAAAGCTTGCCTTGAGGGA | TGCGGCCGCA | AGGGAGTGCTTCAAGTAGTG | IKLALRDAAARECFK* |
| 584 | G | TAAAGCTTGCCTTGAGGGAG | TGCGGCCGCA | GGGAGTGCTTCAAGTAGTGT | KLALRECGRRECFK*C |
| 585 | T | AAAGCTTGCCTTGAGGGAGT | TGCGGCCGCA | GGAGTGCTTCAAGTAGTGTG | KLALRELRPQECFK*C |
| 586 | G | AAGCTTGCCTTGAGGGAGTG | TGCGGCCGCA | GAGTGCTTCAAGTAGTGTGT | KLALRECAAAECFK*C |
| 587 | C | AGCTTGCCTTGAGGGAGTGC | TGCGGCCGCA | AGTGCTTCAAGTAGTGTGTG | LALRECCGRKCFK*CV |
| 588 | T | GCTTGCCTTGAGGGAGTGCT | TGCGGCCGCA | GTGCTTCAAGTAGTGTGTGC | LALRECLRPQCFK*CV |
| 589 | T | CTTGCCTTGAGGGAGTGCTT | TGCGGCCGCA | TGCTTCAAGTAGTGTGTGCC | LALRECFAAACFK*CV |
| 590 | C | TTGCCTTGAGGGAGTGCTTC | TGCGGCCGCA | GCTTCAAGTAGTGTGTGCCC | ALRECFCGRSFK*CVP |
| 591 | A | TGCCTTGAGGGAGTGCTTCA | TGCGGCCGCA | CTTCAAGTAGTGTGTGCCCG | ALRECFMRPHFK*CVP |
| 592 | A | GCCTTGAGGGAGTGCTTCAA | TGCGGCCGCA | TTCAAGTAGTGTGTGCCCGT | ALRECFNAAAFK*CVP |
| 593 | G | CCTTGAGGGAGTGCTTCAAG | TGCGGCCGCA | TCAAGTAGTGTGTGCCCGTC | LRECFKCGRIK*CVPV |
| 594 | T | CTTGAGGGAGTGCTTCAAGT | TGCGGCCGCA | CAAGTAGTGTGTGCCCGTCT | LRECFKLRPHK*CVPV |
| 595 | A | TTGAGGGAGTGCTTCAAGTA | TGCGGCCGCA | AAGTAGTGTGTGCCCGTCTG | LRECFKYAAAK*CVPV |
| 596 | G | TGAGGGAGTGCTTCAAGTAG | TGCGGCCGCA | AGTAGTGTGTGCCCGTCTGT | RECFK*CGRK*CVPVC |

| | | | | | |
|---|---|---|---|---|---|
| 597 | T | GAGGGAGTGCTTCAAGTAGT | TGCGGCCGCA | GTAGTGTGTGCCCGTCTGTT | RECFK*LRPQ*CVPVC |
| 598 | G | AGGGAGTGCTTCAAGTAGTG | TGCGGCCGCA | TAGTGTGTGCCCGTCTGTTG | RECFK*CAAA*CVPVC |
| 599 | T | GGGAGTGCTTCAAGTAGTGT | TGCGGCCGCA | AGTGTGTGCCCGTCTGTTGT | ECFK*CCGRKCVPVCC |
| 600 | G | GGAGTGCTTCAAGTAGTGTG | TGCGGCCGCA | GTGTGTGCCCGTCTGTTGTG | ECFK*CVRPQCVPVCC |
| 601 | T | GAGTGCTTCAAGTAGTGTGT | TGCGGCCGCA | TGTGTGCCCGTCTGTTGTGA | ECFK*CVAAACVPVCC |
| 602 | G | AGTGCTTCAAGTAGTGTGTG | TGCGGCCGCA | GTGTGCCCGTCTGTTGTGAC | CFK*CVCGRSVPVCCD |
| 603 | C | GTGCTTCAAGTAGTGTGTGC | TGCGGCCGCA | TGTGCCCGTCTGTTGTGACT | CFK*CVLRPHVPVCCD |
| 604 | C | TGCTTCAAGTAGTGTGTGCC | TGCGGCCGCA | GTGCCCGTCTGTTGTGACTC | CFK*CVPAAAVPVCCD |
| 605 | C | GCTTCAAGTAGTGTGTGCCC | TGCGGCCGCA | TGCCCGTCTGTTGTGACTCT | FK*CVPCGRMPVCCDS |
| 606 | G | CTTCAAGTAGTGTGTGCCCG | TGCGGCCGCA | GCCCGTCTGTTGTGACTCTG | FK*CVPVRPQPVCCDS |
| 607 | T | TTCAAGTAGTGTGTGCCCGT | TGCGGCCGCA | CCCGTCTGTTGTGACTCTGG | FK*CVPVAAAPVCCDS |
| 608 | C | TCAAGTAGTGTGTGCCCGTC | TGCGGCCGCA | CCGTCTGTTGTGACTCTGGT | K*CVPVCGRTVCCDSG |
| 609 | T | CAAGTAGTGTGTGCCCGTCT | TGCGGCCGCA | CGTCTGTTGTGACTCTGGTA | K*CVPVLRPHVCCDSG |
| 610 | G | AAGTAGTGTGTGCCCGTCTG | TGCGGCCGCA | GTCTGTTGTGACTCTGGTAA | K*CVPVCAAAVCCDSG |
| 611 | T | AGTAGTGTGTGCCCGTCTGT | TGCGGCCGCA | TCTGTTGTGACTCTGGTAAC | *CVPVCCGRICCDSGN |
| 612 | T | GTAGTGTGTGCCCGTCTGTT | TGCGGCCGCA | CTGTTGTGACTCTGGTAACT | *CVPVCLRPHCCDSGN |
| 613 | G | TAGTGTGTGCCCGTCTGTTG | TGCGGCCGCA | TGTTGTGACTCTGGTAACTA | *CVPVCCAAACCDSGN |
| 614 | T | AGTGTGTGCCCGTCTGTTGT | TGCGGCCGCA | GTTGTGACTCTGGTAACTAG | CVPVCCCGRSCDSGN* |
| 615 | G | GTGTGTGCCCGTCTGTTGTG | TGCGGCCGCA | TTGTGACTCTGGTAACTAGA | CVPVCCVRPHCDSGN* |
| 616 | A | TGTGTGCCCGTCTGTTGTGA | TGCGGCCGCA | TGTGACTCTGGTAACTAGAG | CVPVCCDAAACDSGN* |
| 617 | C | GTGTGCCCGTCTGTTGTGAC | TGCGGCCGCA | GTGACTCTGGTAACTAGAGA | VPVCCDCGRSDSGN*R |
| 618 | T | TGTGCCCGTCTGTTGTGACT | TGCGGCCGCA | TGACTCTGGTAACTAGAGAT | VPVCCDLRPHDSGN*R |
| 619 | C | GTGCCCGTCTGTTGTGACTC | TGCGGCCGCA | GACTCTGGTAACTAGAGATC | VPVCCDSAAADSGN*R |
| 620 | T | TGCCCGTCTGTTGTGACTCT | TGCGGCCGCA | ACTCTGGTAACTAGAGATCC | PVCCDSCGRNSGN*RS |
| 621 | G | GCCCGTCTGTTGTGACTCTG | TGCGGCCGCA | CTCTGGTAACTAGAGATCCC | PVCCDSVRPHSGN*RS |
| 622 | G | CCCGTCTGTTGTGACTCTGG | TGCGGCCGCA | TCTGGTAACTAGAGATCCCT | PVCCDSGAAASGN*RS |
| 623 | T | CCGTCTGTTGTGACTCTGGT | TGCGGCCGCA | CTGGTAACTAGAGATCCCTC | VCCDSGCGRTGN*RSL |
| 624 | A | CGTCTGTTGTGACTCTGGTA | TGCGGCCGCA | TGGTAACTAGAGATCCCTCA | VCCDSGMRPHGN*RSL |
| 625 | A | GTCTGTTGTGACTCTGGTAA | TGCGGCCGCA | GGTAACTAGAGATCCCTCAG | VCCDSGNAAAGN*RSL |
| 626 | C | TCTGTTGTGACTCTGGTAAC | TGCGGCCGCA | GTAACTAGAGATCCCTCAGA | CCDSGNCGRSN*RSLR |
| 627 | T | CTGTTGTGACTCTGGTAACT | TGCGGCCGCA | TAACTAGAGATCCCTCAGAC | CCDSGNLRPHN*RSLR |
| 628 | A | TGTTGTGACTCTGGTAACTA | TGCGGCCGCA | AACTAGAGATCCCTCAGACC | CCDSGNYAAAN*RSLR |
| 629 | G | GTTGTGACTCTGGTAACTAG | TGCGGCCGCA | ACTAGAGATCCCTCAGACCC | CDSGN*CGRN*RSLRP |
| 630 | A | TTGTGACTCTGGTAACTAGA | TGCGGCCGCA | CTAGAGATCCCTCAGACCCT | CDSGN*MRPH*RSLRP |
| 631 | G | TGTGACTCTGGTAACTAGAG | TGCGGCCGCA | TAGAGATCCCTCAGACCCTT | CDSGN*SAAA*RSLRP |
| 632 | A | GTGACTCTGGTAACTAGAGA | TGCGGCCGCA | AGAGATCCCTCAGACCCTTT | DSGN*RCGRKRSLRPF |
| 633 | T | TGACTCTGGTAACTAGAGAT | TGCGGCCGCA | GAGATCCCTCAGACCCTTTT | DSGN*RLRPQRSLRPF |
| 634 | C | GACTCTGGTAACTAGAGATC | TGCGGCCGCA | AGATCCCTCAGACCCTTTTA | DSGN*RSAAARSLRPF |
| 635 | C | ACTCTGGTAACTAGAGATCC | TGCGGCCGCA | GATCCCTCAGACCCTTTTAG | SGN*RSCGRRSLRPF* |
| 636 | C | CTCTGGTAACTAGAGATCCC | TGCGGCCGCA | ATCCCTCAGACCCTTTTAGT | SGN*RSLRPQSLRPF* |
| 637 | T | TCTGGTAACTAGAGATCCCT | TGCGGCCGCA | TCCCTCAGACCCTTTTAGTC | SGN*RSLAAASLRPF* |
| 638 | C | CTGGTAACTAGAGATCCCTC | TGCGGCCGCA | CCCTCAGACCCTTTTAGTCA | GN*RSLCGRTLRPF*S |
| 639 | A | TGGTAACTAGAGATCCCTCA | TGCGGCCGCA | CCTCAGACCCTTTTAGTCAG | GN*RSLMRPHLRPF*S |
| 640 | G | GGTAACTAGAGATCCCTCAG | TGCGGCCGCA | CTCAGACCCTTTTAGTCAGT | GN*RSLSAAALRPF*S |
| 641 | A | GTAACTAGAGATCCCTCAGA | TGCGGCCGCA | TCAGACCCTTTTAGTCAGTG | N*RSLRCGRIRPF*SV |
| 642 | C | TAACTAGAGATCCCTCAGAC | TGCGGCCGCA | CAGACCCTTTTAGTCAGTGT | N*RSLRLRPHRPF*SV |
| 643 | C | AACTAGAGATCCCTCAGACC | TGCGGCCGCA | AGACCCTTTTAGTCAGTGTG | N*RSLRPAAARPF*SV |
| 644 | C | ACTAGAGATCCCTCAGACCC | TGCGGCCGCA | GACCCTTTTAGTCAGTGTGG | *RSLRPCGRRPF*SVW |
| 645 | T | CTAGAGATCCCTCAGACCCT | TGCGGCCGCA | ACCCTTTTAGTCAGTGTGGA | *RSLRPLRPQPF*SVW |
| 646 | T | TAGAGATCCCTCAGACCCTT | TGCGGCCGCA | CCCTTTTAGTCAGTGTGGAA | *RSLRPFAAAPF*SVW |
| 647 | T | AGAGATCCCTCAGACCCTTT | TGCGGCCGCA | CCTTTTAGTCAGTGTGGAAA | RSLRPFCGRTF*SVWK |
| 648 | T | GAGATCCCTCAGACCCTTTT | TGCGGCCGCA | CTTTTAGTCAGTGTGGAAAA | RSLRPFLRPHF*SVWK |
| 649 | A | AGATCCCTCAGACCCTTTTA | TGCGGCCGCA | TTTTAGTCAGTGTGGAAAAT | RSLRPFYAAAF*SVWK |
| 650 | G | GATCCCTCAGACCCTTTTAG | TGCGGCCGCA | TTTAGTCAGTGTGGAAAATC | SLRPF*CGRI*SVWKI |
| 651 | T | ATCCCTCAGACCCTTTTAGT | TGCGGCCGCA | TTAGTCAGTGTGGAAAATCT | SLRPF*LRPH*SVWKI |
| 652 | C | TCCCTCAGACCCTTTTAGTC | TGCGGCCGCA | TAGTCAGTGTGGAAAATCTC | SLRPF*SAAA*SVWKI |
| 653 | A | CCCTCAGACCCTTTTAGTCA | TGCGGCCGCA | AGTCAGTGTGGAAAATCTCT | LRPF*SCGRKSVWKIS |
| 654 | G | CCTCAGACCCTTTTAGTCAG | TGCGGCCGCA | GTCAGTGTGGAAAATCTCTA | LRPF*SVRPQSVWKIS |
| 655 | T | CTCAGACCCTTTTAGTCAGT | TGCGGCCGCA | TCAGTGTGGAAAATCTCTAG | LRPF*SVAAASVWKIS |
| 656 | G | TCAGACCCTTTTAGTCAGTG | TGCGGCCGCA | CAGTGTGGAAAATCTCTAGC | RPF*SVCGRTVWKISS |
| 657 | T | CAGACCCTTTTAGTCAGTGT | TGCGGCCGCA | AGTGTGGAAAATCTCTAGCA | RPF*SVLRPQVWKISS |
| 658 | G | AGACCCTTTTAGTCAGTGTG | TGCGGCCGCA | GTGTGGAAAATCTCTAGCAG | RPF*SVCAAAVWKISS |
| 659 | G | GACCCTTTTAGTCAGTGTGG | TGCGGCCGCA | TGTGGAAAATCTCTAGCAGT | PF*SVWCGRMWKISSS |
| 660 | A | ACCCTTTTAGTCAGTGTGGA | TGCGGCCGCA | GTGGAAAATCTCTAGCAGTG | PF*SVWMRPQWKISSS |
| 661 | A | CCCTTTTAGTCAGTGTGGAA | TGCGGCCGCA | TGGAAAATCTCTAGCAGTGG | PF*SVWNAAAWKISSS |
| 662 | A | CCTTTTAGTCAGTGTGGAAA | TGCGGCCGCA | GGAAAATCTCTAGCAGTGGC | F*SVWKCGRRKISSSG |
| 663 | A | CTTTTAGTCAGTGTGGAAAA | TGCGGCCGCA | GAAAATCTCTAGCAGTGGCG | F*SVWKMRPQKISSSG |
| 664 | T | TTTTAGTCAGTGTGGAAAAT | TGCGGCCGCA | AAAATCTCTAGCAGTGGCGC | F*SVWKIAAAKISSSG |
| 665 | C | TTTAGTCAGTGTGGAAAATC | TGCGGCCGCA | AAATCTCTAGCAGTGGCGCC | *SVWKICGRKISSSGA |
| 666 | T | TTAGTCAGTGTGGAAAATCT | TGCGGCCGCA | AATCTCTAGCAGTGGCGCCC | *SVWKILRPQISSSGA |
| 667 | C | TAGTCAGTGTGGAAAATCTC | TGCGGCCGCA | ATCTCTAGCAGTGGCGCCCG | *SVWKISAAAISSSGA |
| 668 | T | AGTCAGTGTGGAAAATCTCT | TGCGGCCGCA | TCTCTAGCAGTGGCGCCCGA | SVWKISCGRISSSGAR |
| 669 | A | GTCAGTGTGGAAAATCTCTA | TGCGGCCGCA | CTCTAGCAGTGGCGCCCGAA | SVWKISMRPHSSSGAR |
| 670 | G | TCAGTGTGGAAAATCTCTAG | TGCGGCCGCA | TCTAGCAGTGGCGCCCGAAC | SVWKISSAAASSSGAR |
| 671 | C | CAGTGTGGAAAATCTCTAGC | TGCGGCCGCA | CTAGCAGTGGCGCCCGAACA | VWKISSCGRTSSGART |
| 672 | A | AGTGTGGAAAATCTCTAGCA | TGCGGCCGCA | TAGCAGTGGCGCCCGAACAG | VWKISSMRPHSSGART |
| 673 | G | GTGTGGAAAATCTCTAGCAG | TGCGGCCGCA | AGCAGTGGCGCCCGAACAGG | VWKISSSAAASSGART |
| 674 | T | TGTGGAAAATCTCTAGCAGT | TGCGGCCGCA | GCAGTGGCGCCCGAACAGGG | WKISSSCGRSSGARTG |
| 675 | G | GTGGAAAATCTCTAGCAGTG | TGCGGCCGCA | CAGTGGCGCCCGAACAGGGA | WKISSSVRPHSGARTG |
| 676 | G | TGGAAAATCTCTAGCAGTGG | TGCGGCCGCA | AGTGGCGCCCGAACAGGGAC | WKISSSGAAASGARTG |
| 677 | C | GGAAAATCTCTAGCAGTGGC | TGCGGCCGCA | GTGGCGCCCGAACAGGGACT | KISSSGCGRSGARTGT |
| 678 | G | GAAAATCTCTAGCAGTGGCG | TGCGGCCGCA | TGGCGCCCGAACAGGGACTT | KISSSGVRPHGARTGT |
| 679 | C | AAAATCTCTAGCAGTGGCGC | TGCGGCCGCA | GGCGCCCGAACAGGGACTTG | KISSSGAAAAGARTGT |
| 680 | C | AAATCTCTAGCAGTGGCGCC | TGCGGCCGCA | GCGCCCGAACAGGGACTTGA | ISSSGACGRSARTGT* |
| 681 | C | AATCTCTAGCAGTGGCGCCC | TGCGGCCGCA | CGCCCGAACAGGGACTTGAA | ISSSGALRPHARTGT* |
| 682 | G | ATCTCTAGCAGTGGCGCCCG | TGCGGCCGCA | GCCCGAACAGGGACTTGAAA | ISSSGARAAAARTGT* |

| | | | | | |
|---|---|---|---|---|---|
| 683 | A | TCTCTAGCAGTGGCGCCCGA | TGCGGCCGCA | CCCGAACAGGGACTTGAAAG | SSSGARCGRTRTGT*K |
| 684 | A | CTCTAGCAGTGGCGCCCGAA | TGCGGCCGCA | CCGAACAGGGACTTGAAAGC | SSSGARMRPHRTGT*K |
| 685 | C | TCTAGCAGTGGCGCCCGAAC | TGCGGCCGCA | CGAACAGGGACTTGAAAGCG | SSSGARTAAARTGT*K |
| 686 | A | CTAGCAGTGGCGCCCGAACA | TGCGGCCGCA | GAACAGGGACTTGAAAGCGA | SSGARTCGRRTGT*KR |
| 687 | G | TAGCAGTGGCGCCCGAACAG | TGCGGCCGCA | AACAGGGACTTGAAAGCGAA | SSGARTVRPQTGT*KR |
| 688 | G | AGCAGTGGCGCCCGAACAGG | TGCGGCCGCA | ACAGGGACTTGAAAGCGAAA | SSGARTGAAATGT*KR |
| 689 | G | GCAGTGGCGCCCGAACAGGG | TGCGGCCGCA | CAGGGACTTGAAAGCGAAAG | SGARTGCGRTGT*KRK |
| 690 | A | CAGTGGCGCCCGAACAGGGA | TGCGGCCGCA | AGGGACTTGAAAGCGAAAGT | SGARTGMRPQGT*KRK |
| 691 | C | AGTGGCGCCCGAACAGGGAC | TGCGGCCGCA | GGGACTTGAAAGCGAAAGTA | SGARTGTAAAGT*KRK |
| 692 | T | GTGGCGCCCGAACAGGGACT | TGCGGCCGCA | GGACTTGAAAGCGAAAGTAA | GARTGTCGRRT*KRK* |
| 693 | T | TGGCGCCCGAACAGGGACTT | TGCGGCCGCA | GACTTGAAAGCGAAAGTAAA | GARTGTLRPQT*KRK* |
| 694 | G | GGCGCCCGAACAGGGACTTG | TGCGGCCGCA | ACTTGAAAGCGAAAGTAAAG | GARTGTCAAAT*KRK* |
| 695 | A | GCGCCCGAACAGGGACTTGA | TGCGGCCGCA | CTTGAAAGCGAAAGTAAAGC | ARTGT*CGRT*KRK*S |
| 696 | A | CGCCCGAACAGGGACTTGAA | TGCGGCCGCA | TTGAAAGCGAAAGTAAAGCC | ARTGT*MRPH*KRK*S |
| 697 | A | GCCCGAACAGGGACTTGAAA | TGCGGCCGCA | TGAAAGCGAAAGTAAAGCCA | ARTGT*NAAA*KRK*S |
| 698 | G | CCCGAACAGGGACTTGAAAG | TGCGGCCGCA | GAAAGCGAAAGTAAAGCCAG | RTGT*KCGRRKRK*SQ |
| 699 | C | CCGAACAGGGACTTGAAAGC | TGCGGCCGCA | AAAGCGAAAGTAAAGCCAGA | RTGT*KLRPQKRK*SQ |
| 700 | G | CGAACAGGGACTTGAAAGCG | TGCGGCCGCA | AAGCGAAAGTAAAGCCAGAG | RTGT*KRAAAKRK*SQ |
| 701 | A | GAACAGGGACTTGAAAGCGA | TGCGGCCGCA | AGCGAAAGTAAAGCCAGAGG | TGT*KRCGRKRK*SQR |
| 702 | A | AACAGGGACTTGAAAGCGAA | TGCGGCCGCA | GCGAAAGTAAAGCCAGAGGA | TGT*KRMRPQRK*SQR |
| 703 | A | ACAGGGACTTGAAAGCGAAA | TGCGGCCGCA | CGAAAGTAAAGCCAGAGGAG | TGT*KRNAAARK*SQR |
| 704 | G | CAGGGACTTGAAAGCGAAAG | TGCGGCCGCA | GAAAGTAAAGCCAGAGGAGA | GT*KRKCGRRK*SQRR |
| 705 | T | AGGGACTTGAAAGCGAAAGT | TGCGGCCGCA | AAAGTAAAGCCAGAGGAGAT | GT*KRKLRPQK*SQRR |
| 706 | A | GGGACTTGAAAGCGAAAGTA | TGCGGCCGCA | AAGTAAAGCCAGAGGAGATC | GT*KRKYAAAK*SQRR |
| 707 | A | GGACTTGAAAGCGAAAGTAA | TGCGGCCGCA | AGTAAAGCCAGAGGAGATCT | T*KRK*CGRK*SQRRS |
| 708 | A | GACTTGAAAGCGAAAGTAAA | TGCGGCCGCA | GTAAAGCCAGAGGAGATCTC | T*KRK*MRPQ*SQRRS |
| 709 | G | ACTTGAAAGCGAAAGTAAAG | TGCGGCCGCA | TAAAGCCAGAGGAGATCTCT | T*KRK*SAAA*SQRRS |
| 710 | C | CTTGAAAGCGAAAGTAAAGC | TGCGGCCGCA | AAAGCCAGAGGAGATCTCTC | *KRK*SCGRKSQRRSL |
| 711 | C | TTGAAAGCGAAAGTAAAGCC | TGCGGCCGCA | AAGCCAGAGGAGATCTCTCG | *KRK*SLRPQSQRRSL |
| 712 | A | TGAAAGCGAAAGTAAAGCCA | TGCGGCCGCA | AGCCAGAGGAGATCTCTCGA | *KRK*SHAAASQRRSL |
| 713 | G | GAAAGCGAAAGTAAAGCCAG | TGCGGCCGCA | GCCAGAGGAGATCTCTCGAC | KRK*SQCGRSQRRSLD |
| 714 | A | AAAGCGAAAGTAAAGCCAGA | TGCGGCCGCA | CCAGAGGAGATCTCTCGACG | KRK*SQMRPHQRRSLD |
| 715 | G | AAGCGAAAGTAAAGCCAGAG | TGCGGCCGCA | CAGAGGAGATCTCTCGACGC | KRK*SQSAAAQRRSLD |
| 716 | G | AGCGAAAGTAAAGCCAGAGG | TGCGGCCGCA | AGAGGAGATCTCTCGACGCA | RK*SQRCGRKRRSLDA |
| 717 | A | GCGAAAGTAAAGCCAGAGGA | TGCGGCCGCA | GAGGAGATCTCTCGACGCAG | RK*SQRMRPQRRSLDA |
| 718 | G | CGAAAGTAAAGCCAGAGGAG | TGCGGCCGCA | AGGAGATCTCTCGACGCAGG | RK*SQRSAAARRSLDA |
| 719 | A | GAAAGTAAAGCCAGAGGAGA | TGCGGCCGCA | GGAGATCTCTCGACGCAGGA | K*SQRRCGRRRSLDAG |
| 720 | T | AAAGTAAAGCCAGAGGAGAT | TGCGGCCGCA | GAGATCTCTCGACGCAGGAC | K*SQRRLRPQRSLDAG |
| 721 | C | AAGTAAAGCCAGAGGAGATC | TGCGGCCGCA | AGATCTCTCGACGCAGGACT | K*SQRRSAAARSLDAG |
| 722 | T | AGTAAAGCCAGAGGAGATCT | TGCGGCCGCA | GATCTCTCGACGCAGGACTC | *SQRRSCGRRSLDAGL |
| 723 | C | GTAAAGCCAGAGGAGATCTC | TGCGGCCGCA | ATCTCTCGACGCAGGACTCG | *SQRRSLRPQSLDAGL |
| 724 | T | TAAAGCCAGAGGAGATCTCT | TGCGGCCGCA | TCTCTCGACGCAGGACTCGG | *SQRRSLAAASLDAGL |
| 725 | C | AAAGCCAGAGGAGATCTCTC | TGCGGCCGCA | CTCTCGACGCAGGACTCGGC | SQRRSLCGRTLDAGLG |
| 726 | G | AAGCCAGAGGAGATCTCTCG | TGCGGCCGCA | TCTCGACGCAGGACTCGGCT | SQRRSLVRPHLDAGLG |
| 727 | A | AGCCAGAGGAGATCTCTCGA | TGCGGCCGCA | CTCGACGCAGGACTCGGCTT | SQRRSLDAAALDAGLG |
| 728 | C | GCCAGAGGAGATCTCTCGAC | TGCGGCCGCA | TCGACGCAGGACTCGGCTTG | QRRSLDCGRIDAGLGL |
| 729 | G | CCAGAGGAGATCTCTCGACG | TGCGGCCGCA | CGACGCAGGACTCGGCTTGC | QRRSLDVRPHDAGLGL |
| 730 | C | CAGAGGAGATCTCTCGACGC | TGCGGCCGCA | GACGCAGGACTCGGCTTGCT | QRRSLDAAAADAGLGL |
| 731 | A | AGAGGAGATCTCTCGACGCA | TGCGGCCGCA | ACGCAGGACTCGGCTTGCTG | RRSLDACGRNAGLGLL |
| 732 | G | GAGGAGATCTCTCGACGCAG | TGCGGCCGCA | CGCAGGACTCGGCTTGCTGA | RRSLDAVRPHAGLGLL |
| 733 | G | AGGAGATCTCTCGACGCAGG | TGCGGCCGCA | GCAGGACTCGGCTTGCTGAA | RRSLDAGAAAAGLGLL |
| 734 | A | GGAGATCTCTCGACGCAGGA | TGCGGCCGCA | CAGGACTCGGCTTGCTGAAG | RSLDAGCGRTGLGLLK |
| 735 | C | GAGATCTCTCGACGCAGGAC | TGCGGCCGCA | AGGACTCGGCTTGCTGAAGC | RSLDAGLRPQGLGLLK |
| 736 | T | AGATCTCTCGACGCAGGACT | TGCGGCCGCA | GGACTCGGCTTGCTGAAGCG | RSLDAGLAAAGLGLLK |
| 737 | C | GATCTCTCGACGCAGGACTC | TGCGGCCGCA | GACTCGGCTTGCTGAAGCGC | SLDAGLCGRRLGLLKR |
| 738 | G | ATCTCTCGACGCAGGACTCG | TGCGGCCGCA | ACTCGGCTTGCTGAAGCGCG | SLDAGLVRPQLGLLKR |
| 739 | G | TCTCTCGACGCAGGACTCGG | TGCGGCCGCA | CTCGGCTTGCTGAAGCGCGC | SLDAGLGAAALGLLKR |
| 740 | C | CTCTCGACGCAGGACTCGGC | TGCGGCCGCA | TCGGCTTGCTGAAGCGCGCA | LDAGLGCGRIGLLKRA |
| 741 | T | TCTCGACGCAGGACTCGGCT | TGCGGCCGCA | CGGCTTGCTGAAGCGCGCAC | LDAGLGLRPHGLLKRA |
| 742 | T | CTCGACGCAGGACTCGGCTT | TGCGGCCGCA | GGCTTGCTGAAGCGCGCACG | LDAGLGFAAAGLLKRA |
| 743 | G | TCGACGCAGGACTCGGCTTG | TGCGGCCGCA | GCTTGCTGAAGCGCGCACGG | DAGLGLCGRSLLKRAR |
| 744 | C | CGACGCAGGACTCGGCTTGC | TGCGGCCGCA | CTTGCTGAAGCGCGCACGGC | DAGLGLLRPHLLKRAR |
| 745 | T | GACGCAGGACTCGGCTTGCT | TGCGGCCGCA | TTGCTGAAGCGCGCACGGCA | DAGLGLLAAALLKRAR |
| 746 | G | ACGCAGGACTCGGCTTGCTG | TGCGGCCGCA | TGCTGAAGCGCGCACGGCAA | AGLGLLCGRMLKRARQ |
| 747 | A | CGCAGGACTCGGCTTGCTGA | TGCGGCCGCA | GCTGAAGCGCGCACGGCAAG | AGLGLLMRPQLKRARQ |
| 748 | A | GCAGGACTCGGCTTGCTGAA | TGCGGCCGCA | CTGAAGCGCGCACGGCAAGA | AGLGLLNAAALKRARQ |
| 749 | G | CAGGACTCGGCTTGCTGAAG | TGCGGCCGCA | TGAAGCGCGCACGGCAAGAG | GLGLLKCGRMKRARQE |
| 750 | C | AGGACTCGGCTTGCTGAAGC | TGCGGCCGCA | GAAGCGCGCACGGCAAGAGG | GLGLLKLRPQKRARQE |
| 751 | G | GGACTCGGCTTGCTGAAGCG | TGCGGCCGCA | AAGCGCGCACGGCAAGAGGC | GLGLLKRAAAKRARQE |
| 752 | C | GACTCGGCTTGCTGAAGCGC | TGCGGCCGCA | AGCGCGCACGGCAAGAGGCG | LGLLKRCGRKRARQEA |
| 753 | G | ACTCGGCTTGCTGAAGCGCG | TGCGGCCGCA | GCGCGCACGGCAAGAGGCGA | LGLLKRVRPQRARQEA |
| 754 | C | CTCGGCTTGCTGAAGCGCGC | TGCGGCCGCA | CGCGCACGGCAAGAGGCGAG | LGLLKRAAAARARQEA |
| 755 | A | TCGGCTTGCTGAAGCGCGCA | TGCGGCCGCA | GCGCACGGCAAGAGGCGAGG | GLLKRACGRSARQEAR |
| 756 | C | CGGCTTGCTGAAGCGCGCAC | TGCGGCCGCA | CGCACGGCAAGAGGCGAGGG | GLLKRALRPHARQEAR |
| 757 | G | GGCTTGCTGAAGCGCGCACG | TGCGGCCGCA | GCACGGCAAGAGGCGAGGGG | GLLKRARAAAARQEAR |
| 758 | G | GCTTGCTGAAGCGCGCACGG | TGCGGCCGCA | CACGGCAAGAGGCGAGGGGC | LLKRARCGRTRQEARG |
| 759 | C | CTTGCTGAAGCGCGCACGGC | TGCGGCCGCA | ACGGCAAGAGGCGAGGGGCG | LLKRARLRPQRQEARG |
| 760 | A | TTGCTGAAGCGCGCACGGCA | TGCGGCCGCA | CGGCAAGAGGCGAGGGGCGG | LLKRARHAAARQEARG |
| 761 | A | TGCTGAAGCGCGCACGGCAA | TGCGGCCGCA | GGCAAGAGGCGAGGGGCGGC | LKRARQCGRRQEARGG |
| 762 | G | GCTGAAGCGCGCACGGCAAG | TGCGGCCGCA | GCAAGAGGCGAGGGGCGGCG | LKRARQVRPQQEARGG |
| 763 | A | CTGAAGCGCGCACGGCAAGA | TGCGGCCGCA | CAAGAGGCGAGGGGCGGCGA | LKRARQDAAAQEARGG |
| 764 | G | TGAAGCGCGCACGGCAAGAG | TGCGGCCGCA | AAGAGGCGAGGGGCGGCGAC | KRARQECGRKEARGGD |
| 765 | G | GAAGCGCGCACGGCAAGAGG | TGCGGCCGCA | AGAGGCGAGGGGCGGCGACT | KRARQEVRPQEARGGD |
| 766 | C | AAGCGCGCACGGCAAGAGGC | TGCGGCCGCA | GAGGCGAGGGGCGGCGACTG | KRARQEAAAAEARGGD |
| 767 | G | AGCGCGCACGGCAAGAGGCG | TGCGGCCGCA | AGGCGAGGGGCGGCGACTGG | RARQEACGRKARGGDW |
| 768 | A | GCGCGCACGGCAAGAGGCGA | TGCGGCCGCA | GGCGAGGGGCGGCGACTGGT | RARQEAMRPQARGGDW |

|  |  |  |  |  |  |
|---|---|---|---|---|---|
| 769 | G | CGCGCACGGCAAGAGGCGAG | TGCGGCCGCA | GCGAGGGGCGGCGACTGGTG | RARQEASAAAARGGDW |
| 770 | G | GCGCACGGCAAGAGGCGAGG | TGCGGCCGCA | CGAGGGGCGGCGACTGGTGA | ARQEARCGRTRGGDW* |
| 771 | G | CGCACGGCAAGAGGCGAGGG | TGCGGCCGCA | GAGGGGCGGCGACTGGTGAG | ARQEARVRPQRGGDW* |
| 772 | G | GCACGGCAAGAGGCGAGGGG | TGCGGCCGCA | AGGGGCGGCGACTGGTGAGT | ARQEARGAAARGGDW* |
| 773 | C | CACGGCAAGAGGCGAGGGGC | TGCGGCCGCA | GGGGCGGCGACTGGTGAGTA | RQEARGCGRRGGDW*V |
| 774 | G | ACGGCAAGAGGCGAGGGGCG | TGCGGCCGCA | GGGCGGCGACTGGTGAGTAC | RQEARGVRPQGGDW*V |
| 775 | G | CGGCAAGAGGCGAGGGGCGG | TGCGGCCGCA | GGCGGCGACTGGTGAGTACG | RQEARGGAAAGGDW*V |
| 776 | C | GGCAAGAGGCGAGGGGCGGC | TGCGGCCGCA | GCGGCGACTGGTGAGTACGC | QEARGGCGRSGDW*VR |
| 777 | G | GCAAGAGGCGAGGGGCGGCG | TGCGGCCGCA | CGGCGACTGGTGAGTACGCC | QEARGGVRPHGDW*VR |
| 778 | A | CAAGAGGCGAGGGGCGGCGA | TGCGGCCGCA | GGCGACTGGTGAGTACGCCA | QEARGGDAAAGDW*VR |
| 779 | C | AAGAGGCGAGGGGCGGCGAC | TGCGGCCGCA | GCGACTGGTGAGTACGCCAA | EARGGDCGRSDW*VRQ |
| 780 | T | AGAGGCGAGGGGCGGCGACT | TGCGGCCGCA | CGACTGGTGAGTACGCCAAA | EARGGDLRPHDW*VRQ |
| 781 | G | GAGGCGAGGGGCGGCGACTG | TGCGGCCGCA | GACTGGTGAGTACGCCAAAA | EARGGDCAAADW*VRQ |
| 782 | G | AGGCGAGGGGCGGCGACTGG | TGCGGCCGCA | ACTGGTGAGTACGCCAAAAA | ARGGDWCGRNW*VRQK |
| 783 | T | GGCGAGGGGCGGCGACTGGT | TGCGGCCGCA | CTGGTGAGTACGCCAAAAAT | ARGGDWLRPHW*VRQK |
| 784 | G | GCGAGGGGCGGCGACTGGTG | TGCGGCCGCA | TGGTGAGTACGCCAAAAATT | ARGGDWCAAAW*VRQK |
| 785 | A | CGAGGGGCGGCGACTGGTGA | TGCGGCCGCA | GGTGAGTACGCCAAAAATTT | RGGDW*CGRR*VRQKF |
| 786 | G | GAGGGGCGGCGACTGGTGAG | TGCGGCCGCA | GTGAGTACGCCAAAAATTTT | RGGDW*VRPQ*VRQKF |
| 787 | T | AGGGGCGGCGACTGGTGAGT | TGCGGCCGCA | TGAGTACGCCAAAAATTTTG | RGGDW*VAAA*VRQKF |
| 788 | A | GGGGCGGCGACTGGTGAGTA | TGCGGCCGCA | GAGTACGCCAAAAATTTTGA | GGDW*VCGRRVRQKF* |
| 789 | C | GGGCGGCGACTGGTGAGTAC | TGCGGCCGCA | AGTACGCCAAAAATTTTGAC | GGDW*VLRPQVRQKF* |
| 790 | G | GGCGGCGACTGGTGAGTACG | TGCGGCCGCA | GTACGCCAAAAATTTTGACT | GGDW*VRAAAVRQKF* |
| 791 | C | GCGGCGACTGGTGAGTACGC | TGCGGCCGCA | TACGCCAAAAATTTTGACTA | GDW*VRCGRIRQKF*L |
| 792 | C | CGGCGACTGGTGAGTACGCC | TGCGGCCGCA | ACGCCAAAAATTTTGACTAG | GDW*VRLRPQRQKF*L |
| 793 | A | GGCGACTGGTGAGTACGCCA | TGCGGCCGCA | CGCCAAAAATTTTGACTAGC | GDW*VRHAAARQKF*L |
| 794 | A | GCGACTGGTGAGTACGCCAA | TGCGGCCGCA | GCCAAAAATTTTGACTAGCG | DW*VRQCGRSQKF*LA |
| 795 | A | CGACTGGTGAGTACGCCAAA | TGCGGCCGCA | CCAAAAATTTTGACTAGCGG | DW*VRQMRPHQKF*LA |
| 796 | A | GACTGGTGAGTACGCCAAAA | TGCGGCCGCA | CAAAAATTTTGACTAGCGGA | DW*VRQNAAAQKF*LA |
| 797 | A | ACTGGTGAGTACGCCAAAAA | TGCGGCCGCA | AAAAATTTTGACTAGCGGAG | W*VRQKCGRKKF*LAE |
| 798 | T | CTGGTGAGTACGCCAAAAAT | TGCGGCCGCA | AAAATTTTGACTAGCGGAGG | W*VRQKLRPQKF*LAE |
| 799 | T | TGGTGAGTACGCCAAAAATT | TGCGGCCGCA | AAATTTTGACTAGCGGAGGC | W*VRQKFAAAKF*LAE |
| 800 | T | GGTGAGTACGCCAAAAATTT | TGCGGCCGCA | AATTTTGACTAGCGGAGGCT | *VRQKFCGRKF*LAEA |
| 801 | T | GTGAGTACGCCAAAAATTTT | TGCGGCCGCA | ATTTTGACTAGCGGAGGCTA | *VRQKFLRPQF*LAEA |
| 802 | G | TGAGTACGCCAAAAATTTTG | TGCGGCCGCA | TTTTGACTAGCGGAGGCTAG | *VRQKFCAAAF*LAEA |
| 803 | A | GAGTACGCCAAAAATTTTGA | TGCGGCCGCA | TTTGACTAGCGGAGGCTAGA | VRQKF*CGRI*LAEAR |
| 804 | C | AGTACGCCAAAAATTTTGAC | TGCGGCCGCA | TTGACTAGCGGAGGCTAGAA | VRQKF*LRPH*LAEAR |
| 805 | T | GTACGCCAAAAATTTTGACT | TGCGGCCGCA | TGACTAGCGGAGGCTAGAAG | VRQKF*LAAA*LAEAR |
| 806 | A | TACGCCAAAAATTTTGACTA | TGCGGCCGCA | GACTAGCGGAGGCTAGAAGG | RQKF*LCGRRLAEARR |
| 807 | G | ACGCCAAAAATTTTGACTAG | TGCGGCCGCA | ACTAGCGGAGGCTAGAAGGA | RQKF*LVRPQLAEARR |
| 808 | C | CGCCAAAAATTTTGACTAGC | TGCGGCCGCA | CTAGCGGAGGCTAGAAGGAG | RQKF*LAAAALAEARR |
| 809 | G | GCCAAAAATTTTGACTAGCG | TGCGGCCGCA | TAGCGGAGGCTAGAAGGAGA | QKF*LACGRIAEARRR |
| 810 | G | CCAAAAATTTTGACTAGCGG | TGCGGCCGCA | AGCGGAGGCTAGAAGGAGAG | QKF*LAVRPQAEARRR |
| 811 | A | CAAAAATTTTGACTAGCGGA | TGCGGCCGCA | GCGGAGGCTAGAAGGAGAGA | QKF*LADAAAAEARRR |
| 812 | G | AAAAATTTTGACTAGCGGAG | TGCGGCCGCA | CGGAGGCTAGAAGGAGAGAG | KF*LAECGRTEARRRE |
| 813 | G | AAAATTTTGACTAGCGGAGG | TGCGGCCGCA | GGAGGCTAGAAGGAGAGAGA | KF*LAEVRPQEARRRE |
| 814 | C | AAATTTTGACTAGCGGAGGC | TGCGGCCGCA | GAGGCTAGAAGGAGAGAGAT | KF*LAEAAAAEARRRE |
| 815 | T | AATTTTGACTAGCGGAGGCT | TGCGGCCGCA | AGGCTAGAAGGAGAGAGATG | F*LAEACGRKARRREM |
| 816 | A | ATTTTGACTAGCGGAGGCTA | TGCGGCCGCA | GGCTAGAAGGAGAGAGATGG | F*LAEAMRPQARRREM |
| 817 | G | TTTTGACTAGCGGAGGCTAG | TGCGGCCGCA | GCTAGAAGGAGAGAGATGGG | F*LAEASAAAARRREM |
| 818 | A | TTTGACTAGCGGAGGCTAGA | TGCGGCCGCA | CTAGAAGGAGAGAGATGGGT | *LAEARCGRTRRREMG |
| 819 | A | TTGACTAGCGGAGGCTAGAA | TGCGGCCGCA | TAGAAGGAGAGAGATGGGTG | *LAEARMRPHRRREMG |
| 820 | G | TGACTAGCGGAGGCTAGAAG | TGCGGCCGCA | AGAAGGAGAGAGATGGGTGC | *LAEARSAAARRREMG |
| 821 | G | GACTAGCGGAGGCTAGAAGG | TGCGGCCGCA | GAAGGAGAGAGATGGGTGCG | LAEARRCGRRRREMGA |
| 822 | A | ACTAGCGGAGGCTAGAAGGA | TGCGGCCGCA | AAGGAGAGAGATGGGTGCGA | LAEARRMRPQRREMGA |
| 823 | G | CTAGCGGAGGCTAGAAGGAG | TGCGGCCGCA | AGGAGAGAGATGGGTGCGAG | LAEARRSAAARREMGA |
| 824 | A | TAGCGGAGGCTAGAAGGAGA | TGCGGCCGCA | GGAGAGAGATGGGTGCGAGA | AEARRRCGRRREMGAR |
| 825 | G | AGCGGAGGCTAGAAGGAGAG | TGCGGCCGCA | GAGAGAGATGGGTGCGAGAG | AEARRRVRPQREMGAR |
| 826 | A | GCGGAGGCTAGAAGGAGAGA | TGCGGCCGCA | AGAGAGATGGGTGCGAGAGC | AEARRRDAAAREMGAR |
| 827 | G | CGGAGGCTAGAAGGAGAGAG | TGCGGCCGCA | GAGAGATGGGTGCGAGAGCG | EARRRECGRREMGARA |
| 828 | A | GGAGGCTAGAAGGAGAGAGA | TGCGGCCGCA | AGAGATGGGTGCGAGAGCGT | EARRREMRPQEMGARA |
| 829 | T | GAGGCTAGAAGGAGAGAGAT | TGCGGCCGCA | GAGATGGGTGCGAGAGCGTC | EARRREIAAAEMGARA |
| 830 | G | AGGCTAGAAGGAGAGAGATG | TGCGGCCGCA | AGATGGGTGCGAGAGCGTCG | ARRREMCGRKMGARAS |
| 831 | G | GGCTAGAAGGAGAGAGATGG | TGCGGCCGCA | GATGGGTGCGAGAGCGTCGG | ARRREMVRPQMGARAS |
| 832 | G | GCTAGAAGGAGAGAGATGGG | TGCGGCCGCA | ATGGGTGCGAGAGCGTCGGT | ARRREMGAAAMGARAS |
| 833 | T | CTAGAAGGAGAGAGATGGGT | TGCGGCCGCA | TGGGTGCGAGAGCGTCGGTA | RRREMGCGRMGARASV |
| 834 | G | TAGAAGGAGAGAGATGGGTG | TGCGGCCGCA | GGGTGCGAGAGCGTCGGTAT | RRREMGVRPQGARASV |
| 835 | C | AGAAGGAGAGAGATGGGTGC | TGCGGCCGCA | GGTGCGAGAGCGTCGGTATT | RRREMGAAAAGARASV |
| 836 | G | GAAGGAGAGAGATGGGTGCG | TGCGGCCGCA | GTGCGAGAGCGTCGGTATTA | RREMGACGRSARASVL |
| 837 | A | AAGGAGAGAGATGGGTGCGA | TGCGGCCGCA | TGCGAGAGCGTCGGTATTAA | RREMGAMRPHARASVL |
| 838 | G | AGGAGAGAGATGGGTGCGAG | TGCGGCCGCA | GCGAGAGCGTCGGTATTAAG | RREMGASAAAARASVL |
| 839 | A | GGAGAGAGATGGGTGCGAGA | TGCGGCCGCA | CGAGAGCGTCGGTATTAAGC | REMGARCGRTRASVLS |
| 840 | G | GAGAGAGATGGGTGCGAGAG | TGCGGCCGCA | GAGAGCGTCGGTATTAAGCG | REMGARVRPQRASVLS |
| 841 | C | AGAGAGATGGGTGCGAGAGC | TGCGGCCGCA | AGAGCGTCGGTATTAAGCGG | REMGARAAAARASVLS |
| 842 | G | GAGAGATGGGTGCGAGAGCG | TGCGGCCGCA | GAGCGTCGGTATTAAGCGGG | EMGARACGRRASVLSG |
| 843 | T | AGAGATGGGTGCGAGAGCGT | TGCGGCCGCA | AGCGTCGGTATTAAGCGGGG | EMGARALRPQASVLSG |
| 844 | C | GAGATGGGTGCGAGAGCGTC | TGCGGCCGCA | GCGTCGGTATTAAGCGGGGG | EMGARASAAAASVLSG |
| 845 | G | AGATGGGTGCGAGAGCGTCG | TGCGGCCGCA | CGTCGGTATTAAGCGGGGGA | MGARASCGRTSVLSGG |
| 846 | G | GATGGGTGCGAGAGCGTCGG | TGCGGCCGCA | GTCGGTATTAAGCGGGGGAG | MGARASVRPQSVLSGG |
| 847 | T | ATGGGTGCGAGAGCGTCGGT | TGCGGCCGCA | TCGGTATTAAGCGGGGGAGA | MGARASVAAASVLSGG |
| 848 | A | TGGGTGCGAGAGCGTCGGTA | TGCGGCCGCA | CGGTATTAAGCGGGGGAGAA | GARASVCGRTVLSGGE |
| 849 | T | GGGTGCGAGAGCGTCGGTAT | TGCGGCCGCA | GGTATTAAGCGGGGGAGAAT | GARASVLRPQVLSGGE |
| 850 | T | GGTGCGAGAGCGTCGGTATT | TGCGGCCGCA | GTATTAAGCGGGGGAGAATT | GARASVFAAAVLSGGE |
| 851 | A | GTGCGAGAGCGTCGGTATTA | TGCGGCCGCA | TATTAAGCGGGGGAGAATTA | ARASVLCGRILSGGEL |
| 852 | A | TGCGAGAGCGTCGGTATTAA | TGCGGCCGCA | ATTAAGCGGGGGAGAATTAG | ARASVLMRPQLSGGEL |
| 853 | G | GCGAGAGCGTCGGTATTAAG | TGCGGCCGCA | TTAAGCGGGGGAGAATTAGA | ARASVLSAAALSGGEL |
| 854 | C | CGAGAGCGTCGGTATTAAGC | TGCGGCCGCA | TAAGCGGGGGAGAATTAGAT | RASVLSCGRISGGELD |

| | | | | | |
|---|---|---|---|---|---|
| 855 | G | GAGAGCGTCGGTATTAAGCG | TGCGGCCGCA | AAGCGGGGGAGAATTAGATA | RASVLSVRPQSGGELD |
| 856 | G | AGAGCGTCGGTATTAAGCGG | TGCGGCCGCA | AGCGGGGGAGAATTAGATAA | RASVLSGAAASGGELD |
| 857 | G | GAGCGTCGGTATTAAGCGGG | TGCGGCCGCA | GCGGGGGAGAATTAGATAAA | ASVLSGCGRSGGELDK |
| 858 | G | AGCGTCGGTATTAAGCGGGG | TGCGGCCGCA | CGGGGGAGAATTAGATAAAT | ASVLSGVRPHGGELDK |
| 859 | G | GCGTCGGTATTAAGCGGGGG | TGCGGCCGCA | GGGGGAGAATTAGATAAATG | ASVLSGGAAAGGELDK |
| 860 | A | CGTCGGTATTAAGCGGGGGA | TGCGGCCGCA | GGGGAGAATTAGATAAATGG | SVLSGGCGRRGELDKW |
| 861 | G | GTCGGTATTAAGCGGGGGAG | TGCGGCCGCA | GGGAGAATTAGATAAATGGG | SVLSGGVRPQGELDKW |
| 862 | A | TCGGTATTAAGCGGGGGAGA | TGCGGCCGCA | GGAGAATTAGATAAATGGGA | SVLSGGDAAAGELDKW |
| 863 | A | CGGTATTAAGCGGGGGAGAA | TGCGGCCGCA | GAGAATTAGATAAATGGGAA | VLSGGECGRRELDKWE |
| 864 | T | GGTATTAAGCGGGGGAGAAT | TGCGGCCGCA | AGAATTAGATAAATGGGAAA | VLSGGELRPQELDKWE |
| 865 | T | GTATTAAGCGGGGGAGAATT | TGCGGCCGCA | GAATTAGATAAATGGGAAAA | VLSGGEFAAAELDKWE |
| 866 | A | TATTAAGCGGGGGAGAATTA | TGCGGCCGCA | AATTAGATAAATGGGAAAAA | LSGGELCGRKLDKWEK |
| 867 | G | ATTAAGCGGGGGAGAATTAG | TGCGGCCGCA | ATTAGATAAATGGGAAAAAA | LSGGELVRPQLDKWEK |
| 868 | A | TTAAGCGGGGGAGAATTAGA | TGCGGCCGCA | TTAGATAAATGGGAAAAAAT | LSGGELDAAALDKWEK |
| 869 | T | TAAGCGGGGGAGAATTAGAT | TGCGGCCGCA | TAGATAAATGGGAAAAAATT | SGGELDCGRIDKWEKI |
| 870 | A | AAGCGGGGGAGAATTAGATA | TGCGGCCGCA | AGATAAATGGGAAAAAATTC | SGGELDMRPQDKWEKI |
| 871 | A | AGCGGGGGAGAATTAGATAA | TGCGGCCGCA | GATAAATGGGAAAAAATTCG | SGGELDNAAADKWEKI |
| 872 | A | GCGGGGGAGAATTAGATAAA | TGCGGCCGCA | ATAAATGGGAAAAAATTCGG | GGELDKCGRNKWEKIR |
| 873 | T | CGGGGGAGAATTAGATAAAT | TGCGGCCGCA | TAAATGGGAAAAAATTCGGT | GGELDKLRPHKWEKIR |
| 874 | G | GGGGGAGAATTAGATAAATG | TGCGGCCGCA | AAATGGGAAAAAATTCGGTT | GGELDKCAAAKWEKIR |
| 875 | G | GGGGAGAATTAGATAAATGG | TGCGGCCGCA | AATGGGAAAAAATTCGGTTA | GELDKWCGRKWEKIRL |
| 876 | G | GGGAGAATTAGATAAATGGG | TGCGGCCGCA | ATGGGAAAAAATTCGGTTAA | GELDKWVRPQWEKIRL |
| 877 | A | GGAGAATTAGATAAATGGGA | TGCGGCCGCA | TGGGAAAAAATTCGGTTAAG | GELDKWDAAAWEKIRL |
| 878 | A | GAGAATTAGATAAATGGGAA | TGCGGCCGCA | GGGAAAAAATTCGGTTAAGG | ELDKWECGRREKIRLR |
| 879 | A | AGAATTAGATAAATGGGAAA | TGCGGCCGCA | GGAAAAAATTCGGTTAAGGC | ELDKWEMRPQEKIRLR |
| 880 | A | GAATTAGATAAATGGGAAAA | TGCGGCCGCA | GAAAAAATTCGGTTAAGGCC | ELDKWENAAAEKIRLR |
| 881 | A | AATTAGATAAATGGGAAAAA | TGCGGCCGCA | AAAAAATTCGGTTAAGGCCA | LDKWEKCGRKKIRLRP |
| 882 | A | ATTAGATAAATGGGAAAAAA | TGCGGCCGCA | AAAAATTCGGTTAAGGCCAG | LDKWEKMRPQKIRLRP |
| 883 | T | TTAGATAAATGGGAAAAAAT | TGCGGCCGCA | AAAATTCGGTTAAGGCCAGG | LDKWEKIAAAKIRLRP |
| 884 | T | TAGATAAATGGGAAAAAATT | TGCGGCCGCA | AAATTCGGTTAAGGCCAGGG | DKWEKICGRKIRLRPG |
| 885 | C | AGATAAATGGGAAAAAATTC | TGCGGCCGCA | AATTCGGTTAAGGCCAGGGG | DKWEKILRPQIRLRPG |
| 886 | G | GATAAATGGGAAAAAATTCG | TGCGGCCGCA | ATTCGGTTAAGGCCAGGGGG | DKWEKIRAAAIRLRPG |
| 887 | G | ATAAATGGGAAAAAATTCGG | TGCGGCCGCA | TTCGGTTAAGGCCAGGGGGA | KWEKIRCGRIRLRPGG |
| 888 | T | TAAATGGGAAAAAATTCGGT | TGCGGCCGCA | TCGGTTAAGGCCAGGGGGAA | KWEKIRLRPHRLRPGG |
| 889 | T | AAATGGGAAAAAATTCGGTT | TGCGGCCGCA | CGGTTAAGGCCAGGGGGAAA | KWEKIRFAAARLRPGG |
| 890 | A | AATGGGAAAAAATTCGGTTA | TGCGGCCGCA | GGTTAAGGCCAGGGGGAAAG | WEKIRLCGRRLRPGGK |
| 891 | A | ATGGGAAAAAATTCGGTTAA | TGCGGCCGCA | GTTAAGGCCAGGGGGAAAGA | WEKIRLMRPQLRPGGK |
| 892 | G | TGGGAAAAAATTCGGTTAAG | TGCGGCCGCA | TTAAGGCCAGGGGGAAAGAA | WEKIRLSAAALRPGGK |
| 893 | G | GGGAAAAAATTCGGTTAAGG | TGCGGCCGCA | TAAGGCCAGGGGGAAAGAAA | EKIRLRCGRIRPGGKK |
| 894 | C | GGAAAAAATTCGGTTAAGGC | TGCGGCCGCA | AAGGCCAGGGGGAAAGAAAC | EKIRLRLRPQRPGGKK |
| 895 | C | GAAAAAATTCGGTTAAGGCC | TGCGGCCGCA | AGGCCAGGGGGAAAGAAACA | EKIRLRPAAARPGGKK |
| 896 | A | AAAAAATTCGGTTAAGGCCA | TGCGGCCGCA | GGCCAGGGGGAAAGAAACAA | KIRLRPCGRRPGGKKQ |
| 897 | G | AAAAATTCGGTTAAGGCCAG | TGCGGCCGCA | GCCAGGGGGAAAGAAACAAT | KIRLRPVRPQPGGKKQ |
| 898 | G | AAAATTCGGTTAAGGCCAGG | TGCGGCCGCA | CCAGGGGGAAAGAAACAATA | KIRLRPGAAAPGGKKQ |
| 899 | G | AAATTCGGTTAAGGCCAGGG | TGCGGCCGCA | CAGGGGGAAAGAAACAATAT | IRLRPGCGRTGGKKQY |
| 900 | G | AATTCGGTTAAGGCCAGGGG | TGCGGCCGCA | AGGGGGAAAGAAACAATATA | IRLRPGVRPQGGKKQY |
| 901 | G | ATTCGGTTAAGGCCAGGGGG | TGCGGCCGCA | GGGGGAAAGAAACAATATAA | IRLRPGGAAAGGKKQY |
| 902 | A | TTCGGTTAAGGCCAGGGGGA | TGCGGCCGCA | GGGGAAAGAAACAATATAAA | RLRPGGCGRRGKKQYK |
| 903 | A | TCGGTTAAGGCCAGGGGGAA | TGCGGCCGCA | GGGAAAGAAACAATATAAAC | RLRPGGMRPQGKKQYK |
| 904 | A | CGGTTAAGGCCAGGGGGAAA | TGCGGCCGCA | GGAAAGAAACAATATAAACT | RLRPGGNAAAGKKQYK |
| 905 | G | GGTTAAGGCCAGGGGGAAAG | TGCGGCCGCA | GAAAGAAACAATATAAACTA | LRPGGKCGRRKKQYKL |
| 906 | A | GTTAAGGCCAGGGGGAAAGA | TGCGGCCGCA | AAAGAAACAATATAAACTAA | LRPGGKMRPQKKQYKL |
| 907 | A | TTAAGGCCAGGGGGAAAGAA | TGCGGCCGCA | AAGAAACAATATAAACTAAA | LRPGGKNAAAKKQYKL |
| 908 | A | TAAGGCCAGGGGGAAAGAAA | TGCGGCCGCA | AGAAACAATATAAACTAAAA | RPGGKKCGRKKQYKLK |
| 909 | C | AAGGCCAGGGGGAAAGAAAC | TGCGGCCGCA | GAAACAATATAAACTAAAAC | RPGGKKLRPQKQYKLK |
| 910 | A | AGGCCAGGGGGAAAGAAACA | TGCGGCCGCA | AAACAATATAAACTAAAACA | RPGGKKHAAAKQYKLK |
| 911 | A | GGCCAGGGGGAAAGAAACAA | TGCGGCCGCA | AACAATATAAACTAAAACAT | PGGKKQCGRKQYKLKH |
| 912 | T | GCCAGGGGGAAAGAAACAAT | TGCGGCCGCA | ACAATATAAACTAAAACATA | PGGKKQLRPQQYKLKH |
| 913 | A | CCAGGGGGAAAGAAACAATA | TGCGGCCGCA | CAATATAAACTAAAACATAT | PGGKKQYAAAQYKLKH |
| 914 | T | CAGGGGGAAAGAAACAATAT | TGCGGCCGCA | AATATAAACTAAAACATATA | GGKKQYCGRKYKLKHI |
| 915 | A | AGGGGGAAAGAAACAATATA | TGCGGCCGCA | ATATAAACTAAAACATATAG | GGKKQYMRPQYKLKHI |
| 916 | A | GGGGGAAAGAAACAATATAA | TGCGGCCGCA | TATAAACTAAAACATATAGT | GGKKQYNAAAYKLKHI |
| 917 | A | GGGGAAAGAAACAATATAAA | TGCGGCCGCA | ATAAACTAAAACATATAGTA | GKKQYKCGRNKLKHIV |
| 918 | C | GGGAAAGAAACAATATAAAC | TGCGGCCGCA | TAAACTAAAACATATAGTAT | GKKQYKLRPHKLKHIV |
| 919 | T | GGAAAGAAACAATATAAACT | TGCGGCCGCA | AAACTAAAACATATAGTATG | GKKQYKLAAAKLKHIV |
| 920 | A | GAAAGAAACAATATAAACTA | TGCGGCCGCA | AACTAAAACATATAGTATGG | KKQYKLCGRKLKHIVW |
| 921 | A | AAAGAAACAATATAAACTAA | TGCGGCCGCA | ACTAAAACATATAGTATGGG | KKQYKLMRPQLKHIVW |
| 922 | A | AAGAAACAATATAAACTAAA | TGCGGCCGCA | CTAAAACATATAGTATGGGC | KKQYKLNAAALKHIVW |
| 923 | A | AGAAACAATATAAACTAAAA | TGCGGCCGCA | TAAAACATATAGTATGGGCA | KQYKLKCGRIKHIVWA |
| 924 | C | GAAACAATATAAACTAAAAC | TGCGGCCGCA | AAAACATATAGTATGGGCAA | KQYKLKLRPQKHIVWA |
| 925 | A | AAACAATATAAACTAAAACA | TGCGGCCGCA | AAACATATAGTATGGGCAAG | KQYKLKHAAAKHIVWA |
| 926 | T | AACAATATAAACTAAAACAT | TGCGGCCGCA | AACATATAGTATGGGCAAGC | QYKLKHCGRKHIVWAS |
| 927 | A | ACAATATAAACTAAAACATA | TGCGGCCGCA | ACATATAGTATGGGCAAGCA | QYKLKHMRPQHIVWAS |
| 928 | T | CAATATAAACTAAAACATAT | TGCGGCCGCA | CATATAGTATGGGCAAGCAG | QYKLKHIAAAHIVWAS |
| 929 | A | AATATAAACTAAAACATATA | TGCGGCCGCA | ATATAGTATGGGCAAGCAGG | YKLKHICGRNIVWASR |
| 930 | G | ATATAAACTAAAACATATAG | TGCGGCCGCA | TATAGTATGGGCAAGCAGGG | YKLKHIVRPHIVWASR |
| 931 | T | TATAAACTAAAACATATAGT | TGCGGCCGCA | ATAGTATGGGCAAGCAGGGA | YKLKHIVAAAIVWASR |
| 932 | A | ATAAACTAAAACATATAGTA | TGCGGCCGCA | TAGTATGGGCAAGCAGGGAG | KLKHIVCGRIVWASRE |
| 933 | T | TAAACTAAAACATATAGTAT | TGCGGCCGCA | AGTATGGGCAAGCAGGGAGC | KLKHIVLRPQVWASRE |
| 934 | G | AAACTAAAACATATAGTATG | TGCGGCCGCA | GTATGGGCAAGCAGGGAGCT | KLKHIVCAAAVWASRE |
| 935 | G | AACTAAAACATATAGTATGG | TGCGGCCGCA | TATGGGCAAGCAGGGAGCTA | LKHIVWCGRIWASREL |
| 936 | G | ACTAAAACATATAGTATGGG | TGCGGCCGCA | ATGGGCAAGCAGGGAGCTAG | LKHIVWVRPQWASREL |
| 937 | C | CTAAAACATATAGTATGGGC | TGCGGCCGCA | TGGGCAAGCAGGGAGCTAGA | LKHIVWAAAAWASREL |
| 938 | A | TAAAACATATAGTATGGGCA | TGCGGCCGCA | GGGCAAGCAGGGAGCTAGAA | KHIVWACGRRASRELE |
| 939 | A | AAAACATATAGTATGGGCAA | TGCGGCCGCA | GGCAAGCAGGGAGCTAGAAC | KHIVWAMRPQASRELE |
| 940 | G | AAACATATAGTATGGGCAAG | TGCGGCCGCA | GCAAGCAGGGAGCTAGAACG | KHIVWASAAAASRELE |

| | | | | |
|---|---|---|---|---|
| 941 | C | AACATATAGTATGGGCAAGC | TGCGGCCGCA | CAAGCAGGGAGCTAGAACGA | HIVWASCGRTSRELER |
| 942 | A | ACATATAGTATGGGCAAGCA | TGCGGCCGCA | AAGCAGGGAGCTAGAACGAT | HIVWASMRPQSRELER |
| 943 | G | CATATAGTATGGGCAAGCAG | TGCGGCCGCA | AGCAGGGAGCTAGAACGATT | HIVWASSAAASRELER |
| 944 | G | ATATAGTATGGGCAAGCAGG | TGCGGCCGCA | GCAGGGAGCTAGAACGATTC | IVWASRCGRSRELERF |
| 945 | G | TATAGTATGGGCAAGCAGGG | TGCGGCCGCA | CAGGGAGCTAGAACGATTCG | IVWASRVRPHRELERF |
| 946 | A | ATAGTATGGGCAAGCAGGGA | TGCGGCCGCA | AGGGAGCTAGAACGATTCGC | IVWASRDAAARELERF |
| 947 | G | TAGTATGGGCAAGCAGGGAG | TGCGGCCGCA | GGGAGCTAGAACGATTCGCA | VWASRECGRRELERFA |
| 948 | C | AGTATGGGCAAGCAGGGAGC | TGCGGCCGCA | GGAGCTAGAACGATTCGCAG | VWASRELRPQELERFA |
| 949 | T | GTATGGGCAAGCAGGGAGCT | TGCGGCCGCA | GAGCTAGAACGATTCGCAGT | VWASRELAAAELERFA |
| 950 | A | TATGGGCAAGCAGGGAGCTA | TGCGGCCGCA | AGCTAGAACGATTCGCAGTT | WASRELCGRKLERFAV |
| 951 | G | ATGGGCAAGCAGGGAGCTAG | TGCGGCCGCA | GCTAGAACGATTCGCAGTTA | WASRELVRPQLERFAV |
| 952 | A | TGGGCAAGCAGGGAGCTAGA | TGCGGCCGCA | CTAGAACGATTCGCAGTTAA | WASRELDAAALERFAV |
| 953 | A | GGGCAAGCAGGGAGCTAGAA | TGCGGCCGCA | TAGAACGATTCGCAGTTAAT | ASRELECGRIERFAVN |
| 954 | C | GGCAAGCAGGGAGCTAGAAC | TGCGGCCGCA | AGAACGATTCGCAGTTAATC | ASRELELRPQERFAVN |
| 955 | G | GCAAGCAGGGAGCTAGAACG | TGCGGCCGCA | GAACGATTCGCAGTTAATCC | ASRELERAAAERFAVN |
| 956 | A | CAAGCAGGGAGCTAGAACGA | TGCGGCCGCA | AACGATTCGCAGTTAATCCT | SRELERCGRKRFAVNP |
| 957 | T | AAGCAGGGAGCTAGAACGAT | TGCGGCCGCA | ACGATTCGCAGTTAATCCTG | SRELERLRPQRFAVNP |
| 958 | T | AGCAGGGAGCTAGAACGATT | TGCGGCCGCA | CGATTCGCAGTTAATCCTGG | SRELERFAAARFAVNP |
| 959 | C | GCAGGGAGCTAGAACGATTC | TGCGGCCGCA | GATTCGCAGTTAATCCTGGC | RELERFCGRRFAVNPG |
| 960 | G | CAGGGAGCTAGAACGATTCG | TGCGGCCGCA | ATTCGCAGTTAATCCTGGCC | RELERFVRPQFAVNPG |
| 961 | C | AGGGAGCTAGAACGATTCGC | TGCGGCCGCA | TTCGCAGTTAATCCTGGCCT | RELERFAAAAFAVNPG |
| 962 | A | GGGAGCTAGAACGATTCGCA | TGCGGCCGCA | TCGCAGTTAATCCTGGCCTT | ELERFACGRIAVNPGL |
| 963 | G | GGAGCTAGAACGATTCGCAG | TGCGGCCGCA | CGCAGTTAATCCTGGCCTTT | ELERFAVRPHAVNPGL |
| 964 | T | GAGCTAGAACGATTCGCAGT | TGCGGCCGCA | GCAGTTAATCCTGGCCTTTT | ELERFAVAAAAVNPGL |
| 965 | T | AGCTAGAACGATTCGCAGTT | TGCGGCCGCA | CAGTTAATCCTGGCCTTTTA | LERFAVCGRTVNPGLL |
| 966 | A | GCTAGAACGATTCGCAGTTA | TGCGGCCGCA | AGTTAATCCTGGCCTTTTAG | LERFAVMRPQVNPGLL |
| 967 | A | CTAGAACGATTCGCAGTTAA | TGCGGCCGCA | GTTAATCCTGGCCTTTTAGA | LERFAVNAAAVNPGLL |
| 968 | T | TAGAACGATTCGCAGTTAAT | TGCGGCCGCA | TTAATCCTGGCCTTTTAGAG | ERFAVNCGRINPGLLE |
| 969 | C | AGAACGATTCGCAGTTAATC | TGCGGCCGCA | TAATCCTGGCCTTTTAGAGA | ERFAVNLRPHNPGLLE |
| 970 | C | GAACGATTCGCAGTTAATCC | TGCGGCCGCA | AATCCTGGCCTTTTAGAGAC | ERFAVNPAAANPGLLE |
| 971 | T | AACGATTCGCAGTTAATCCT | TGCGGCCGCA | ATCCTGGCCTTTTAGAGACA | RFAVNPCGRNPGLLET |
| 972 | G | ACGATTCGCAGTTAATCCTG | TGCGGCCGCA | TCCTGGCCTTTTAGAGACAT | RFAVNPVRPHPGLLET |
| 973 | G | CGATTCGCAGTTAATCCTGG | TGCGGCCGCA | CCTGGCCTTTTAGAGACATC | RFAVNPGAAAPGLLET |
| 974 | C | GATTCGCAGTTAATCCTGGC | TGCGGCCGCA | CTGGCCTTTTAGAGACATCA | FAVNPGCGRTGLLETS |
| 975 | C | ATTCGCAGTTAATCCTGGCC | TGCGGCCGCA | TGGCCTTTTAGAGACATCAG | FAVNPGLRPHGLLETS |
| 976 | T | TTCGCAGTTAATCCTGGCCT | TGCGGCCGCA | GGCCTTTTAGAGACATCAGA | FAVNPGLAAAGLLETS |
| 977 | T | TCGCAGTTAATCCTGGCCTT | TGCGGCCGCA | GCCTTTTAGAGACATCAGAA | AVNPGLCGRSLLETSE |
| 978 | T | CGCAGTTAATCCTGGCCTTT | TGCGGCCGCA | CCTTTTAGAGACATCAGAAG | AVNPGLLRPHLLETSE |
| 979 | T | GCAGTTAATCCTGGCCTTTT | TGCGGCCGCA | CTTTTAGAGACATCAGAAGG | AVNPGLFAAALLETSE |
| 980 | A | CAGTTAATCCTGGCCTTTTA | TGCGGCCGCA | TTTTAGAGACATCAGAAGGC | VNPGLLCGRILETSEG |
| 981 | G | AGTTAATCCTGGCCTTTTAG | TGCGGCCGCA | TTTAGAGACATCAGAAGGCT | VNPGLLVRPHLETSEG |
| 982 | A | GTTAATCCTGGCCTTTTAGA | TGCGGCCGCA | TTAGAGACATCAGAAGGCTG | VNPGLLDAAALETSEG |
| 983 | G | TTAATCCTGGCCTTTTAGAG | TGCGGCCGCA | TAGAGACATCAGAAGGCTGT | NPGLLECGRIETSEGC |
| 984 | A | TAATCCTGGCCTTTTAGAGA | TGCGGCCGCA | AGAGACATCAGAAGGCTGTA | NPGLLEMRPQETSEGC |
| 985 | C | AATCCTGGCCTTTTAGAGAC | TGCGGCCGCA | GAGACATCAGAAGGCTGTAG | NPGLLETAAAETSEGC |
| 986 | A | ATCCTGGCCTTTTAGAGACA | TGCGGCCGCA | AGACATCAGAAGGCTGTAGA | PGLLETCGRKTSEGCR |
| 987 | T | TCCTGGCCTTTTAGAGACAT | TGCGGCCGCA | GACATCAGAAGGCTGTAGAC | PGLLETLRPQTSEGCR |
| 988 | C | CCTGGCCTTTTAGAGACATC | TGCGGCCGCA | ACATCAGAAGGCTGTAGACA | PGLLETSAAATSEGCR |
| 989 | A | CTGGCCTTTTAGAGACATCA | TGCGGCCGCA | CATCAGAAGGCTGTAGACAA | GLLETSCGRTSEGCRQ |
| 990 | G | TGGCCTTTTAGAGACATCAG | TGCGGCCGCA | ATCAGAAGGCTGTAGACAAA | GLLETSVRPQSEGCRQ |
| 991 | A | GGCCTTTTAGAGACATCAGA | TGCGGCCGCA | TCAGAAGGCTGTAGACAAAT | GLLETSDAAASEGCRQ |
| 992 | A | GCCTTTTAGAGACATCAGAA | TGCGGCCGCA | CAGAAGGCTGTAGACAAATA | LLETSECGRTEGCRQI |
| 993 | G | CCTTTTAGAGACATCAGAAG | TGCGGCCGCA | AGAAGGCTGTAGACAAATAC | LLETSEVRPQEGCRQI |
| 994 | G | CTTTTAGAGACATCAGAAGG | TGCGGCCGCA | GAAGGCTGTAGACAAATACT | LLETSEGAAAEGCRQI |
| 995 | C | TTTTAGAGACATCAGAAGGC | TGCGGCCGCA | AAGGCTGTAGACAAATACTG | LETSEGCGRKGCRQIL |
| 996 | T | TTTAGAGACATCAGAAGGCT | TGCGGCCGCA | AGGCTGTAGACAAATACTGG | LETSEGLRPQGCRQIL |
| 997 | G | TTAGAGACATCAGAAGGCTG | TGCGGCCGCA | GGCTGTAGACAAATACTGGG | LETSEGCAAAGCRQIL |
| 998 | T | TAGAGACATCAGAAGGCTGT | TGCGGCCGCA | GCTGTAGACAAATACTGGGA | ETSEGCCGRSCRQILG |
| 999 | A | AGAGACATCAGAAGGCTGTA | TGCGGCCGCA | CTGTAGACAAATACTGGGAC | ETSEGCMRPHCRQILG |
| 1000 | G | GAGACATCAGAAGGCTGTAG | TGCGGCCGCA | TGTAGACAAATACTGGGACA | ETSEGCSAAACRQILG |
| 1001 | A | AGACATCAGAAGGCTGTAGA | TGCGGCCGCA | GTAGACAAATACTGGGACAG | TSEGCRCGRSRQILGQ |
| 1002 | C | GACATCAGAAGGCTGTAGAC | TGCGGCCGCA | TAGACAAATACTGGGACAGC | TSEGCRLRPHRQILGQ |
| 1003 | A | ACATCAGAAGGCTGTAGACA | TGCGGCCGCA | AGACAAATACTGGGACAGCT | TSEGCRHAAARQILGQ |
| 1004 | A | CATCAGAAGGCTGTAGACAA | TGCGGCCGCA | GACAAATACTGGGACAGCTA | SEGCRQCGRRQILGQL |
| 1005 | A | ATCAGAAGGCTGTAGACAAA | TGCGGCCGCA | ACAAATACTGGGACAGCTAC | SEGCRQMRPQQILGQL |
| 1006 | T | TCAGAAGGCTGTAGACAAAT | TGCGGCCGCA | CAAATACTGGGACAGCTACA | SEGCRQIAAAQILGQL |
| 1007 | A | CAGAAGGCTGTAGACAAATA | TGCGGCCGCA | AAATACTGGGACAGCTACAA | EGCRQICGRKILGQLQ |
| 1008 | C | AGAAGGCTGTAGACAAATAC | TGCGGCCGCA | AATACTGGGACAGCTACAAC | EGCRQILRPQILGQLQ |
| 1009 | T | GAAGGCTGTAGACAAATACT | TGCGGCCGCA | ATACTGGGACAGCTACAACC | EGCRQILAAAILGQLQ |
| 1010 | G | AAGGCTGTAGACAAATACTG | TGCGGCCGCA | TACTGGGACAGCTACAACCA | GCRQILCGRILGQLQP |
| 1011 | G | AGGCTGTAGACAAATACTGG | TGCGGCCGCA | ACTGGGACAGCTACAACCAT | GCRQILVRPQLGQLQP |
| 1012 | G | GGCTGTAGACAAATACTGGG | TGCGGCCGCA | CTGGGACAGCTACAACCATC | GCRQILGAAALGQLQP |
| 1013 | A | GCTGTAGACAAATACTGGGA | TGCGGCCGCA | TGGGACAGCTACAACCATCC | CRQILGCGRMGQLQPS |
| 1014 | C | CTGTAGACAAATACTGGGAC | TGCGGCCGCA | GGGACAGCTACAACCATCCC | CRQILGLRPQGQLQPS |
| 1015 | A | TGTAGACAAATACTGGGACA | TGCGGCCGCA | GGACAGCTACAACCATCCCT | CRQILGHAAAGQLQPS |
| 1016 | G | GTAGACAAATACTGGGACAG | TGCGGCCGCA | GACAGCTACAACCATCCCTT | RQILGQCGRRQLQPSL |
| 1017 | C | TAGACAAATACTGGGACAGC | TGCGGCCGCA | ACAGCTACAACCATCCCTTC | RQILGQLRPQQLQPSL |
| 1018 | T | AGACAAATACTGGGACAGCT | TGCGGCCGCA | CAGCTACAACCATCCCTTCA | RQILGQLAAAQLQPSL |
| 1019 | A | GACAAATACTGGGACAGCTA | TGCGGCCGCA | AGCTACAACCATCCCTTCAG | QILGQLCGRKLQPSLQ |
| 1020 | C | ACAAATACTGGGACAGCTAC | TGCGGCCGCA | GCTACAACCATCCCTTCAGA | QILGQLLRPQLQPSLQ |
| 1021 | A | CAAATACTGGGACAGCTACA | TGCGGCCGCA | CTACAACCATCCCTTCAGAC | QILGQLHAAALQPSLQ |
| 1022 | A | AAATACTGGGACAGCTACAA | TGCGGCCGCA | TACAACCATCCCTTCAGACA | ILGQLQCGRIQPSLQT |
| 1023 | C | AATACTGGGACAGCTACAAC | TGCGGCCGCA | ACAACCATCCCTTCAGACAG | ILGQLQLRPQQPSLQT |
| 1024 | C | ATACTGGGACAGCTACAACC | TGCGGCCGCA | CAACCATCCCTTCAGACAGG | ILGQLQPAAAQPSLQT |
| 1025 | A | TACTGGGACAGCTACAACCA | TGCGGCCGCA | AACCATCCCTTCAGACAGGA | LGQLQPCGRKPSLQTG |
| 1026 | T | ACTGGGACAGCTACAACCAT | TGCGGCCGCA | ACCATCCCTTCAGACAGGAT | LGQLQPLRPQPSLQTG |

| | | | | |
|---|---|---|---|---|
| 1027 | C | CTGGGACAGCTACAACCATC | TGCGGCCGCA | CCATCCCTTCAGACAGGATC | LGQLQPSAAAPSLQTG |
| 1028 | C | TGGGACAGCTACAACCATCC | TGCGGCCGCA | CATCCCTTCAGACAGGATCA | GQLQPSCGRTSLQTGS |
| 1029 | C | GGGACAGCTACAACCATCCC | TGCGGCCGCA | ATCCCTTCAGACAGGATCAG | GQLQPSLRPQSLQTGS |
| 1030 | T | GGACAGCTACAACCATCCCT | TGCGGCCGCA | TCCCTTCAGACAGGATCAGA | GQLQPSLAAASLQTGS |
| 1031 | T | GACAGCTACAACCATCCCTT | TGCGGCCGCA | CCCTTCAGACAGGATCAGAA | QLQPSLCGRTLQTGSE |
| 1032 | C | ACAGCTACAACCATCCCTTC | TGCGGCCGCA | CCTTCAGACAGGATCAGAAG | QLQPSLLRPHLQTGSE |
| 1033 | A | CAGCTACAACCATCCCTTCA | TGCGGCCGCA | CTTCAGACAGGATCAGAAGA | QLQPSLHAAALQTGSE |
| 1034 | G | AGCTACAACCATCCCTTCAG | TGCGGCCGCA | TTCAGACAGGATCAGAAGAA | LQPSLQCGRIQTGSEE |
| 1035 | A | GCTACAACCATCCCTTCAGA | TGCGGCCGCA | TCAGACAGGATCAGAAGAAC | LQPSLQMRPHQTGSEE |
| 1036 | C | CTACAACCATCCCTTCAGAC | TGCGGCCGCA | CAGACAGGATCAGAAGAACT | LQPSLQTAAAQTGSEE |
| 1037 | A | TACAACCATCCCTTCAGACA | TGCGGCCGCA | AGACAGGATCAGAAGAACTT | QPSLQTCGRKTGSEEL |
| 1038 | G | ACAACCATCCCTTCAGACAG | TGCGGCCGCA | GACAGGATCAGAAGAACTTA | QPSLQTVRPQTGSEEL |
| 1039 | G | CAACCATCCCTTCAGACAGG | TGCGGCCGCA | ACAGGATCAGAAGAACTTAG | QPSLQTGAAATGSEEL |
| 1040 | A | AACCATCCCTTCAGACAGGA | TGCGGCCGCA | CAGGATCAGAAGAACTTAGA | PSLQTGCGRTGSEELR |
| 1041 | T | ACCATCCCTTCAGACAGGAT | TGCGGCCGCA | AGGATCAGAAGAACTTAGAT | PSLQTGLRPQGSEELR |
| 1042 | C | CCATCCCTTCAGACAGGATC | TGCGGCCGCA | GGATCAGAAGAACTTAGATC | PSLQTGSAAAGSEELR |
| 1043 | A | CATCCCTTCAGACAGGATCA | TGCGGCCGCA | GATCAGAAGAACTTAGATCA | SLQTGSCGRRSEELRS |
| 1044 | G | ATCCCTTCAGACAGGATCAG | TGCGGCCGCA | ATCAGAAGAACTTAGATCAT | SLQTGSVRPQSEELRS |
| 1045 | A | TCCCTTCAGACAGGATCAGA | TGCGGCCGCA | TCAGAAGAACTTAGATCATT | SLQTGSDAAASEELRS |
| 1046 | A | CCCTTCAGACAGGATCAGAA | TGCGGCCGCA | CAGAAGAACTTAGATCATTA | LQTGSECGRTEELRSL |
| 1047 | G | CCTTCAGACAGGATCAGAAG | TGCGGCCGCA | AGAAGAACTTAGATCATTAT | LQTGSEVRPQEELRSL |
| 1048 | A | CTTCAGACAGGATCAGAAGA | TGCGGCCGCA | GAAGAACTTAGATCATTATA | LQTGSEDAAAEELRSL |
| 1049 | A | TTCAGACAGGATCAGAAGAA | TGCGGCCGCA | AAGAACTTAGATCATTATAT | QTGSEECGRKELRSLY |
| 1050 | C | TCAGACAGGATCAGAAGAAC | TGCGGCCGCA | AGAACTTAGATCATTATATA | QTGSEELRPQELRSLY |
| 1051 | T | CAGACAGGATCAGAAGAACT | TGCGGCCGCA | GAACTTAGATCATTATATAA | QTGSEELAAAELRSLY |
| 1052 | T | AGACAGGATCAGAAGAACTT | TGCGGCCGCA | AACTTAGATCATTATATAAT | TGSEELCGRKLRSLYN |
| 1053 | A | GACAGGATCAGAAGAACTTA | TGCGGCCGCA | ACTTAGATCATTATATAATA | TGSEELMRPQLRSLYN |
| 1054 | G | ACAGGATCAGAAGAACTTAG | TGCGGCCGCA | CTTAGATCATTATATAATAC | TGSEELSAAALRSLYN |
| 1055 | A | CAGGATCAGAAGAACTTAGA | TGCGGCCGCA | TTAGATCATTATATAATACA | GSEELRCGRIRSLYNT |
| 1056 | T | AGGATCAGAAGAACTTAGAT | TGCGGCCGCA | TAGATCATTATATAATACAA | GSEELRLRPHRSLYNT |
| 1057 | C | GGATCAGAAGAACTTAGATC | TGCGGCCGCA | AGATCATTATATAATACAAT | GSEELRSAAARSLYNT |
| 1058 | A | GATCAGAAGAACTTAGATCA | TGCGGCCGCA | GATCATTATATAATACAATA | SEELRSCGRRSLYNTI |
| 1059 | T | ATCAGAAGAACTTAGATCAT | TGCGGCCGCA | ATCATTATATAATACAATAG | SEELRSLRPQSLYNTI |
| 1060 | T | TCAGAAGAACTTAGATCATT | TGCGGCCGCA | TCATTATATAATACAATAGC | SEELRSFAAASLYNTI |
| 1061 | A | CAGAAGAACTTAGATCATTA | TGCGGCCGCA | CATTATATAATACAATAGCA | EELRSLCGRTLYNTIA |
| 1062 | T | AGAAGAACTTAGATCATTAT | TGCGGCCGCA | ATTATATAATACAATAGCAG | EELRSLLRPQLYNTIA |
| 1063 | A | GAAGAACTTAGATCATTATA | TGCGGCCGCA | TTATATAATACAATAGCAGT | EELRSLYAAALYNTIA |
| 1064 | T | AAGAACTTAGATCATTATAT | TGCGGCCGCA | TATATAATACAATAGCAGTC | ELRSLYCGRIYNTIAV |
| 1065 | A | AGAACTTAGATCATTATATA | TGCGGCCGCA | ATATAATACAATAGCAGTCC | ELRSLYMRPQYNTIAV |
| 1066 | A | GAACTTAGATCATTATATAA | TGCGGCCGCA | TATAATACAATAGCAGTCCT | ELRSLYNAAAYNTIAV |
| 1067 | T | AACTTAGATCATTATATAAT | TGCGGCCGCA | ATAATACAATAGCAGTCCTC | LRSLYNCGRNNTIAVL |
| 1068 | A | ACTTAGATCATTATATAATA | TGCGGCCGCA | TAATACAATAGCAGTCCTCT | LRSLYNMRPHNTIAVL |
| 1069 | C | CTTAGATCATTATATAATAC | TGCGGCCGCA | AATACAATAGCAGTCCTCTA | LRSLYNTAAANTIAVL |
| 1070 | A | TTAGATCATTATATAATACA | TGCGGCCGCA | ATACAATAGCAGTCCTCTAT | RSLYNTCGRNTIAVLY |
| 1071 | A | TAGATCATTATATAATACAA | TGCGGCCGCA | TACAATAGCAGTCCTCTATT | RSLYNTMRPHTIAVLY |
| 1072 | T | AGATCATTATATAATACAAT | TGCGGCCGCA | ACAATAGCAGTCCTCTATTG | RSLYNTIAAATIAVLY |
| 1073 | A | GATCATTATATAATACAATA | TGCGGCCGCA | CAATAGCAGTCCTCTATTGT | SLYNTICGRTIAVLYC |
| 1074 | G | ATCATTATATAATACAATAG | TGCGGCCGCA | AATAGCAGTCCTCTATTGTG | SLYNTIVRPQIAVLYC |
| 1075 | C | TCATTATATAATACAATAGC | TGCGGCCGCA | ATAGCAGTCCTCTATTGTGT | SLYNTIAAAAIAVLYC |
| 1076 | A | CATTATATAATACAATAGCA | TGCGGCCGCA | TAGCAGTCCTCTATTGTGTG | LYNTIACGRIAVLYCV |
| 1077 | G | ATTATATAATACAATAGCAG | TGCGGCCGCA | AGCAGTCCTCTATTGTGTGC | LYNTIAVRPQAVLYCV |
| 1078 | T | TTATATAATACAATAGCAGT | TGCGGCCGCA | GCAGTCCTCTATTGTGTGCA | LYNTIAVAAAAVLYCV |
| 1079 | C | TATATAATACAATAGCAGTC | TGCGGCCGCA | CAGTCCTCTATTGTGTGCAT | YNTIAVCGRTVLYCVH |
| 1080 | C | ATATAATACAATAGCAGTCC | TGCGGCCGCA | AGTCCTCTATTGTGTGCATC | YNTIAVLRPQVLYCVH |
| 1081 | T | TATAATACAATAGCAGTCCT | TGCGGCCGCA | GTCCTCTATTGTGTGCATCA | YNTIAVLAAAVLYCVH |
| 1082 | C | ATAATACAATAGCAGTCCTC | TGCGGCCGCA | TCCTCTATTGTGTGCATCAA | NTIAVLCGRILYCVHQ |
| 1083 | T | TAATACAATAGCAGTCCTCT | TGCGGCCGCA | CCTCTATTGTGTGCATCAAA | NTIAVLLRPHLYCVHQ |
| 1084 | A | AATACAATAGCAGTCCTCTA | TGCGGCCGCA | CTCTATTGTGTGCATCAAAG | NTIAVLYAAALYCVHQ |
| 1085 | T | ATACAATAGCAGTCCTCTAT | TGCGGCCGCA | TCTATTGTGTGCATCAAAGG | TIAVLYCGRIYCVHQR |
| 1086 | T | TACAATAGCAGTCCTCTATT | TGCGGCCGCA | CTATTGTGTGCATCAAAGGA | TIAVLYLRPHYCVHQR |
| 1087 | G | ACAATAGCAGTCCTCTATTG | TGCGGCCGCA | TATTGTGTGCATCAAAGGAT | TIAVLYCAAAYCVHQR |
| 1088 | T | CAATAGCAGTCCTCTATTGT | TGCGGCCGCA | ATTGTGTGCATCAAAGGATA | IAVLYCCGRNCVHQRI |
| 1089 | G | AATAGCAGTCCTCTATTGTG | TGCGGCCGCA | TTGTGTGCATCAAAGGATAG | IAVLYCVRPHCVHQRI |
| 1090 | T | ATAGCAGTCCTCTATTGTGT | TGCGGCCGCA | TGTGTGCATCAAAGGATAGA | IAVLYCVAAACVHQRI |
| 1091 | G | TAGCAGTCCTCTATTGTGTG | TGCGGCCGCA | GTGTGCATCAAAGGATAGAT | AVLYCVCGRSVHQRID |
| 1092 | C | AGCAGTCCTCTATTGTGTGC | TGCGGCCGCA | TGTGCATCAAAGGATAGATG | AVLYCVLRPHVHQRID |
| 1093 | A | GCAGTCCTCTATTGTGTGCA | TGCGGCCGCA | GTGCATCAAAGGATAGATGT | AVLYCVHAAAVHQRID |
| 1094 | T | CAGTCCTCTATTGTGTGCAT | TGCGGCCGCA | TGCATCAAAGGATAGATGTA | VLYCVHCGRMHQRIDV |
| 1095 | C | AGTCCTCTATTGTGTGCATC | TGCGGCCGCA | GCATCAAAGGATAGATGTAA | VLYCVHLRPQHQRIDV |
| 1096 | A | GTCCTCTATTGTGTGCATCA | TGCGGCCGCA | CATCAAAGGATAGATGTAAA | VLYCVHHAAAHQRIDV |
| 1097 | A | TCCTCTATTGTGTGCATCAA | TGCGGCCGCA | ATCAAAGGATAGATGTAAAA | LYCVHQCGRNQRIDVK |
| 1098 | A | CCTCTATTGTGTGCATCAAA | TGCGGCCGCA | TCAAAGGATAGATGTAAAAG | LYCVHQMRPHQRIDVK |
| 1099 | G | CTCTATTGTGTGCATCAAAG | TGCGGCCGCA | CAAAGGATAGATGTAAAAGA | LYCVHQSAAAQRIDVK |
| 1100 | G | TCTATTGTGTGCATCAAAGG | TGCGGCCGCA | AAAGGATAGATGTAAAAGAC | YCVHQRCGRKRIDVKD |
| 1101 | A | CTATTGTGTGCATCAAAGGA | TGCGGCCGCA | AAGGATAGATGTAAAAGACA | YCVHQRMRPQRIDVKD |
| 1102 | T | TATTGTGTGCATCAAAGGAT | TGCGGCCGCA | AGGATAGATGTAAAAGACAC | YCVHQRIAAARIDVKD |
| 1103 | A | ATTGTGTGCATCAAAGGATA | TGCGGCCGCA | GGATAGATGTAAAAGACACC | CVHQRICGRRIDVKDT |
| 1104 | G | TTGTGTGCATCAAAGGATAG | TGCGGCCGCA | GATAGATGTAAAAGACACCA | CVHQRIVRPQIDVKDT |
| 1105 | A | TGTGTGCATCAAAGGATAGA | TGCGGCCGCA | ATAGATGTAAAAGACACCAA | CVHQRIDAAAIDVKDT |
| 1106 | T | GTGTGCATCAAAGGATAGAT | TGCGGCCGCA | TAGATGTAAAAGACACCAAG | VHQRIDCGRIDVKDTK |
| 1107 | G | TGTGCATCAAAGGATAGATG | TGCGGCCGCA | AGATGTAAAAGACACCAAGG | VHQRIDVRPQDVKDTK |
| 1108 | T | GTGCATCAAAGGATAGATGT | TGCGGCCGCA | GATGTAAAAGACACCAAGGA | VHQRIDVAAADVKDTK |
| 1109 | A | TGCATCAAAGGATAGATGTA | TGCGGCCGCA | ATGTAAAAGACACCAAGGAA | HQRIDVCGRNVKDTKE |
| 1110 | A | GCATCAAAGGATAGATGTAA | TGCGGCCGCA | TGTAAAAGACACCAAGGAAG | HQRIDVMRPHVKDTKE |
| 1111 | A | CATCAAAGGATAGATGTAAA | TGCGGCCGCA | GTAAAAGACACCAAGGAAGC | HQRIDVNAAAVKDTKE |
| 1112 | A | ATCAAAGGATAGATGTAAAA | TGCGGCCGCA | TAAAAGACACCAAGGAAGCC | QRIDVKCGRIKDTKEA |

| | | | | | |
|---|---|---|---|---|---|
| 1113 | G | TCAAAGGATAGATGTAAAAG | TGCGGCCGCA | AAAAGACACCAAGGAAGCCT | QRIDVKVRPQKDTKEA |
| 1114 | A | CAAAGGATAGATGTAAAAGA | TGCGGCCGCA | AAAGACACCAAGGAAGCCTT | QRIDVKDAAAKDTKEA |
| 1115 | C | AAAGGATAGATGTAAAAGAC | TGCGGCCGCA | AAGACACCAAGGAAGCCTTA | RIDVKDCGRKDTKEAL |
| 1116 | A | AAGGATAGATGTAAAAGACA | TGCGGCCGCA | AGACACCAAGGAAGCCTTAG | RIDVKDMRPQDTKEAL |
| 1117 | C | AGGATAGATGTAAAAGACAC | TGCGGCCGCA | GACACCAAGGAAGCCTTAGA | RIDVKDTAAADTKEAL |
| 1118 | C | GGATAGATGTAAAAGACACC | TGCGGCCGCA | ACACCAAGGAAGCCTTAGAT | IDVKDTCGRNTKEALD |
| 1119 | A | GATAGATGTAAAAGACACCA | TGCGGCCGCA | CACCAAGGAAGCCTTAGATA | IDVKDTMRPHTKEALD |
| 1120 | A | ATAGATGTAAAAGACACCAA | TGCGGCCGCA | ACCAAGGAAGCCTTAGATAA | IDVKDTNAAATKEALD |
| 1121 | G | TAGATGTAAAAGACACCAAG | TGCGGCCGCA | CCAAGGAAGCCTTAGATAAG | DVKDTKCGRTKEALDK |
| 1122 | G | AGATGTAAAAGACACCAAGG | TGCGGCCGCA | CAAGGAAGCCTTAGATAAGA | DVKDTKVRPHKEALDK |
| 1123 | A | GATGTAAAAGACACCAAGGA | TGCGGCCGCA | AAGGAAGCCTTAGATAAGAT | DVKDTKDAAAKEALDK |
| 1124 | A | ATGTAAAAGACACCAAGGAA | TGCGGCCGCA | AGGAAGCCTTAGATAAGATA | VKDTKECGRKEALDKI |
| 1125 | G | TGTAAAAGACACCAAGGAAG | TGCGGCCGCA | GGAAGCCTTAGATAAGATAG | VKDTKEVRPQEALDKI |
| 1126 | C | GTAAAAGACACCAAGGAAGC | TGCGGCCGCA | GAAGCCTTAGATAAGATAGA | VKDTKEAAAAEALDKI |
| 1127 | C | TAAAAGACACCAAGGAAGCC | TGCGGCCGCA | AAGCCTTAGATAAGATAGAG | KDTKEACGRKALDKIE |
| 1128 | T | AAAAGACACCAAGGAAGCCT | TGCGGCCGCA | AGCCTTAGATAAGATAGAGG | KDTKEALRPQALDKIE |
| 1129 | T | AAAGACACCAAGGAAGCCTT | TGCGGCCGCA | GCCTTAGATAAGATAGAGGA | KDTKEAFAAAALDKIE |
| 1130 | A | AAGACACCAAGGAAGCCTTA | TGCGGCCGCA | CCTTAGATAAGATAGAGGAA | DTKEALCGRTLDKIEE |
| 1131 | G | AGACACCAAGGAAGCCTTAG | TGCGGCCGCA | CTTAGATAAGATAGAGGAAG | DTKEALVRPHLDKIEE |
| 1132 | A | GACACCAAGGAAGCCTTAGA | TGCGGCCGCA | TTAGATAAGATAGAGGAAGA | DTKEALDAAALDKIEE |
| 1133 | T | ACACCAAGGAAGCCTTAGAT | TGCGGCCGCA | TAGATAAGATAGAGGAAGAG | TKEALDCGRIDKIEEE |
| 1134 | A | CACCAAGGAAGCCTTAGATA | TGCGGCCGCA | AGATAAGATAGAGGAAGAGC | TKEALDMRPQDKIEEE |
| 1135 | A | ACCAAGGAAGCCTTAGATAA | TGCGGCCGCA | GATAAGATAGAGGAAGAGCA | TKEALDNAAADKIEEE |
| 1136 | G | CCAAGGAAGCCTTAGATAAG | TGCGGCCGCA | ATAAGATAGAGGAAGAGCAA | KEALDKCGRNKIEEEQ |
| 1137 | A | CAAGGAAGCCTTAGATAAGA | TGCGGCCGCA | TAAGATAGAGGAAGAGCAAA | KEALDKMRPHKIEEEQ |
| 1138 | T | AAGGAAGCCTTAGATAAGAT | TGCGGCCGCA | AAGATAGAGGAAGAGCAAAA | KEALDKIAAAKIEEEQ |
| 1139 | A | AGGAAGCCTTAGATAAGATA | TGCGGCCGCA | AGATAGAGGAAGAGCAAAAC | EALDKICGRKIEEEQN |
| 1140 | G | GGAAGCCTTAGATAAGATAG | TGCGGCCGCA | GATAGAGGAAGAGCAAAACA | EALDKIVRPQIEEEQN |
| 1141 | A | GAAGCCTTAGATAAGATAGA | TGCGGCCGCA | ATAGAGGAAGAGCAAAACAA | EALDKIDAAAIEEEQN |
| 1142 | G | AAGCCTTAGATAAGATAGAG | TGCGGCCGCA | TAGAGGAAGAGCAAAACAAA | ALDKIECGRIEEEQNK |
| 1143 | G | AGCCTTAGATAAGATAGAGG | TGCGGCCGCA | AGAGGAAGAGCAAAACAAAA | ALDKIEVRPQEEEQNK |
| 1144 | A | GCCTTAGATAAGATAGAGGA | TGCGGCCGCA | GAGGAAGAGCAAAACAAAAG | ALDKIEDAAAEEEQNK |
| 1145 | A | CCTTAGATAAGATAGAGGAA | TGCGGCCGCA | AGGAAGAGCAAAACAAAAGT | LDKIEECGRKEEQNKS |
| 1146 | G | CTTAGATAAGATAGAGGAAG | TGCGGCCGCA | GGAAGAGCAAAACAAAAGTA | LDKIEEVRPQEEQNKS |
| 1147 | A | TTAGATAAGATAGAGGAAGA | TGCGGCCGCA | GAAGAGCAAAACAAAAGTAA | LDKIEEDAAAEEQNKS |
| 1148 | G | TAGATAAGATAGAGGAAGAG | TGCGGCCGCA | AAGAGCAAAACAAAAGTAAG | DKIEEECGRKEQNKSK |
| 1149 | C | AGATAAGATAGAGGAAGAGC | TGCGGCCGCA | AGAGCAAAACAAAAGTAAGA | DKIEEELRPQEQNKSK |
| 1150 | A | GATAAGATAGAGGAAGAGCA | TGCGGCCGCA | GAGCAAAACAAAAGTAAGAA | DKIEEEHAAAEQNKSK |
| 1151 | A | ATAAGATAGAGGAAGAGCAA | TGCGGCCGCA | AGCAAAACAAAAGTAAGAAA | KIEEEQCGRKQNKSKK |
| 1152 | A | TAAGATAGAGGAAGAGCAAA | TGCGGCCGCA | GCAAAACAAAAGTAAGAAAA | KIEEEQMRPQQNKSKK |
| 1153 | A | AAGATAGAGGAAGAGCAAAA | TGCGGCCGCA | CAAAACAAAAGTAAGAAAAA | KIEEEQNAAAQNKSKK |
| 1154 | C | AGATAGAGGAAGAGCAAAAC | TGCGGCCGCA | AAAACAAAAGTAAGAAAAAG | IEEEQNCGRKNKSKKK |
| 1155 | A | GATAGAGGAAGAGCAAAACA | TGCGGCCGCA | AAACAAAAGTAAGAAAAAGG | IEEEQNMRPQNKSKKK |
| 1156 | A | ATAGAGGAAGAGCAAAACAA | TGCGGCCGCA | AACAAAAGTAAGAAAAAGGC | IEEEQNNAAANKSKKK |
| 1157 | A | TAGAGGAAGAGCAAAACAAA | TGCGGCCGCA | ACAAAAGTAAGAAAAAGGCA | EEEQNKCGRNKSKKKA |
| 1158 | A | AGAGGAAGAGCAAAACAAAA | TGCGGCCGCA | CAAAAGTAAGAAAAAGGCAC | EEEQNKMRPHKSKKKA |
| 1159 | G | GAGGAAGAGCAAAACAAAAG | TGCGGCCGCA | AAAAGTAAGAAAAAGGCACA | EEEQNKSAAAKSKKKA |
| 1160 | T | AGGAAGAGCAAAACAAAAGT | TGCGGCCGCA | AAAGTAAGAAAAAGGCACAG | EEQNKSCGRKSKKKAQ |
| 1161 | A | GGAAGAGCAAAACAAAAGTA | TGCGGCCGCA | AAGTAAGAAAAAGGCACAGC | EEQNKSMRPQSKKKAQ |
| 1162 | A | GAAGAGCAAAACAAAAGTAA | TGCGGCCGCA | AGTAAGAAAAAGGCACAGCA | EEQNKSNAAASKKKAQ |
| 1163 | G | AAGAGCAAAACAAAAGTAAG | TGCGGCCGCA | GTAAGAAAAAGGCACAGCAA | EQNKSKCGRSKKKAQQ |
| 1164 | A | AGAGCAAAACAAAAGTAAGA | TGCGGCCGCA | TAAGAAAAAGGCACAGCAAG | EQNKSKMRPHKKKAQQ |
| 1165 | A | GAGCAAAACAAAAGTAAGAA | TGCGGCCGCA | AAGAAAAAGGCACAGCAAGC | EQNKSKNAAAKKKAQQ |
| 1166 | A | AGCAAAACAAAAGTAAGAAA | TGCGGCCGCA | AGAAAAAGGCACAGCAAGCA | QNKSKKCGRKKKAQQA |
| 1167 | A | GCAAAACAAAAGTAAGAAAA | TGCGGCCGCA | GAAAAAGGCACAGCAAGCAG | QNKSKKMRPQKKAQQA |
| 1168 | A | CAAAACAAAAGTAAGAAAAA | TGCGGCCGCA | AAAAAGGCACAGCAAGCAGC | QNKSKKNAAAKKAQQA |
| 1169 | G | AAAACAAAAGTAAGAAAAAG | TGCGGCCGCA | AAAAGGCACAGCAAGCAGCA | NKSKKKCGRKKAQQAA |
| 1170 | G | AAACAAAAGTAAGAAAAAGG | TGCGGCCGCA | AAAGGCACAGCAAGCAGCAG | NKSKKKVRPQKAQQAA |
| 1171 | C | AACAAAAGTAAGAAAAAGGC | TGCGGCCGCA | AAGGCACAGCAAGCAGCAGC | NKSKKKAAAAKAQQAA |
| 1172 | A | ACAAAAGTAAGAAAAAGGCA | TGCGGCCGCA | AGGCACAGCAAGCAGCAGCT | KSKKKACGRKAQQAAA |
| 1173 | C | CAAAAGTAAGAAAAAGGCAC | TGCGGCCGCA | GGCACAGCAAGCAGCAGCTG | KSKKKALRPQAQQAAA |
| 1174 | A | AAAAGTAAGAAAAAGGCACA | TGCGGCCGCA | GCACAGCAAGCAGCAGCTGA | KSKKKAHAAAAQQAAA |
| 1175 | G | AAAGTAAGAAAAAGGCACAG | TGCGGCCGCA | CACAGCAAGCAGCAGCTGAC | SKKKAQCGRTQQAAAD |
| 1176 | C | AAGTAAGAAAAAGGCACAGC | TGCGGCCGCA | ACAGCAAGCAGCAGCTGACA | SKKKAQLRPQQQAAAD |
| 1177 | A | AGTAAGAAAAAGGCACAGCA | TGCGGCCGCA | CAGCAAGCAGCAGCTGACAC | SKKKAQHAAAQQAAAD |
| 1178 | A | GTAAGAAAAAGGCACAGCAA | TGCGGCCGCA | AGCAAGCAGCAGCTGACACA | KKKAQQCGRKQAAADT |
| 1179 | G | TAAGAAAAAGGCACAGCAAG | TGCGGCCGCA | GCAAGCAGCAGCTGACACAG | KKKAQQVRPQQAAADT |
| 1180 | C | AAGAAAAAGGCACAGCAAGC | TGCGGCCGCA | CAAGCAGCAGCTGACACAGG | KKKAQQAAAAQAAADT |
| 1181 | A | AGAAAAAGGCACAGCAAGCA | TGCGGCCGCA | AAGCAGCAGCTGACACAGGA | KKAQQACGRKAAADTG |
| 1182 | G | GAAAAAGGCACAGCAAGCAG | TGCGGCCGCA | AGCAGCAGCTGACACAGGAA | KKAQQAVRPQAAADTG |
| 1183 | C | AAAAAGGCACAGCAAGCAGC | TGCGGCCGCA | GCAGCAGCTGACACAGGAAA | KKAQQAAAAAAAADTG |
| 1184 | A | AAAAGGCACAGCAAGCAGCA | TGCGGCCGCA | CAGCAGCTGACACAGGAAAC | KAQQAACGRTAADTGN |
| 1185 | G | AAAGGCACAGCAAGCAGCAG | TGCGGCCGCA | AGCAGCTGACACAGGAAACA | KAQQAAVRPQAADTGN |
| 1186 | C | AAGGCACAGCAAGCAGCAGC | TGCGGCCGCA | GCAGCTGACACAGGAAACAA | KAQQAAAAAAAADTGN |
| 1187 | T | AGGCACAGCAAGCAGCAGCT | TGCGGCCGCA | CAGCTGACACAGGAAACAAC | AQQAAACGRTADTGNN |
| 1188 | G | GGCACAGCAAGCAGCAGCTG | TGCGGCCGCA | AGCTGACACAGGAAACAACA | AQQAAAVRPQADTGNN |
| 1189 | A | GCACAGCAAGCAGCAGCTGA | TGCGGCCGCA | GCTGACACAGGAAACAACAG | AQQAAADAAAADTGNN |
| 1190 | C | CACAGCAAGCAGCAGCTGAC | TGCGGCCGCA | CTGACACAGGAAACAACAGC | QQAAADCGRTDTGNNS |
| 1191 | A | ACAGCAAGCAGCAGCTGACA | TGCGGCCGCA | TGACACAGGAAACAACAGCC | QQAAADMRPHDTGNNS |
| 1192 | C | CAGCAAGCAGCAGCTGACAC | TGCGGCCGCA | GACACAGGAAACAACAGCCA | QQAAADTAAADTGNNS |
| 1193 | A | AGCAAGCAGCAGCTGACACA | TGCGGCCGCA | ACACAGGAAACAACAGCCAG | QAAADTCGRNTGNNSQ |
| 1194 | G | GCAAGCAGCAGCTGACACAG | TGCGGCCGCA | CACAGGAAACAACAGCCAGG | QAAADTVRPHTGNNSQ |
| 1195 | G | CAAGCAGCAGCTGACACAGG | TGCGGCCGCA | ACAGGAAACAACAGCCAGGT | QAAADTGAAATGNNSQ |
| 1196 | A | AAGCAGCAGCTGACACAGGA | TGCGGCCGCA | CAGGAAACAACAGCCAGGTC | AAADTGCGRTGNNSQV |
| 1197 | A | AGCAGCAGCTGACACAGGAA | TGCGGCCGCA | AGGAAACAACAGCCAGGTCA | AAADTGMRPQGNNSQV |
| 1198 | A | GCAGCAGCTGACACAGGAAA | TGCGGCCGCA | GGAAACAACAGCCAGGTCAG | AAADTGNAAAGNNSQV |

| | | | | | |
|---|---|---|---|---|---|
| 1199 | C | CAGCAGCTGACACAGGAAAC | TGCGGCCGCA | GAAACAACAGCCAGGTCAGC | AADTGNCGRRNNSQVS |
| 1200 | A | AGCAGCTGACACAGGAAACA | TGCGGCCGCA | AAACAACAGCCAGGTCAGCC | AADTGNMRPQNNSQVS |
| 1201 | A | GCAGCTGACACAGGAAACAA | TGCGGCCGCA | AACAACAGCCAGGTCAGCCA | AADTGNNAAANNSQVS |
| 1202 | C | CAGCTGACACAGGAAACAAC | TGCGGCCGCA | ACAACAGCCAGGTCAGCCAA | ADTGNNCGRNNSQVSQ |
| 1203 | A | AGCTGACACAGGAAACAACA | TGCGGCCGCA | CAACAGCCAGGTCAGCCAAA | ADTGNNMRPHNSQVSQ |
| 1204 | G | GCTGACACAGGAAACAACAG | TGCGGCCGCA | AACAGCCAGGTCAGCCAAAA | ADTGNNSAAANSQVSQ |
| 1205 | C | CTGACACAGGAAACAACAGC | TGCGGCCGCA | ACAGCCAGGTCAGCCAAAAT | DTGNNSCGRNSQVSQN |
| 1206 | C | TGACACAGGAAACAACAGCC | TGCGGCCGCA | CAGCCAGGTCAGCCAAAATT | DTGNNSLRPHSQVSQN |
| 1207 | A | GACACAGGAAACAACAGCCA | TGCGGCCGCA | AGCCAGGTCAGCCAAAATTA | DTGNNSHAAASQVSQN |
| 1208 | G | ACACAGGAAACAACAGCCAG | TGCGGCCGCA | GCCAGGTCAGCCAAAATTAC | TGNNSQCGRSQVSQNY |
| 1209 | G | CACAGGAAACAACAGCCAGG | TGCGGCCGCA | CCAGGTCAGCCAAAATTACC | TGNNSQVRPHQVSQNY |
| 1210 | T | ACAGGAAACAACAGCCAGGT | TGCGGCCGCA | CAGGTCAGCCAAAATTACCC | TGNNSQVAAAQVSQNY |
| 1211 | C | CAGGAAACAACAGCCAGGTC | TGCGGCCGCA | AGGTCAGCCAAAATTACCCT | GNNSQVCGRKVSQNYP |
| 1212 | A | AGGAAACAACAGCCAGGTCA | TGCGGCCGCA | GGTCAGCCAAAATTACCCTA | GNNSQVMRPQVSQNYP |
| 1213 | G | GGAAACAACAGCCAGGTCAG | TGCGGCCGCA | GTCAGCCAAAATTACCCTAT | GNNSQVSAAAVSQNYP |
| 1214 | C | GAAACAACAGCCAGGTCAGC | TGCGGCCGCA | TCAGCCAAAATTACCCTATA | NNSQVSCGRISQNYPI |
| 1215 | C | AAACAACAGCCAGGTCAGCC | TGCGGCCGCA | CAGCCAAAATTACCCTATAG | NNSQVSLRPHSQNYPI |
| 1216 | A | AACAACAGCCAGGTCAGCCA | TGCGGCCGCA | AGCCAAAATTACCCTATAGT | NNSQVSHAAASQNYPI |
| 1217 | A | ACAACAGCCAGGTCAGCCAA | TGCGGCCGCA | GCCAAAATTACCCTATAGTC | NSQVSQCGRSQNYPIV |
| 1218 | A | CAACAGCCAGGTCAGCCAAA | TGCGGCCGCA | CCAAAATTACCCTATAGTCC | NSQVSQMRPHQNYPIV |
| 1219 | A | AACAGCCAGGTCAGCCAAAA | TGCGGCCGCA | CAAAATTACCCTATAGTCCA | NSQVSQNAAAQNYPIV |
| 1220 | T | ACAGCCAGGTCAGCCAAAAT | TGCGGCCGCA | AAAATTACCCTATAGTCCAG | SQVSQNCGRKNYPIVQ |
| 1221 | T | CAGCCAGGTCAGCCAAAATT | TGCGGCCGCA | AAATTACCCTATAGTCCAGA | SQVSQNLRPQNYPIVQ |
| 1222 | A | AGCCAGGTCAGCCAAAATTA | TGCGGCCGCA | AATTACCCTATAGTCCAGAA | SQVSQNYAAANYPIVQ |
| 1223 | C | GCCAGGTCAGCCAAAATTAC | TGCGGCCGCA | ATTACCCTATAGTCCAGAAC | QVSQNYCGRNYPIVQN |
| 1224 | C | CCAGGTCAGCCAAAATTACC | TGCGGCCGCA | TTACCCTATAGTCCAGAACC | QVSQNYLRPHYPIVQN |
| 1225 | C | CAGGTCAGCCAAAATTACCC | TGCGGCCGCA | TACCCTATAGTCCAGAACCT | QVSQNYPAAAYPIVQN |
| 1226 | T | AGGTCAGCCAAAATTACCCT | TGCGGCCGCA | ACCCTATAGTCCAGAACCTC | VSQNYPCGRNPIVQNL |
| 1227 | A | GGTCAGCCAAAATTACCCTA | TGCGGCCGCA | CCCTATAGTCCAGAACCTCC | VSQNYPMRPHPIVQNL |
| 1228 | T | GTCAGCCAAAATTACCCTAT | TGCGGCCGCA | CCTATAGTCCAGAACCTCCA | VSQNYPIAAAPIVQNL |
| 1229 | A | TCAGCCAAAATTACCCTATA | TGCGGCCGCA | CTATAGTCCAGAACCTCCAG | SQNYPICGRTIVQNLQ |
| 1230 | G | CAGCCAAAATTACCCTATAG | TGCGGCCGCA | TATAGTCCAGAACCTCCAGG | SQNYPIVRPHIVQNLQ |
| 1231 | T | AGCCAAAATTACCCTATAGT | TGCGGCCGCA | ATAGTCCAGAACCTCCAGGG | SQNYPIVAAAIVQNLQ |
| 1232 | C | GCCAAAATTACCCTATAGTC | TGCGGCCGCA | TAGTCCAGAACCTCCAGGGG | QNYPIVCGRIVQNLQG |
| 1233 | C | CCAAAATTACCCTATAGTCC | TGCGGCCGCA | AGTCCAGAACCTCCAGGGGC | QNYPIVLRPQVQNLQG |
| 1234 | A | CAAAATTACCCTATAGTCCA | TGCGGCCGCA | GTCCAGAACCTCCAGGGGCA | QNYPIVHAAAVQNLQG |
| 1235 | G | AAAATTACCCTATAGTCCAG | TGCGGCCGCA | TCCAGAACCTCCAGGGGCAA | NYPIVQCGRIQNLQGQ |
| 1236 | A | AAATTACCCTATAGTCCAGA | TGCGGCCGCA | CCAGAACCTCCAGGGGCAAA | NYPIVQMRPHQNLQGQ |
| 1237 | A | AATTACCCTATAGTCCAGAA | TGCGGCCGCA | CAGAACCTCCAGGGGCAAAT | NYPIVQNAAAQNLQGQ |
| 1238 | C | ATTACCCTATAGTCCAGAAC | TGCGGCCGCA | AGAACCTCCAGGGGCAAATG | YPIVQNCGRKNLQGQM |
| 1239 | C | TTACCCTATAGTCCAGAACC | TGCGGCCGCA | GAACCTCCAGGGGCAAATGG | YPIVQNLRPQNLQGQM |
| 1240 | T | TACCCTATAGTCCAGAACCT | TGCGGCCGCA | AACCTCCAGGGGCAAATGGT | YPIVQNLAAANLQGQM |
| 1241 | C | ACCCTATAGTCCAGAACCTC | TGCGGCCGCA | ACCTCCAGGGGCAAATGGTA | PIVQNLCGRNLQGQMV |
| 1242 | C | CCCTATAGTCCAGAACCTCC | TGCGGCCGCA | CCTCCAGGGGCAAATGGTAC | PIVQNLLRPHLQGQMV |
| 1243 | A | CCTATAGTCCAGAACCTCCA | TGCGGCCGCA | CTCCAGGGGCAAATGGTACA | PIVQNLHAAALQGQMV |
| 1244 | G | CTATAGTCCAGAACCTCCAG | TGCGGCCGCA | TCCAGGGGCAAATGGTACAT | IVQNLQCGRIQGQMVH |
| 1245 | G | TATAGTCCAGAACCTCCAGG | TGCGGCCGCA | CCAGGGGCAAATGGTACATC | IVQNLQVRPHQGQMVH |
| 1246 | G | ATAGTCCAGAACCTCCAGGG | TGCGGCCGCA | CAGGGGCAAATGGTACATCA | IVQNLQGAAAQGQMVH |
| 1247 | G | TAGTCCAGAACCTCCAGGGG | TGCGGCCGCA | AGGGGCAAATGGTACATCAG | VQNLQGCGRKGQMVHQ |
| 1248 | C | AGTCCAGAACCTCCAGGGGC | TGCGGCCGCA | GGGGCAAATGGTACATCAGG | VQNLQGLRPQGQMVHQ |
| 1249 | A | GTCCAGAACCTCCAGGGGCA | TGCGGCCGCA | GGGCAAATGGTACATCAGGC | VQNLQGHAAAGQMVHQ |
| 1250 | A | TCCAGAACCTCCAGGGGCAA | TGCGGCCGCA | GGCAAATGGTACATCAGGCC | QNLQGQCGRRQMVHQA |
| 1251 | A | CCAGAACCTCCAGGGGCAAA | TGCGGCCGCA | GCAAATGGTACATCAGGCCA | QNLQGQMRPQQMVHQA |
| 1252 | T | CAGAACCTCCAGGGGCAAAT | TGCGGCCGCA | CAAATGGTACATCAGGCCAT | QNLQGQIAAAQMVHQA |
| 1253 | G | AGAACCTCCAGGGGCAAATG | TGCGGCCGCA | AAATGGTACATCAGGCCATA | NLQGQMCGRKMVHQAI |
| 1254 | G | GAACCTCCAGGGGCAAATGG | TGCGGCCGCA | AATGGTACATCAGGCCATAT | NLQGQMVRPQMVHQAI |
| 1255 | T | AACCTCCAGGGGCAAATGGT | TGCGGCCGCA | ATGGTACATCAGGCCATATC | NLQGQMVAAAMVHQAI |
| 1256 | A | ACCTCCAGGGGCAAATGGTA | TGCGGCCGCA | TGGTACATCAGGCCATATCA | LQGQMVCGRMVHQAIS |
| 1257 | C | CCTCCAGGGGCAAATGGTAC | TGCGGCCGCA | GGTACATCAGGCCATATCAC | LQGQMVLRPQVHQAIS |
| 1258 | A | CTCCAGGGGCAAATGGTACA | TGCGGCCGCA | GTACATCAGGCCATATCACC | LQGQMVHAAAVHQAIS |
| 1259 | T | TCCAGGGGCAAATGGTACAT | TGCGGCCGCA | TACATCAGGCCATATCACCT | QGQMVHCGRIHQAISP |
| 1260 | C | CCAGGGGCAAATGGTACATC | TGCGGCCGCA | ACATCAGGCCATATCACCTA | QGQMVHLRPQHQAISP |
| 1261 | A | CAGGGGCAAATGGTACATCA | TGCGGCCGCA | CATCAGGCCATATCACCTAG | QGQMVHHAAAHQAISP |
| 1262 | G | AGGGGCAAATGGTACATCAG | TGCGGCCGCA | ATCAGGCCATATCACCTAGA | GQMVHQCGRNQAISPR |
| 1263 | G | GGGGCAAATGGTACATCAGG | TGCGGCCGCA | TCAGGCCATATCACCTAGAA | GQMVHQVRPHQAISPR |
| 1264 | C | GGGCAAATGGTACATCAGGC | TGCGGCCGCA | CAGGCCATATCACCTAGAAC | GQMVHQAAAAQAISPR |
| 1265 | C | GGCAAATGGTACATCAGGCC | TGCGGCCGCA | AGGCCATATCACCTAGAACT | QMVHQACGRKAISPRT |
| 1266 | A | GCAAATGGTACATCAGGCCA | TGCGGCCGCA | GGCCATATCACCTAGAACTT | QMVHQAMRPQAISPRT |
| 1267 | T | CAAATGGTACATCAGGCCAT | TGCGGCCGCA | GCCATATCACCTAGAACTTT | QMVHQAIAAAAISPRT |
| 1268 | A | AAATGGTACATCAGGCCATA | TGCGGCCGCA | CCATATCACCTAGAACTTTA | MVHQAICGRTISPRTL |
| 1269 | T | AATGGTACATCAGGCCATAT | TGCGGCCGCA | CATATCACCTAGAACTTTAA | MVHQAILRPHISPRTL |
| 1270 | C | ATGGTACATCAGGCCATATC | TGCGGCCGCA | ATATCACCTAGAACTTTAAA | MVHQAISAAAISPRTL |
| 1271 | A | TGGTACATCAGGCCATATCA | TGCGGCCGCA | TATCACCTAGAACTTTAAAT | VHQAISCGRISPRTLN |
| 1272 | C | GGTACATCAGGCCATATCAC | TGCGGCCGCA | ATCACCTAGAACTTTAAATG | VHQAISLRPQSPRTLN |
| 1273 | C | GTACATCAGGCCATATCACC | TGCGGCCGCA | TCACCTAGAACTTTAAATGC | VHQAISPAAASPRTLN |
| 1274 | T | TACATCAGGCCATATCACCT | TGCGGCCGCA | CACCTAGAACTTTAAATGCA | HQAISPCGRTPRTLNA |
| 1275 | A | ACATCAGGCCATATCACCTA | TGCGGCCGCA | ACCTAGAACTTTAAATGCAT | HQAISPMRPQPRTLNA |
| 1276 | G | CATCAGGCCATATCACCTAG | TGCGGCCGCA | CCTAGAACTTTAAATGCATG | HQAISPSAAAPRTLNA |
| 1277 | A | ATCAGGCCATATCACCTAGA | TGCGGCCGCA | CTAGAACTTTAAATGCATGG | QAISPRCGRTRTLNAW |
| 1278 | A | TCAGGCCATATCACCTAGAA | TGCGGCCGCA | TAGAACTTTAAATGCATGGG | QAISPRMRPHRTLNAW |
| 1279 | C | CAGGCCATATCACCTAGAAC | TGCGGCCGCA | AGAACTTTAAATGCATGGGT | QAISPRTAAARTLNAW |
| 1280 | T | AGGCCATATCACCTAGAACT | TGCGGCCGCA | GAACTTTAAATGCATGGGTA | AISPRTCGRRTLNAWV |
| 1281 | T | GGCCATATCACCTAGAACTT | TGCGGCCGCA | AACTTTAAATGCATGGGTAA | AISPRTLRPQTLNAWV |
| 1282 | T | GCCATATCACCTAGAACTTT | TGCGGCCGCA | ACTTTAAATGCATGGGTAAA | AISPRTFAAATLNAWV |
| 1283 | A | CCATATCACCTAGAACTTTA | TGCGGCCGCA | CTTTAAATGCATGGGTAAAA | ISPRTLCGRTLNAWVK |
| 1284 | A | CATATCACCTAGAACTTTAA | TGCGGCCGCA | TTTAAATGCATGGGTAAAAG | ISPRTLMRPHLNAWVK |

| | | | | | |
|---|---|---|---|---|---|
| 1285 | A | ATATCACCTAGAACTTTAAA | TGCGGCCGCA | TTAAATGCATGGGTAAAAGT | ISPRTLNAAALNAWVK |
| 1286 | T | TATCACCTAGAACTTTAAAT | TGCGGCCGCA | TAAATGCATGGGTAAAAGTA | SPRTLNCGRINAWVKV |
| 1287 | G | ATCACCTAGAACTTTAAATG | TGCGGCCGCA | AAATGCATGGGTAAAAGTAG | SPRTLNVRPQNAWVKV |
| 1288 | C | TCACCTAGAACTTTAAATGC | TGCGGCCGCA | AATGCATGGGTAAAAGTAGT | SPRTLNAAAANAWVKV |
| 1289 | A | CACCTAGAACTTTAAATGCA | TGCGGCCGCA | ATGCATGGGTAAAAGTAGTA | PRTLNACGRNAWVKVV |
| 1290 | T | ACCTAGAACTTTAAATGCAT | TGCGGCCGCA | TGCATGGGTAAAAGTAGTAG | PRTLNALRPHAWVKVV |
| 1291 | G | CCTAGAACTTTAAATGCATG | TGCGGCCGCA | GCATGGGTAAAAGTAGTAGA | PRTLNACAAAAWVKVV |
| 1292 | G | CTAGAACTTTAAATGCATGG | TGCGGCCGCA | CATGGGTAAAAGTAGTAGAA | RTLNAWCGRTWVKVVE |
| 1293 | G | TAGAACTTTAAATGCATGGG | TGCGGCCGCA | ATGGGTAAAAGTAGTAGAAG | RTLNAWVRPQWVKVVE |
| 1294 | T | AGAACTTTAAATGCATGGGT | TGCGGCCGCA | TGGGTAAAAGTAGTAGAAGA | RTLNAWVAAAWVKVVE |
| 1295 | A | GAACTTTAAATGCATGGGTA | TGCGGCCGCA | GGGTAAAAGTAGTAGAAGAG | TLNAWVCGRRVKVVEE |
| 1296 | A | AACTTTAAATGCATGGGTAA | TGCGGCCGCA | GGTAAAAGTAGTAGAAGAGA | TLNAWVMRPQVKVVEE |
| 1297 | A | ACTTTAAATGCATGGGTAAA | TGCGGCCGCA | GTAAAAGTAGTAGAAGAGAA | TLNAWVNAAAVKVVEE |
| 1298 | A | CTTTAAATGCATGGGTAAAA | TGCGGCCGCA | TAAAAGTAGTAGAAGAGAAG | LNAWVKCGRIKVVEEK |
| 1299 | G | TTTAAATGCATGGGTAAAAG | TGCGGCCGCA | AAAAGTAGTAGAAGAGAAGG | LNAWVKVRPQKVVEEK |
| 1300 | T | TTAAATGCATGGGTAAAAGT | TGCGGCCGCA | AAAGTAGTAGAAGAGAAGGC | LNAWVKVAAAKVVEEK |
| 1301 | A | TAAATGCATGGGTAAAAGTA | TGCGGCCGCA | AAGTAGTAGAAGAGAAGGCT | NAWVKVCGRKVVEEKA |
| 1302 | G | AAATGCATGGGTAAAAGTAG | TGCGGCCGCA | AGTAGTAGAAGAGAAGGCTT | NAWVKVVRPQVVEEKA |
| 1303 | T | AATGCATGGGTAAAAGTAGT | TGCGGCCGCA | GTAGTAGAAGAGAAGGCTTT | NAWVKVVAAAVVEEKA |
| 1304 | A | ATGCATGGGTAAAAGTAGTA | TGCGGCCGCA | TAGTAGAAGAGAAGGCTTTC | AWVKVVCGRIVEEKAF |
| 1305 | G | TGCATGGGTAAAAGTAGTAG | TGCGGCCGCA | AGTAGAAGAGAAGGCTTTCA | AWVKVVVRPQVEEKAF |
| 1306 | A | GCATGGGTAAAAGTAGTAGA | TGCGGCCGCA | GTAGAAGAGAAGGCTTTCAG | AWVKVVDAAAVEEKAF |
| 1307 | A | CATGGGTAAAAGTAGTAGAA | TGCGGCCGCA | TAGAAGAGAAGGCTTTCAGC | WVKVVECGRIEEKAFS |
| 1308 | G | ATGGGTAAAAGTAGTAGAAG | TGCGGCCGCA | AGAAGAGAAGGCTTTCAGCC | WVKVVEVRPQEEKAFS |
| 1309 | A | TGGGTAAAAGTAGTAGAAGA | TGCGGCCGCA | GAAGAGAAGGCTTTCAGCCC | WVKVVEDAAAEEKAFS |
| 1310 | G | GGGTAAAAGTAGTAGAAGAG | TGCGGCCGCA | AAGAGAAGGCTTTCAGCCCA | VKVVEECGRKEKAFSP |
| 1311 | A | GGTAAAAGTAGTAGAAGAGA | TGCGGCCGCA | AGAGAAGGCTTTCAGCCCAG | VKVVEEMRPQEKAFSP |
| 1312 | A | GTAAAAGTAGTAGAAGAGAA | TGCGGCCGCA | GAGAAGGCTTTCAGCCCAGA | VKVVEENAAAEKAFSP |
| 1313 | G | TAAAAGTAGTAGAAGAGAAG | TGCGGCCGCA | AGAAGGCTTTCAGCCCAGAA | KVVEEKCGRKKAFSPE |
| 1314 | G | AAAAGTAGTAGAAGAGAAGG | TGCGGCCGCA | GAAGGCTTTCAGCCCAGAAG | KVVEEKVRPQKAFSPE |
| 1315 | C | AAAGTAGTAGAAGAGAAGGC | TGCGGCCGCA | AAGGCTTTCAGCCCAGAAGT | KVVEEKAAAAKAFSPE |
| 1316 | T | AAGTAGTAGAAGAGAAGGCT | TGCGGCCGCA | AGGCTTTCAGCCCAGAAGTA | VVEEKACGRKAFSPEV |
| 1317 | T | AGTAGTAGAAGAGAAGGCTT | TGCGGCCGCA | GGCTTTCAGCCCAGAAGTAA | VVEEKALRPQAFSPEV |
| 1318 | T | GTAGTAGAAGAGAAGGCTTT | TGCGGCCGCA | GCTTTCAGCCCAGAAGTAAT | VVEEKAFAAAAFSPEV |
| 1319 | C | TAGTAGAAGAGAAGGCTTTC | TGCGGCCGCA | CTTTCAGCCCAGAAGTAATA | VEEKAFCGRTFSPEVI |
| 1320 | A | AGTAGAAGAGAAGGCTTTCA | TGCGGCCGCA | TTTCAGCCCAGAAGTAATAC | VEEKAFMRPHFSPEVI |
| 1321 | G | GTAGAAGAGAAGGCTTTCAG | TGCGGCCGCA | TTCAGCCCAGAAGTAATACC | VEEKAFSAAAFSPEVI |
| 1322 | C | TAGAAGAGAAGGCTTTCAGC | TGCGGCCGCA | TCAGCCCAGAAGTAATACCC | EEKAFSCGRISPEVIP |
| 1323 | C | AGAAGAGAAGGCTTTCAGCC | TGCGGCCGCA | CAGCCCAGAAGTAATACCCA | EEKAFSLRPHSPEVIP |
| 1324 | C | GAAGAGAAGGCTTTCAGCCC | TGCGGCCGCA | AGCCCAGAAGTAATACCCAT | EEKAFSPAAASPEVIP |
| 1325 | A | AAGAGAAGGCTTTCAGCCCA | TGCGGCCGCA | GCCCAGAAGTAATACCCATG | EKAFSPCGRSPEVIPM |
| 1326 | G | AGAGAAGGCTTTCAGCCCAG | TGCGGCCGCA | CCCAGAAGTAATACCCATGT | EKAFSPVRPHPEVIPM |
| 1327 | A | GAGAAGGCTTTCAGCCCAGA | TGCGGCCGCA | CCAGAAGTAATACCCATGTT | EKAFSPDAAAPEVIPM |
| 1328 | A | AGAAGGCTTTCAGCCCAGAA | TGCGGCCGCA | CAGAAGTAATACCCATGTTT | KAFSPECGRTEVIPMF |
| 1329 | G | GAAGGCTTTCAGCCCAGAAG | TGCGGCCGCA | AGAAGTAATACCCATGTTTT | KAFSPEVRPQEVIPMF |
| 1330 | T | AAGGCTTTCAGCCCAGAAGT | TGCGGCCGCA | GAAGTAATACCCATGTTTTC | KAFSPEVAAAEVIPMF |
| 1331 | A | AGGCTTTCAGCCCAGAAGTA | TGCGGCCGCA | AAGTAATACCCATGTTTTCA | AFSPEVCGRKVIPMFS |
| 1332 | A | GGCTTTCAGCCCAGAAGTAA | TGCGGCCGCA | AGTAATACCCATGTTTTCAG | AFSPEVMRPQVIPMFS |
| 1333 | T | GCTTTCAGCCCAGAAGTAAT | TGCGGCCGCA | GTAATACCCATGTTTTCAGC | AFSPEVIAAAVIPMFS |
| 1334 | A | CTTTCAGCCCAGAAGTAATA | TGCGGCCGCA | TAATACCCATGTTTTCAGCA | FSPEVICGRIIPMFSA |
| 1335 | C | TTTCAGCCCAGAAGTAATAC | TGCGGCCGCA | AATACCCATGTTTTCAGCAT | FSPEVILRPQIPMFSA |
| 1336 | C | TTCAGCCCAGAAGTAATACC | TGCGGCCGCA | ATACCCATGTTTTCAGCATT | FSPEVIPAAAIPMFSA |
| 1337 | C | TCAGCCCAGAAGTAATACCC | TGCGGCCGCA | TACCCATGTTTTCAGCATTA | SPEVIPCGRIPMFSAL |
| 1338 | A | CAGCCCAGAAGTAATACCCA | TGCGGCCGCA | ACCCATGTTTTCAGCATTAT | SPEVIPMRPQPMFSAL |
| 1339 | T | AGCCCAGAAGTAATACCCAT | TGCGGCCGCA | CCCATGTTTTCAGCATTATC | SPEVIPIAAAPMFSAL |
| 1340 | G | GCCCAGAAGTAATACCCATG | TGCGGCCGCA | CCATGTTTTCAGCATTATCA | PEVIPMCGRTMFSALS |
| 1341 | T | CCCAGAAGTAATACCCATGT | TGCGGCCGCA | CATGTTTTCAGCATTATCAG | PEVIPMLRPHMFSALS |
| 1342 | T | CCAGAAGTAATACCCATGTT | TGCGGCCGCA | ATGTTTTCAGCATTATCAGA | PEVIPMFAAAMFSALS |
| 1343 | T | CAGAAGTAATACCCATGTTT | TGCGGCCGCA | TGTTTTCAGCATTATCAGAA | EVIPMFCGRMFSALSE |
| 1344 | T | AGAAGTAATACCCATGTTTT | TGCGGCCGCA | GTTTTCAGCATTATCAGAAG | EVIPMFLRPQFSALSE |
| 1345 | C | GAAGTAATACCCATGTTTTC | TGCGGCCGCA | TTTTCAGCATTATCAGAAGG | EVIPMFSAAAFSALSE |
| 1346 | A | AAGTAATACCCATGTTTTCA | TGCGGCCGCA | TTTCAGCATTATCAGAAGGA | VIPMFSCGRISALSEG |
| 1347 | G | AGTAATACCCATGTTTTCAG | TGCGGCCGCA | TTCAGCATTATCAGAAGGAG | VIPMFSVRPHSALSEG |
| 1348 | C | GTAATACCCATGTTTTCAGC | TGCGGCCGCA | TCAGCATTATCAGAAGGAGC | VIPMFSAAAASALSEG |
| 1349 | A | TAATACCCATGTTTTCAGCA | TGCGGCCGCA | CAGCATTATCAGAAGGAGCC | IPMFSACGRTALSEGA |
| 1350 | T | AATACCCATGTTTTCAGCAT | TGCGGCCGCA | AGCATTATCAGAAGGAGCCA | IPMFSALRPQALSEGA |
| 1351 | T | ATACCCATGTTTTCAGCATT | TGCGGCCGCA | GCATTATCAGAAGGAGCCAC | IPMFSAFAAAALSEGA |
| 1352 | A | TACCCATGTTTTCAGCATTA | TGCGGCCGCA | CATTATCAGAAGGAGCCACC | PMFSALCGRTLSEGAT |
| 1353 | T | ACCCATGTTTTCAGCATTAT | TGCGGCCGCA | ATTATCAGAAGGAGCCACCC | PMFSALLRPQLSEGAT |
| 1354 | C | CCCATGTTTTCAGCATTATC | TGCGGCCGCA | TTATCAGAAGGAGCCACCCC | PMFSALSAAALSEGAT |
| 1355 | A | CCATGTTTTCAGCATTATCA | TGCGGCCGCA | TATCAGAAGGAGCCACCCCA | MFSALSCGRISEGATP |
| 1356 | G | CATGTTTTCAGCATTATCAG | TGCGGCCGCA | ATCAGAAGGAGCCACCCCAC | MFSALSVRPQSEGATP |
| 1357 | A | ATGTTTTCAGCATTATCAGA | TGCGGCCGCA | TCAGAAGGAGCCACCCCACA | MFSALSDAAASEGATP |
| 1358 | A | TGTTTTCAGCATTATCAGAA | TGCGGCCGCA | CAGAAGGAGCCACCCCACAA | FSALSECGRTEGATPQ |
| 1359 | G | GTTTTCAGCATTATCAGAAG | TGCGGCCGCA | AGAAGGAGCCACCCCACAAG | FSALSEVRPQEGATPQ |
| 1360 | G | TTTTCAGCATTATCAGAAGG | TGCGGCCGCA | GAAGGAGCCACCCCACAAGA | FSALSEGAAAEGATPQ |
| 1361 | A | TTTCAGCATTATCAGAAGGA | TGCGGCCGCA | AAGGAGCCACCCCACAAGAT | SALSEGCGRKGATPQD |
| 1362 | G | TTCAGCATTATCAGAAGGAG | TGCGGCCGCA | AGGAGCCACCCCACAAGATT | SALSEGVRPQGATPQD |
| 1363 | C | TCAGCATTATCAGAAGGAGC | TGCGGCCGCA | GGAGCCACCCCACAAGATTT | SALSEGAAAAGATPQD |
| 1364 | C | CAGCATTATCAGAAGGAGCC | TGCGGCCGCA | GAGCCACCCCACAAGATTTA | ALSEGACGRRATPQDL |
| 1365 | A | AGCATTATCAGAAGGAGCCA | TGCGGCCGCA | AGCCACCCCACAAGATTTAA | ALSEGAMRPQATPQDL |
| 1366 | C | GCATTATCAGAAGGAGCCAC | TGCGGCCGCA | GCCACCCCACAAGATTTAAA | ALSEGATAAAATPQDL |
| 1367 | C | CATTATCAGAAGGAGCCACC | TGCGGCCGCA | CCACCCCACAAGATTTAAAT | LSEGATCGRTTPQDLN |
| 1368 | C | ATTATCAGAAGGAGCCACCC | TGCGGCCGCA | CACCCCACAAGATTTAAATA | LSEGATLRPHTPQDLN |
| 1369 | C | TTATCAGAAGGAGCCACCCC | TGCGGCCGCA | ACCCCACAAGATTTAAATAC | LSEGATPAAATPQDLN |
| 1370 | A | TATCAGAAGGAGCCACCCCA | TGCGGCCGCA | CCCCACAAGATTTAAATACC | SEGATPCGRTPQDLNT |

| | | | | | |
|---|---|---|---|---|---|
| 1371 | C | ATCAGAAGGAGCCACCCCAC | TGCGGCCGCA | CCCACAAGATTTAAATACCA | SEGATPLRPHPQDLNT |
| 1372 | A | TCAGAAGGAGCCACCCCACA | TGCGGCCGCA | CCACAAGATTTAAATACCAT | SEGATPHAAAPQDLNT |
| 1373 | A | CAGAAGGAGCCACCCCACAA | TGCGGCCGCA | CACAAGATTTAAATACCATG | EGATPQCGRTQDLNTM |
| 1374 | G | AGAAGGAGCCACCCCACAAG | TGCGGCCGCA | ACAAGATTTAAATACCATGC | EGATPQVRPQQDLNTM |
| 1375 | A | GAAGGAGCCACCCCACAAGA | TGCGGCCGCA | CAAGATTTAAATACCATGCT | EGATPQDAAAQDLNTM |
| 1376 | T | AAGGAGCCACCCCACAAGAT | TGCGGCCGCA | AAGATTTAAATACCATGCTA | GATPQDCGRKDLNTML |
| 1377 | T | AGGAGCCACCCCACAAGATT | TGCGGCCGCA | AGATTTAAATACCATGCTAA | GATPQDLRPQDLNTML |
| 1378 | T | GGAGCCACCCCACAAGATTT | TGCGGCCGCA | GATTTAAATACCATGCTAAA | GATPQDFAAADLNTML |
| 1379 | A | GAGCCACCCCACAAGATTTA | TGCGGCCGCA | ATTTAAATACCATGCTAAAC | ATPQDLCGRNLNTMLN |
| 1380 | A | AGCCACCCCACAAGATTTAA | TGCGGCCGCA | TTTAAATACCATGCTAAACA | ATPQDLMRPHLNTMLN |
| 1381 | A | GCCACCCCACAAGATTTAAA | TGCGGCCGCA | TTAAATACCATGCTAAACAC | ATPQDLNAAALNTMLN |
| 1382 | T | CCACCCCACAAGATTTAAAT | TGCGGCCGCA | TAAATACCATGCTAAACACA | TPQDLNCGRINTMLNT |
| 1383 | A | CACCCCACAAGATTTAAATA | TGCGGCCGCA | AAATACCATGCTAAACACAG | TPQDLNMRPQNTMLNT |
| 1384 | C | ACCCCACAAGATTTAAATAC | TGCGGCCGCA | AATACCATGCTAAACACAGT | TPQDLNTAAANTMLNT |
| 1385 | C | CCCCACAAGATTTAAATACC | TGCGGCCGCA | ATACCATGCTAAACACAGTG | PQDLNTCGRNTMLNTV |
| 1386 | A | CCCACAAGATTTAAATACCA | TGCGGCCGCA | TACCATGCTAAACACAGTGG | PQDLNTMRPHTMLNTV |
| 1387 | T | CCACAAGATTTAAATACCAT | TGCGGCCGCA | ACCATGCTAAACACAGTGGG | PQDLNTIAAATMLNTV |
| 1388 | G | CACAAGATTTAAATACCATG | TGCGGCCGCA | CCATGCTAAACACAGTGGGG | QDLNTMCGRTMLNTVG |
| 1389 | C | ACAAGATTTAAATACCATGC | TGCGGCCGCA | CATGCTAAACACAGTGGGGG | QDLNTMLRPHMLNTVG |
| 1390 | T | CAAGATTTAAATACCATGCT | TGCGGCCGCA | ATGCTAAACACAGTGGGGGG | QDLNTMLAAAMLNTVG |
| 1391 | A | AAGATTTAAATACCATGCTA | TGCGGCCGCA | TGCTAAACACAGTGGGGGGA | DLNTMLCGRMLNTVGG |
| 1392 | A | AGATTTAAATACCATGCTAA | TGCGGCCGCA | GCTAAACACAGTGGGGGGAC | DLNTMLMRPQLNTVGG |
| 1393 | A | GATTTAAATACCATGCTAAA | TGCGGCCGCA | CTAAACACAGTGGGGGGACA | DLNTMLNAAALNTVGG |
| 1394 | C | ATTTAAATACCATGCTAAAC | TGCGGCCGCA | TAAACACAGTGGGGGGACAT | LNTMLNCGRINTVGGH |
| 1395 | A | TTTAAATACCATGCTAAACA | TGCGGCCGCA | AAACACAGTGGGGGGACATC | LNTMLNMRPQNTVGGH |
| 1396 | C | TTAAATACCATGCTAAACAC | TGCGGCCGCA | AACACAGTGGGGGGACATCA | LNTMLNTAAANTVGGH |
| 1397 | A | TAAATACCATGCTAAACACA | TGCGGCCGCA | ACACAGTGGGGGGACATCAA | NTMLNTCGRNTVGGHQ |
| 1398 | G | AAATACCATGCTAAACACAG | TGCGGCCGCA | CACAGTGGGGGGACATCAAG | NTMLNTVRPHTVGGHQ |
| 1399 | T | AATACCATGCTAAACACAGT | TGCGGCCGCA | ACAGTGGGGGGACATCAAGC | NTMLNTVAAATVGGHQ |
| 1400 | G | ATACCATGCTAAACACAGTG | TGCGGCCGCA | CAGTGGGGGGACATCAAGCA | TMLNTVCGRTVGGHQA |
| 1401 | G | TACCATGCTAAACACAGTGG | TGCGGCCGCA | AGTGGGGGGACATCAAGCAG | TMLNTVVRPQVGGHQA |
| 1402 | G | ACCATGCTAAACACAGTGGG | TGCGGCCGCA | GTGGGGGGACATCAAGCAGC | TMLNTVGAAAVGGHQA |
| 1403 | G | CCATGCTAAACACAGTGGGG | TGCGGCCGCA | TGGGGGGACATCAAGCAGCC | MLNTVGCGRMGGHQAA |
| 1404 | G | CATGCTAAACACAGTGGGGG | TGCGGCCGCA | GGGGGGACATCAAGCAGCCA | MLNTVGVRPQGGHQAA |
| 1405 | G | ATGCTAAACACAGTGGGGGG | TGCGGCCGCA | GGGGGACATCAAGCAGCCAT | MLNTVGGAAAGGHQAA |
| 1406 | A | TGCTAAACACAGTGGGGGGA | TGCGGCCGCA | GGGGACATCAAGCAGCCATG | LNTVGGCGRRGHQAAM |
| 1407 | C | GCTAAACACAGTGGGGGGAC | TGCGGCCGCA | GGGACATCAAGCAGCCATGC | LNTVGGLRPQGHQAAM |
| 1408 | A | CTAAACACAGTGGGGGGACA | TGCGGCCGCA | GGACATCAAGCAGCCATGCA | LNTVGGHAAAGHQAAM |
| 1409 | T | TAAACACAGTGGGGGGACAT | TGCGGCCGCA | GACATCAAGCAGCCATGCAA | NTVGGHCGRRHQAAMQ |
| 1410 | C | AAACACAGTGGGGGGACATC | TGCGGCCGCA | ACATCAAGCAGCCATGCAAA | NTVGGHLRPQHQAAMQ |
| 1411 | A | AACACAGTGGGGGGACATCA | TGCGGCCGCA | CATCAAGCAGCCATGCAAAT | NTVGGHHAAAHQAAMQ |
| 1412 | A | ACACAGTGGGGGGACATCAA | TGCGGCCGCA | ATCAAGCAGCCATGCAAATG | TVGGHQCGRNQAAMQM |
| 1413 | G | CACAGTGGGGGGACATCAAG | TGCGGCCGCA | TCAAGCAGCCATGCAAATGT | TVGGHQVRPHQAAMQM |
| 1414 | C | ACAGTGGGGGGACATCAAGC | TGCGGCCGCA | CAAGCAGCCATGCAAATGTT | TVGGHQAAAAQAAMQM |
| 1415 | A | CAGTGGGGGGACATCAAGCA | TGCGGCCGCA | AAGCAGCCATGCAAATGTTA | VGGHQACGRKAAMQML |
| 1416 | G | AGTGGGGGGACATCAAGCAG | TGCGGCCGCA | AGCAGCCATGCAAATGTTAA | VGGHQAVRPQAAMQML |
| 1417 | C | GTGGGGGGACATCAAGCAGC | TGCGGCCGCA | GCAGCCATGCAAATGTTAAA | VGGHQAAAAAAAMQML |
| 1418 | C | TGGGGGGACATCAAGCAGCC | TGCGGCCGCA | CAGCCATGCAAATGTTAAAA | GGHQAACGRTAMQMLK |
| 1419 | A | GGGGGGACATCAAGCAGCCA | TGCGGCCGCA | AGCCATGCAAATGTTAAAAG | GGHQAAMRPQAMQMLK |
| 1420 | T | GGGGGACATCAAGCAGCCAT | TGCGGCCGCA | GCCATGCAAATGTTAAAAGA | GGHQAAIAAAAMQMLK |
| 1421 | G | GGGGACATCAAGCAGCCATG | TGCGGCCGCA | CCATGCAAATGTTAAAAGAG | GHQAAMCGRTMQMLKE |
| 1422 | C | GGGACATCAAGCAGCCATGC | TGCGGCCGCA | CATGCAAATGTTAAAAGAGA | GHQAAMLRPHMQMLKE |
| 1423 | A | GGACATCAAGCAGCCATGCA | TGCGGCCGCA | ATGCAAATGTTAAAAGAGAC | GHQAAMHAAAMQMLKE |
| 1424 | A | GACATCAAGCAGCCATGCAA | TGCGGCCGCA | TGCAAATGTTAAAAGAGACC | HQAAMQCGRMQMLKET |
| 1425 | A | ACATCAAGCAGCCATGCAAA | TGCGGCCGCA | GCAAATGTTAAAAGAGACCA | HQAAMQMRPQQMLKET |
| 1426 | T | CATCAAGCAGCCATGCAAAT | TGCGGCCGCA | CAAATGTTAAAAGAGACCAT | HQAAMQIAAAQMLKET |
| 1427 | G | ATCAAGCAGCCATGCAAATG | TGCGGCCGCA | AAATGTTAAAAGAGACCATC | QAAMQMCGRKMLKETI |
| 1428 | T | TCAAGCAGCCATGCAAATGT | TGCGGCCGCA | AATGTTAAAAGAGACCATCA | QAAMQMLRPQMLKETI |
| 1429 | T | CAAGCAGCCATGCAAATGTT | TGCGGCCGCA | ATGTTAAAAGAGACCATCAA | QAAMQMFAAAMLKETI |
| 1430 | A | AAGCAGCCATGCAAATGTTA | TGCGGCCGCA | TGTTAAAAGAGACCATCAAT | AAMQMLCGRMLKETIN |
| 1431 | A | AGCAGCCATGCAAATGTTAA | TGCGGCCGCA | GTTAAAAGAGACCATCAATG | AAMQMLMRPQLKETIN |
| 1432 | A | GCAGCCATGCAAATGTTAAA | TGCGGCCGCA | TTAAAAGAGACCATCAATGA | AAMQMLNAAALKETIN |
| 1433 | A | CAGCCATGCAAATGTTAAAA | TGCGGCCGCA | TAAAAGAGACCATCAATGAG | AMQMLKCGRIKETINE |
| 1434 | G | AGCCATGCAAATGTTAAAAG | TGCGGCCGCA | AAAAGAGACCATCAATGAGG | AMQMLKVRPQKETINE |
| 1435 | A | GCCATGCAAATGTTAAAAGA | TGCGGCCGCA | AAAGAGACCATCAATGAGGA | AMQMLKDAAAKETINE |
| 1436 | G | CCATGCAAATGTTAAAAGAG | TGCGGCCGCA | AAGAGACCATCAATGAGGAA | MQMLKECGRKETINEE |
| 1437 | A | CATGCAAATGTTAAAAGAGA | TGCGGCCGCA | AGAGACCATCAATGAGGAAG | MQMLKEMRPQETINEE |
| 1438 | C | ATGCAAATGTTAAAAGAGAC | TGCGGCCGCA | GAGACCATCAATGAGGAAGC | MQMLKETAAAETINEE |
| 1439 | C | TGCAAATGTTAAAAGAGACC | TGCGGCCGCA | AGACCATCAATGAGGAAGCT | QMLKETCGRKTINEEA |
| 1440 | A | GCAAATGTTAAAAGAGACCA | TGCGGCCGCA | GACCATCAATGAGGAAGCTG | QMLKETMRPQTINEEA |
| 1441 | T | CAAATGTTAAAAGAGACCAT | TGCGGCCGCA | ACCATCAATGAGGAAGCTGC | QMLKETIAAATINEEA |
| 1442 | C | AAATGTTAAAAGAGACCATC | TGCGGCCGCA | CCATCAATGAGGAAGCTGCA | MLKETICGRTINEEAA |
| 1443 | A | AATGTTAAAAGAGACCATCA | TGCGGCCGCA | CATCAATGAGGAAGCTGCAG | MLKETIMRPHINEEAA |
| 1444 | A | ATGTTAAAAGAGACCATCAA | TGCGGCCGCA | ATCAATGAGGAAGCTGCAGA | MLKETINAAAINEEAA |
| 1445 | T | TGTTAAAAGAGACCATCAAT | TGCGGCCGCA | TCAATGAGGAAGCTGCAGAA | LKETINCGRINEEAAE |
| 1446 | G | GTTAAAAGAGACCATCAATG | TGCGGCCGCA | CAATGAGGAAGCTGCAGAAT | LKETINVRPHNEEAAE |
| 1447 | A | TTAAAAGAGACCATCAATGA | TGCGGCCGCA | AATGAGGAAGCTGCAGAATG | LKETINDAAANEEAAE |
| 1448 | G | TAAAAGAGACCATCAATGAG | TGCGGCCGCA | ATGAGGAAGCTGCAGAATGG | KETINECGRNEEAAEW |
| 1449 | G | AAAAGAGACCATCAATGAGG | TGCGGCCGCA | TGAGGAAGCTGCAGAATGGG | KETINEVRPHEEAAEW |
| 1450 | A | AAAGAGACCATCAATGAGGA | TGCGGCCGCA | GAGGAAGCTGCAGAATGGGA | KETINEDAAAEEAAEW |
| 1451 | A | AAGAGACCATCAATGAGGAA | TGCGGCCGCA | AGGAAGCTGCAGAATGGGAT | ETINEECGRKEAAEWD |
| 1452 | G | AGAGACCATCAATGAGGAAG | TGCGGCCGCA | GGAAGCTGCAGAATGGGATA | ETINEEVRPQEAAEWD |
| 1453 | C | GAGACCATCAATGAGGAAGC | TGCGGCCGCA | GAAGCTGCAGAATGGGATAG | ETINEEAAAAEAAEWD |
| 1454 | T | AGACCATCAATGAGGAAGCT | TGCGGCCGCA | AAGCTGCAGAATGGGATAGA | TINEEACGRKAAEWDR |
| 1455 | G | GACCATCAATGAGGAAGCTG | TGCGGCCGCA | AGCTGCAGAATGGGATAGAT | TINEEAVRPQAAEWDR |
| 1456 | C | ACCATCAATGAGGAAGCTGC | TGCGGCCGCA | GCTGCAGAATGGGATAGATT | TINEEAAAAAAAEWDR |

| | | | | | |
|---|---|---|---|---|---|
| 1457 | A | CCATCAATGAGGAAGCTGCA | TGCGGCCGCA | CTGCAGAATGGGATAGATTG | INEEAACGRTAEWDRL |
| 1458 | G | CATCAATGAGGAAGCTGCAG | TGCGGCCGCA | TGCAGAATGGGATAGATTGC | INEEAAVRPHAEWDRL |
| 1459 | A | ATCAATGAGGAAGCTGCAGA | TGCGGCCGCA | GCAGAATGGGATAGATTGCA | INEEAADAAAAEWDRL |
| 1460 | A | TCAATGAGGAAGCTGCAGAA | TGCGGCCGCA | CAGAATGGGATAGATTGCAT | NEEAAECGRTEWDRLH |
| 1461 | T | CAATGAGGAAGCTGCAGAAT | TGCGGCCGCA | AGAATGGGATAGATTGCATC | NEEAAELRPQEWDRLH |
| 1462 | G | AATGAGGAAGCTGCAGAATG | TGCGGCCGCA | GAATGGGATAGATTGCATCC | NEEAAECAAAEWDRLH |
| 1463 | G | ATGAGGAAGCTGCAGAATGG | TGCGGCCGCA | AATGGGATAGATTGCATCCA | EEAAEWCGRKWDRLHP |
| 1464 | G | TGAGGAAGCTGCAGAATGGG | TGCGGCCGCA | ATGGGATAGATTGCATCCAG | EEAAEWVRPQWDRLHP |
| 1465 | A | GAGGAAGCTGCAGAATGGGA | TGCGGCCGCA | TGGGATAGATTGCATCCAGT | EEAAEWDAAAWDRLHP |
| 1466 | T | AGGAAGCTGCAGAATGGGAT | TGCGGCCGCA | GGGATAGATTGCATCCAGTG | EAAEWDCGRRDRLHPV |
| 1467 | A | GGAAGCTGCAGAATGGGATA | TGCGGCCGCA | GGATAGATTGCATCCAGTGC | EAAEWDMRPQDRLHPV |
| 1468 | G | GAAGCTGCAGAATGGGATAG | TGCGGCCGCA | GATAGATTGCATCCAGTGCA | EAAEWDSAAADRLHPV |
| 1469 | A | AAGCTGCAGAATGGGATAGA | TGCGGCCGCA | ATAGATTGCATCCAGTGCAT | AAEWDRCGRNRLHPVH |
| 1470 | T | AGCTGCAGAATGGGATAGAT | TGCGGCCGCA | TAGATTGCATCCAGTGCATG | AAEWDRLRPHRLHPVH |
| 1471 | T | GCTGCAGAATGGGATAGATT | TGCGGCCGCA | AGATTGCATCCAGTGCATGC | AAEWDRFAAARLHPVH |
| 1472 | G | CTGCAGAATGGGATAGATTG | TGCGGCCGCA | GATTGCATCCAGTGCATGCA | AEWDRLCGRRLHPVHA |
| 1473 | C | TGCAGAATGGGATAGATTGC | TGCGGCCGCA | ATTGCATCCAGTGCATGCAG | AEWDRLLRPQLHPVHA |
| 1474 | A | GCAGAATGGGATAGATTGCA | TGCGGCCGCA | TTGCATCCAGTGCATGCAGG | AEWDRLHAAALHPVHA |
| 1475 | T | CAGAATGGGATAGATTGCAT | TGCGGCCGCA | TGCATCCAGTGCATGCAGGG | EWDRLHCGRMHPVHAG |
| 1476 | C | AGAATGGGATAGATTGCATC | TGCGGCCGCA | GCATCCAGTGCATGCAGGGC | EWDRLHLRPQHPVHAG |
| 1477 | C | GAATGGGATAGATTGCATCC | TGCGGCCGCA | CATCCAGTGCATGCAGGGCC | EWDRLHPAAAHPVHAG |
| 1478 | A | AATGGGATAGATTGCATCCA | TGCGGCCGCA | ATCCAGTGCATGCAGGGCCT | WDRLHPCGRNPVHAGP |
| 1479 | G | ATGGGATAGATTGCATCCAG | TGCGGCCGCA | TCCAGTGCATGCAGGGCCTA | WDRLHPVRPHPVHAGP |
| 1480 | T | TGGGATAGATTGCATCCAGT | TGCGGCCGCA | CCAGTGCATGCAGGGCCTAT | WDRLHPVAAAPVHAGP |
| 1481 | G | GGGATAGATTGCATCCAGTG | TGCGGCCGCA | CAGTGCATGCAGGGCCTATT | DRLHPVCGRTVHAGPI |
| 1482 | C | GGATAGATTGCATCCAGTGC | TGCGGCCGCA | AGTGCATGCAGGGCCTATTG | DRLHPVLRPQVHAGPI |
| 1483 | A | GATAGATTGCATCCAGTGCA | TGCGGCCGCA | GTGCATGCAGGGCCTATTGC | DRLHPVHAAAVHAGPI |
| 1484 | T | ATAGATTGCATCCAGTGCAT | TGCGGCCGCA | TGCATGCAGGGCCTATTGCA | RLHPVHCGRMHAGPIA |
| 1485 | G | TAGATTGCATCCAGTGCATG | TGCGGCCGCA | GCATGCAGGGCCTATTGCAC | RLHPVHVRPQHAGPIA |
| 1486 | C | AGATTGCATCCAGTGCATGC | TGCGGCCGCA | CATGCAGGGCCTATTGCACC | RLHPVHAAAAHAGPIA |
| 1487 | A | GATTGCATCCAGTGCATGCA | TGCGGCCGCA | ATGCAGGGCCTATTGCACCA | LHPVHACGRNAGPIAP |
| 1488 | G | ATTGCATCCAGTGCATGCAG | TGCGGCCGCA | TGCAGGGCCTATTGCACCAG | LHPVHAVRPHAGPIAP |
| 1489 | G | TTGCATCCAGTGCATGCAGG | TGCGGCCGCA | GCAGGGCCTATTGCACCAGG | LHPVHAGAAAAGPIAP |
| 1490 | G | TGCATCCAGTGCATGCAGGG | TGCGGCCGCA | CAGGGCCTATTGCACCAGGC | HPVHAGCGRTGPIAPG |
| 1491 | C | GCATCCAGTGCATGCAGGGC | TGCGGCCGCA | AGGGCCTATTGCACCAGGCC | HPVHAGLRPQGPIAPG |
| 1492 | C | CATCCAGTGCATGCAGGGCC | TGCGGCCGCA | GGGCCTATTGCACCAGGCCA | HPVHAGPAAAGPIAPG |
| 1493 | T | ATCCAGTGCATGCAGGGCCT | TGCGGCCGCA | GGCCTATTGCACCAGGCCAG | PVHAGPCGRRPIAPGQ |
| 1494 | A | TCCAGTGCATGCAGGGCCTA | TGCGGCCGCA | GCCTATTGCACCAGGCCAGA | PVHAGPMRPQPIAPGQ |
| 1495 | T | CCAGTGCATGCAGGGCCTAT | TGCGGCCGCA | CCTATTGCACCAGGCCAGAT | PVHAGPIAAAPIAPGQ |
| 1496 | T | CAGTGCATGCAGGGCCTATT | TGCGGCCGCA | CTATTGCACCAGGCCAGATG | VHAGPICGRTIAPGQM |
| 1497 | G | AGTGCATGCAGGGCCTATTG | TGCGGCCGCA | TATTGCACCAGGCCAGATGA | VHAGPIVRPHIAPGQM |
| 1498 | C | GTGCATGCAGGGCCTATTGC | TGCGGCCGCA | ATTGCACCAGGCCAGATGAG | VHAGPIAAAAIAPGQM |
| 1499 | A | TGCATGCAGGGCCTATTGCA | TGCGGCCGCA | TTGCACCAGGCCAGATGAGA | HAGPIACGRIAPGQMR |
| 1500 | C | GCATGCAGGGCCTATTGCAC | TGCGGCCGCA | TGCACCAGGCCAGATGAGAG | HAGPIALRPHAPGQMR |
| 1501 | C | CATGCAGGGCCTATTGCACC | TGCGGCCGCA | GCACCAGGCCAGATGAGAGA | HAGPIAPAAAAPGQMR |
| 1502 | A | ATGCAGGGCCTATTGCACCA | TGCGGCCGCA | CACCAGGCCAGATGAGAGAA | AGPIAPCGRTPGQMRE |
| 1503 | G | TGCAGGGCCTATTGCACCAG | TGCGGCCGCA | ACCAGGCCAGATGAGAGAAC | AGPIAPVRPQPGQMRE |
| 1504 | G | GCAGGGCCTATTGCACCAGG | TGCGGCCGCA | CCAGGCCAGATGAGAGAACC | AGPIAPGAAAPGQMRE |
| 1505 | C | CAGGGCCTATTGCACCAGGC | TGCGGCCGCA | CAGGCCAGATGAGAGAACCA | GPIAPGCGRTGQMREP |
| 1506 | C | AGGGCCTATTGCACCAGGCC | TGCGGCCGCA | AGGCCAGATGAGAGAACCAA | GPIAPGLRPQGQMREP |
| 1507 | A | GGGCCTATTGCACCAGGCCA | TGCGGCCGCA | GGCCAGATGAGAGAACCAAG | GPIAPGHAAAGQMREP |
| 1508 | G | GGCCTATTGCACCAGGCCAG | TGCGGCCGCA | GCCAGATGAGAGAACCAAGG | PIAPGQCGRSQMREPR |
| 1509 | A | GCCTATTGCACCAGGCCAGA | TGCGGCCGCA | CCAGATGAGAGAACCAAGGG | PIAPGQMRPHQMREPR |
| 1510 | T | CCTATTGCACCAGGCCAGAT | TGCGGCCGCA | CAGATGAGAGAACCAAGGGG | PIAPGQIAAAQMREPR |
| 1511 | G | CTATTGCACCAGGCCAGATG | TGCGGCCGCA | AGATGAGAGAACCAAGGGGA | IAPGQMCGRKMREPRG |
| 1512 | A | TATTGCACCAGGCCAGATGA | TGCGGCCGCA | GATGAGAGAACCAAGGGGAA | IAPGQMMRPQMREPRG |
| 1513 | G | ATTGCACCAGGCCAGATGAG | TGCGGCCGCA | ATGAGAGAACCAAGGGGAAG | IAPGQMSAAAMREPRG |
| 1514 | A | TTGCACCAGGCCAGATGAGA | TGCGGCCGCA | TGAGAGAACCAAGGGGAAGT | APGQMRCGRMREPRGS |
| 1515 | G | TGCACCAGGCCAGATGAGAG | TGCGGCCGCA | GAGAGAACCAAGGGGAAGTG | APGQMRVRPQREPRGS |
| 1516 | A | GCACCAGGCCAGATGAGAGA | TGCGGCCGCA | AGAGAACCAAGGGGAAGTGA | APGQMRDAAAREPRGS |
| 1517 | A | CACCAGGCCAGATGAGAGAA | TGCGGCCGCA | GAGAACCAAGGGGAAGTGAC | PGQMRECGRREPRGSD |
| 1518 | C | ACCAGGCCAGATGAGAGAAC | TGCGGCCGCA | AGAACCAAGGGGAAGTGACA | PGQMRELRPQEPRGSD |
| 1519 | C | CCAGGCCAGATGAGAGAACC | TGCGGCCGCA | GAACCAAGGGGAAGTGACAT | PGQMREPAAAEPRGSD |
| 1520 | A | CAGGCCAGATGAGAGAACCA | TGCGGCCGCA | AACCAAGGGGAAGTGACATA | GQMREPCGRKPRGSDI |
| 1521 | A | AGGCCAGATGAGAGAACCAA | TGCGGCCGCA | ACCAAGGGGAAGTGACATAG | GQMREPMRPQPRGSDI |
| 1522 | G | GGCCAGATGAGAGAACCAAG | TGCGGCCGCA | CCAAGGGGAAGTGACATAGC | GQMREPSAAAPRGSDI |
| 1523 | G | GCCAGATGAGAGAACCAAGG | TGCGGCCGCA | CAAGGGGAAGTGACATAGCA | QMREPRCGRTRGSDIA |
| 1524 | G | CCAGATGAGAGAACCAAGGG | TGCGGCCGCA | AAGGGGAAGTGACATAGCAG | QMREPRVRPQRGSDIA |
| 1525 | G | CAGATGAGAGAACCAAGGGG | TGCGGCCGCA | AGGGGAAGTGACATAGCAGG | QMREPRGAAARGSDIA |
| 1526 | A | AGATGAGAGAACCAAGGGGA | TGCGGCCGCA | GGGGAAGTGACATAGCAGGA | MREPRGCGRRGSDIAG |
| 1527 | A | GATGAGAGAACCAAGGGGAA | TGCGGCCGCA | GGGAAGTGACATAGCAGGAA | MREPRGMRPQGSDIAG |
| 1528 | G | ATGAGAGAACCAAGGGGAAG | TGCGGCCGCA | GGAAGTGACATAGCAGGAAC | MREPRGSAAAGSDIAG |
| 1529 | T | TGAGAGAACCAAGGGGAAGT | TGCGGCCGCA | GAAGTGACATAGCAGGAACT | REPRGSCGRRSDIAGT |
| 1530 | G | GAGAGAACCAAGGGGAAGTG | TGCGGCCGCA | AAGTGACATAGCAGGAACTA | REPRGSVRPQSDIAGT |
| 1531 | A | AGAGAACCAAGGGGAAGTGA | TGCGGCCGCA | AGTGACATAGCAGGAACTAC | REPRGSDAAASDIAGT |
| 1532 | C | GAGAACCAAGGGGAAGTGAC | TGCGGCCGCA | GTGACATAGCAGGAACTACT | EPRGSDCGRSDIAGTT |
| 1533 | A | AGAACCAAGGGGAAGTGACA | TGCGGCCGCA | TGACATAGCAGGAACTACTA | EPRGSDMRPHDIAGTT |
| 1534 | T | GAACCAAGGGGAAGTGACAT | TGCGGCCGCA | GACATAGCAGGAACTACTAG | EPRGSDIAAADIAGTT |
| 1535 | A | AACCAAGGGGAAGTGACATA | TGCGGCCGCA | ACATAGCAGGAACTACTAGT | PRGSDICGRNIAGTTS |
| 1536 | G | ACCAAGGGGAAGTGACATAG | TGCGGCCGCA | CATAGCAGGAACTACTAGT | PRGSDIVRPHIAGTTS |
| 1537 | C | CCAAGGGGAAGTGACATAGC | TGCGGCCGCA | ATAGCAGGAACTACTAGT | PRGSDIAAAAIAGTTS |
| 1538 | A | CAAGGGGAAGTGACATAGCA | TGCGGCCGCA | TAGCAGGAACTACTAGT | RGSDIACGRIAGTTS |
| 1539 | G | AAGGGGAAGTGACATAGCAG | TGCGGCCGCA | AGCAGGAACTACTAGT | RGSDIAVRPQAGTTS |
| 1540 | G | AGGGGAAGTGACATAGCAGG | TGCGGCCGCA | GCAGGAACTACTAGT | RGSDIAGAAAAGTTS |
| 1541 | A | GGGGAAGTGACATAGCAGGA | TGCGGCCGCA | CAGGAACTACTAGT | GSDIAGCGRTGTTS |
| 1542 | A | GGGAAGTGACATAGCAGGAA | TGCGGCCGCA | AGGAACTACTAGT | GSDIAGMRPQGTTS |

|  |  |  |  |  |  |
|---|---|---|---|---|---|
| 1543 | C | GGAAGTGACATAGCAGGAAC | TGCGGCCGCA | GGAACTACTAGT | GSDIAGTAAAGTTS |
| 1544 | T | GAAGTGACATAGCAGGAACT | TGCGGCCGCA | GAACTACTAGT | SDIAGTCGRRTTS |
| 1545 | A | AAGTGACATAGCAGGAACTA | TGCGGCCGCA | AACTACTAGT | SDIAGTMRPQTTS |
| 1546 | C | AGTGACATAGCAGGAACTAC | TGCGGCCGCA | ACTACTAGT | SDIAGTTAAATTS |
| 1547 | T | GTGACATAGCAGGAACTACT | TGCGGCCGCA | CTACTAGT | DIAGTTCGRTTS |
| 1548 | A | TGACATAGCAGGAACTACTA | TGCGGCCGCA | TACTAGT | DIAGTTMRPHTS |
| 1549 | G | GACATAGCAGGAACTACTAG | TGCGGCCGCA | ACTAGT | DIAGTTSAAATS |
| 1550 | T | ACATAGCAGGAACTACTAGT | TGCGGCCGCA | CTAGT | IAGTTSCGRTS |

## Appendix C. Oligonucleotides used for genetic footprinting.

| Oligo Name | Sequence | $T_m$ (°C) |
|---|---|---|
| HIV1 | 5'-ACATGTAGCCCCAGTTCTACTTACACC | 80 |
| HIV37 | 5'-TGGAAGGGCTAATTCACTCCCAAAG | 74 |
| HIV251 | 5'-GAGCCTGCATGGAATGGATG | 62 |
| HIV270r | 5'-CATCCATTCCATGCAGGCTC | 62 |
| HIV361 | 5'-ACTGCTGACATCGAGCTTGC | 62 |
| HIV400r | 5'-CCAGCGGAAAGTCCCTTGATGC | 66 |
| HIV492r | 5'-CCCAGTACAGGCAAAAAGCAGC | 65 |
| HIV493 | 5'-TCTCTCTGGTTAGACCAGATCTG | 63 |
| HIV501 | 5'-GTTAGACCAGATCTGAGCCTGGG | 66 |
| HIV521 | 5'-GGGAGCTCTCTGGCTAACTAGGG | 68 |
| HIV591r | 5'-TGAAGCACTCCCTCAAGGCAAGC | 66 |
| HIV592 | 5'-AGTAGTGTGTGCCCGTCTGTTG | 65 |
| HIV644r | 5'-GGGTCTGAGGGATCTCTAGTTACC | 74 |
| HIV672r | 5'-TGCTAGAGATTTTCCACACTGAC | 66 |
| HIV751 | 5'-GCGCACGGCAAGAGGCGAGG | 71 |
| HIV770r | 5'-CCTCGCCTCTTGCCGTGCGC | 71 |
| HIV905r | 5'-CTTTCCCCCTGGCCTTAACCG | 68 |
| HIV1027 | 5'-CCCTTCAGACAGGATCAGAAGAAC | 65 |
| HIV1224 | 5'-CCTATAGTCCAGAACCTCCAG | 64 |
| HIV1244r | 5'-CTGGAGGTTCTGGACTATAGG | 64 |
| HIV1539 | 5'-GGAACTACTAGTACCCTTCAGG | 66 |
| HIV1573r | 5'-CATCCTATTTGTTCCTGAAGGG | 64 |

## Appendix D.

## Documentation for software developed for genetic footprinting

# Help for Louise and Marc's Footprinting Utilities

Top

## Overview

The Footprinting Utilities are a set of tools to gather, manipulate, and present quantitative data from scanned gels using Excel and Matlab. From a set of footprinting gels and the sequence of the mutagenized DNA, you will be able to quantitatively assess band intensities, normalize data gathered from different gels, consolidate data from many spreadsheets into a single spreadsheet, and color code this data.

Top

## Scanning a Gel

Your gel must be scanned:

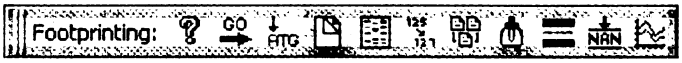
- at 300 DPI.
- without auto levelling (contrast and brightness at 50%, or 125), and no enhancement (e.g. use DeskScan's "Black and White Photo," NOT "Sharp Black and White Picture").
- in 256 gray scale.

Save the gel image as a TIFF file. Include no more than 8 characters in the filename and make sure that the file has a ".tif" extension.

The gel should be scanned vertically. Matlab takes care of rotating the image to better fit the screen. Since gels are often larger than the window of the scanner, for better image quality, a weight should be put on top of the scanner's lid to properly push the gel against the glass. It is a good idea to crop the gel image as much as possible, as it will speed up the program. In cases where you have extremely large gel images, it may be worth saving half of the gel (i.e. the top half) in one file and the other half (i.e. the bottom half) in another file in the interest of speed. Using a gel image that is twice as big may slow down the program much more than two-fold.

Top

## Excel Utilities



**Tools:** The Excel utilities enable you to:
- Create Excel sheets to hold your data.
- Start the Matlab Quantitation Utility.
- Correct mistakes in band assignments from previous Matlab Quantitation Utility sessions.
- Consolidate data from many spreadsheets into a single spreadsheet.
- Color-code numerical data in Excel spreadsheets.
- Format data for the Normalization Utility.

**Excel Sheets:** A Footprinting spreadsheet is composed of 2 sheets. Sheet 1 is **the summary sheet**. It displays the sequence of the target gene vertically. The position of each nucleotide is indicated in the column to the left of the target gene sequence. To the right of each nucleotide in the target sequence is written the structure of a mutant at that position, both in nucleotide form and in peptide form. Sheet 2 is **the data sheet.** It contains only the position numbers and target sequence until Matlab sends more data. Do not rename the sheets since Matlab sends the data to Sheet 2. Footprinting spreadsheets can be generated, saved, and reopened at a later time to add data using Matlab. Data can be entered to the same spreadsheet over several sessions. Simply open the appropriate spreadsheet before starting the Matlab Quantitation Utility at the beginning of each session and save your data at the end of each session. A Footprinting spreadsheet must be open before using the Matlab Quantitation Utility.

▬ **Help:** Launch a HTML browser with this help file.

▬ **Make Footprinting Sheets:** To create a Footprinting spreadsheet, first open a blank Excel Workbook, then click on the icon. It will take you through the entire process using several prompts. You may choose to start the Matlab Quantitation Utility directly after creating this spreadsheet in order to select bands on a gel image.

▬ **Select Peptide Reading Frame:** This icon allows you to change the translation frame for the peptides displayed on the summary sheet.

▬ **Make Data Sheet:** This icon allows you to create only the second sheet (the data sheet) in case you don't want the summary sheet. You can enter data using Matlab Utilities in this data sheet.

▬ **Start the Matlab Quantitation Utility:** This icon starts Matlab Quantitation Utility. You will be using the Matlab Quantitation Utility to select the bands you want to quantitate and to send the resulting data to the data sheet. Each time you click on this icon, you will start a new Matlab session. It is a good idea to have only one Matlab session open at a given time, so close the current session before opening a new one. Matlab Quantitation Utility are explained more fully below. The Matlab Quantitation Utility can be started directly from Matlab by typing "foot" in the Matlab Command Window.

▬ **Change Number:** If you want to change the number of a band after it has been entered by Matlab, open the appropriate data sheet, select the cell with the nucleotide number you want to change from and click on this icon. You will be asked to enter the nucleotide number to which you want the data to move. This operation will move the data on the Excel data sheet from the old nucleotide position to the new position, as well as change the label of the band on the gel image.

▬ **Regroup Data:** This icon allows you to import data from many spreadsheets into one spreadsheet. Open the destination spreadsheet (a new, blank sheet) and click on the icon. You will be asked to select a source file (Excel spreadsheet). The rows holding data will be imported into the destination spreadsheet. If you want to import data from several source files, do not click anywhere except on this icon to repeat the operation. This tool will only work with unmodified data sheets generated by Matlab Quantitation Utility as the source files.

▬ **Paste Column:** This icon allows you to import columns from many spreadsheets into one spreadsheet. For example, a given nucleic acid sample (called

George) would typically be analyzed using several different primer pairs. The data for George would therefore end up on several data sheets. If you made sure that the data for George was always entered in a specific column (e.g. column B), you could use this tool to consolidate all of the data for George into a single spreadsheet. Open the destination spreadsheet and click anywhere in the column into which you wish to import data. Any data in this column will be erased. You will be asked which column number you wish to copy. In Excel, column headings are letters, so you must convert the letter of the desired column into a number (e.g. A=1, B=2, C=3,...). Next, you will be asked how many columns you wish to skip between pastes. If you wish to paste data into consecutive columns, enter "0" here. Finally, select the spreadsheet from which you wish to copy a column of data. You can select several source spreadsheets in succession. The same column number will be accessed for each spreadsheet. When you have finished, click the 'Cancel' button. This tool will work using any type of Excel spreadsheet as the source file.

**Color Code Numerical Data:** This icon color codes cell values on a 56-shade grayscale. White is assigned to the number 1 and black is assigned to the numbers 100 and higher. You may scale your data as you wish to fit this scale. Highlight the cells holding the data you wish to color code and click on the icon. Unfortunately, due to a Microsoft bug, using this tool will change the entire color scheme of the current spreadsheet!

**Formating for the Normalization Utility:** To use the Matlab Normalization Utility, you must get your data into the proper format. Paste the values for your data into the upper left-hand corner of a blank Excel worksheet (to paste only values and not formulae, use Edit | Paste Special... | "as Text"). Eliminate any non-data information (e.g. data labels, nucleotide position numbers, column headings) -- the normalization program will try to normalize anything you give it. Your data should be organized such that a given column contains data from a single gel and a given row contains data for a given nucleotide position. Highlight the region of the spreadsheet where you have data. Hit the "NaN" function icon. NaN ("Not a Number") will appear in all the blank cells in the area where you have your data. Save your worksheet as "Text -- Tab-delimited." A ".txt" extension should automatically appear on your file. This specific extension is required by the Matlab Normalization Utility. Put the files you want to normalize into a dedicated folder. Do not put extraneous files into this folder -- the normalization utility will try to normalize them. You will not mess up your files, but the utility will crash.

**Start the Matlab Normalization Utility:** This icon starts the Matlab Normalization Utility. The Matlab Normalization Utility can be started directly from Matlab by typing "normalize" in the Matlab Command Window.

Top

## The Matlab Quantitation Utility

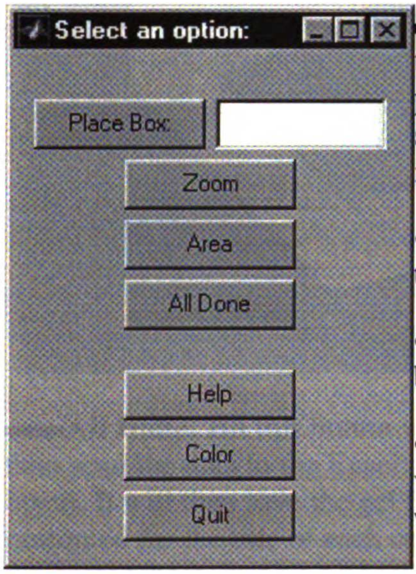
Number of Bands per Set
Main Dialog Screen
Figure Menu

Create or open the appropriate spreadsheet in Excel and click on this icon ▦.

## Number of Bands per Set

Enter the number of bands that you wish quantitate per nucleotide position, or hit Cancel to simply view the gel (the Zoom will then automatically be turned on).

## Main Dialog Screen



**Place Box:** Enter the nucleotide position number in the edit box. Then press 'Return', or click 'Place Box'. The pointer becomes a long crosshair, indicating that you can select the first box of a set. Click on the band you are interested in. Two boxes and a number will appear:

- The first box around the band will remain; it indicates the area where data is taken.
- The second box shows the area used to determine the background and will disappear.
- A number will temporarily appear; this number is the value of the data read (with background subtracted). It will disappear when you select the next band.

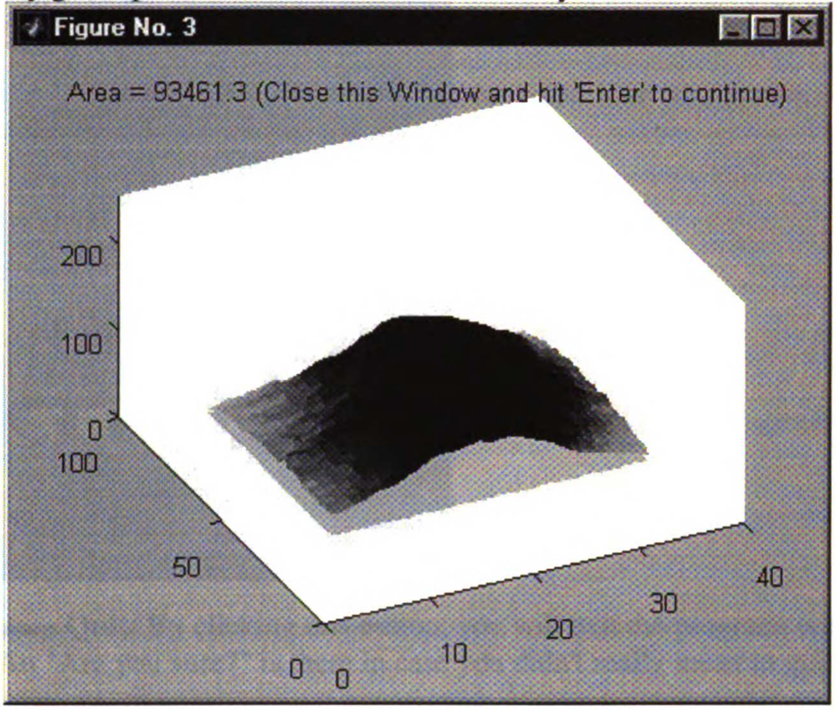
At this point you may decide to keep this data by selecting the next band, or delete it (giving you a chance to re-select it) by pressing 'Enter'.

When you reach the last band at a given position, you will receive a "last band" prompt. You can either keep the data by clicking anywhere on the figure, or delete it and re-select it by pressing 'Enter'.

Every time you hit the 'Enter' key, you will delete the data from one band, starting from the most recent band selected and proceeding backwards in time. However, you can only delete data at the nucleotide position where you are placing bands. Once you make the final click after the "last band" prompt, you cannot go back and delete data for that position using the 'Enter' key. You can always "Quit" without saving and start all over again.

**Zoom:** When the zoom is enabled, the cursor icon changes and you can Left-click to zoom in, Right-click to zoom out, or drag a box around the area you want to zoom. When you have zoomed to the level of magnification you want to work with, press 'Enter'.

**Area:** You probably won't be using this tool. It calculates the volume that lies between the plane defined by the background and the surface defined by the band. At any given point, the elevation is the intensity.

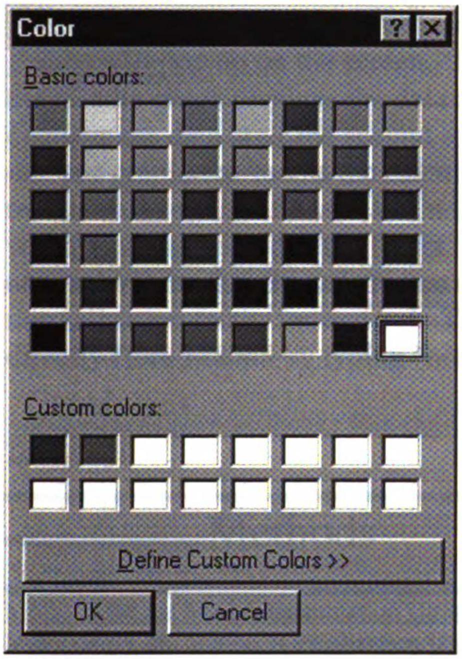


Area = 93461.3 (Close this Window and hit 'Enter' to continue)

**All Done:** Hit this button when you are finished entering data for your gel. It will then send the data to the Excel spreadsheet (the correct spreadsheet should already be open). It will then save the gel file with the boxes and position numbers. You will be prompted for the title of each column (if no titles are desired, or the titles are already in the spreadsheet, just hit 'Cancel'). Make sure to save the Excel spreadsheet with your new data before closing it.

**Help:** This will start a Web Browser if one is not already running and display this document.

**Color:** Click here to set the color you wish for the boxes and the text on the gel.



**Quit:** By clicking this button, you will exit the program, without saving anything. An "Are you sure?" is there in case you didn't really mean to quit without saving.

### Figure Menu

**Print:** Click here to print the current gel image, including boxed and labelled bands. To resume or finish your quantitation session, find the "Select an option" button on the taskbar at the bottom of your screen and click on it to reactivate the Main Dialog Screen.

**Colormap:** Click here to open another window displaying the current gel image next to a colorbar showing the color scheme used to pseudocolor the gel image. To resume or finish your quantitation session, close this colorbar window, find the "Select an option" button on the taskbar at the bottom of your screen, and click on it to reactivate the Main Dialog Screen.

Top

## The Matlab Normalization Utility

This tool allows you to merge data for the same nucleic acid sample between different gels or different exposures of the same gel.

First, make sure your data is in the proper format and in a dedicated folder, as described in the section on Formating for the Normalization Utility. Then start the Matlab Normalization Utility. You will be asked to select any file in the dedicated folder. The Normalization Utility will proceed to normalize every file in that folder. The program will run for a while, perhaps a long while if you are comparing many gels (we have normalized up to sixty gels in one .txt file). When the program is finished, a series of graphs will pop up on the screen. The top graph displays the data before normalization, the middle graph displays the data after one round of normalization, and the bottom graph displays the data after the second round of normalization. The highlighted blue curves, one per graph, represents the calculated average of the curves in that graph. Your real results will consist of a series of normalization factors, and are saved in ".res" files which will appear in the dedicated folder. To get your normalized data, you multiply the original values (i.e. the values you read off of the gel image) for a given gel by the normalization factor for that gel. The Matlab Normalization Utility can be started directly from Matlab by typing "normalize" in the Matlab Command Window.

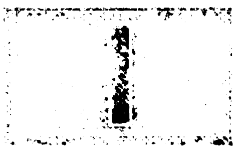
Top

## Methods

### Methods for the Matlab Quantitation Utility

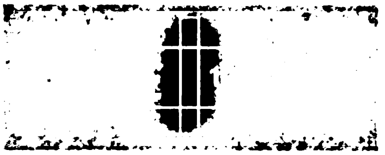
**Pseudocolors:** The scanned gel is a large 2-dimensional matrix, where each element of the matrix represents a pixel location and holds a number between 1 and 256 indicating the intensity of the gray. For pseudocolor, the image (= matrix) is searched for the biggest of the 10 smallest values of gray and the smallest of the 10 biggest values of gray. This gives us a good low and high boundaries for the gray present in the image. Then, a custom-made colormap (= set of 256 different colors) is stretched to fit exactly between the two boundaries. The use of this tailored pseudocoloring scheme helps when selecting bands on the gel. To see the colormap of a given gel, start the Matlab Utility and at the "Number of Bands per Set" prompt hit 'Cancel'. The menu should then have a 'ColorMap' option. Values that are outside the dynamic range of the film are colored bright red to indicate saturation.

**Data Values:**



For each band, an area around the band (60 X 10 pixels) is considered. In this area, the darkest 50 pixels are averaged to give the raw reading. The background value (see below) is then subtracted to give the data. It is a number between 1 (light) and 256 (dark). This value is temporarily displayed on the gel as the user selects bands.

**Background Subtraction:**



For background subtraction, a longer and narrower area than the one used for collecting the data is considered (30 X 160 pixels). This area is divided into 16 vertical strips (30 X 10 pixels). For each strip, the darkest 50 pixels are averaged. The lowest of these 16 averages is considered the value of the background. The background value is subtracted from the value read in the Data Area (above).

## Methods for the Matlab Normalization Utility

You may wish to merge data for the same nucleic acid sample between different gels or different exposures of the same gel. As you might expect, the normalization algorithm works better if you have more positions in common between gels. The goal of the algorithm is to minimize the weighted sum of the coefficients of variation for each position. First, the Normalization Utility pre-processes the data in three ways:

- It eliminates values that exceed the "maxgrey" value. Values that are beyond the dynamic range of the film are meaningless.
- It eliminates nucleotide positions that contain only one value.
- It eliminates gels that contain only one value.

Next, the average value at each nucleotide position is calculated. These average values (the Starting Averages for round 1) are kept throughout the first round of normalization. At a given nucleotide position, a weighted coefficient of variation is calculated using the values for the data points as well as the Starting Average. The Starting Average is assigned a weight that is equal to the total number of data points at this position. For example, suppose you have three data points (x1, x2, x3) at position 637, with a Starting Average of AV. x1, x2, and x3 are each given a weight of 1, while AV is given a weight of 3. A weighted sum of the weighted coefficients of variation is taken. The weighted coeffient of variation at a given position is assigned a weight according to the number of data points (not including the synthetic "average value" data points) present at that position. A position with two data points is given a weight of 2, a position with three data points is given a weight of 3, and so on. The Normalization Utility tries by iteration to minimize this weighted sum by adjusting each gel. The adjustment is achieved by multiplying the data from each gel by a factor (a different factor for each gel, but the same factor for all data points within a gel). We stop the iteration process when the variation in the weighted sum is less than the "precision" value or after a defined number ("iteration") of iterations has been performed.

A new set of average values is calculated using the normalization factors from the first round of normalization. These average values (the Starting Averages for round 2) are kept throughout the second round of normalization. The second round of normalization is carried out exactly like the first round.

Your results will consist of a series of normalization factors, one for each gel (note that if a gel was eliminated during pre-processing, it will not receive a normalization factor). To get your normalized data, you multiply the original values (i.e. the values you read off of the gel image) for a given gel by the normalization factor for that gel.

The default values for "**maxgrey**", "**precision**", and "**iteration**" are 110, 0.001, and 10000, respectively. You can modify these values. For example, type "maxgrey = 120" in the Matlab Command Window to set "maxgrey" to 120. Your modifications will not be saved between Matlab sessions. To verify the current values for these properties, type "maxgrey", "precision", or "iteration" in the Matlab Command Window (case-sensitive).

Top

# Appendix E.
# Code for software developed for genetic footprinting

---

CODE FOR MATLAB FOORPRINTING UTILITIES

---

FILE FOOT.M

---

```
function Foot
%
%  Utility for gathering data on scanned gels for foot printing.
%


% Initialize
global Dir

try
   cd(Dir.GelDir)
catch
   h = msgbox('The Default Gel Directory path is wrong.  Edit
''Startup.m'' to correct the path.');
   drawnow
   waitfor(h)
end
clear S
global S
S.Data = [];
S.Image = [];
O = NaN;
ZoomPointer = [ ...
O O O 1 1 1 1 1 O O O O O O O O;...
O 1 1 2 2 2 2 2 1 1 O O O O O O;...
O 1 2 2 1 1 1 2 2 1 O O O O O O;...
1 2 2 1 O O O 1 2 2 1 O O O O O;...
1 2 1 O O O O O 1 2 1 O O O O O;...
1 2 1 O O 1 O O 1 2 1 O O O O O;...
1 2 1 O O O O O 1 2 1 O O O O O;...
1 2 2 1 O O O 1 2 2 1 O O O O O;...
1 1 2 2 1 1 1 2 2 1 O O O O O O;...
O 1 1 2 2 2 2 2 1 1 1 O O O O O;...
O O O 1 1 1 1 1 O 1 1 1 O O O O;...
O O O O O O O O O O 1 1 1 O O O;...
O O O O O O O O O O O 1 1 1 O O;...
O O O O O O O O O O O O 1 1 1 O;...
O O O O O O O O O O O O O 1 1 1;...
O O O O O O O O O O O O O O 1 1];

% Open Gel
% Get a file
```

```
[ImFile, ImPath]  = uigetfile('*.tif');
if max(ImFile == 0)
   return
end
warning off
ImFile = [ImPath, ImFile];
im = imread(ImFile, 'tiff');
warning on
% Put it horizontal and show it
S.Image = rot90(im);
imshow(S.Image)

S.FigH = gcf;
S.AxeH = gca;

SetBestColorMap

% Add print to menu, and add setup for pretty printing
set(S.FigH, 'NumberTitle', 'off')
set(S.FigH, 'Name', ImFile)
set(S.FigH, 'Color',[1 1 1], ...
   'PaperOrientation', 'Landscape')
set(S.FigH, 'PaperUnits', 'inches');
PaperSize = get(S.FigH, 'PaperSize');
PaperPos=[0.3 0.3 (PaperSize(1)-.3) (PaperSize(2)-0.3)];
set(S.FigH, 'PaperPosition', PaperPos);
h1 = uimenu('Parent',S.FigH, ...
   'Callback','printdlg', ...
   'Label','&Print', ...
   'Tag','MnPrint');
h1 = uimenu('Parent',S.FigH, ...
   'Callback','Colors', ...
   'Label','&ColorMap', ...
   'Tag','MnColors');

% Try to load previous lines if needed
[LineFile,rem] = strtok(ImFile, '.');
LineFile = strcat(LineFile, '.fig');
EvalStr = ['hgload( ''',LineFile, ''');'];
eval(EvalStr, '');

% Get general info about gel
figure(S.FigH)
drawnow
% Play sound
try
   [a,b,c] = wavread('Utopia Windows Start.wav');
   playsnd(a,b,c);
catch
end

List = {'2','3','4','5','6','7','8','9',...
      '10','11','12','13','14','15','16','17','18','19',...
      '20','21','22','23','24','25','26','27','28','29',...
      '30','31','32','33','34','35','36','37','38','39','40'};
[NumCol,v] = listdlg('PromptString','How many lines per set?',...
   'SelectionMode','single',...
   'ListString',List);
S.NumCol = NumCol+1;
if v == 0
   set(S.FigH, 'Pointer', 'custom', 'PointerShapeCData', ZoomPointer,
'PointerShapeHotSpot', [6 6])
   zoom
   return
end

i = 0;
while(1)
   drawnow
   S.Ret = 'PBQuit';
   DlgH = FgWhatNext;
   CenterFigure(S.FigH, DlgH);
   waitfor(DlgH)
   if S.Ret == 'PBDone'
      % Get out
      break
   elseif S.Ret == 'PBQuit'
      % Make sure and quit
      ButtonName=questdlg('Do you really want to quit WITHOUT saving
?', 'Are you nuts?', ...
                         'Quit','No','No');
      if strcmp(ButtonName, 'Quit')
         set(S.FigH, 'Pointer', 'arrow')
         return
      end
   elseif S.Ret == 'PBColo'
      TheColor = uisetcolor;
      if size(TheColor,2) == 3
         Child = get(gca, 'Children');
         for i = 1 : size(Child)-1
            set(Child(i), 'Color', TheColor)
         end
      end
   elseif S.Ret == 'PBArea'
      GetArea
   elseif S.Ret == 'PBHelp'
      % show help
      set(S.FigH, 'Pointer', 'watch')
      try
         web(Dir.HelpPage)
      catch
         h = msgbox('The Help Page path is wrong.  Edit ''Startup.m''
to correct the path.');
         drawnow
         waitfor(h)
      end
      set(S.FigH, 'Pointer', 'arrow')
   elseif S.Ret == 'PBZoom'
      % Zoom
      set(S.FigH, 'Name', [ImFile, ' - Zoom'])
      set(S.FigH, 'Pointer', 'custom', 'PointerShapeCData',
ZoomPointer, 'PointerShapeHotSpot', [6 6])
      zoom on
      Key = 0;
   while Key == 0
      Key = waitforbuttonpress;
   end
   zoom off
   set(S.FigH, 'Pointer', 'arrow')
   set(S.FigH, 'Name', ImFile)
elseif S.Ret == 'EdNumb'
   % Do it!
   i = i+1;
   iMax = i;
   S.Data(i).Place = str2num(S.EdNumber);
   set(S.FigH, 'Name', [ImFile, ' - ', S.EdNumber])

   j = 1;
   LastAction = 'Put';
   while j <= S.NumCol+1
      [x, y] = ginput(1);
      if isempty(x)
         if (j > 1)
            if (j == S.NumCol+1)
               % Remove 2 boxes and text
               NumChildren = 10;
            elseif (LastAction ~= 'Rmv')
               % Remove 2 boxes
               NumChildren = 9;
            else
               % Remove 1 box
               NumChildren = 4;
            end
            % Last was an error, erase (Enter was hit)
            Children = get(S.AxeH, 'children');
            delete(Children(1: NumChildren))
            LastAction = 'Rmv';
            j = j -1;
         else
            S.Data(i) = [];
            i = i -1;
            break
         end
      else
         if (j ~= S.NumCol+1)
            if (j ~= 1)
               % Remove line across from previous
               if (LastAction ~= 'Rmv')
                  Children = get(S.AxeH, 'children');
                  delete(Children(1:5))
               end
            end
            LastAction = 'Put';
            if (j == S.NumCol)
               text(x+2, y-30, S.EdNumber, ...
                  'Color', 'w', ...
                  'Rotation', 90, ...
                  'FontWeight', 'bold');
            end
            % Start getting data
            x = round(x);
```

```
                y = round(y);
                S.Data(i).BoxVal(j) = BoxTopAverage(S.Image, x, y);
                text(x+10, y, num2str(S.Data(i).BoxVal(j)), ...
                      'Color', 'w', ...
                      'FontWeight', 'bold');
                if (S.Data(i).BoxVal(j) < 0.0)
                   h = msgbox('Value < 0 !!!');
                   drawnow
                   waitfor(h)
                   drawnow
                   S.Data(i).BoxVal(j) = 0;
                end
             else
                % Remove line across from previous
                Children = get(S.AxeH, 'children');
                delete(Children(1:5))
             end
             j = j+1;
             % Tell that this was the last one to give a chance to
delete it
             if j == S.NumCol+1
                h = msgbox('Last one.','','custom',S.Image,S.ColorMap);
                drawnow
                pause(1)
                try
                   delete(h)
                catch
                end
                drawnow
             end
          end
       end
   end
   [SFile,rem] = strtok(ImFile, '.');
   SFile = strcat(SFile, '.mat');
   save(SFile, 'S')
end

% Save the picture under bitmap format with colormap
%[BmpFile,rem] = strtok(ImFile, '.');
%BmpFile = strcat(BmpFile, '.bmp');
%imwrite(S.Image, S.ColorMap, BmpFile, 'bmp')

if ~isempty(S.Data)
   %Check that spreadsheet is open
   try
      Channel = ddeinit('excel', 'book1.xls:Sheet2');
      DNAStart = ddereq(Channel, 'r3c1:r3c1');
      ddeterm(Channel);
   catch
      msg = sprintf('Your Excel Spreadsheet doesn''t seem to be open.
\nOpen it FIRST and THEN press OK.');
      h = msgbox(msg, 'Error', 'Error');
      drawnow
      waitfor(h)
   end
   % Save the S structure
   [SFile,rem] = strtok(ImFile, '.');
   SFile = strcat(SFile, '.mat');
   save(SFile, 'S')

   % Save the lines
   ChildH = get(S.AxeH, 'Children');
   LinesH = ChildH(1 : size(ChildH, 1)-1);
   hgsave(LinesH, LineFile);

   % Send it to excel
   % Get the right page in Excel
   Channel = ddeinit('excel', 'book1.xls:Sheet2');
   DNAStart = ddereq(Channel, 'r3c1:r3c1');
   ddeterm(Channel);
   % Put file name at top of spreadsheet
   Channel = ddeinit('excel', 'book1.xls:Sheet2');
   ddepoke(Channel, 'r1c4:r1c4', ImFile);
   ddeterm(Channel);

   % Send the data, first
   for i = 1 : iMax
      disp(S.Data(i).Place)
      for j = 1 : S.NumCol
         Row = num2str(S.Data(i).Place - DNAStart + 3);
         Col = num2str(2 + (4*(j-1)) + 2);
         ColPlus1 = num2str(2 + (4*(j-1)) + 2 + 1);
         CellStr = ['r', Row, 'c', Col,':r', Row, 'c', Col];
         Channel = ddeinit('excel', 'book1.xls:Sheet2');
         ddepoke(Channel, CellStr, S.Data(i).BoxVal(j));
         ddeterm(Channel);
      end
   end
   % Then format Spread sheet colunm
   for i = 1 : S.NumCol
      Tmp = inputdlg(['Enter the title for column ',num2str(i), ' on
the Excel Spreadsheet'],...
         'Cool Title', 1);
      if isempty(Tmp)
         Tmp = '';
      elseif isempty(Tmp{1,:})
         Tmp = '';
      else
         Tmp = Tmp{1,:};
      end
      Channel = ddeinit('excel', 'book1.xls:Sheet2');
      ddepoke(Channel, 'r1c1:r1c1', i);
      ddeterm(Channel);
      Channel = ddeinit('excel', 'book1.xls:Sheet2');
      ddepoke(Channel, 'r1c2:r1c2', Tmp);
      ddeterm(Channel);
      Channel = ddeinit('excel', 'book1.xls:Sheet2');
      ddeexec(Channel, '[run("''Foot Printing.xls''!FormatGelCol")]');
      ddeterm(Channel);
   end
end
warning off
```

```
[im, map] = imread('face.tif');
warning on
msgbox('All done','Clara dit:','custom',im, map)
set(S.FigH, 'Name', ImFile)
```

## FILE STARTUP.M

```
iptsetpref('ImshowBorder', 'tight')
iptsetpref('ImshowTruesize', 'manual')
set(0, 'DefaultFigureMenuBar','none')
set(0, 'DefaultFigurePosition',[2, 70, 1022, 657])
set(0, 'DefaultFigureInvertHardCopy', 'on')
global S
global Dir

% Edit following if you change the directories
Dir.GelDir = 'd:\data';
Dir.HelpPage = 'd:\foot printing\help\foothelp.htm';
Dir.CodeDir = 'd:\foot printing\code matlab';


try
   cd(Dir.CodeDir);
catch
   h = msgbox('The Footprinting Code path is wrong.  Edit ''Startup.m''
to correct the path.');
   drawnow
   waitfor(h)
end

disp(' ')
disp('  Type ''Foot'' to start the Footprinting utility.');
disp(' ')

Setup % for gel curve fitting
```

## FILE SETUP.M

```
global V
global sV
global Curves
global Points

precision = 1.e-3;
iteration = 10000;
maxgrey = 110;
```

FILE FGWHATNEXT.M

```
function fig = FgWhatNext()
% This is the machine-generated representation of a Handle Graphics
object
% and its children.  Note that handle values may change when these
objects
% are re-created. This may cause problems with any callbacks written to
% depend on the value of the handle at the time the object was saved.
%
% To reopen this object, just type the name of the M-file at the MATLAB
% prompt. The M-file and its associated MAT-file must be on your path.

load FgWhatNext

h0 = figure('Color',[0.8 0.8 0.8], ...
    'Colormap',mat0, ...
    'MenuBar','none', ...
    'Name','Select an option:', ...
    'NumberTitle','off', ...
    'PointerShapeCData',mat1, ...
    'Position',[503 205 195 254], ...
    'Tag','Fig1');
h1 = uicontrol('Parent',h0, ...
    'Units','points', ...
    'BackgroundColor',[1 1 1], ...
    'Callback','FgWhatNextGUI EdNumber', ...
    'HorizontalAlignment','left', ...
    'ListboxTop',0, ...
    'Position',[75 153.75 63.75 18.75], ...
    'Style','edit', ...
    'Tag','EdNumber');
h1 = uicontrol('Parent',h0, ...
    'Units','points', ...
    'Callback','FgWhatNextGUI PBNumber', ...
    'ListboxTop',0, ...
    'Position',[7.5 153.75 63.75 18.75], ...
    'String','Place Box:', ...
    'Tag','PBNumber');
h1 = uicontrol('Parent',h0, ...
    'Units','points', ...
    'Callback','FgWhatNextGUI PBZoom', ...
    'ListboxTop',0, ...
    'Position',[41.25 131.25 63.75 18.75], ...
    'String','Zoom', ...
    'Tag','PBZoom');
h1 = uicontrol('Parent',h0, ...
    'Units','points', ...
    'Callback','FgWhatNextGUI PBDone', ...
    'ListboxTop',0, ...
    'Position',[41.25 86.25 63.75 18.75], ...
    'String','All Done', ...
    'Tag','PBDone');
h1 = uicontrol('Parent',h0, ...
   'Units','points', ...
   'Callback','FgWhatNextGUI PBQuit', ...
   'ListboxTop',0, ...
   'Position',[41.25 7.5 63.75 18.75], ...
   'String','Quit', ...
   'Tag','PBQuit');
h1 = uicontrol('Parent',h0, ...
   'Units','points', ...
   'Callback','FgWhatNextGUI PBHelp', ...
   'ListboxTop',0, ...
   'Position',[41.25 52.5 63.75 18.75], ...
   'String','Help', ...
   'Tag','PBHelp');
h1 = uicontrol('Parent',h0, ...
   'Units','points', ...
   'Callback','FgWhatNextGUI PBColor', ...
   'ListboxTop',0, ...
   'Position',[41.25 30 63.75 18.75], ...
   'String','Color', ...
   'Tag','PBColor');
h1 = uicontrol('Parent',h0, ...
   'Units','points', ...
   'Callback','FgWhatNextGUI PBArea', ...
   'ListboxTop',0, ...
   'Position',[41.25 108.75 63.75 18.75], ...
   'String','Area', ...
   'Tag','PBArea');
if nargout > 0, fig = h0; end
```

## FILE FGWHATNEXTGUI.M

```
function FgWhatNextGUI(action)
global S
%
%  Callback's for FgWhatNext GUI page
%

FigH = gcf;


switch action

case {'EdNumber', 'PBNumber'}
   S.EdNumber = get(findobj('tag', 'EdNumber'), 'string');
   if isempty(S.EdNumber) | max(isletter(S.EdNumber))
      MsgBox('Please enter a number.');
      set(findobj('tag', 'EdNumber'), 'string', '');
   else
      S.Ret = 'EdNumb';
      delete(FigH)
   end

case 'PBZoom'
```

```
   S.Ret = 'PBZoom';
   delete(FigH)

case 'PBDone'
   S.Ret = 'PBDone';
   delete(FigH)

case 'PBArea'
   S.Ret = 'PBArea';
   delete(FigH)

case 'PBColor'
   S.Ret = 'PBColo';
   delete(FigH)

case 'PBHelp'
   S.Ret = 'PBHelp';
   delete(FigH)

case 'PBQuit'
   S.Ret = 'PBQuit';
   delete(FigH)


otherwise
   msgbox('ERROR...')
   S.Ret = 'ERROR_';
end
```

FILE WAITSCREEN.M

```
function fig = WaitScreen()
% This is the machine-generated representation of a Handle Graphics
object
% and its children.  Note that handle values may change when these
objects
% are re-created. This may cause problems with any callbacks written to
% depend on the value of the handle at the time the object was saved.
%
% To reopen this object, just type the name of the M-file at the MATLAB
% prompt. The M-file and its associated MAT-file must be on your path.

load WaitScreen

h0 = figure('Color',[0.8 0.8 0.8], ...
   'Colormap',mat0, ...
   'MenuBar','none', ...
   'Name','Please wait while MATLAB updates its data...', ...
   'NextPlot','replacechildren', ...
   'NumberTitle','off', ...
   'PointerShapeCData',mat1, ...
   'Position',[273 265 377 257], ...
   'Tag','Fig1');
```

```
h1 = axes('Parent',h0, ...
   'Box','on', ...
   'CameraUpVector',[0 1 0], ...
   'Color',[1 1 1], ...
   'ColorOrder',mat2, ...
   'DataAspectRatioMode','manual', ...
   'Layer','top', ...
   'Position',[0 0 1 1], ...
   'Tag','Axes1', ...
   'TickDir','out', ...
   'TickDirMode','manual', ...
   'Visible','off', ...
   'WarpToFill','off', ...
   'XColor',[0 0 0], ...
   'XLim',[0.5 417.5], ...
   'XLimMode','manual', ...
   'YColor',[0 0 0], ...
   'YDir','reverse', ...
   'YLim',[0.5 284.5], ...
   'YLimMode','manual', ...
   'ZColor',[0 0 0]);
h2 = image('Parent',h1, ...
   'BusyAction','cancel', ...
   'CData',mat3, ...
   'Interruptible','off', ...
   'Tag','Axes1Image1', ...
   'XData',[1 417], ...
   'YData',[1 284]);
h2 = text('Parent',h1, ...
   'Color',[0 0 0], ...
   'HandleVisibility','off', ...
   'HorizontalAlignment','center', ...
   'Position',[207.890625 -7.265625 2523.127598358505], ...
   'Tag','Axes1Text4', ...
   'VerticalAlignment','bottom');
set(get(h2,'Parent'),'Title',h2);
h2 = text('Parent',h1, ...
   'Color',[0 0 0], ...
   'HandleVisibility','off', ...
   'HorizontalAlignment','center', ...
   'Position',[207.890625 315.5625 2523.127598358505], ...
   'Tag','Axes1Text3', ...
   'VerticalAlignment','cap');
set(get(h2,'Parent'),'XLabel',h2);
h2 = text('Parent',h1, ...
   'Color',[0 0 0], ...
   'HandleVisibility','off', ...
   'HorizontalAlignment','center', ...
   'Position',[-39.50000000000001 143.609375 2523.127598358505], ...
   'Rotation',90, ...
   'Tag','Axes1Text2', ...
   'VerticalAlignment','baseline');
set(get(h2,'Parent'),'YLabel',h2);
h2 = text('Parent',h1, ...
   'Color',[0 0 0], ...
   'HandleVisibility','off', ...
   'HorizontalAlignment','right', ...
   'Position',[-0.67187500000001 1.609375 2523.127598358505], ...
   'Tag','Axes1Text1', ...
   'Visible','off');
set(get(h2,'Parent'),'ZLabel',h2);
if nargout > 0, fig = h0; end
```

## FILE CENTERFIGURE.M

```
function CenterFigure(MainFgH, NewFgH)
global S
%
%  Centers the new figure in the previous(MainFgH) figure
%   or center on screen if no previous figure(MainFgH = 0)
%

MainUnits = get(MainFgH, 'Units');
NewUnits = get(NewFgH, 'Units');
set(NewFgH, 'Units', MainUnits);
if (MainFgH == 0)
   % Center on screen
   MainPos = get(MainFgH, 'ScreenSize');
else
   MainPos = get(MainFgH, 'Position');
end
NewPos = get(NewFgH, 'Position');
set(NewFgH, 'Position', ...
   [MainPos(1) + (MainPos(3)/2) - (NewPos(3)/2) ...
   MainPos(2) + (MainPos(4)/2) - (NewPos(4)/2) ...
   NewPos(3) ...
   NewPos(4)])
set(NewFgH, 'Units', NewUnits);
```

## FILE ZOOM.M

This Matlab file (Revision: 5.34 Date: 1997/12/02 21:08:55) was modified:

```
Line 226:
if isempty(state),
      %ML added
      LocPointer = get(fig, 'Pointer');
      %ML end Added
      state = uisuspend(fig);
      setuprop(fig,'ZOOMFigureState',state);
   end
   %ML changed
   %set(fig,'windowbuttondownfcn','zoom down', ...
   %   'windowbuttonupfcn','ones;', ...
   %   'windowbuttonmotionfcn','','buttondownfcn','', ...
   %   'interruptible','on');
   set(fig,'windowbuttondownfcn','zoom down', ...
      'windowbuttonupfcn','ones;', ...
      'windowbuttonmotionfcn','','buttondownfcn','', ...
      'Pointer',LocPointer, ...
      'interruptible','on');
   %ML end Changed

Line 371:
%
% Actual zoom operation
%
%ML added
LocPointer = get(fig, 'Pointer');
set(fig, 'Pointer', 'watch')
%ML end Changed

Line 445 (end of function):
%ML Added
drawnow
set(fig, 'Pointer', LocPointer)
%ML end added
```

## FILE SETBESTCOLORMAP.M

```
function SetBestColorMap
global S
%
%  Sets the best color map based on the picture
%

%Min = double(min(min(S.Image)));
%Max = double(max(max(S.Image)));
Tmp = sort(double(min(S.Image)));
Min = Tmp(5);
Tmp = sort(double(max(S.Image)));
Max = Tmp(size(Tmp,2)-5);
Total = Max-Min;

Half = round(Total/2);
OtherHalf = Total - Half;

OneQuarter = round(OtherHalf/2);
OtherQuarter = OtherHalf - OneQuarter;

Bleu2Black = [];
for i =1 : OtherQuarter
   Bleu2Black(i) = (i-1)*(.5625/OtherQuarter);
end
Bleu2Black = Bleu2Black';
Bleu2Black = [zeros(OtherQuarter,1), zeros(OtherQuarter,1),
Bleu2Black];
```

```
Satur = 10;
TmpMap = bone(OneQuarter-Satur);
for i =1 : OneQuarter-Satur
   Black2White(i,:) = TmpMap(OneQuarter-Satur-i+1,:);
end

BeforePad = [ones(Min, 1), ones(Min, 1), ones(Min, 1)];
AfterPad = [ones(256-Max, 1)*.5, zeros(256-Max, 1), zeros(256-Max, 1)];
Saturation = [ones(Satur, 1)*1, ones(Satur, 1)*.0, ones(Satur, 1)*0];
%for i = 1 : Satur
%   Saturation(i,:) = [1, ((i-1)/(Satur-1))^2, ((i-1)/(Satur-1))^2];
%end

GelMap = [BeforePad; Saturation; Black2White; Bleu2Black; jet(Half);
AfterPad];

set(S.FigH, 'ColorMap', GelMap)

S.ColorMap = GelMap;
```

## FILE COLORS.M

```
global S
figure
imshow(S.Image)
colormap(S.ColorMap)
colorbar
```

## FILE RECT.M

```
function rect(x, y)
%
%  draw a rectangle from x(1),y(1) to x(2),y(2)
%

line([x(1), x(1)],[y(1), y(2)], 'Color', 'w')
line([x(1), x(2)],[y(2), y(2)], 'Color', 'w')
line([x(2), x(2)],[y(2), y(1)], 'Color', 'w')
line([x(2), x(1)],[y(1), y(1)], 'Color', 'w')
```

## FILE BOXTOPAVERAGE.M

```
function BoxRet  = BoxTopAverage(im, x,y)
```

```
global S
%
% Return the average of the 10 most dark points in a rectangle around
x,y
%  minus the background color (== 5 brigthest points on line accross)

% Average and box
HalfH = 30;
HalfW = 5;
rect([x+HalfW, x-HalfW], [y+HalfH, y-HalfH])

Tmp = im(y-HalfH : y+HalfH, x-HalfW : x+HalfW);
Tmp = reshape(Tmp, size(Tmp,1)*size(Tmp,2),1);
Tmp = sort(double(Tmp));
BoxAverage = 255 -(sum(Tmp(1:50))/50);

% Background and line
HalfW = 80;
HalfH = 15;
XMinus = x-HalfW;
XPlus  = x+HalfW;
YMinus = y-HalfH;
YPlus  = y+HalfH;
Size1 = size(im,1);
Size2 = size(im,2);
if XMinus < 0
   XPlus = XPlus - XMinus;
   XMinus = 1;
elseif XPlus > Size2
   XMinus = XMinus - (XPlus - Size2);
   XPlus = Size2;
end
if YMinus < 0
   YPlus = XPlus - XMinus;
   YMinus = 1;
elseif YPlus > Size1
   YMinus = YMinus - (YPlus - Size1);
   YPlus = Size1;
end

rect([XPlus, XMinus], [YPlus, YMinus])
Incr = HalfW*2/16;
for i =  1 : 16
   Tmp1(i).Tmp = im(YMinus : YPlus, XMinus +((i-1)*Incr) : XMinus
+(i*Incr));
   Tmp1(i).Tmp = reshape(Tmp1(i).Tmp,
size(Tmp1(i).Tmp,1)*size(Tmp1(i).Tmp,2),1);
   Tmp1(i).Tmp = sort(double(Tmp1(i).Tmp));
   BackGround(i) = 255 -(sum(Tmp1(i).Tmp(1:50))/50);
end

BackGround = min(BackGround);

BoxRet = BoxAverage - BackGround;
```

FILE GETAREA.M

```
function GetArea
global S
%
% Calculates the integral under the curve
%

%clear c3
%clear P3
%clear P1
%clear BackGround
%clear Tmp1

[x,y] = ginput(1);


% Average and box
HalfH = 30;
HalfW = 5;
%rect([x+HalfW, x-HalfW], [y+HalfH, y-HalfH])
warning off
Tmp = S.Image(y-HalfH : y+HalfH, x-HalfW : x+HalfW);
Tmp = reshape(Tmp, size(Tmp,1)*size(Tmp,2),1);
Tmp = sort(double(Tmp));
BoxAverage = 255 -(sum(Tmp(1:50))/50);
warning on
% Background and line
HalfW = 80;
HalfH = 15;
XMinus = x-HalfW;
XPlus  = x+HalfW;
YMinus = y-HalfH;
YPlus  = y+HalfH;
Size1 = size(S.Image,1);
Size2 = size(S.Image,2);
if XMinus < 0
   XPlus = XPlus - XMinus;
   XMinus = 1;
elseif XPlus > Size2
   XMinus = XMinus - (XPlus - Size2);
   XPlus = Size2;
end
if YMinus < 0
   YPlus = XPlus - XMinus;
   YMinus = 1;
elseif YPlus > Size1
   YMinus = YMinus - (YPlus - Size1);
   YPlus = Size1;
end

warning off
%rect([XPlus, XMinus], [YPlus, YMinus])
Incr = HalfW*2/16;
```

```
for i =  1 : 16
   Tmp1(i).Tmp = S.Image(YMinus : YPlus, XMinus +((i-1)*Incr) : XMinus
+(i*Incr));
   Tmp1(i).Tmp = reshape(Tmp1(i).Tmp,
size(Tmp1(i).Tmp,1)*size(Tmp1(i).Tmp,2),1);
   Tmp1(i).Tmp = sort(double(Tmp1(i).Tmp));
   BackGround(i) = 255 -(sum(Tmp1(i).Tmp(1:50))/50);
end
warning on
BackGround = min(BackGround);

BoxRet = BoxAverage - BackGround;


ii = 0;
xx = 15;
x = round(x);
y = round(y);
TheArea = 0;
for i = y-30 : y+30
   ii = ii + 1;
   c3t(ii).Data = 255-double(S.Image(i, x-xx:x+xx));
   c3tB(ii).Data = 255-double(S.Image(i, x-xx:x+xx))-BackGround;
   for j = 1 : size(c3tB(ii).Data,2)
      if c3tB(ii).Data(j) < 0
         c3tB(ii).Data(j) = 0;
      end
   end
   TheArea = TheArea + trapz(c3tB(ii).Data);
end
rect([x+xx, x-xx], [y+30, y-30])
TheArea
for i = 1 : ii
   P3(i,:) = c3t(i).Data;
end
scrsz = get(0,'ScreenSize');
figure('Position',[20 20 500 500])
surface(P3, 'linestyle', 'none')
set(gca, 'CLim', [1, 256])
set(gca, 'ZLim', [0,255])
title(['Area = ', num2str(TheArea), ' (Close this Window and hit
''Enter'' to continue)']);
try
   for i = 1 : size(S.ColorMap)
      TmpMap(i,:) = S.ColorMap(257-i,:);
   end
   set(gcf, 'ColorMap', TmpMap)
catch
end
view([-26, 46])
hold
Pl = ones(size(P3,1),size(P3, 2))*BackGround;
surface(Pl, 'linestyle', 'none')

pause
Children = get(S.AxeH, 'children');
delete(Children(1:4))
```

## FILE CHANGENUMBER.M

```
function ChangeNumber
global S
%
%
%

hh = WaitScreen;
drawnow

Data=1:64;Data=(Data'*Data)/64;

FigH = figure;
set(FigH, 'visible', 'off')
set(FigH, 'Pointer', 'watch')

%Get the file name
Channel = ddeinit('excel', 'book1.xls:Sheet2');
GelFile = ddereq(Channel, 'r1c4:r1c4', [1,1]);
ddeterm(Channel);
%Get the Old Number
Channel = ddeinit('excel', 'book1.xls:Sheet2');
OldNumber = ddereq(Channel, 'r1c1:r1c1');
ddeterm(Channel);
%Get the New Number
Channel = ddeinit('excel', 'book1.xls:Sheet2');
NewNumber = ddereq(Channel, 'r1c2:r1c2');
ddeterm(Channel);

% Try to load previous lines if needed
[LineFile,rem] = strtok(GelFile, '.');
LineFile = strcat(LineFile, '.fig');
Child = hgload(LineFile);

% Check that you are not overwriting a set of data
for i = 1 : size(Child)
   if strcmp(get(Child(i), 'Type'), 'text')
      Num = str2num(get(Child(i), 'String'));
      if Num == NewNumber
         close(hh);
         ButtonName=questdlg(['Do you really want to overwrite the ',
num2str(NewNumber), ' box ?'], ...
            'Yo!!', ...
            'Yes', 'No', 'No');
         switch ButtonName,
         case 'Yes',
            hh = WaitScreen
            % Keep on going
            break
         case 'No',
            % Clean worksheet
            Channel = ddeinit('excel', 'book1.xls:Sheet2');
            ddepoke(Channel, 'r1c1:r1c1', '');
            ddeterm(Channel);
            Channel = ddeinit('excel', 'book1.xls:Sheet2');
            ddepoke(Channel, 'r1c2:r1c2', '');
            ddeterm(Channel);
            % End program
            exit
            return
         end
         break
      end
   end
end


Changed = 0;
for i = 1 : size(Child)
   if strcmp(get(Child(i), 'Type'), 'text')
      Num = str2num(get(Child(i), 'String'));
      if Num == OldNumber
         set(Child(i), 'String', num2str(NewNumber))
         Changed = 1;
         break
      end
   end
end

if Changed == 1
   hgsave(Child, LineFile);
   % Play sound
   try
      [a,b,c] = wavread('Utopia Critical Stop.wav');
      playsnd(a,b,c);
   catch
   end
   h = msgbox(['Changed : ', num2str(OldNumber), ' to ',
num2str(NewNumber), '.'],'Yo!!', 'custom', Data, hot(64));
   close(hh);
else
   % Play sound
   try
      [a,b,c] = wavread('Robotz Error.wav');
      playsnd(a,b,c);
   catch
   end
   h = msgbox(['Did not find ', num2str(OldNumber), ' in file ',
LineFile, '.'],'Yo!!', 'custom', Data, hot(64));
   close(hh);
end

% Clean worksheet
Channel = ddeinit('excel', 'book1.xls:Sheet2');
ddepoke(Channel, 'r1c1:r1c1', '');
ddeterm(Channel);
Channel = ddeinit('excel', 'book1.xls:Sheet2');
ddepoke(Channel, 'r1c2:r1c2', '');
```

```
ddeterm(Channel);

uiwait(h);

exit
```

## FILE FOOTHELP.M

```
%
%  Start Netscape with help file
%
global Dir

web(Dir.HelpPage)

exit
```

## FILE NORMALIZE.M

```
global V
global File

cd('D:\Data\excel sheets\source')
[File, Path] = uigetfile('*.*', 'Select any file in the Directory');
D = dir(Path);
n = size(D, 1);
for i = 3 : n
   File = strcat(Path, D(i).name);
   disp(['Working on ', File])
   drawnow
   V = load(File);
   V = V';
   h = msgbox('Keep Ctrl-C down for 10 sec. to stop.');
   drawnow
   %pause(10)
   try
      delete(h)
   catch
   end
   drawnow
   NormCode
end
warning off
[im, map] = imread('face.tif');
warning on
msgbox('All done','Clara dit:','custom',im, map)
```

## FILE NORMFUNC.M

```
function y = NormalizeFunc(v)
%
%  Just if you try to read this code...
%     a(p) is the average at position p
%     V(c,p) is the value of curve c, position p
%     sV is a binary representation of V (NaN or Value => 0 or 1)
%     v is the coef to move curves up or down (changed to minimize this
function)
%

global V
global sV
global a
global Curves
global Points



Means = zeros(Points,1);
Sigma = zeros(Points,1);
n = sum(sV);
%vv = zeros(Points,Points);
for p = 1 : Points
   % Mean
   Means(p) = (sum(v'.*V(:,p)) + n(p)*a(p)) / (2*n(p));
end
for p = 1 : Points
   % Sigma
   Sigma(p) = sum((v'.*V(:,p)-ones(1)*Means(p)).*(v'.*V(:,p)-
ones(1)*Means(p))) ...
      + n(p)*(a(p)-Means(p))*(a(p)-Means(p));
   Sigma(p) = sqrt(Sigma(p)/(2*n(p)-1)) / Means(p);
end
y = sum(n'.*Sigma);
```

## FILE NORMCODE.M

```
global V
global sV
global File
global Curves
global Points

precision
iteration
maxgrey

tic
```

```
%Change NaN to zero, just in case
for p = 1 : size(V(1,:),2)
   for c = 1 : size(V(:,1),1)
      if isnan(V(c,p));
         V(c,p) = 0;
      end
   end
end

% remove anything bigger than maxgrey
for p = 1 : size(V(1,:),2)
   for c = 1 : size(V(:,1),1)
      if V(c,p) >= maxgrey;
         sprintf('V(%d,%d) = %d = 0', c,p,V(c,p));
         V(c,p) = 0;
      end
   end
end

% remove any DNA points that has only one point
for p = 1 : size(V(1,:),2)
   n = 0;
   for c = 1 : size(V(:,1),1)
      if V(c,p) > 0;
         n = n +1;
      end
   end
   if n == 1
      V(:,p) = 0;
   end
end

% remove any curve that has only one point
for c = 1 : size(V(:,1),1)
   n = 0;
   for p = 1 : size(V(1,:),2)
      if V(c,p) > 0;
         n = n +1;
      end
   end
   if n == 1
      V(c,:) = 0;
   end
end

% Clean V of 0 Column
V(:,all((V'==0)')) = [];

% Clean V of 0 Row
V(all((V==0)'),:) = [];

% Create sV = NaN or not (binary matrix)
sV = V>0;

% Average at each point
clear global a;
clear a1;
clear a2;
global a
for p = 1 : size(V(1,:),2)
   a(p) = 0;
   n(p) = 0;
   for c = 1 : size(V(:,1),1)
      if V(c,p)>0
         n(p) = n(p) + 1;
         a(p) = a(p) + V(c,p);
      end
   end
   a(p) = a(p)/n(p);
end

% Coef to get each curve to average
clear global v;
clear v1;
clear v2;
global v
for c = 1 : size(V(:,1),1)
   v(c) = 0;
   n(c) = 0;
   for p = 1 : size(V(1,:),2)
      if V(c,p) > 0
         n(c) = n(c) + 1;
         v(c) = v(c) + a(p)/V(c,p);
      end
   end
   v(c) = v(c)/n(c);
end
a1 = a;
v1 = v;

v
clear vv
Curves = size(V(:,1),1);
Points = size(V(1,:),2);

options = [0, precision, precision, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
iteration];
[Res1, Opt] = fmins('NormFunc', v, options, a);
Res1
Opt(10)

V1 = V;
for i = 1 : size(V(:,1),1)
   V1(i,:) = Res1(i).*V(i,:);
end

% Average at each point
for p = 1 : size(V1(1,:),2)
   a(p) = 0;
   n(p) = 0;
   for c = 1 : size(V1(:,1),1)
      if V1(c,p)>0
         n(p) = n(p) + 1;
         a(p) = a(p) + V1(c,p);
      end
   end
   a(p) = a(p)/n(p);
end

a2 = a;

%v
options = [0, precision, precision, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
iteration];
[Res2, Opt] = fmins('NormFunc', Res1, options, a);
Res2
Opt(10)

toc

for p = 1 : size(V(1,:),2)
   for c = 1 : size(V(:,1),1)
      if V(c,p) == 0;
         V(c,p) = NaN;
      end
   end
end
for p = 1 : size(V1(1,:),2)
   for c = 1 : size(V1(:,1),1)
      if V1(c,p) == 0;
         V1(c,p) = NaN;
      end
   end
end

% Plot
figure
subplot(3,1,1)
plot(V','*-')
warning off;
title(File);
warning on;
hold on;
plot(a1, 'o-','LineWidth',1.5)
hold off;

subplot(3,1,2)
plot(a1, 'o-','LineWidth',1.5)
hold on;
plot(V1','*-')
hold off;

subplot(3,1,3)
V2 = V;
for i = 1 : size(V(:,1),1)
   V2(i,:) = Res2(i).*V(i,:);
end
plot(a2, 'o-','LineWidth',1.5)
hold on;
plot(V2','*-')
hold off;
```

```
% set printing
set(gcf, 'Color',[1 1 1], ...
   'PaperOrientation', 'Landscape')
set(gcf, 'PaperUnits', 'inches');
PaperSize = get(gcf, 'PaperSize');
PaperPos=[0.3 0.3 (PaperSize(1)-.3) (PaperSize(2)-0.3)];
set(gcf, 'PaperPosition', PaperPos);
h1 = uimenu('Parent',gcf, ...
   'Callback','printdlg', ...
   'Label','&Print', ...
   'Tag','MnPrint');


% Save to file
NewFile = strcat(File, '.res');
fid = fopen(NewFile, 'wt');
fprintf(fid, '%s\n', File);
for i = 1 : size(V(:,1),1)
   fprintf(fid, '%s\n', num2str(Res2(i)));
end
fclose(fid);
```