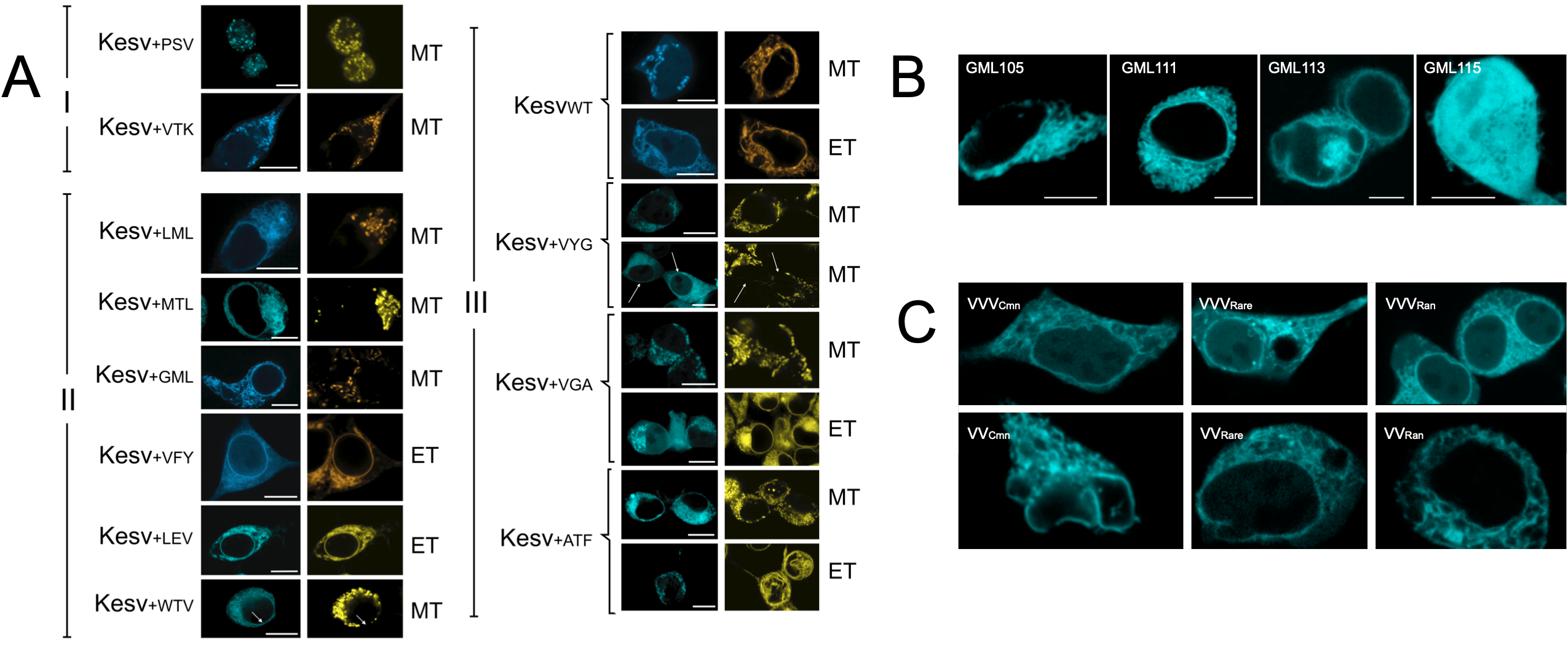
**Supporting information**

**Combination of hydrophobicity and codon usage bias determines sorting of model K+ channel protein to either mitochondria or endoplasmic reticulum.**

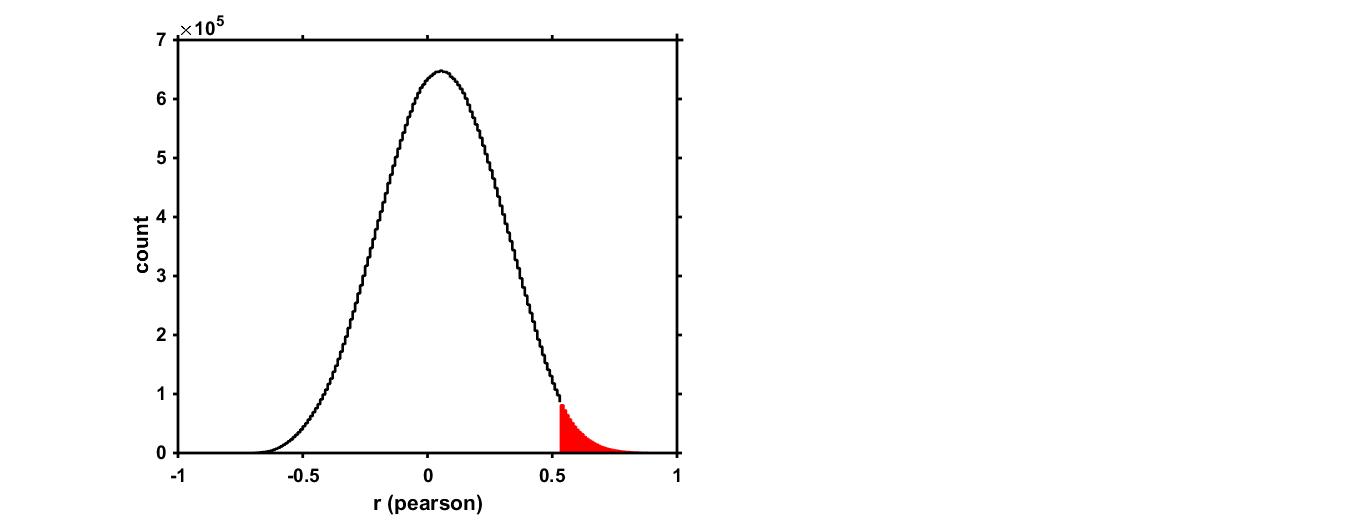
Anja J. Engel1, Steffen Paech², Markus Langhans², James L. Van Etten3, Anna Moroni4, Gerhard Thiel1, Oliver Rauh1.

**Figure S1**



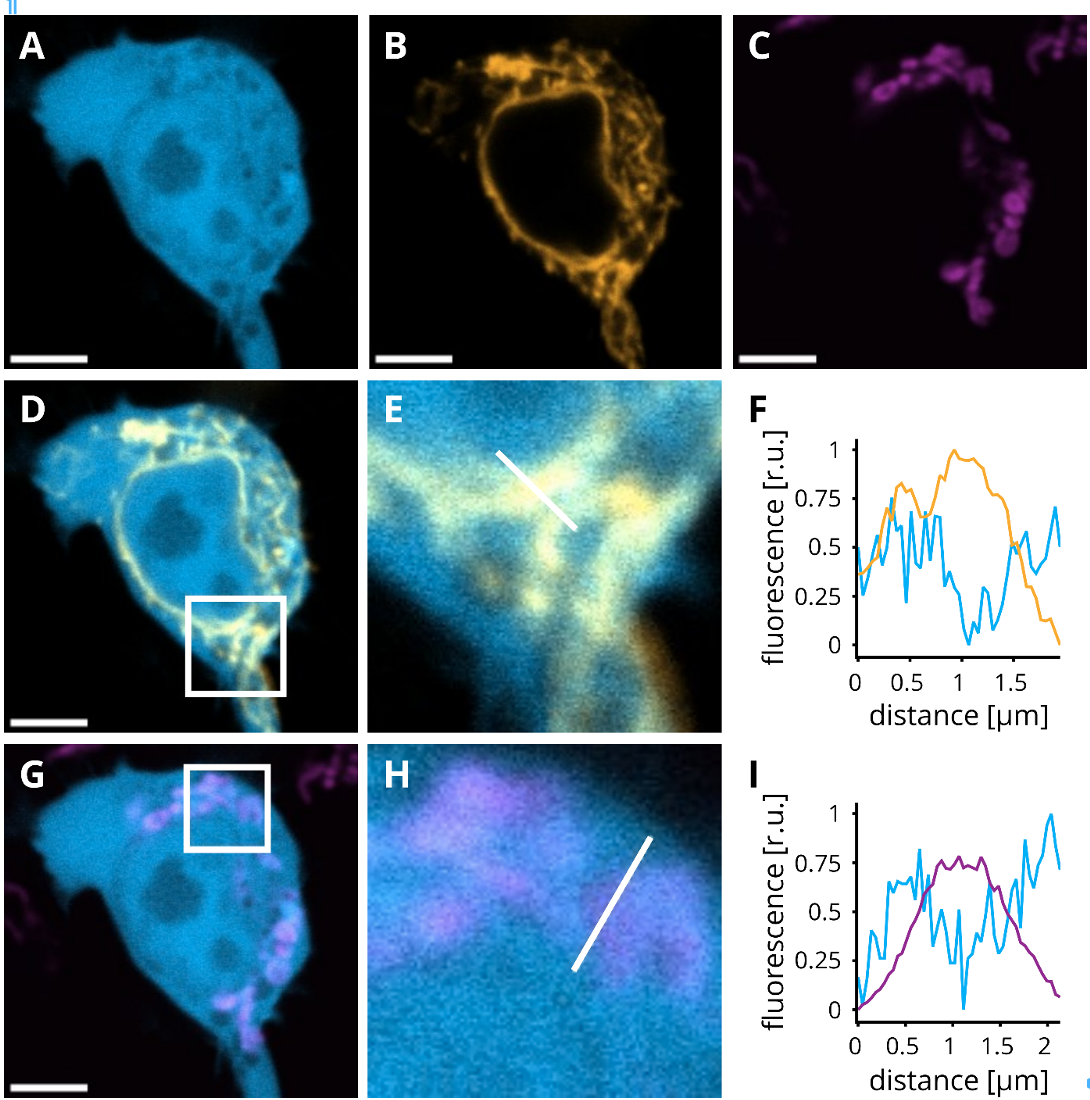
**Figure 1: Sorting destiny of Kesv with insertion of AA triplet in ct-TMD. (A)** Exemplary fluorescent images of HEK293 cells transfected with Kesv+3 in which the constructs are positively sorted according to classification of Fig. 2A to the mitochondria (type I), to the ER (type II) or to either of the two destinations (type III). Images in left panels show fluorescence (cyan) from GFP tagged Kesv+3 constructs given on the left of images. The same cells were stained (yellow) with either ER-TrackerTM Red (ET) or MitoTrackerTM Red (MT) as indicated on the right of images. In the case of Kesv+WTV and Kesv+VYG arrows indicate the presence of staining of the perinuclear ring in the GFP channel and the absence of staining in the MitoTrackerTM Red channel. All scale bares 10 µm. **(B).** Exemplary fluorescent images (cyan) of HEK293 cells transfected with GFP tagged Kesv+3 constructs containing the AA triplets given in individual panels. Scale bars 10 µm. **(C)** Exemplary fluorescent images (cyan) from HEK293 cells transfected with GFP tagged Kesv+GML in which GML triplet was inserted in positions indicated in images. All scale bars 10 µm.

**Figure S2**



**Figure S2: Effect of randomization of CUB values on the Pearson correlation coefficient for Kesv+3 sorting to the ER as function of the product of hydrophobicity and codon usage bias (HyP x CUB).** To address the question if a Pearson correlation coefficient r(pearson) of 0.5459 for sorting of Kesv+3 constructs with randomly chosen codons to the ER as function of HyP x CUB is a coincidence, the CUB values of all 11 constructs from Fig. 1 were randomized before multiplication with the HyP values. The absolute frequency distribution of the calculated correlation coefficients of all possible combinations (11! = 39,916,800) is shown as histogram with a bin width of 0.01. The distribution shows a mean value of 0.0466 with a standard deviation of 0.2349. A correlation coefficient greater than the one of the original data set (r = 0.5459) is found for only 1.36% of all possible combinations (red area).

**Figure S3**



**Figure S3: Fluorescent images of HEK293 cell transfected with Kesv+QQQ in position 113**. Fluorescent images of an exemplary HEK293 cell transfected **A**: Kesv+QQQ::eGFP; **B**: ER marker HDEL::mCherry, **C**: MitoTrackerTM DeepRed; Overlay images of A and B in **D** and A and C in **G.** Image in **E** and **H** are blow ups from white frames in **D** and **G** respectively. **F and I**: Line Plots with normalized fluorescence intensity along lines in **E** and **H** respectively show coincidence of a minimum in GFP signal with a maximum of yellow ER marker  **(E)** or turquoise mitochondria dye **(H)**. Scale bars: 5 µm.

**Table S1: Prediction of ct-TMD fold and length for Kesv wt and insertion mutants Kesv+3.** Start and end of ct-TMD with corresponding number of amino acids (AA) forming the ct-TMD2 as well as the proposed fold, was obtained from structural prediction algorithm DeepTMHMM [Hallgren et al. 2022]. Fold of ct-TMD as well as its length was also obtained from structural predictions by AlphaFold [Jumper et al. 2021]. The length of the respective TMDs in Å was measured as distance in the predicted structure between first and last alpha C-atom in this domain with ChimeraX [Pettersen et al. 2021].

**Prediction from DeepTMHMM Prediction from AlphaF**old

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Protein | AA for start of ct-TMD | AA for end of ct-TMD | Length of ct-TMD in AA | Predicted domain fold | Length of ct-TMD in Å | Predicted domain fold |
| Kesv wt | 101 | 118 | 17 | 𝛼-helix | 25.701 | 𝛼-helix |
| Kesv+ATF | 101 | 118 | 17 | 𝛼-helix | 26.157 | 𝛼-helix |
| Kesv+GML | 101 | 121 | 20 | 𝛼-helix | 29.853 | 𝛼-helix |
| Kesv+LEV | 101 | 118 | 17 | 𝛼-helix | 25.891 | 𝛼-helix |
| Kesv+LML | 101 | 118 | 17 | 𝛼-helix | 25.982 | 𝛼-helix |
| Kesv+LVA | 101 | 118 | 17 | 𝛼-helix | 26.116 | 𝛼-helix |
| Kesv+LVA | 101 | 118 | 17 | 𝛼-helix | 26.115 | 𝛼-helix |
| Kesv+MTL | 101 | 118 | 17 | 𝛼-helix | 26.082 | 𝛼-helix |
| Kesv+PSV | 101 | 118 | 17 | 𝛼-helix | 26.824 | 𝛼-helix |
| Kesv+SKA | 101 | 121 | 20 | 𝛼-helix | 29.633 | 𝛼-helix |
| Kesv+VFY | 101 | 118 | 17 | 𝛼-helix | 26.053 | 𝛼-helix |
| Kesv+VGA | 101 | 118 | 17 | 𝛼-helix | 25.993 | 𝛼-helix |
| Kesv+VTK | 101 | 118 | 17 | 𝛼-helix | 26.122 | 𝛼-helix |
| Kesv+VYG | 101 | 118 | 17 | 𝛼-helix | 25.807 | 𝛼-helix |
| Kesv+WTF | 101 | 118 | 17 | 𝛼-helix | 26.013 | 𝛼-helix |
| Kesv+WTV | 101 | 118 | 17 | 𝛼-helix | 25.998 | 𝛼-helix |
| Kesv+AAA | 101 | 118 | 17 | 𝛼-helix | 26.073 | 𝛼-helix |
| Kesv+EEE | 101 | 118 | 17 | 𝛼-helix | 26.017 | 𝛼-helix |
| Kesv+III | 101 | 118 | 17 | 𝛼-helix | 26.047 | 𝛼-helix |
| Kesv+LLL | 101 | 118 | 17 | 𝛼-helix | 25.723 | 𝛼-helix |
| Kesv+QQQ | 101 | 118 | 17 | 𝛼-helix | 26.085 | 𝛼-helix |
| Kesv+RRR | 101 | 118 | 17 | 𝛼-helix | 26.146 | 𝛼-helix |
| Kesv+SSS | 101 | 121 | 20 | 𝛼-helix | 29.841 | 𝛼-helix |
| Kesv+TTT | 101 | 121 | 20 | 𝛼-helix | 29.700 | 𝛼-helix |
| Kesv+WWW | 101 | 118 | 17 | 𝛼-helix | 25.896 | 𝛼-helix |
| Kesv+VVV | 101 | 118 | 17 | 𝛼-helix | 26.14 | 𝛼-helix |
| Kesv+VV | 101 | 120 | 19 | 𝛼-helix | 27.104 | 𝛼-helix |
| Kesv+V | 101 | 116 | 15 | 𝛼-helix | 22.192 | 𝛼-helix |
| Kesv+KVT | 101 | 121 | 20 | 𝛼-helix | 29.663 | 𝛼-helix |
| Kesv+VKT | 101 | 121 | 20 | 𝛼-helix | 29.630 | 𝛼-helix |
| Kesv+EKE | 101 | 118 | 17 | 𝛼-helix | 25.984 | 𝛼-helix |
| Kesv+KEE | 101 | 118 | 17 | 𝛼-helix | 26.058 | 𝛼-helix |
| Kesv+EEK | 101 | 118 | 17 | 𝛼-helix | 26.000 | 𝛼-helix |

**Referees**

Hallgren, J., Tsirigos, K.D., Pedersen, M.D., et al. (2022). DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *bioRxiv*, DOI: 10.1101/2022.04.08.487609.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* 30(1), 70-82.

**Table S2. List of codons for insertion of one to three amino acids in ct-TMD of Kesv.** Synonymous codons for amino acids (in one letter code) were chosen according to the following criteria:

1. Codons which are most common in mammalian cells (Cmn)
2. Codons which are the least common (Rar)
3. Random choice of codons (Rand)

**Amino acids Codon**

ATFRare GCG ACG TTT

ATFRan  GCG ACC TTC

ATFCmn GCC ACC TTC

GMLRare GGT ATG TTA

GMLRan GGG ATG TTG

GMLCmn GGC ATG CTG

LEVRare TTA GAA GTA

LEVRan  TTG GAG GTG

LEVCmn CTG GAG GTG

LMLRare TTA ATG TTA

LMLRan  TTG ATG TTG

LMLCmn CTG ATG CTG

LVARare TTA GTA GCG

LVARan  TTG GTG GCG

LVACmn CTG GTG GCC

MTLRare ATG ACG TTA

MTLRan  ATG ACC TTG

MTLCmn ATG ACC CTG

PSVRare CCG TCG GTA

PSVRan  CCC TCT GTG

PSVCmn CCC AGC GTG

SKARare TCA AAA GCG

SKARan  TCC AAG GCG

SKACmn ACG AAG GCC

VFYRare GTA TTT TAT

VFYRan  GTG TTC TAC

VFYCmn GTG TTC TAC

VGARare GTA GGT GCG

VGARan GTG GGC GCG

VGACmn GTG GGC GCC

VTKRare GTA ACG AAA

VTKRan  GTG ACC AAG

VTKCmn GTG ACC AAG

VYGRare GTA TAT GGT

VYGRan GTG TAC GGC

VYGCmn GTG TAC GGC

WTFRare TGG ACG TTT

WTFRan TGG ACT TTC

WTFCmn TGG ACC TTC

WTVRare TGG ACG GTA

WTVRan TGG ACC GTG

WTVCmn TGG ACC GTG

AAARare GCG GCG GCG

AAACmn GCC GCC GCC

EEERare GAA GAA GAA

EEECmn GAG GAG GAG

IIIRare ATA ATA ATA

IIICmn ATC ATC ATC

LLLRare CTA CTA CTA

LLLCmn CTG CTG CTG

QQQRare CAA CAA CAA

QQQCmn CAG CAG CAG

RRRRare CGA CGA CGA

RRRCmn CGG CGG CGG

SSSRare TCG TCG TCG

SSSCmn AGC AGC AGC

TTTRare ACG ACG ACG

TTTRan  ACC ACC ACC

TTTCmn ACC ACC ACC

VVVRare GTA GTA GTA

VVVRan  GTG GTG GTG

VVVCmn GTG GTG GTG

WWW TGG TGG TGG

VRare GTA

VRan GTC

VCmn GTG

VVRare GTA GTA

VVRan  GTA GTG

VVCmn GTG GTG

KVTCmn AAG GTG ACC

VKTCmn GTG AAG ACC

EKECmn GAG AAG GAG

KEECmn AAG GAG GAG

EEKCmn GAG GAG AAG