

SUPPORTING DATA

TFAM DETECTS CO-EVOLUTION OF tRNA IDENTITY RULES AFTER LATERAL TRANSFER OF HISTIDYL-TRNA SYNTHETASE

By David Ardell and Siv Andersson

Supplementary Methods: Construction of MSDB and cross-validation comparison of TFAM to other methods.

Supplementary Dataset 1: Alignment of data used to make Figure 4 in PHYLIP format.

Supplementary Figure 1: BIONJ tree of recent genomically annotated HisRS and HisRS-like sequences for the phylogenomic analysis used to make Figure 4.

Supplementary Figure 2: Maximum likelihood tree of 251 sequences combining PSI-Blast queried with *Agrobacterium tumefaciens* HisRS with genomic HisRS sequences. The figure is meant to be examined electronically as a PDF file. PROTML RELL values are shown at nodes, and GenBank gi numbers are shown at the end of the description lines.

Supplemental Methods. Creation of Modified Sprinzl Database. The 1999 version of the Sprinzl database (820 sequences) was back-translated into DNA sequences using a Perl script called SPRINZL2FAS accounting for posttranscriptional modification codes and then compressed into 663 unique sequences. Of these 663, the 5 sel-cys tRNAs were removed and three Ala tRNAs that lack the G3-U70 base-pair which is extremely important for Ala identity (cite Martinis and Schimmel). All other Ala tDNAs had a G3-U70 pair. The exceptional removed Alas were RA1540 from *Bacillus subtilis*, ref. (H.ISHIKURA, K.MURAO, Y.YAMADA, EMBO-FEBS MEETING, STRASBOURG, JULY 1980). A search of *Bacillus subtilis* tRNAs from the genome (tRNAscan-SE) does not indicate any tRNAs with this sequence. The discrepancy could come from a sequencing error in the tRNA, in the genome (less likely) or a strange mutant. Similarly lacking the G-U pair against genome evidence were RA1661 (R.J.WILLIAMS, W.NAGEL, B.ROE, B.DUDOCK (1974) BIOCHEM. BIOPHYS. RES. COMM. 60, 1215-1221) from *E. coli*, and DA1730 (A.WILMOTTE, J.NEEFS, R.DEWACHTER (1994) MICROBIOLOGY 140, 2159-2164) from TRICHODESMIUM SP

The remaining 655 tDNAs were split into pregenomic (415 tDNAs) and postgenomic (240 tDNAs) datasets. Sprinzl annotations in the pregenomic dataset were screened in three ways: 1) internal consistency by exact sequence matching, 2) tRNAscan-SE-based anticodon identification, 3) internal consistency by profile-scoring with TFAM. Using criteria 1 and 2, 1 back-translated tDNA was a duplicate of another tDNA in the pre-genomic set with conflicting identity annotations. Using tRNAscan-SE to identify anticodons and disambiguate the identities, DC1230 and DK1230 were assigned Lys identity. TRNAscan-SE found two more incorrect annotations: DI1821 with an ala anticodon and DE1201 with a phe anticodon. TFAM was then used, with this corrected pregenomic dataset as both training and test data, which confirmed these annotation changes, indicating that these sequences have typical bacterial identity elements as well as anticodons. In addition, TFAM suggested these additional changes, which were inspected, and changed or not changed for the following reasons: RM1580 as an initiator (identical in

sequence to RX1581, confirmed with TFAM as an initiator, has bacterial conserved C1:A72 basepair for initiator identity); DM1231 as an initiator (initiator by TFAM, has conserved C1:A72 basepair); DM1660 = DM1750 = RM1660 (“=” indicates “identical in sequence to”, annotated as Ile by TFAM, left as Met because of RNA evidence in SprinzlDB); DN1230 (Thr identity in TFAM, but confirmed AIn anticodon by tRNAscan-SE); DV1230, DV1480, DV1540 = RV1540, DV1720 and DV1570 (Ala by TFAM, but lacks G3-U70 basepair, so left as Val), DV1661 = RV1660 and DV1662 = RV1661 (Ile in TFAM, left as Val because of RNA evidence). After these changes, a resubstitution analysis with TFAM converged (results were the same as the last round).

The postgenomic data was then analyzed with tRNAscan-SE (prokaryotic tRNA covariance model), with TFAM using the converged pregenomic annotation training data, and with BLAST. To use BLAST, the pregenomic data was made into a database with **formatdb**. Postgenomic data was BLASTed against this database with **blastall** and parameters to increase sensitivity (**-q -1 -r 1 -W 7 -G 2 -E 1**). The results of this analysis were compared with the output of TFAM and tRNAscan-SE. In the cases in which the results of all three analyses did not agree, the tRNAscan-SE annotation was preferred except when the tRNA in question had a CAU (Met) or UCA (STOP) anticodon or was annotated as a Pseudo-tRNA by tRNAscan-SE, in which case the BLAST annotation was preferred.

Supplementary Table 2

	Accuracy
HMMer Local	(638/655) 97.4%
HMMer Global	(637/655) 97.3%
TFAM	(636/655) 97.1%
COVE	(636/655) 97.1%
Infernal	(623/655) 95.1%

Results of “leave-one-out” cross-validation analysis comparing the ability of different methods to recover the annotations assigned in the Modified Sprinzl Database (“Accuracy”) generated as

described in the preceding section.

Parameters for analysis:

HMMer Local: Separate HMM models were made for each identity class and each leave one-out-dataset. Models were built using **hmmbuild -s -nucleic** from training data separately prealigned with **covea** and the TRNA2-prok.cm covariance model (Lowe and Eddy, 1996). Data scored against all models using **hmmsearch -A0**. The model with highest score was taken to be the annotation.

HMMer Global: Separate HMM models were made for each identity class and each leave one-out-dataset. Models were built using **hmmbuild -g -nucleic** from training data separately prealigned with **covea** and the TRNA2-prok.cm covariance model (Lowe and Eddy, 1996). Data scored against all models using **hmmsearch -A0**. The model with highest score was taken to be the annotation.

TFAM: Coveam models were made for each leave-one-out dataset as in the text, and ran with **tfam -m**

COVE: Separate covariance models were made for each identity class and each leave one-out-dataset. Models were built using **coveb** from training data separately prealigned with **covea** and the TRNA2-prok.cm covariance model (Lowe and Eddy, 1996). Data scored against all models using **coves**.

Infernal: Separate covariance models were made for each identity class and each leave one-out-dataset. Models were built using **cmbuild** from training data separately prealigned with **covea** and the TRNA2-prok.cm covariance model (Lowe and Eddy, 1996). Data scored against all models using **cmscore**.

Results: The TFAM methodology performs as well as COVE but is faster, because of the incorporation on COVEMF into TFAM. TFAM performs

better than Infernal and slightly less well than HMMer. The advantage of TFAM over HMMer is that TFAM outputs alignments of all data tRNAs to a models that are themselves mutually aligned. This can then be used to analyze tRNA identity determinants such as in the accompanying article (Freyhult et al.)

```

77 437
15612180 ITPKVLSGFK DRLPKDAIQK AQLLAKVSVV FQSFQFVPIE TPHLEYAQL LPDASS---D
34483510 ITPRTLSGFK DRLPQEMAK SRLLRVTSEV FEGFGFVPIE TPHELYAEIL VKQGD5---E
34397963 QKPS1PKGMR DFPVEMSRK NYIFDITREV YELFGFNQIE TPTMEMSLT MKGYGE---E
29347250 AKPS1PKGTR DFPVEMAKR NYIFDITREV YHLYGQQIE TPTMEMSLT MKGYGD---E
23137435 EKPS1PKGTR DFGPAQMLKR QYLLQITREI FTKYGFPIE TPAMENSLV TKGYGD---E
Bhae-RS TKARLPRGFV DRTSAQLYAI EVMIAQIRQV YELYGFEALE TPIFEYTDVL --GKFLP---D
Bana-RS TKARLPRGFV DRTSAQLYAI ETMIAQIREV YELYGFEALE TPIFEYTDVL --GKFLP---D
Atsh-RS LKARLPRGFV DRSAADIHAT NEMWDKIRRV YELYGFDPIE TPLFEYTDAL --GKFLP---D
Smti-RS LRARLPRGFV DRSAADIRAV DEMIAKIREV YERYGFDPIE TPLFEYTDAL --GKFLP---D
Bs30-RS MKARLPRGFV DRVPDDLRAA EKMMATIREV YDLYGFEPVE TPLVEYTDAL --GKFLP---D
Mlti-RS TKARLPRGFA DRSAEDIRAV EKMMATIRSV YELYGFEPAD QPLIEYTDAL --GKFLP---D
Bjum-RS LKARLPRGLE DRDPAAIRAT REMVEKIRAV YELYGFEPVE TPAMEYTDAL --GKFLP---D
Apix_1 GRPRPPRGFR DFPPEVMILR KRLLSRLREV FESYGFDPLO TPAVEYWETL SGKYGEE---A
Bmis_2 MKARLPRGFV DRVPDDLRAA EKMMATIREV YDLYGFEPVE TPLVEYTDAL GKFLPDQ-DR
Rp09_1 LRARLPRGLA DRGPAEIAAT RAMVEKIREV YERYGFEPVE TPAFEYTDAL GKFLPDQ-DR
SpS3_1_3 PKAETPKGFR DYFGAETVQR TEMLRDIAGV YHRYGFEALE SSVAVETVEAL GKFLPDVD-R
GoH1_1 SGLQPPRGTH DLIGETMRRH HHVVTARRV LRTYGFEEWS TPFIETDRVF SRTLGETSDV
ZmM4 -----TGR NSGFRITAVF ARSLGETTDV
Weer_1 MTNQTVRGTK DLLDFEWYKF KYIEQTANRI SLYGFLPAQ TPICEFTEVF TKTLGDS5DI
WeRS_1 MTNQTVRGTK DLLDFEWYRF KYIQIANTRI SLYGFLPAQ TPICEFTEVF MTKLGD5SDI
Erel_1 NKLREVRGTK DLLDDELRYF QYIQLHAQTI ASRYGFIPOV TPIIEFTEVF TKLGNTS-DI
Ames_1 GKLPVGRGTR DLLPEECYKF WHIRDVAHDI GERYGFVPIE TPIFEQDQVF LKTLGDS5DI
Ccus-RS PEARAPRGFA DKRARDLRAE RAILLAESAV YERYGFEALD TGAFAYADAL --GKFLP---D
14600873 GRPRPPRGFR DFPPEVMILR KRLLSRLREV FESYGFDPLO TPAVEYWETL SGKYGEE---A
22972087 APTQNVKGMR DHPADMMMLR QYIWNLTTRI SESHGFEPLQ TPIIEYAEIL EGKLG---D
15838813 IKPRTPPGVL ELLPREQIAF QRMLDVIIRN YERFGFLPVE TPVFELSDVL LTKSGG---E
21231254 IKPRTPPGIM ELLPREQIAF QRMLDVIIRN YERFGFLPVE TPVFELSDVL LTKSGG---E
15895997 IKPSILPGFM ELLPKEQLVF NDIIVSKITGV YEONGFLPMO TPIIEKEEVL LAKSAG---E
28211584 VKPSILPGFM ELLPADQIVF NKMKDITRHN YEKFGFIPLO TPIIEKEEVL LAKSGG---E
17232504 INFSTPSGFP EFLPSEKRLLE LYLLDITRIRV YESYGFPTIE TPAVERLEVL QAK-GN---Q
23464645 AGKASISGFP EWLPSERVVE QRVIDTLRKL FELNGFIEI TRAVETGASL LKK--G---E
23019288 IRPPTISGFP EWTPTIRSVL LRWLDHIRRG FERYGFSVSE TPSVEVLVDL LSK--G---E
32470802 IQPRTLKGRF DYLAAMIPRQ EQLMOTAREV FRSFGFAPID TPTLEHLEIL TKGSGD---E
15639628 VSPVKLKGFR DLLPDEEIER ALLVEKLTVA LRQMGFVPID TPALYEYDEL LTKSEG---D
15594480 MDIKTLKGFK DYLPKDSLIR IHIVRQIFSV LNSYNFDLID TPVLEYSDDL LKKSAG---E
18976636 --MIERVKGTR DFLPEEMVKR RWFVERIRER FETYGFKEVL TPTMEYKFL ELRSGE---E
Meja8406 VMFQKPRGTR DFLPEEMKKR RFVENKLRER FERYGYKEIL TPTFESFELI AKKTGE---E
ArFu8631 MKIERPRGTR DFLPEEMERR REIEKRMRKI AESFGYREVA TPTFYELELF TRKSGE---G
15678272 MDISRPRGTR DFLFAEMRNR KEVEGLRRT FETYGHEIK TPIFEELKLF TVKSGE---E
15231337 IDWNPVKGTR DFPPEDMMLR NMLFNHFKRV SRYGYEEDV PVLETEALF IRKAGE---E
16081268 LQIEKIRGFR DFPYEDMDVE KFIFKTAEEA AEAFFRRID FPSLEYLDLY RIKSGE---E
13540985 MQIERIKGFR DFPYEDMIEI KFIFTEASSV AESYGFRRID FPSLEYLDLY RIKSGE---E
15790871 --MYDRKLGFR DFPYPAEMGPR RAAIDAIEDA AASYGFREVQ TPAMERTEMY VDKSGA---E
24212754 LTTAPYKGRF DFPYEDMMLR NMMFVSMRET VLSFGYEEYD GPILSFDLY KAKSGEE---I
6996014 FVLKTPKGRF DYSPRQMAVR EKVFDVIIRC FKRHGAEVID TPVFEKTEL MKGYGE---D
15029520 FIITKPKGRF DLSPQMVVVR EKILDVIVSC FKRHGAGQMD TPAFELKTEL TKGTE---D
27683037 ISCFKRHGAK GLDTPAFELK EMLTEKYEDN FG--LMDLE DQGGELLSLR YDLTVSFR-R
24643061 FTLKTPKGRF DYGPQQMTRL QGVLDKIVQV FKRHGGEAID TPVFEKTEL TKGYGE---D
4103154 -----GTR DFGREQMAIR ERAF5IITSV FKMHGVALD TPVFELETL MKGYGE---D
6325290 VSLKTPKGRF DWADSDMVR EAI5FSTL5GL FKHGGVTID TPVFELETL AGKYGE---D
32413441 LQIKTPKGRF DWHPDMLLR EQLFQTVSDV FKRHGQPLD TPVFELETL AGKYGE---D
19074312 TSLKTPKGTM DESPRQCFLH EEIIRIRRV FVLYGAVPIS TPTFEMKSL TNKYGE---D
21401251 MEIRNVKGTK DYLPDEEQLR SKIKRACEDT FERYGCKPLE TPTLNMVELM SYKYGGG-DE
Aaus-RS MNIQSVRGFH DILGDKAKKF RKISDTARKI LKLYNFEIIL LPVVEYAEFL QRSVGETTIDV
18310916 MAIQKQKGTK DMLPNDAIKV HYIEEKLRIK SAEYGIERIR TPTMEATELF KRQVGETTIDV
21402447 MSIQTPRGTK DILPGTVELW QYIEGQAREI CRRYNYKEIR TPIFEHELFL LRQVGETTIDV
Bsis-RS MGYNTPRGTK DILPGESDRW QFVEQIMRDT CRTYQYKEIR TPIFEHELFL ARGVGETTIDV
Ec12-RS KNIAQIRGMN DYLPGETAIW QRIEGTILKNV LGSYGYSEIR LPIVEQPTFL KRAIGEVTDV
SP03-RS GAIQAIRGTR DILPPETNYW QWVEIAKSI LDRALYQEIR TPIFEQTSLF ERGIGEATDV
22969029 AALQPVRGTH DYLPDEARSH RHVEETARTV AERFGFGEII TPIFEFDVDF ARTLGD5SDV
Rpi1-RS ENLQPLRGMK DLLPNDYQIH NYIINKARDV GALYGYQMS TPILYTKVF NRSMGESSDV
Rci1-RS EKLQPLRGMK DLLPDDYKWH DYIINKARDV GVLGYQMS TPILYTKVF NRSMGESSDV
Q8R880 --MIYLPDGVQ DFLPEEYEFK RNIEDKFREV FKSFGYKEIM PPTFEYSENF SHLFD5---D
BaAnth-RS2 WKRANPNGTR DYLFEECTLI EEEVQLKRRT FLERGYEIER TPTIEFYDVF AFQSRP---I
034459 FMFEKPHGMR DTLPLGYETK KKVRRSLTDL IDKWGYQFME TPTIEFYDVF GVQSAIE---D
067223 MKTSLISGER FYSLEESNLV KEIFKAVKVI FEAFGYDINL LSHFPEYEFQ ELTFGE---D
Q55267 MVHQPPAGTR DLLPQDVTQK RWIESRLQQV FQWQYQRII TPTLERLDTL VAGGAV---D
Q9RUE3 STSALPAGVR DVLPAAEWERR EHLRSHLSD LRSWGYRGVD LPALELADPA HPQGNH---D
Q9HUM5 DRWLLPDGIE EVLPPEAARI EAARRVLDL FHRWYEFVY TPHEIYELSL LTGAGQ---D
Q9JV26 QTWQLPEHIA DVLPNARQL ESAREQLLAL FRVHGYELVQ PPLMEYHSL LTHIDAG---L
RhodRub-Z ARALLPNGLR DVLPDAAFE VATVERLVSV FAANGYERVK TPLIEFEGL LTAGSR---A
Bmis-Z ----- --MTMVGSR TSPFNALRVE LNAREALVE IPLIQPADPF LDMAGED---D
Atsh-Z ----- --MPLIDM PEFAGELLE EFAARRTSRVN TPVIQPAEPF LDMAGED---D
Smti-Z ----- --MPLINL PAFAGDLLAD FERRNTRVD TPVIQPAEPF LDMAGED---D
Mlti-Z ----- --MTRSRF PATAANITKL FAARRNTHAVE VAILQPADPF LDMAGED---D
Ccus-Z --MRCARSSG VEVLRSRISIP APALDAIRAP LLDAGAVLTD APVLQPLGLL LDLAGEA---D
RhodSphe-Z AAGGSRRP6GA GHALVTKAAV RAAEALFAA FQAAGALPVE ADILLPAGTL LDLYGED---D

IQKELYRFKD HGDRDVALRF DLTVPARFV SLHHQMPFKR YAIQNVFRGE RAQKGRYREF
IQKELYRFDD HGGDRDVALRF DQTVPLARFI SQHRNLPFKR YAIQNVFRGE RAQKGRYREF
GDRLLFKVLN SGDCEKGLRY DLTVPFARYV VMHR-FPFKR YQIQPVWRAD RPQKGRYREF
GDKLLFKIQN SGDCEKGLRY DLTVPFARYV VMHR-FPFKR YQIQPVWRAD RPQKGRYREF
GDQLYKILN SGDSKALRY DLTVPFARHV VMNR-FPFKR YQIQPVWRAD RPQKGRYREF
EDRPNAGVFS LQDQWMSLRY DLTAPLARYF SENF-KPYRS YRLGFVFRNE KPGGGRFRQF
EDRPNAGVFS LQDQWMSLRY DLTAPLARYF AENF-KPYRS YRLGFVFRNE KPGGGRFRQF
SDRPNAGVFS LQDQWMSLRY DLTAPLARHV AENF-LPYRT YRAGYVFRNE KPGGGRFRQF
SDRPNAGVFS LQDQWMSLRY DLTAPLARHV AENF-LPYRT YRAGYVFRNE KPGGGRFRQF
QDRPNAGVFS FQDQWMSLRY DLTAPLARYV AENF-KPYRS YRNGWVFRNE KPGGGRFRQF
QDRPNAGVFS FQDQWMSLRY DLTAPARFV AENF-KPYRS YRSGWVFRNE KPGGGRFRQF
QDRPNAGVFS FQDQWMSLRY DLTAPLARYV GERYGLPYRS YRNGWVFRNE KPGGGRFRQF
ESRLIWRFQD PWGRNYALRY DLTVPARFV ASNPLPFKR YHIGKVVWRHE EPQKGRYREF
PNEGVSFQD DDEQWMSLRY DLTAPLARYV AENFETPYRS YRNGWVFRNE KPGGGRFRQF
PNEGVSFLQD DDEQWMSLRY DLTAPLARYV AENFDQPHRS YRFGWVFRNE KPGGGRFRQF
PNEGVSFAWQE SEGDMALRY DLTAPLARYV AQHRNDPYRR YAMGPVWRNE KPGGGRYRQF
VSKEMYTFED RGGESITLRF EGTAACRACL VTNGLTPQVK FYHGSMPFRYE RPQKGRYRQF
VSKEMYSPFD RSGEQLTRP EFTAGLTRAY LSTGWQPCIK ATHGMPFRYE RPQKGRYRQF
ITKEMYSFND KGGKSITLRF EFTAIVRLL IEKKLQPIKL FSTGPAFRYE RPQKGRQRF
INKEMYTFYD KGGKSITLRF EFTAIVRLL VEKKLQPIKL FSTGPAFRYE RPQKGRQRF
VTKEMYNFKD KAGEDITLRF EFTAIVRLL ISKNLTPVKL FSSGVPFRYE RPQKCRQRF
IGKEMYSPFD RGGDVLVLRP ELTAAVARML ICERLTPARL FTFGPVFRYE RPQKCRQRF
SDRPNAGVFA LQDQWMSLRY DLTAPLARFA AQNW-KPFKR YAFGPVWRNE KPGGGRFRQF
EEAESRLIWR FQDRNYALRY DLTVPARFV ASNP-MPFKR YHIGKVVWRHE EPQKGRYREF
EEK---LIYR FEDRRALRY DQTVPLARVT AQYAGLWRR YAYGPSYRGE RPARGRYREF
TERQVYVQS TGPELALRF DLTVPARFV AEYEHFPRR YQMQRYRGE RAQGRFRREF
TERQVYVQS TGAPELALRF DLTVPARFV AEHEHFPRR YQMQRYRGE RAQGRFRREF

```

TEKQVYRIDN E--RKQVLR DLTPVPSRFA AQYMSFFPKR YQLGKVRGE RNQKGRYREF
TEKQIYRFNK G--TDLSLRF DLTPVPLARYV SQHFSFFPRR YHISKVFRGE RNQKGRFREF
GDNIIYGLPE ILPEARALKF DQTVPLAAYI ARHLNFPFAR YQMDVVFRGE RAKDGRFRQF
TSKEIYLLSR LQEERLGLHF DLTPVLSRYV VEHSGFFPKR WQIQKVVWVRGE RPOQGRFREF
TSQEIYTLQR LQAARLGLHF DLTPVFARYV AQHFNFPFKR YQIQRVWVRGE RPOQGRFREF
TDRQLYSFED NNGRPPVGMRF DLTPVLRFA AQHIGTFPKR YHIAPVWVRGE KPOQGRYREF
TEKQMFVFDV KGGDRVALRF DLTPVPLARFV ATHYAFPFKR YHFAKVVRGE KPQMGYREF
TEKQIYRFNK NNGRDSVSMRF DLTPVPLARFV ATNLSLPPFR SQFGKVRGE NSQKGRYREF
VVKQLYAFKO KGGDRVSLRP DMTSSVARYL VNMFKPKIKW YYIANMFRYE EPQSGRYREF
IRKQLYVFKO HGGREMLALRP EMTSPVVRFY LNELKKLRLR YYFANCFRYE RPQAGRYREF
IIEEMVYFKO KSGRDLALRP ELTAPVMMRF VNECSKPLRF YYFANCFRYE RPQKGRYREF
VWNIYHFD KGGRELALRP ELTAPVARYL MNEMQKPKIKM YYFGSCFRYE RPQAGRYREF
IRDQLYCFED RGNRRVALRP ELTPSLARLV IQGKGLPLKW FAIGQCVRYE RMTGRRRREH
LLQQTYSFVD KGGREVTLIP EATPSTVRML TSKRDLRPLW YSFPKVMRYE EPQAGRYREF
LLNQYYSFVD RGGREVTLIP EATPSTVRML TSKRDKPVWV YSFPKVMRYE EPQAGRYREF
IVDELYSFVD QGGROVALTP ELTPVARMV VAKQKPKIKM YSSRPFVRYE EPQAGRYREF
VERQLYDFID KGERRVAIRP EMITPLARMV AGNLRKPVWV FSIPLWRYE QPKGRLREH
DSKLYDLDK QGGELLSLRY DLTPVFARYL AMNKLTKIKR YHIAKVYRRD NPAMGRYREF
DSGLMYDLK QGGELLSLRY DLTPVFARYL AMNKLTKIKR YHIAKVYRRD NPAMGRYREF
MQEPDLPK SRGLAENRG TLTPVFARYL AMNKLTKIKR YHIAKVYRRD NPAMGRYREF
DSKLYDLDK QGGELLSLRY DLTPVFARYL AMNKLTKIKR YHIAKVYRRD NPAMGRYREF
DSKLYDLDK QGGELLSLRY DLTPVFARYL AMNKLTKIKR YHIAKVYRRD NPAMGRYREF
DSKLYDLDK QGGELLSLRY DLTPVFARYL AMNKLTKIKR YHIAKVYRRD NPAMGRYREF
DSKLYDLDK QGGELLSLRY DLTPVFARYL AMNKLTKIKR YHIAKVYRRD NPAMGRYREF
DTKLTYDLDK QGGELLSLRY DLTPVFARYL AMNKLTKIKR YHIAKVYRRD NPAMGRYREF
ILKEIYTLQD QGKRDALALRP DLTPVFARYL AMNKLTKIKR YHIAKVYRRD NPAMGRYREF
VQKEMFVFD RKGKRLALRP EGTAGTVRAF IQHKLPLYK FYEGPMFRYE RPQAGRYREF
VQKEMTYFED KGGRSITLKP EGTAPVRAF IENSLQPTKM FYFTPCFRYE KMQKGRLEF
VQKEMYSFQD RGRSLTLRP EGTAPVRSY VENKMQPTKL YYIGQMFRYE RPQAGRYREF
VQKEMTYFED RGRSLTLRP EGTAAVRAF NENKLOPTKL YYVGMFRYE RPQAGRYREF
VEKEMTYFED RNDGSLTLRP EGTAGCVRAF IEHGLQEQRL WYIGPMFRYE RPQAGRYREF
VQKEMYSFQD RGRDITLKP EGTAGVRAF IEQNLQVQL WYIGPMFRYE RPQAGRYREF
VTKEMTYFED RSGDSLTLRP ENTAGVARAF ISGGLVPVKL FYRGMFRYE RPQKGRLEF
MSKEIYFVD KSNNAVALRP EFTSGIRSF ISNGLLPLKF FSTGPVFRYD RPQAGRYREF
ISKEIYFVD KSNDFVALRP EFTACIIRSL ISNRLPLKF FSTGPVFRYD RPQAGRYREF
--NSMYFFD KKGILALALRP DVTAQIARIV STKLEYPLK CVYANVRYE DTQGMFRYE
DEEKMYFFD EKGRIVLALRP DMITPLARV GTQRCTPLKV TYSGNVFRAN ESLAGKYNEI
-EQQLFKLD QDGKTLVLRP DMTQPARVA ASKLLHPLRV GYAANVFRQ EREGGRPAEF
GSKAITFKD SYTESFGLRL DFTTQVVRTI SHLRNLPERV YFGKVFSLD RRGF---EK
QRSVAIVQVS DEESGLLRP ELTASIAARA VTRLALPLRL YYLANVFRPA FQDGLQOREL
----AFKLD SGGVVALRSL EYTTALGRVL GTHFPPLRL YQGGRLWLR QTSEGRLEF
LDRITLVTD PASRLMGFRA DITPQVARMV AHSRGPVSR CYAGSVLHAQ PRAL--TSRSP
SLKTLVTD LSGRQLGIRA DITPQVARMV AHSRGPVSR CYAGSVLHAQ SDGLNMRP
TAGQFRVMD PITRMGLRA DITIQUARLA ASRLARPLRL TYSQVLRVK GTQL--PERQF
LRRRIFL TEN ENGSCLLRP EFTIPVCRNH IALNATPKRY AYLGEVFRQ RDGA---AEF
LRRRIFL TEN ETGESCLLRP EFTIPVCLRH IETATPKRY SYLGEVFRQ REGA---SEF
LRRRIFL TEN ETGESCLLRP EFTIPVCLRH IETATPKRY AYLGEVFRQ REGS---SEF
LRRRIFL TEN ETGTQLLRP EFTIPVCLDH ISSQATPRY SYLGEVFRQ REGG---NEF
MRSRLVFSV EAGEACLRL DFTVAVARQH FEGGGAGRR RYEGKAFRVA PRGSDRAEEF
IRARAYVTD PLRGMMLRP DFTVAVARQH MAEGAEPARY TYMGVFRQ EHAGPRAEY

TQCDPFIGS ESLVDAEII QVIAASLAL DLEDFCVSIN HRKILNGICE YFGAQ---VN
TQCDPFIGS ESIGSDAEII QVYASLAL GIERFTISMN NRKILNGICD YFGKEQ--TS
YQCDGFIGS DLSLNEVELI QITSEVFRLL GIR--TRILL NRKILSGIAE VVD--EADRLT
YQCDADVGS DLSLNEVELM QIVDTVSRF NIR--VCIKIN NRKILSGIAE IIG--EADKIV
MQCDADVGT DLSLCEAEIV LMIHEAFAA GITDFVVKIN NRKILSGIAD AIG--KEGSEA
MQFDADIVGT PTVAADAEIC MMAADSLQKL GFQHYVIRLN NRKILDAVLE KVGEOQRRL
MQFDADIVGT PTVAADAEIC MMAADSLQKL GFQHYVIRLN NRKILDAVLE QIGKAEKRL
MQFDADIVGA AGVQADAEMC MMAADTMEAL GIARYVIRVN NRKVLGVME AIGDNAGRRL
MQFDADIVGA AGVQADAEMC MMAADTMEAL GIARYVIRVN NRKVLGVLE AIGQNTNRL
MQFDADIVGA PNVSADAEMC MMAADTLERL GIQRYAIRVN NRKVLGVLD AIGGNAAKRL
MQFDADIVGT PGVAADAEMA MMAADVMEAL GIKRYVIRVN NRKVLGVLE AIGENIGRRL
MQFDADIVGS ATPAADAEIC MMAADTMEAL GVQRYVVKVN NRKVLGVLE AIGENARRL
YQCDADIVGS PYPEADAIEV SLAVDSIDAL GLPGFRVRVN DRLLLAGVFE EELG--FDNPL
MQFDADIVGA PNVSADAEMC MMAADTLERL GIQRYAIRVN NRKVLGVLD AIGLNAAKRL
MQFDADIVGA PTAADAEMC MMAADTMEAL GIPRYVVKVN NRKVLGVLE SIGLNAGRRL
YQCDADIVGS ASMAADAIEC AMLSDTLETV GIPRYVVKVN NRKVLGVLE TMLDEGQRD
HQIDAEIIGA ESPLRDAEII AMAQDILREL GLGD----H VTLNLTLDG LASR--QAWRE
HQIDAEIIGA AEPEADVELL SLADQLLESL GIKQ----H VTLNLTLDG SASR--EAWRA
HQINFEVFGI EDPKADIELI SLAQHLTEF GINK----N VKLEINSLDG GETI--KYRE
HQINFEVFGI ENPKADIELI SLAHLTEF NISK----N VRLEINSLDG SETI--AKYKE
HQINFEVFGS DSPADVEMI SLCYNISEL KLSL----R IKLEINFLDG QETI--SYKI
HQINFEVFGI G--CTADAEIM ALAYDILGAL NLRVQLEIN SLGNQDSILG YRNS-----
IQCDADIVGS ARPEADAIEI AMAVEGLEAA GLPRAVLKIN NRKLNGLLT AA--DSAQKL
YQCDADIVGS PYPEADAIEV SLAVDSIDAL GLPGFRVRVN DRLLLAGVFE EE--DNP---L
YQFDLIVGS ASPLADAIEV AMLCEALAT GFPGFVTLN HRQIIGGIAR VSGEAG---
YQCDIDVIG QSIRYDAEVL AVIHAVFSEL GIGDFQVQLN NRKVLGRFLE SQGRDGLQL
YQCDIDVIG QSIRYDAEVL AVIHAVFSEL GIGDFQVQLN NRKVLGRFLE SLGAEGLQL
YQCDVDVGN GSIKNDAEII NMASKALRKI GLDSYKQIS NRKILTVGLE GLNTN---MQ
YQCDIDVIG GSIINDAEIP SIIYQFKEL GFEDFTIRIN NRKVLGNFFE GTDED--RK
YQCDIDVGR ESLLYDAQMP AITTEIFEAV NIGDFVIRIN NRKVLTFGFQ SLNSETQ--IK
VQADIDVIGA GPDHYEVELP LVMVEELRAY GLPKATVHAN NRKLVSEFYR GLGTD---VE
TQCDIDVIN DPLHFAELP RIVHEVLTP DIPAWLNIN NRKVLQGFYE GLGTD---PL
YQCDPDTIGT TSELADIEAV CVIDALRAI GIDAFITISIN NRAILTGLLE SLG--LADKTT
TQCDPDIAGS DSVCAFDEL KSIK--MLYMA GAH--IHVA HRGLFDRFLR ALS--LSDQAE
MQFDPIVGE DTFRGADEL SVVYEIIFLN FIEGFIHYS HIGLNSFFE KLG--LKEKSI
WQAGVELIGS DKVEADAIEV ALFVESYLAT GLREFTVNIQ DRILLDEFK MLG--VKDDI
WQMGCELIGS KEPLADAIEV NLAMDGLINI GLD--FDVHIG HLGVLKGVLE KFNVSSEEEV
WQFVELIGS ESYLADAIEV ILADKILKDV GVN--FSLEIG HVGIMRHLK PIG--EDRAS
WQFCELIGS RSPLAIEV SLASESLHEL GLDGEVHIG HLGILRILG REEIPSELQD
YQNMIDVIG PQVTAIEV SSVITFFKRI GITAVGKVS SRKVLQELK KYGVPEDLFG
YQFNADIFGS DSPEADAIEV ALASSIDLRL GLQDYIEIRIN SRKIMEIIG GMT--SSDFP
YQFNADIFGP SNEEADAIEI SLASGILDGL GLSGYEIRVN SRIMMEDILN GMG--IADPY
YQTNLDIFGT SDPRADAIEV AVAADGLTSL GLTGFVFRV HRDILGGLLA AFD--ATVDE
WQLNVDLFGV DSHRAELI LADSLKFK GAPYQIKVS HRKLLDSFLK NSLLNGDQVH
YQCDPIAGN FPMIPDAECL KIMCEISSL QLDGFLKVN DRRLDGMFA ICGVSDSKFR
YQCDPIAGN FPMIPDAECL KIMCEISSL QLDGFLKVN DRRLDGMFA VCGVPESKFR
YQCDPIAGN FPMIPDAECL KIMCEISSL QLDGFLKVN DRRLDGMFA VCGVPESKFR
YQCDPIAGN YPMLPDAECV KIVSEILDITL DIGDYIKLN HRQLDGMFQ ACGVPDSFR
YQCDPIAGN Y--MEPDEVEL KVLTELDK DIVVEIKLN HRKLLDGMLE ICGVPAEKFR
YQCDPIAGN F--MVPDEVEL SILVEGLTSL GIKDFIKLN HRKLLDGMFQ IAGVKDEDR
YQCDPIAGN Y--MIPDAIEV RIIVEFDAL KLG--ITIKLN HRKLLDGLFA VAGVPAEKFR
YQCDPIAGN C--MMADEAVV SCVDRILSMF GIGYEIRKVN DRRLTSLLE VVGVSCESYG
IQCDVDIVGV E--VMAEALM SMAFELQTL NLEV--TIQYN NRKLLNGILQ AINIPTELTS
HQIAGEVFGV A--HADAIEI KIVYDILQAL GIKGVVEIN SLGCKK----DREAYRE
HYGIEVFGS QEASDAIEI SLVMALEDTF GIKGLSLNIN SLGCKPKRAK FNE-----

VQFGIEAIGS NDPAIDAEVI ALAVEFYRGM GLKN----- IKVVLNSLGD AAS-RQAHRD
YQFGIEAIGS KDPAIDAEVM ALAM-IYEKA GLEN----- VKLVINSLGD QDS-RKSYRE
HQLGCEVFLG QGPDIDAEIL MLTA-WWRAL GISEVTLELN SIGSLEARAN YRD-----
HQLGVEVLGS ADPRADVEVI ALGTDLKAL GLSN----- LSLALNSVGN GGD-RQRYRE
HQVGVLELGA E-PLADVEVI ALGSQLLGAL VGGG----R TILHLNSLGD PES-REAYRA
HQLNVEYLGA K-AITDAETV KLAVIDLKAL EIEE-TLELN SLGCHESTRIV YQQ-----
HQLNVEYLGA K-AITDADTL KLAVIDLKAL EIEQ-TLELN SLGCNESRSV YQQ-----
TQAGVELIGT NHEESDAEVI ALSIEALKST GLKDFKIDIG HAEEVFGSIVR NMLNGNEDVN
VQSGIEVIGI D-VRAEIECV ISVIQSLKQL RVQSFITIEIG QVQLYKIVK KLSIHEEEK
EQVGVLELGD GTTSADAEVI ALVV-ALKNA GLASFKAITG HAGIADALFV EVL-NVERAD
LQTGVLELIGE KSLILADFEVI QVLTFLKSL GLKDLKVLIS HAGIVRKVAD ERE-----
FQAGVELLGV GGTLADAEVL HVLADALAEI FGGQWHLVWG EASLTRSLLQ PFP--KDLRE
NQVGAELIGV T-VQADAELL ALAHAALGQA GVQA-QLEVQ FPGFVDAALT DAGLPGPVRA
IQLGAELYGD PSPASDVEVI SLMLELEMA EVPDVHMDLG HVGIYRGLAR AAGLSGEVEQ
LQAGAEMYGF A-IRGDIETI DLMKSKMIA DMGKVLVSLG HIGIFRALSD AAHLDAQSQA
TQAGIELIGA GGVDAIAESV LLTVEALQA GLPRVTVDLN CPPLVPALCR ALGLSEDLG
LQAGIEDLGA AEAASDARSL ADAL-CVKAT APDALEIVLG DQSVFAGMLK ALGLPQGNRK
YQAGIEDLGD TIAAADARVV IDATAILQRL LPGR LAVTLG DQVFEAVVA ALGLPLGWQK
YQAGIEDLGD PTAAADARVV GDAM-FVLAN RLPGLKVTLG DQSVFEAVIA ACGLPGWQK
FQAGIEDLGD RTAQADARSV ADAH-ALLSL VLPGLAVTLG DQGFIEAVLA ALGLPRGNRM
LQVGEAFET G-VAADAEMA ALAWASAVAG GREDLSVLVG DVSLFSAFVE SLDLAPPLAA
LQVGFELFDR T-EAADAEVF ALIAELLAPL GLQA---ATG DIGILMAAVR GLSTTEKRAA

EVLRIVDKLE KIGLNGVEEE LKKECDLDSN TIKDLEMMVQ IKQNEFFEKI AYLKDYENEL
DVLRIIDKLD KIGEEVVIKE LVEELGSLQE SARGILGFTS LKQ-EFFAKV ASYGLSDSL
DITVAIDKLD KIGLQKVNDE LRT--GFEE AIERLRPFIE MKGDEKDLR KSALASSETG
DITVAIDKLD KIGLENVNAE LKE--GISDE AIAKLOPIIL LSGTEKLATL KSVLAASSETG
ELCVAIDKLD KIGKEAVLEE LKT--GFTDS QLTRLLPVFD LKGNIEFPL RTILAQSVMG
TVLRAIDKLD KFGLEGVRL LGKGRELKD EIDGILSLLT IEVGETFDAL RKIVDQSEEG
TVLRAIDKLD KFGLEGVRL LGKGRQLKQ EIDSVALLT GEAEETLDAL RHVVGHHIQG
NVLRAIDKLD KFGPEGVRL LGPGRGLGDE QIEKLVFPVQ IK--ASADDL AKLVAGTSKQ
TVLRAIDKLD KFGPEGVRL LGEGRGLNEE QIEKLVFPVQ IT--RSADL AALVAGTARG
NVLRAIDKLD KFGPEGVRL LGKGRQLPEA AIEKVLAFTA AGGAQTIANL QAVVAGNAEQ
TVLRAIDKLD KLGPGEVRL LGPGRGLNG QAEAVLQATA RN--AQAAQ DSVSDVNATY
TVLRAIDKLD KFPADAEVRL LGPGRNLSAA EADVLAITK PRD-EAIAAA EAYLAKSEVG
PIYRAIDKLD KIGVEGVRRE LERL-LPGS TVERIMEIIS WRGREVLESV RRGVGSNEKV
NVLRAIDKLD KFGPEGVRL LGKGRQLPEA AIEKVLAFTA AGGAQTIANL QAVVAGNAEQ
TVLRAIDKLD KFAPEIKKL LGPGRNLSDD QSELILRATS PSFIIDTVEL LRSTADNETL
AVLRTIDKLD KVGEDGVREL LGKGRNLSDD QAGPVLAFT SRRGATTVAN RAAVAGSQIG
ALVAYFRDHA DALSEESRVR LEANPLILDS KSEQDRKLLD AAPA-FADYL NDDSRQF---
ALIEYFGHFH DLSLQESQER LTRNPLILDS KDPNDKIAE KAP--IIDDY LTHEAAS---
ALILYFTKYQ NDLSSESKNR LTKNPLILDS KDEKDKIIS DAPK--ISNY YTKESD---
LVLPIYFKFQ NDLSSESKNR LTKNPLILDS KDEKDKIIS DAPK--ISDY YTKESD---
HLVEYLNKYQ KDLSEDSQRR LVNPLILDS KSPEDCKILL NAPS--ISDF YTONSQN---
--LLKYFEKHE HALSEDSRRL LQTNPLILDS KDRGDIALLC GAPV-IADFY DDESNMT---
AVLRAVDKLD RLGVDGVRL LGEGRGLKQ AVDQVDFVQ AGSTATLDNI ARVAVGSAEQ
PIYRAIDKLD KIGVEGVRRE LER--GLPGS TVERIMEIIS WRGREVLESV RRGVGSNEKV
GYRAIDKLD KIGADGVRAE LTK--GVTPA AADQILELVQ IDGTEVDAL AQKLAADERA
AVLREVDKLD KRGVLDVRL LGQ--GFGIP AAQVENILTF VATRSHADAL VRLDALTEML
AVLREVDKLD KRGAEYVRT LAGE-GFGIA AAQVDKILAF VAVRSHADAL AQLQALEVTL
EVMILDKYD KITTEQFLSE LNKL-IGEE KAKVISKVIK ISGSEVENL KKVETKNEMEL
GLVRAIDKIE KIGEDGVKRE LQEL-GLSDS KIDKIFINFI IKGTVEIKSL KLDIENDVDF
SCISIIDNLE KIGEAQVLE KEKE-GINPE QTKIIFVVK IDGSDVLDK KHLSQLTEQF
GLVREIDKLD KIGADAVARL LTTETCGATEA QARAELAE LTAACEAHGI VKDSEAYTLA
AVIRAVDKLH KIGADAVREI LIDQAGLSSD QAAACLELAE IRGSTVDDAV AKLVGSHPTL
PVLRSDLKLG KIGREKTANE MVES-GVTAE QADAVLRLAE CDGDSILASL PEITGGNETA
HILRIIDKRA KMAPHVLTQA LESLCPVRVQ KIMTYVSAGE VDGVHTLSAI ETLTGG--V
FILRNIDKID KIG--IDKV KEALLEIEKE AVDSILSLVS LQG-DKIQAL KSLIG---D
GLMRIIDKLD KLPQEDFINA LKEFG-LDKN GIEKLELILN IKGKDEVPL AEELFTSEVA
KIRRIIDKED YDNLKIVLQ ILG----EE KKEIFIFILK FKGSVELDEL KEILKDF-KS
KVMRLIDKLD REGLESYLAE IRVN----EE LRDKIFSLLE LKGDSEVIEA KEIIDY---
RIMGVLDKLD VELEALLGE VEVS----GE SRELLMKLIG TQGGVELDDV EYLYGGY-ES
RVCIIDKIE KIPIDEIKKE LGFTG-ISED ATEQLLQVLS VKS---LDDL EDIIGGA---
SVFSIIDRYH KISREEFVQD LRSAG-IGED GVSMIADLCS GTRG--IDEM ARITGKS---
TVFSIVDRFH KVSKEFVQD LMSVG-LSEE NSNTIYRMCES EASE---GGI SGLFGSG-KV
AAIRTVDKRE KIERAAYLDD LAAAG-LSYD QAEQFDDLLD GDD---LDAL VSFAGTD--V
GVSLLDKKS KISPEAFETE MKPFL---- NFKEQFSLIE TYLN-NLETV SKIPGID---
TICSSVDKLD KVSWEVKNE MVGKGLAPE VADRIGDYQ QHGSLVEQL LQDPKLSKQA
AICSSVDKLD KMAWKVRHE MVVK-LAPE VADRIGDYQ CHGSLVEQM FQDPKLSKQA
TICSSVDKLD KMSWEDVRL MVAGK-LAPE VADRIGDYQ CHGSLVEEL FKDPKLSQLA
TICSAVDKLD KSPWADVRKE MVDEKGLDE AADRIGEYVR LSGGELVEQL LANEKLN-NA
TVCSVDKLD KLTFEEVKKE LVEEKGVSN TAENIGSLVK TKGPEVLEL RKEGSKFEGS
KISSAVDKLD KSPWEAVKKE MTEEKQSEE TADKIGEVYK LINGSEIHAV SADANITEKA
TISSAVDKLD KMPWADVRKE MVDEKGLPEE VADKIGEVYQ NAGNDIINYI KADADLPDDI
TVCSVDKLD KMSWEDIGRE LVLKG-LSEE QVRTTKYMT MGGKDLGLL KADPYGGERC
DVLSDKIE KIGIDGVKRD VLREG-ISEE MADTICNTVL SCLQ--IADF KEAFNNP-LV
ALLNYLTVGK EELCSDCISR MDRNP-LRVL DCKVETCKVA VREAKMIDFL CDECHEH---
ALKQYLVKENY DNLCECTCKR FEKNPRIIDC KEKRCKEIVK EAPS-ILDYI CECSGDH---
ALIAHFEPRI GEFCSDCQSR LEKNPRIIDC KDRNHLMG TAPS-ITEYL NEDSAVY---
ALVXHFEPRI EEFCSDCQSR LHTNPRILDC KDRNHLMK SAPS-ILTYL NEESAAY---
ALVAFLEQHK EKLDDEDCKRR MYTNPRVLD KNEVQALLN DAPA--DYL DEESREH---
ALYALTPFK AELDPSQDR LERNPRILDS KAKRTQETVQ DAPS-ILDHL GVDQRH---
ALVSYLSGHR DLSLRSLER LERNPRVLD KDEGDRAVVA GAPR-LIDHL NEASRAF---
KLVEYLNDFK DQLSGESRLR LKNPRILDS KSEIDQKIIA HAP--ILSEY HTNESKK---
KLVEYLNDFK DQLSEESKIR LSKNPRILDS KSETDQKIIA NAP--VSEY YTDSEKE---
LLRELEQKN QSAIEDFIKQ KEIKR---E EAKLRELP LFGGELLEK KKENFKE---
FLRTYIESKN YASLSNFIRD KKLDR--CDE TVKLELKLPR LFGGELIEA EKLAASN-EM
LRRFLYEKN YVGYREHVKS LPLSS---I DKSRLLELLE LRGGEVCGRA EEIVDSA--G
ELLRAFERN LEALKEVLG----- --NDIKFVVE VSKNEVNLFL EKYDLEK---
KVRQATAQFD RVTLESPLLE SQR----- --DRALLLHD LRQGVFALK QQLTLTP-LE
ALHDAIRKS GADDLARQ HGVS-----E VTRTLHSLTE LYGGEVLEA RVLARGV--A
QLFDALQRKA VDEVEALTD LPAEL----- --GMLRALAE LCGGDALQEQ RARLAAAADV
TLLALMQDKD TETVEAQVKA WKLDG----M WAKAFSLPR LYGGEVLEA RGRPLD---
ELRKLLEQKD FTAAAAAGD AAG----- ALSALSGTV PAER-VLAAL DLPGEEA---
KLLRSFGDAH SMDLALAEI GTQRR-----D PLPESLAVL AEGD--IGLA R-----
RLVQAFGDMA QLDALLESLV HPKQ-----MTG LDRVAVGLLA TGEVLDVYL DTMVQET---
RLTHAFGDQK QLERLLAALA DPKSPGVFVH DVERVAALGM LEDERLVAHI GETMKAT---
RLARAFGSAP MLQAALADLA NPPRN-----G QLSGEVAALV LDGD--LDGL STHIAGG---
RLTRSFAPR QLKAELDGDV AAPTG----- --EQSRIAA LLAG-----L PEAEAS---
ALLRHWRPT RFRALLDRFA GRTPV----- -EARLRLDR LAGA-AVALM EAAGPAI---

KGQIQDLERL YQLLDLQIS QNLYKIDFSI ARGLGYTYGI VYETTLNDKS LGSVCSGGRY
REGMGEMRAL YAILDTLEMD RACYKINFSI ARGLGYTYGI VYETTLNDAPS IGSVCSGGRY
LKGIEELTYI LDKVERVQLH S-ELMVDISL ARGLSYYTGA ILEVKALDAE MGSITGGGRY
MGVVEESEFI LGTLETMGLK N-EIELDLTL ARGLNYYTGA IFEVKALDVQ IGSITGGGRY
LKGIDLEKVK FTLIEKSSGLG NASIETDLTL ARGLSYYTGA IFEVKLVNVQ MGSVCSGGRY

LEGVRELEEM QAIFAENDT- QNRIKIDPSV VRGLEYYTGP VFEAALLFV FGSVGGGGRY
LEGVRELEEM QAIFAANGY- QNRIKIDPSV VRGLEYYTGP VFEATLLFV FGSIGGGGGRY
EEGVDELNTI GALVSGAGYD ATRIKIDPSV VRGLEYYTGP VYEAELTFV FGSVGGGGRY
AEGVNELENTI RGLVLSAGYE ADRIKIDPSV VRGLEYYTGP VFEAELQFV FGSVGGGGRY
EEGVQELADM QALFVAGGY- EGRVKIDPSV VRGLEYYTGP VFEAELLFV FGSVGGGGRY
QEGVGELEATI EALVRAAGYD EDRIAMDRSV VRGLEYYTGP VFEAELLAVR FGSVGGGGRY
QAGVSELEEI AKLVMASGYG ADRIKIDPSV VRGLEYYTGP VYEVLLVLR FGSVGGGGRY
REALDHLHEEM LSLADDDR- ---VELDMSL VRGLDYTTGP ILEVLLDEPR TGSVAGGGRY
EEGVQELADM QALFVAGGY- EGRVKIDPSV VRGLEYYTGP VFEAELLFV TGSVGGGGRY
KQGCDELAVI ADLLDACGYD DVRVIDPSV VRGLEYYTGP VYEVLLDFT KGSVGGGGRY
VQGIDELEQI GDLLAAGGY- PDRIVIDPSV VRGLDYTTGP VYEAELTFE FGSVGGGGRY
----WDGL RSSLQAFG- ---VDFVNPRI VRGLDYYSHT AFEFVTSKLG AGTVLAGGRY
----FFKAV TTGLDAAG- ---VAMTRDPR VRGLDYRHT AFEFVTDQLG AGTVLAGGRY
----FFEQI LNGLTILDI- ---YTVNKL VRGLDYCHT VFEFVTEDLG AGAVFAGGRY
----FFEQI LNGLTALN- ---IPCTVNSKL VRGLDYCHT VFEFVTEDFG AGAVFAGGRY
----FFKEV LDGLDNL- ---INVINHNI VRGLDYCKT VFEFVTESELG SNSVIAGGRY
----FNGV MQQLDNLG- ---IPYTVNPR VRGLDYCGT VFEFKTSLG SDAVIAGGRY
DEGLLELSRI DAALSSGLIA DDQAFIDPSI VRGLEYYTGP VFEAELLVSV FGSIGGGGGRY
REALDHLHEEM LSLAD- ---DKRVELDMSL VRGLDYTTGP ILEVLL- ---R TGSVAGGGRY
VAAIANLRAV LDGLMAMGVP AEQYRVAPRL ARGLSYTYG VFEAIT- ---WP EGSLLGGGGRY
RQGVAELEREV LTLVNLVLPV EHAYLNFISI ARGLDYTYG VYETSLINPQ IGSICSGGGRY
GEGIAELREV LELVKALGVP ETAYCLNFSI ARGLDYTYG VYETTLTDPQ IGSICSGGGRY
EGGIEVEEVV IKYKLFVGE DSEYAINLKI IRGLDYTYG VYETTLTGES YGSICSGGGRY
KEGVYELEKV VHYIKSFNVP DKNYKIDLTI ARGLDYTYG VYETVLLNPP IGSVCSGGRY
NLGVSELETV ITGVNRLVGP DKRFIDLAI ARGLDYTYG VYETTLIGEA LGSICSGGGRY
RQGLDTLAMI VD- ---EAAAIR PGSVIADLKI ARGLDYTYG VYETFLDGA LGSICSGGGRY
TAGLDELGEV LD- ---ELSDLP SGVSVADLSI ARGLDYTYG VYEAFTNDPG YGSICAGGRY
AAGIERLTOI YRGALASGVS PDRKIDVSI ARGLDYTYG IFETTLDEPG IGSVCSGGRY
SEESTRLRKI YELLCVNIQ 5-SYVFDPSI TRGFDYTYG VCETFLTQPH IGSVCSGGRY
NESIKRVEDV FQHLSLKIQ D-SFNLNLKI SRGLDYTYG VFESEVFN MGSVCSGGRY
KNEINRLYAL IDLLEAYEVK D-WIRLDLGI ARGFYTYSTI VFEAIAPNLG IGSIGGGGGRY
MEAINLEETI LEFVH- ---DKYIINLGI ARGLDYTYG VFEIYKKG- G AKQICGGGGRY
---FGHLESL SALLRDVG- ---VDFTLNLI ARGLDYTYG VFEYCAEGB QKQVCGGGSY
LEALEEFREL VETLHYFG- ---VDDYHVNLI ARGLDYTYG VFEIYVPEGA QKQICGGGTI
-EAIADLKQL FSLAEKFG- Y SEWIOFDASV VRGLAYTYG VFEFVDRKKG LRAICGGGGRY
SEEARMAAV EDLLASYGVK N- ---VRYDFSI VRGLSYTYG VFEAYDRSGQ FRAILGGGGRY
SKAAERLIRT IDILKEYGVK S- ---VKYDFSI VRGLSYTYG VFEAYDKSQ FRAILGGGGRY
RDAVANLRNV LAAETLGV- V RQYCDVSLTT ARGLDYTYG VFEFCFDSTDV GRAVFGGGRY
TNSVSVIQLN FQELGELG- I DKQLLFDPSI IRGFDYTYG IFEVFDNEN RRSLYGGGGRY
LEGLGDLKLL FEYTLFG- I DDKISFDLSL ARGLDYTYG IYEAULLQLG VGSVAAGGGRY
LEGLGDLKLL FEYTLFG- I ADKISFDLSL ARGLDYTYG IYEAULLQLN VGSVAAGGGRY
LQGLGDLKLL FEYTLFG- I ADKISFDLSL ARGLDYTYG IYEAULLQLN VGSVAAGGGRY
VKGLEGMKQL LKYSIFG- L DKRVSFDLSL ARGLDYTYG IYEVLLKAT VGSVAAGGGRY
VVALNELEIL FKALEKAN- A LDRISFDLSL ARGLDYTYG IYEAUVFKGTQ VGSIAAGGGRY
KQGLDDIATL MKYTEAFD- I DSFISFDLSL ARGLDYTYG IYEVVTSVAVG VGSIAAGGGRY
KAGIHDMEEL AAYLDALEVT D-KVSFDLSL ARGLDYTYG IFEVINKPSQ VGSIAAGGGRY
KAAVHMEEL FRLCRILG- C SGSLVMDISI ARGLDYTYG ILEAEYVKS VGSVIGGGRY
ADGVNELQQL QOYLIALG- I NENTIFNPL ARGLTMYGT VYEIFLKDSI TSSIGSGGGRY
---YEKLKNY LKALDIP- ---FRENYNL VRGLDYTYG VFEAVSDEG- -LTLIAGGRY
---FVSKL KAYLDVMG- ---IEYNIDPQI VRGLDYYSKT VFEVIKDG- -LTVCGGGRY
---YDKVQEL LTMMDVP- ---FEKDPNL VRGLDYTYG VFEIMSEAGA ITTLLSGGGRY
---FEKV QOYLNDLG- ---ISYIDPNL VRGLDYTYG AFEIMSNGA ITTLLAGGGRY
---FAGL CKLLESAG- ---IAYTVNQL VRGLDYTYG VFEWVTSNGS QGTVCAGGGRY
---FDQV QQLLNLG- ---IAYQLTPTL VRGLDYTYG AFEIQSSDGA QATVCAGGGRY
---FAGV TGGLEALG- ---IPFEIDPLL VRGLDYTYG AFEFVTDGA QGTVMAGGGRY
---YFDEL QKYLIDLG- ---IKYSVNPRL VRGLDYTYG AFEFVTKKGS QSTILAGGGRY
---YFEEL IQYLIDLG- ---VKYSINPRL VRGLDYTYG AFEFVTKKGS QSTILAGGGRY
EGVLEYLDRV KYILEDFFG- M KEYSFDLGM VQNLNYTYG IFRGVKGL- L GYAITCGGGRY
KMAITRKEI YEAIKELG- Y GSYISIDLGT IQHLDYTYG IFKGYIYEG- -EIVSGGGRY
KSVVDLKLAL WDILEDYGC- C TENVRDLNM VSHMSYTYG IFEVYAEV- V GFVIGSGGGRY
--EKEELLE GLKLEEF- ---IDYAYDLGE VRNFPYTYG IFEVYELK- S GKALAGGGRY
QTLRDRLAQL VELYNASAGQ DSPLLDLSL LRSFDYTYG VFEVYTYG PWLQAQGGRY
EQAVEHLSAV HAAAEQAG- ---VELLDLGV SRRYGYTYG LFRAYVDG- I NQPVLLGGGGRY
QVALNELTEI ADSLAGRFP- GLPLYFDLGE LRGYHYHTG VFAAFVPG- V GQVIAQGGRY
-AVGALGEL QAVCDAFP- ---DCEITHDSE LRVNDYTYG LYAAAD- F HDAVARGGGRY
--HRDALLAL ARKVMTAAP- EVDVTDPLE HKGFYTYG SFALFAQG- A RGLGRGGGY
---MLEAEM EAGISPGAG- RTPVEIARR IEKEDLAHT FPAALDL- ---QFLAIRVSL
---YSTNAS RSPLEIARR- ---LREKL ALAATRLPDE SFELLKQF- ---LALQAPL
---YSTNAG RSPROIARR- ---LKEKVELAA IRLDKEALGV MRAFLALD- ---LPLADA
-----MEQ AGLSASAGR- ---SPTDIARR IEKELRSVR LSNFAA- L KNFLAIHVPL
--VLQELWSL AGIEPVGGR- ---RPAEIARR VEKADAAGK RLDPAQADKV RAFLAVSDRP
-----L REPCEIARR- ---AALIEDAE AAPLGAEEAA VLDDLSL- E APAAALGRL

DHLTKNFSKE NLQGVGASIG IDRLIVALSE MQL-LDE-RS TQAKVLIACM HEE-FSYANR
DNLTKQFSKE RMSGVGASIG VDRLLAALAE LGL-LEG-RG TSAQVLVACM EES-LAYSYSK
DNLTKIFGLE GMSGVGISFG ADRIDVMMQ LDL-FPAEKL TGQTLLVNF SEE-ADSLLP
DNLTKGVFGMA GSVGVGISFG ADRIDVNLQ LEL-YPKEAV NGTELMFVNF GDK-AAFSMS
DNLTKGVFLP NMSGVGISFG IDRIYVMNE LNL-FPSQES NSTRLLICAF DEN-FMHALP
DGLVARFRGE NIPATGFSIG VSRLLAALQN LGK-LPV-KE KTGPPVVLMM DKEPVARYQK
DGLVARFRGE NVPATGFSIG ISRLITALQN LSK-LPV-KK TPGPVVLLMM DREPVAQYQK
DGLVSRFMGQ PVPATGFSIG VSRMLTALKN LGK-LGQ-VK PLAPVLITVM DGDVMGRYQR
DGLVSRFMGQ PVPATGFSIG VSRMLTALKN LGK-LGA-EQ VTAPVVVVM DRDIMGRYQR
DGLVSRFRGE PVPATGFSIG VSRMLTALKN LGK-LDV-SD TVGPVVVLM DKDTLGRYQK
DGLVSRFRGE PVPATGFSIG VSRMLTALKN LGK-LDN-AD VIAPVVVLM DKDTLGRYQK
DGLVSRFRGE PVPATGFSIG VSRQAALTL LGK-LDT-RP EFGPVVTVF DRD-VADYQK
DGLVGMFAGR DIPATGVSIG IERIDAGLE LGV-YSLDVK TVAQVAVVL DRRYRYAYE
DGLVSRFRGE PVPATGFSIG VSRMLTALKN LGKLDVSDTV GPVVVMDK DTQNLSTRYQK
DGLVSRFRGE PVPATGFSIG VSRQAALTL IGQ- ---LGNK PQAGPVVTV FGGIAGYQK
DDLKFRFTQ EVPATGVSIG VDRLLAALAE KGR- ---IRTA ERGPVVVVM DRDRMADYQA
EGLVEQMGPP AIPATGWAGG IERLAMLDS VPE- ---LPGIALVPM GDEAVLKAAS
DGLSLEAGLN LTPAIGWAAG IERLAMLDE PKE- ---DAMNVAIPL GEEAKNPAIS
DNLVSSVGGK HTPAIGFAGG IERIMELINY SPK- ---EERPIYIPI GREAEKHALT
DNLVSSVGGK YTPAIGFAGG IERIMELINY SMK- ---GDRPICLVPI GREAEHALT
DGLVKSMMGGH STPAIGFAAG VERLSALIDY EHK- ---KPKRIVLPI GDDTASYAIK
DKLVASMMGG DVPVAVFAGG VERLASLAAY SHS- ---TRFSVAFPL GEEAACAMR
DDLVARFTGN VTPATGFSIG VSRLLAALRA AGR-EPG-GA ARGPVVVITF DQA-MGEYFS
DGLVGMFAGR DIPATGVSIG IERIDAGLE LGV-YSLDVK TVAQVAVVL DRR- YRYAYE
DHLIGLFAGR NIPTVGLAFG IDRLHVMIE LNL-GPR-PQ TTLAFVTLF GAE-VAESLR
ENLANHYTQS KLPVGVGISIG LTRLFWQLRD AGV-MDGIAE SSVQAMVLM DE- LDDALD
ENLASHYTKS KLPVGVGISIG LTRLFWQLRE AGL- IAGIAE SSVHAMVALM DES- LDDALD
DNLQAQNTYEN VLPVGMVMSIG ITRLFFVLE IGF- IENYNS SLNEKYLIVP IGD-FEYCTK
DNLAEHYTQ KLPVGVGISIG LTRLISRLK AGI- LNTLPP TPAQVVVVM QDE-MPTYLK
EELVGTFIGE KMPGVGISIG LTRLISRLK TAG- AHANRV SPAAVLAVM NEE- RPAANR
DNLASQGNRK -YPGVGLSIG LSRVSYMLH TAG- AHANRV SPAAVLAVM NEE- RPAANR
ENLAGQFIRR SLPGVGISIG LTRIFAKLVS EGR- ITGGRF CPTDVLVLP SDE- RSAALA
DNLAGLYTKQ HLPVGVGISIG LDRLLAALAE LGR- LSG-VS KPCAVFVPPF DKG-RDDYLK

DHLTALYMKD AVSGVGASIG LDRLYAAFQQ LGMSREHVCF VQA----LIF CQDSMDVYQK
DNLVSSFSNQ KISGVGGSFG VDRIKDIIDL EKFSYIKIFV TKARSKVLIV NLDSONYQYE
DNLIEVFGGK PTPATGFAIG IERLIPILEA KGL--IPEVQ IAPDYYVVP I GER--KQVAIR
DNLIIETFGGE PTPAVGFAYG FDRIMNIDD LDI----- EESISLIIPV KDKKIKKSLI
E--LSSLFGGP VTPSTGFAIG FDRVCEACS EAG----- -EKSVMVAVS FKG--EQAFR
N--LVETFGGE RVESTGFAYG FDRMLNMG GM DEE-----NM RMVDVFIPI QDS--VPEAIR
DRLLSTYGGD DFPACGFGG DAVIVELLKE KDL--LPELG QEVENIVCAL DKD--QGAAT
DNLASLMSGE SVPVGFSGM DAVISLILKR ENVQIPREKK SVYICRVGKI NSS---IMNE
DNLAKLFSEQ DIPAVGFGMG DAVISLILKS KGVKAPHAKK SVYICRVGTV KAER---IAS
DDLIERFGGQ PTPAVGFAYG HATLNLLEEN AGVLPDDEFI PDDYVLQVGD TDA---EAAD
DNLIGLFSKE QLSGIGFGLG DVTLKNFLEG HNL--IPNLS REKTIPLPIM DES--FVDTFK
DGLVGMFDPK KVPVGLSIG VERIFSVIEQ RLEALEEKIR TTETQVLVAS AQKLEERLK
DGLVGMFDPH KVPVGLSIG VERIFSVIEQ RMKTKGKVR TTETQVVFAT PQKNLQERLK
DRLVAQFDPH SVPVGLSIG VERIFSVIEQ KMKISCEKVR TTETQVVFAT PQKNLQERLK
DNLVGMFDPK AVPCVGSIG VERIFSVLEA RAAASGLKLR TSDVEVYVAS AHKGHEQRLK
DKLVGMFSNK QVPVGSIG IERVFIMEQ QEKEKNQVIR ATETEVLVSI LGK--LILAAE
DNLNMFSET QIPCVGISFG VERIFSLIKQ RINSSTTIKP TATQVFMVF GGGKLEPERMK
DNLVGMYGKK QIPCVGISFG IDRIFTILKA RGE--NDIKK REVDVFMVF GGGKLEVERMK
DNLNTELVGE STPCVGSFG VSRIFSLIYE EYD-----K ESSTMVYVGA SGG--LDERLS
DNLIIAGFRGM SYPTVGSIFG LDVIYALSQ KET-----IS STADVFIPL GTEL--QCLQ
DYLVEELGGP PTPALGFALG VERLMLLPD EEE----- KEEVYVPIF GDVH--EYALR
DYLVEEVGGP KTPAMGFGLG LERLLILDE EGIEIPEPVR CEVYIGSMGD RAR--LEAMK
NGLVQEGGG EMPGIFGMS IERLIMALK ENIELPIEHS IDCYVVALGE KAKD--HAAK
DGLVEQIGGG EAPGIFGMS IERLAAIDA EKRELPVOKG IDCYVTLGE KAKD--YSVS
DGLVEQLGGR ATPAVGFAMG LERLLVLQA VNPEFKADPV VDIYLVASGA DTQS--AAMA
DGLVAELGGP VTPAVGWAMG LERLIIILQQ MAT--PPAPS PDLYLISKGE KAFP--QALI
DGLIRMGQGS QTPVGVMAAG VERLSMLMAA DRV-----G APPRPLAMIPA GPDDEVALK
DMLSRIMGNY DVHAIGFAAG IERIALMREY KIS----- VIKPVFVLP I GKNNICYALD
DGLAKIMGND DVPAIGFAAG IERIALMREY NIS----- EVKPVFVLP I GKNNICYALE
DKLLKIYKG DLPAVGFALG VERLMLALQ QSK---GEMW KPRRVLVRYK EEEER--RRAEY
DELIGNFGE MLPAVGLAVQ VNQIKVALQE QQE---PYE RKRIDIMIH Y ELNRLAEAEY
NKLGHFDS-- PAPATGFGLR IDRLIEALHM KDE----- -PCEIDAVIF SKEQRAQAI
DNLKSIYK-- EYPATGGAVY LERLLILPK NVE----- -KKDYFVDK TKKR--LGLK
DRLLDVYDPA QGPGIFGSCN IENLQVLLA ANRLPHRPA IDQLVIVPDS EAYG--AALA
ALPGG-----LPGAGFAIG LERLAAMVPA GVP-----S EPETVLALDF AAAT-----
DDIGADFR-- ARPATGFSTL KLSLVTLQA RLD----- QAVSGIWAPE EGAG--LWQ
DGLGGYFGR-- ARPATGFSTL LRSFIRGLPA IER----- -QPVLVDAE DAEA---AHE
PGDLNVTVE--PCCGATLY MDAVLDMVVR PAP----- APRVVLVSTL ARTT-----
DTAAVTLRA--FAADNLD LGAVLQKFEA RADAIAQAGI EMKDIYDAS FGRPLDYTG
PQASQVLGD--FAARAKLK LGALSAFDK RVAALANAGV DLETVTYGAA FGRP-----
PAALHRFAG--KARLK IDDALALFDA RVAALARAGA DPGLMRYRAA FGRP-----
DGAARALET--FAAGALGS LGAALEKFAA RAKAIEAHGL PAEKIRYDAA FGRP-----
DAALSAIAA--LAGPKAEA LNAVLDGWRA RLAGMAQRGV ALERATLAAA FGRA--YYDG
EALARRMPA--LEPSVAR FGRRLAALAA RGI----- AVERLFPDAS HGQS-----

LAESLRQSGI FSEVYEAQK IKKPFYSYANH KGHEFVAVIG EEEFKSETLS LKNMHSQM--
VAERLRTLI KSEVYEAQK PKRSLAYANS KGHPYVAVIG EDELKQGVVT LKDMQWGDQ--
LIQLREQGL RTELYPEPAK IKKQMSYANA ANIPYVALVG ENERAAGQVN LKNMLTGEQQ--
MLAKVRAAGI RAEIFPDAK MKKQMSYANA KSVPFVAVIG ENEMNEGKAM LKNMETGEQN--
MVTKLRAQGI NTELVPDPVK IKKQLSYADD KIPFALLIG EDEMNGLYS LKNLISGEQF--
MMQRLRSAGI PAELYL GASG MKAQMKYADR RQAPCVVIQG SQERESGTIQ IKDLVEGARV--
MMQRLKAGI RAELYL GASG IKAQMKYADR RQAPCVVIQG SQERQDRKIQ IKDLIEGTRI--
FTQALRAEIG RAEMYQGNK FGNQLKYADR LGSPIAIIQG GDERAEVWQ IKDLIEGKRV--
FVQDLRHAGI RAEMYQGNK FGDQLKYADR RGSPIAIIQG GDERASGVVQ IKDLIEGKRV--
MVSRLRAGI RAEMYVGSIG MKAQMKYADR RAAPCVVIQG SQERAGEVQ IKDLVEGKRI--
MVSRLRAGI RSEMYLGGAG MKAQMKYADR RGSPIAIIQG GDERAKGELQ IKDLIEGARV--
MVASLRTAGI RAELYLGNKN MGNQLKYADR RNSPCVVIQG SDEKARGELQ IKDLIEGAKF--
AARILRRGGL NVRIDLRSAS GQVQRKASR LGIPVLAFFG AKEAEGGFLT LYSAAKGERV--
MVSRLRAGI RAEMYVGSIG MKAQMKYADR RAAPCVVIQG SQERAGEVQ IKDLVEGKRI--
MVAELRAGI RAELYLGNPN LQQMKYADK RNSPCVVIQG SDEKQGIQV IKDLILGAQF--
IARVLRAGI RAEIETGRN MKKRMEVVK SGASHAIVLG DEEIAKNIVQ LRDLSSREQQ--
ILSKLRHAGI SADMWRGN MKRMRQKANG WGVRYAVILG EEEELARNEAT VRSMVTGEQQ--
LANELRRNLG YVIEYSGT--LRTRMKKANQ ANAKAALIFG DEELSSKTLK IKNMDTGEQK--
LANELRRNLG YVIEYNGA--LKNRKKASQ ANAKAALIFG NEELSSKTLK IKNMDTGEQK--
LAYELRCKGI HVCWNNYRGS LKNELRKAND TD--IVLIFG DEELQSNTVK VKDMKTGDQQ--
SAYELRGRGI RVLCDGVVEK LKIGLKHADR SGVDLALTLG DEEIAKGEVL CRHMDTLQO--
VVGQLRAAGI AAEVYLTSG MKPQMKYADR RGAPAVVMLG GDEIAAGTVT IKDLDAGRMQ--
AARILRRGGL NVRIDLRSAS GQVQRKASR LGIPVLAFFG AKEAEGGFLT LYSAAKGERV--
LAELRAAGI NTLIALEPGN IGKQFKEADR RGVRFALVLG PDELARGEVQ IKDLQRGEOR--
IARCLRIGGI NTEVQMEAKK VSKQFYASR AGIRFVVLG DDERARGVVA VKDLTREQQF--
IARLRRIGGI NTEVQMEPKK VGKQFYASR AGIRFVVLG DDERARGVVA VKDLVREQQF--
ILNKLVLNKG SAEIYFEEGK LKKKLYYANK LDIKYVILIG EDEVTNKL V IKDMITGEQK--
VANKLRENEI ISELYLEDAG IGKQFYADK LKIPYVILIG EDELKEEKVS VKNMETGNEQ--
VSQQLRQAGL NVITNFKEKQ LGKQFQAADK QGIRFCVITG ADEAAAQKSS LKDLQSGEQV--
IANQLRARGI ATDVAPTAAK LGKQIKYADK LGIPYVWFFA TAAEGAE GDE VKNVITGEOV--
VAAQLRERGF NTEVYHQAQK IGKQIYAVK KDIPVWFFP--FDDGRP--HE VKNLATGEQV--
LASQLRDAGI GVEVYEPKPK LGQQLKYADS QGFAVAIAG GNEWEAGAVQ VKTLATKESV--
LCS---VQ--V ATEVFP--RK LSQQYFAEK KGIRWGFVFE QRNAVVDLV LRDLSTRKDT--
LATRFVRKNI SCEVYFKNKN IKEQIEYALS KEIRFLVFGV QEE--YKEMK VRDLTKKEEL--
IATSLRKSIG KTEIELMGRK LKAMDYANK VGKKTIIIG EKDLERGVVT VRDMESGEQN--
IADKLKAGK IVELEIMGRK LRKALDYANS RGFKKVIVG EKELNEGKVT VKDMITGEQK--
VASMLRERGF TAVVDVGRN LKQMSFASG MGKRYAVILG PDEVKSGRVA IKNLETQEQV--
ITQELRAAGV SADLEVSGRK LRKALSHADN LGARFVLTG EREIQDGKVT VKDMERNSEQE--
VATALRDKGQ TVDLVLESKP LKWVFKRAR VNRRLVIVG KTEWEDGSVS VKVLSGSEQF--
YSRKLREGRM NVTVEIMERG LSAQLKYASA IGADFAVIFG ERDLERGVVT IRNMYTGSQE--
ISKMLRSGSF IVSAEIMDRS LSSQLKYASV EGCEYAVIAG ERDLNNSVT VRDLSTGEQE--
VASALRERGN RVDTDITGRS FGAQMNAYDN VGADTVIVG EQDLADGNVT IKNMSAGEQT--
LSKELRENEI LTEMDSAK IGKQIQAIEK KGYRYVFLG ESEIRTETVQ IKDLISGEQK--
LVSELWDAGI KAEILYKPK LNLQLYCEE AGIPLVAVIG EQELKDGVIK LRSVTSREEV--
LIAELWDSGI KAEMLYKNPK LLTQLHYCES TGIPLVAVIG EQELKEGVK IRSVASREEV--
LIAELWDSGI KAEMLYKNPK LLTQLHYCES EDIPLMVIIG EQEQNEGVK LRSVASREEV--
VLNLWDAGV KAEHSYKLPK LLVQLQHEE HQIPLVAVIG DAELAQGLVK LREVTTRREE--
LVNELW--AGI KAEFLK--R--VNHIKIYATQ TGIPLMVLVGE ESELROGKVK LKDIRANQEE--
VTQLWDAGI EAEYVYKAAN PRKQFADA EK AGCHIAVILG KEEYLEGLR VKRLQGFPAE--
IARQLW--AGI RAEYMAKA--K PKLQAFKFAA EGTPLVILG EEEELKAGQVR LKLLDQOQGN--
VLRILQDANI CSEFYTRSS FESQKQYVAK KKIPIVWVG ESEIAAGSVQ VINALTRKKE--
IAQQLRTTSL KIELELAGRK LRKALYANK ENIPYVILIG EEEICTETVM LRNMKEGSEV--
VADILRKGK VVEYSYRGGG LKQLEFADK LGVYAVIIG EDEVKNOEVT IKDMETGEOR--
LAFNLRSKSI KAEIDHLGKS VKAQMKYANK IGAKYTFVIG DSEIEENKIK IKRMSDGEQV--
VAFDLRAGL AVEKDYLRK MKAQFASDR LKAKFVAVLG EDELKGIIN LKDMATGEQE--
LVYKLRAGI SSEIDYENK MKGQFADK LGARVAVLG ESEVANGTAV VKDLRSGEQT--
LAERLRLPGV KLMTNHGGGN FKKQFARADK WGARVAVLG ESEVANGTAV VKDLRSGEQT--
LAQKLRNQL AVALDSASA FGKQFRADK SGATACLVIG DGEIATGTVQ LKWLADKAQE--
LAHALRQGL FVDGFSG--N LGKRMKRANR INARAAIILG GEERANGTV VRDLDSGEQE--
IVDKLRQNI VSIIDPIG--K IAKRIQRVNL EDAKFIIFIG DEERMNNLK FKDLNQEKEY--

IVDKLRTENI AIIIESLG-K IAKRMQRIFN ENAQFIIFIG DEEQANNLK IKDLKKAEE-
KANVLRKEGN VVEMYTFK-- ---RCENIDL KSFDEVVDVG E-----
LRNLQKDGK KVALSLFS-N LNDTFQFARK NQIVTVVEAK SESLVEYVMK EKWVVQKEG-
YANEERMKN KVLQDLS-- -IENIDQMTK SFANVTYFIG ARKEEQNG--
LADVLEKKG KVGMEIVKRG LEHSLIYAFD KGFKEVLLVE DEIVKVYT-T PKDY-----V
EAQRLQQDQL RVELYLDSTRR PEVVOQAFQR RRGIRIVWVS SGSAPQSE--
---AARAAGL GAELAWTG-D EAELRHYAQA RGLRRWVQGA ELRDVTPA--
AVQRLRRDGG RVVQALPG-- --QDAASARE AGCDRQLALR DGNWQVAP-- ---LAS---
AVEALREQGQ CVVIDYGIHN VSEELAGRLK KTDGVMQVVK R-----
-----RRAL QADGWTV-A LTAEAGAAEA RRLGCGHVD DEGRPVAV--
LVYEIRDASN RQDGLAG-- -----GG RYDRLLTMLG ACEAIPGV--
-LDYYT--GL VFEVVEAG-- -SDSVLAGGG RYDRLLTLLG AQEKIPAV--
-LDYYT--GL VFEIGVEG-- -TPAVLAGGG RYDRLLTLLG AREHIPAV--
-LDYYT--GV VFEIAAQG-- -GERPLAGGG RYDRLLTLLG AKTAIPGV--
YLFETRSASL GDERPVAADG LPAKLG VATD TGAVGFMRP ARAYAGGG--
-SLEYD-DGF VFSFHAAD-- AGLPPVSTGG RYDALTAVLG QGRSIPAI--

-LNCLSFLKA LEIIGEN
-KPCISIEAA AEILLG-
MLPADAATIV AAIV---
LVSVEELIAA LRN---
KVTTFDEVVHK LI-----
TVDEEQVKT VQDILAA
TVDEERLVQA VQDILIS
SVPPEALVAK VREILEH
SVPGEELVAK VRQILAE
TVREDGLVDA VREILDA
TVAESELVAE VVKILAA
SCAEADLVAK VREVLAR
AVPLGEAVKA VERLLPS
TVREDGLVDA VREILDA
SCNEDEMAK VQELLQR
EVARADLVAK VREILAE
EVPVADLARI LSAGRA-
AIALDQIVDR ILSELRN
IIARDNTIEN IY-----
IIAYGNIMEN IHQTLV
NVAVCDLND LHRIV-
TVSISNLGDY VSGMESN
TIPRDQLVDA VRKIIGG
AVPLGEAVKA VERLLPS
SLPRAAAISL LAEALSE
EIPREELAST LQVELEQ
DVARDELAST LQVELEQ
KLNIEEI---
SMSLEEAIKI IKNA---
EVALADLAE IKRRLT-
AADCTSWEPT TVVAQQT
EADPASWNG-
AYSHESPDAL IEAIQAA
RLPAHEVRRR MAEAG--
LLSFEESINL IKCNEKL
EVKIDIVVF MKNALKT
LIGIDELTNF
VVAEEELFSI LQ-----
TLPRDEMVDK IKGRC--
QVKLSDLE--
NVGLDSVVEH LISQAT-
IVTLGDLVPF LEKSTSY
TAPIAAPGA HDAPRVE
SLPRKGLSDT LKKDFQL
DVRREDLVEE IKRRTGQ
AIKRENFVAE IQKRLSE
TVNRESLVAE IQKRLSE
NVKLEDLAAE IRRRQQA
EVPRKDFVEV LKQRLSN
LVSAAIVPI VQEKLSQ
LVPIDDLADE VKKLL-
IVKVEQVMSY LLDPENQ
KVPLSLSNY L-----
RVKLSLSEL---
SLDINEIVNI VK-----
EVALDVFASY VAEKLI-
EVALDEFIHV MKANQKG
AVAQDSVAAH LRTLLG-
TLQLQDLMGN ITELKQR
TIGLAVLEEF LARFL--
IIDFEKLEL LKQ----
IVDFAKALEL LKK----

--ETSCVTFK LR-----

VMKIREFLDL IE-----
AVAVAERATT TC-----
DLKIATEVNA

SIWLDRLQAL AGEKQ--
SLWLDRIKAV RGSDKP-
SLWLDRIEQA IAAGRAE
SVWLDRIEAL REAAP--

GVTRPGLVAR LKGLA--



