

Identification of Conserved Amino Acid Residues of the Salmonella S Chaperone Crl Involved in Crl- S Interactions

Véronique Monteil, Annie Kolb, Jacques d'Alayer, Pierre Beguin, Françoise Norel

▶ To cite this version:

Véronique Monteil, Annie Kolb, Jacques d'Alayer, Pierre Beguin, Françoise Norel. Identification of Conserved Amino Acid Residues of the Salmonella S Chaperone Crl Involved in Crl- S Interactions. Journal of Bacteriology, 2010, 192 (4), pp.1075 - 1087. 10.1128/JB.01197-09. pasteur-01414597

HAL Id: pasteur-01414597 https://pasteur.hal.science/pasteur-01414597

Submitted on 27 Dec 2016 $\,$

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

| 1 | REVISED MANUSCRIPT | | | |
|----|---|--|--|--|
| 2 | | | | |
| 3 | Identification of conserved amino acid residues of the Salmonella σ^s | | | |
| 4 | chaperone Crl involved in Crl- σ^s interactions | | | |
| 5 | | | | |
| 6 | Véronique Monteil ¹ , Annie Kolb ¹ , Jacques D'Alayer ² , Pierre Beguin ³ , and | | | |
| 7 | Françoise Norel ¹ * | | | |
| | | | | |
| 8 | Institut Pasteur, Paris, France (1) Unité de Génétique moléculaire and CNRS URA2172 ; (2) | | | |
| 9 | Plate-Forme d'Analyse et de Microséquençage des Protéines ; (3) PF5 Production de | | | |
| 10 | Protéines Recombinantes et d'Anticorps, 25-28, rue du Dr Roux, 75724 Paris Cedex 15, | | | |
| 11 | France | | | |
| 12 | Running title: Mutagenesis of the σ^{s} chaperone Crl | | | |
| 13 | Key words: Salmonella, crl, rpoS, sigma factor, RNA polymerase | | | |
| 14 | * corresponding author | | | |
| 15 | Institut Pasteur, Unité de Génétique Moléculaire; URA-CNRS 2172; 25 rue du Docteur Roux, | | | |
| 16 | 75724 Paris Cedex 15, France. E.mail: francoise.norel@pasteur.fr; | | | |
| 17 | Phone : 33 140613122; Fax 33 145688960. | | | |

ABSTRACT

1

2 Proteins that bind σ factors typically attenuate σ factor function by restricting its 3 4 access to the RNA polymerase (RNAP) core enzyme. An exception to this general rule is the Crl protein that binds the stationary phase sigma factor σ^{s} (RpoS) and enhances its affinity for 5 the RNAP core enzyme, thereby increasing expression of σ^{s} -dependent genes. Analyses of 6 7 sequenced bacterial genomes revealed that *crl* is less widespread and less conserved at the 8 sequence level than is rpoS. Seventeen residues are conserved in all members of the Crl 9 family. Site-directed mutagenesis of the crl gene from Salmonella enterica serovar 10 Typhimurium and complementation of a Δcrl mutant of *Salmonella* indicated that substitution 11 of the conserved residues Y22, F53, W56 and W82 decreased Crl activity. This conclusion was further confirmed by promoter binding and abortive transcription assays. We also used a 12 13 bacterial two-hybrid system (BACTH) to show that the four substitutions in Crl abolish Crl- σ^{s} interaction and that residues 1 to 71 in σ^{s} are dispensable for Crl binding. In *Escherichia* 14 *coli*, it has been reported that Crl also interacts with the ferric uptake regulator Fur, and that 15 Fur represses crl transcription. However, the Salmonella Crl and Fur proteins did not interact 16 17 in the BACTH system. In addition, a fur mutation did not have any significant effect on the 18 expression level of Crl in Salmonella. These results suggest that the relationship between Crl 19 and Fur is different in Salmonella and E. coli.

INTRODUCTION

1 2

In bacteria, transcription depends on a multisubunit RNA polymerase (RNAP) 3 4 consisting of a catalytically active core enzyme (E) with a subunit structure $\alpha_2\beta\beta'\omega$, that associates with any one of several σ factors to form different holoenzyme (E σ) species. The σ 5 subunit is required for specific promoter binding, and different σ factors direct RNAP to 6 7 different classes of promoters, thereby modulating the gene expression patterns (17). The RNA polymerase holoenzyme containing the σ^{70} subunit is responsible for the transcription of 8 most genes during exponential growth (17). When cells enter stationary phase or are under 9 specific stress conditions (high osmolarity, low pH or high and low temperatures) during 10 exponential growth, σ^{S} , which is encoded by the *rpoS* gene, becomes more abundant, 11 12 associates with the core enzyme, and directs the transcription of genes essential for the 13 general stress response and for stationary phase survival (17, 20, 25).

14 Sigma factors compete for binding to a limited amount of the core polymerase (16, 17, 20, 34). σ^{70} is abundant throughout the growth cycle and has the highest affinity of all sigma 15 factors for E *in vitro* (20). In contrast, levels of σ^{s} only reach about one-third of the σ^{70} levels 16 upon entry into stationary phase, and σ^{s} exhibits the lowest affinity for E of all sigma factors 17 in vitro (20). The cell uses at least two strategies to ensure the switch between σ^{70} and σ^{s} -18 associated RNA polymerases and to allow gene expression to be reprogrammed upon entry 19 into stationary phase. Several factors (Rsd, 6S RNA, ppGpp and DksA) indirectly increase σ^{s} 20 competitiveness by decreasing the ability of σ^{70} to bind to E (25). In addition, the 21 unconventional regulatory protein Crl increases the performance of σ^{S} . 22

23 The *crl* gene product is a regulator of σ^{s} activity in *Escherichia coli* (2, 14, 36, 50) and 24 *Salmonella* (38, 40, 41). In both species, the Crl protein binds σ^{s} (2, 12) and facilitates RNA polymerase holoenzyme $E\sigma^{S}$ formation (12, 50), thereby enhancing σ^{S} effectiveness (10, 14, 27, 38, 40, 41, 50). The Crl protein of *Salmonella* binds σ^{S} with a stoechiometry 1:1 and increases the affinity of σ^{S} for the core enzyme 7-fold (12). In contrast, Crl does not bind σ^{70} and does not modify the affinity of σ^{70} for the core enzyme (12). Lelong *et al.* (28) reported that, in *E. coli* W3110, Crl can interact with the ferric uptake regulator Fur, a key protein for the control of intracellular iron concentration (8 and references therein) and that Fur represses *crl* transcription (28).

8 In an attempt to gain further insight into Crl function, we searched for homologues of 9 Crl in protein databases. Residues conserved in all members of the Crl family were 10 substituted by site-directed mutagenesis experiments to identify residues important for Crl 11 activity in *Salmonella*. We used the bacterial adenylate cyclase two-hybrid system (BACTH, 12 22) to determine whether these residues and specific regions of Crl are required for efficient 13 interaction with σ^{S} . We also used the BACTH system to assess the effect of the Crl mutations 14 on the interaction of Crl with Fur.

MATERIALS AND METHODS

2

1

3 Bacterial strains, plasmids and growth conditions. Strains and plasmids are listed in Table 1. Bacteriophage P22HT105/1int was used to transfer mutations between Salmonella strains 4 5 by transduction (44). Green plates, for screening for P22-infected cells or lysogens, were prepared as described previously (48). Strains were routinely cultured in Luria Bertani 6 medium (LB; 43). Minimal medium was M9 (43) containing 20 mM glucose. For routine 7 monitoring of multicellular behaviour (rdar morphotype), Salmonella strains were grown at 8 28°C on LB agar plates without NaCl (LB0) supplemented with Congo Red (40 μ g ml⁻¹) as 9 described (CR plates, 42). Antibiotics were used at the following concentrations: ampicillin 10 (Ap), 100 μ g ml⁻¹; carbenicillin (Cb), 100 μ g ml⁻¹; chloramphenicol (Cm), 15 μ g ml⁻¹ for the 11 chromosomal resistance gene and 30 μ g ml⁻¹ for the plasmid resistance gene; kanamycin, 12 (Km) 50 μ g ml⁻¹; and tetracycline (Tet) 20 μ g ml⁻¹. 13

14

15 **Oxidative shock survival assay**. Cells were grown to stationary phase in LB (optical density 16 at 600nM (OD_{600}) of 3.5 to 4), washed, resuspended in PBS and mixed with H_2O_2 at a final 17 concentration of 15 mM. Aliquots of bacteria were removed at time intervals, and numbers of 18 colony forming cells were determined on LB plates.

19

20 DNA manipulations and sequence analysis. Standard molecular biology techniques were 21 used (43). Oligonucleotides were obtained from Sigma-Aldrich (France). DNA sequencing 22 was performed by Cogenics (France). DNA and amino acid sequence analyses were conducted using the BLAST programs at the NCBI (National Center for Biotechnology
 Information).

3

Cloning of the rpoS_{T2} allele. We demonstrated that the nucleotide sequence of a PCR-4 5 amplified *rpoS* gene from *Salmonella enterica* serovar Typhi strain T2 ($rpoS_{T2}$) is identical to that in Salmonella ATCC14028 (41) except for a G/A mutation at position 318, which does 6 7 not change the amino acid sequence, and a deletion of 8 nucleotides starting at position 109, 8 which results in a frameshift and appearance of a premature stop codon in $rpoS_{T2}$. Total DNA 9 from S. Typhi T2 was cleaved with BglII and the resulting fragments were cloned into the 10 BamHI site of pACYC184. Recombinant plasmids were then transformed into the E. coli 11 rpoS mutant strain MC1061K carrying an rpoS-dependent spvRAB-lacZ fusion on pSTF4 (26). Selection of Cb^R Cm^R Lac⁺ clones and subsequent examination of their plasmid DNA 12 13 by restriction analysis and DNA sequencing of *rpoS* produced pAC*rpoS*_{T2} containing a 6 kb 14 BglII fragment carrying rpoS_{T2}. A 3.3 kb SphI-ScaI fragment carrying rpoS_{T2} was 15 subsequently cloned into the SphI and SmaI sites of pUC19 to yield pUCrpoS_{T2}.

16

17 **Construction of S. Typhimurium 2922K***rpoS*_{T2}. The *rpoS* allele in *S*. Typhimurium 18 ATCC14028 was replaced by the *rpoS*_{T2} allele using the following strategy (41). $pUCrpoS_{T2}$ 19 contains the *rpoS*_{T2} allele of *S*. Typhi strain T2 and the downstream genes, including a gene 20 encoding a putative decarboxylase (named STY3047 in *S*. Typhi and STM2922 in *S*. 21 Typhimurium) that is not regulated by *rpoS* (1). In $pUCrpoS_{T2}K$, a kanamycin resistance 22 (Km) cartridge is located in the *Sma*I site of STY3047. $pUCrpoS_{T2}K$ was introduced by 23 electroporation into ATCC*rpoS*, where it was unstable. Recombination of the Km cartridge 1 into the host genome with simultaneous loss of pUCrpoS_{T2}K resulted in recombinants that were resistant to kanamycin and sensitive to carbenicillin. A Km^R Cb^S Cm^S recombinant was 2 selected, checked by PCR for the presence of the STM2922::Km mutation and simultaneous 3 4 replacement of the $\Delta rpoS$::Cm mutation by the $rpoS_{T2}$ allele. The $rpoS_{T2}$ and STM2922::Km alleles were then co-transduced to ATCCrpoS. Transductants that were Km^R but Cm^S were 5 selected, and transduction of the STM2922 :: Km mutation and simultaneous replacement of 6 the $\Delta rpoS$:: Cm mutation by the $rpoS_{T2}$ allele were confirmed by PCR. One Km^R Cm^S strain, 7 8 designated 2922KrpoS_{T2}, was also checked by DNA sequencing for the presence of the 9 rpoS_{T2} allele.

10

Determination of the N-terminal sequence of σ^{S}_{T2} , pQE*rpoS*_{T2}, which produces the *rpoS*_{T2} 11 12 gene with a C-terminal His₆ extension under the control of the pQE60 IPTG-inducible promoter, was constructed as follows. Primers specific to the ends of rpoS (5'-13 14 5'-CCGTCAAGAATTCACGGGGTAGGAGCCACCTTATGA-3' and 15 TGATCGAGATCTCTCGCGGAACAGCGCTTCGATATTCA- 3') were used to amplify a 1 kb fragment of S. Typhi T2 total DNA by PCR. EcoRI and BglII restriction sites were 16 17 incorporated at its 5' and 3' ends, respectively. After digestion with EcoRI and BglII, the PCR-amplified fragment was ligated into the EcoRI and BglII sites of pQE60. The nucleotide 18 19 sequence of the rpoS insert in pQE $rpoS_{T2}$ was verified. In this construct, $rpoS_{T2}$ is expressed 20 from its own translation sequences. E. coli JM109 carrying pQErpoS_{T2} was grown in LB 21 containing ampicillin at 37°C to an optical density of 0.9, when 0.7 mM isopropyl-β-D-22 thiogalactoside (IPTG) was added. The cells were harvested after 5 h, resuspended in buffer A 23 (50 mM Na₂HPO₄ pH 8, 300 mM NaCl) and then disrupted in by French Press. The crude cell 24 extract was supplemented with Benzonase (1250 U) and imidazole (10 mM) and centrifuged 25 at 27,000 g for 30 min. The supernatant was added to 1.5 ml Ni-NTA column and washed 1 with buffer A containing 20 mM imidazole. σ^{s}_{T2} -His₆ was eluted with buffer A containing 2 300 mM imidazole. After electrophoresis in a 10% polyacrylamide-SDS gel, the σ^{s}_{T2} -His₆ 3 protein was transferred on a PVDF membrane, stained with Amido Black and excised from 4 the membrane for N-terminal sequence analysis. Amino-terminal Edman degradation was 5 carried out on an Applied Biosystems sequencer.

6

Construction of a chromosomal crl-lac transcriptional fusion. We previously created a 7 8 chromosomal mutation in the crl gene of Salmonella ATCC14028 using PCR-generated 9 linear DNA fragments and the λ Red recombination method (ATCC*crl*, 40). A single-copy crl-lacZY transcriptional gene fusion was constructed in ATCCcrl using conditional plasmids 10 11 containing promoterless lacZY genes and the FLP recognition target (FRT) site as described 12 (11). PCR assays were used to ensure integration of the plasmids in the correct location and to 13 exclude the presence of multiple plasmid integrants (using standard test primers, such as those 14 described in 11). We also used flanking locus-specific primers to amplify junction fragments 15 that were subsequently analysed by DNA sequencing. Isogenic strains were constructed by 16 P22 HT int mediated transduction of the mutations into the appropriate strains.

17

Enzymatic assays. β-galactosidase activity was measured as described by Miller (31) and is
expressed in Miller units.

20

BACTH assays. The bacterial adenylate cyclase-based two hybrid (BACTH) assay is dependent upon the functional reconstitution of the *Bordetella pertussis* adenyl cyclase T18 and T25 subdomains by two interacting partners (22). The resulting cAMP binds to and activates the transcription activator CRP, a positive regulator of t *lac* and *mal* operons involved in lactose and maltose catabolism. Derivatives of pUT18, pKT25 and pKNT25, used

1 in the BACTH assays, were constructed by cloning PCR amplified DNA fragments encoding 2 the protein of interest from S. Typhimurium between the PstI and EcoRI sites of pUT18 and pKNT25 and the XbaI and KpnI sites of pKT25. The DNA encoding Crl was amplified using 3 primers Crl-Pst 5'-CA TGC CTG CAG AAG GAG ATC GCA ATG ACG TTA CCG AGT-4 5 3' and Crl-Eco 5'-T CGA TGA ATT CGA TGC CGA CAG TTT TAC CGG CTC GTC GT-3', restricted by *PstI* and *Eco*RI and cloned into the *PstI* and *Eco*RI sites of both pUT18 and 6 pKNT25 to yield pUTCrl and pKNTCrl, respectively. The DNA encoding RpoS was 7 amplified using primers RpoS-Xba 5'-TCG ACT CTA GAT ATG AGT CAG AAT ACG 8 9 CTG AAA GTT CAT-3' and RpoS-Kpn 5'-AC TTA GGT ACC TTA CTC GCG GAA CAG 10 CGC TTC GAT-3' restricted by XbaI and KpnI and cloned into the XbaI and KpnI sites of pKT25 to yield pKTRpoS. The DNA encoding Fur was amplified using primers Fur-Pst 5'-11 CAT GCC TGC AGC ATG ACT GAC AAC AAT ACC GCA-3' and Fur-Eco 5'-T CGA 12 TGA ATT CGA TTT AGT CGC GTC ATC GTG CGC GT-3' for cloning into the PstI and 13 *Eco*RI sites of pUT18 (yielding pUTFur) and using Fur-Xba 5'- TCG ACT CTA GAC ATG 14 15 ACT GAC AAC AAT ACC GCA TTA AAG A-3' and Fur-Kpn 5'- AC TTA GGT ACC 16 TTA TTT AGT CGC GTC ATC GTG CGC GT-3' for cloning into the XbaI and KpnI sites of pKT25 (yielding pKTFur). pKTRpoS_{$\Delta 12-71$} is a derivative of pKTRpoS with an in frame 17 deletion of an internal HpaI-DraI fragment. pKTRpoS $_{\Delta 1-38}$ contains a DNA fragment 18 amplified with primers RpoST2-Xba 5'- TCG ACT CTA GAC CTG GCT GAA GAA GAG 19 20 CTG TTA TCG CAA-3' and RpoS-Kpn (see above) and cloned into the XbaI and KpnI sites 21 of pKT25. pKTRpoS $_{\Delta 1-71}$ contains a DNA fragment amplified with primers RpoS-Xba5 5'-TCG ACT CTA GAA ACA GCC GAA GAA GAA GTC TAT TTT GCG CGT-3' and RpoS-22 Kpn (see above) and cloned into the XbaI and KpnI sites of pKT25. All plasmids were 23 24 confirmed to be correct by DNA sequencing. The E. coli cya strain BTH101 was transformed 25 with derivatives of plasmids pKT25, pKNT25, and pUT18 encoding proteins fused to the T25

1 and T18 fragments of Bordetella pertussis adenyl cyclase. Co-transformants were plated onto 2 MacConkey maltose plates supplemented with carbenicillin, kanamycin and 0.5 mM IPTG to assess the Mal⁺ phenotype and on LB plates supplemented with 5-bromo-4-chloro-indolyl-β-3 D-galactoside (X-Gal, 40 µg ml⁻¹), carbenicillin, kanamycin and 0.5 mM IPTG to assess the 4 5 Lac⁺ phenotype and β -galactosidase activity. Plates were incubated at 30°C for 3 days and 6 colonies were then collected and their β -galactosidase activities were measured as described by Miller (31). For immunodetection of the Crl-T18 and T25- σ^{s} chimeras, BTH101 7 derivatives were grown to stationary phase (OD600= 3.5) in LB in the presence of 2 mM 8 9 cAMP and 0.5 mM IPTG to induce fully the *lac* promoter on pKT25, pKNT25 and pUT18.

10

11 Construction of plasmids with mutated *crl* genes

12 Site-directed mutagenesis was performed using the QuikChange II Site-directed mutagenesis 13 kit (Stratagene) as recommended by the manufacturer. Site-directed mutagenesis of plasmid 14 pACcrl-1 yielded plasmids pACcrl_{Y22A}, pACcrl_{F35A}, pACcrl_{C28A}, pACcrl_{C37A}, pACcrl_{C41A}, 15 pACcrl_{F53A}, pACcrl_{W56A}, pACcrl_{G74A}, pACcrl_{G80A}, pACcrl_{W82A}, pACcrl_{F103A}, pACcrl_{C28-37}, 16 pACcrl_{C28-41}, pACcrl_{C37-C41} and pACcrl_{C28-37-41} (Table 1). Plasmids pUTCrl_{Y22A}, pUTCrl_{F53A}, 17 pUTCrl_{W56A}, pUTCrl_{W82A} and pUTCrl_{G80A} (Table 1) were obtained by cloning PCR amplified 18 DNA fragments from plasmids pACcrl_{Y22A}, pACcrl_{F53A}, pACcrl_{W56A}, pACcrl_{W82A}, and 19 pACcrl_{G80A}, respectively, between the PstI and EcoRI sites of pUT18 using primers Crl-Pst 20 and Crl-Eco (as described above). Plasmids pUTCrl_{R51A}, pUTCrl_{E52A}, pUTCrl_{G55A}, pUTCrl_{W57A} 21 were obtained by site-directed mutagenesis of plasmid pUTCrl (Table 1). All plasmids were 22 confirmed to be correct by DNA sequencing.

23

Electrophoresis and immunoblot analysis of proteins. Whole-cell extracts were prepared
 and SDS-polyacrylamide gel electrophoresis was carried out as described by Silhavy *et al.*

1 (46). The amount of protein in whole-cell lysates was determined using the DC Protein Assay 2 kit (Bio-Rad). Equal amounts of protein were loaded in each slot. The molecular sizes of the 3 proteins were estimated using molecular size standards (Fermentas, France). Rabbit antibodies against the Crl protein of Salmonella were from Robbe-Saule et al. (40). Rabbit 4 antibodies against the σ^{S} protein of S. enterica serovar Typhimurium were from Coynault et 5 al. (7). Proteins were transferred to Amersham Hybond P membranes (GE Healthcare) and 6 7 incubated with the polyclonal rabbit antibody serum as previously described (7). Bound antibodies were detected using a secondary anti-rabbit antibody linked to peroxidase and the 8 9 Amersham ECL plus western blotting detection system kit (GE Healthcare).

10

Overproduction and purification of His6-Crl variants. A 250 ml of JM109 carrying 11 12 pQEcrl wild-type or mutant was grown in LB medium containing carbenicillin at 30°C to an 13 optical density of 0.6, and then 1 mM IPTG was added. After 30 min, the temperature was 14 lowered to 23° C. The cells were harvested after 4 h, washed, resuspended in 10 ml of buffer 15 A (25 mM Tris pH 8.0, 1 M NaCl, 0.01% NP-40) supplemented with "Complete EDTA free 16 anti-protease" as described by the manufacturer (Roche) and then disrupted in a French press 17 (Aminco). The crude cell extract was centrifuged at 15,000 g for 30 min. The supernatant was 18 adjusted to 3 mM imidazole, added to 1.5 ml of Ni-nitrilotriacetic acid agarose (QIAGEN) 19 and gently mixed for 1 h. The slurry was packed onto an Econo-Pac column (Bio-Rad) and 20 washed with 15 ml of buffer A containing 20 mM imidazole. His₆-Crl was eluted with buffer 21 A containing 250 mM imidazole, dialyzed against buffer B (20 mM Tris-HCl pH 8, 50 mM 22 NaCl, 0.01% NP-40), and added to DNA-cellulose equilibrated in the same buffer (0.3 ml, 23 Sigma D-8515). The DNA-cellulose was removed by centrifugation and the concentrations of 24 the proteins determined by a Bradford assay.

1 Gel retardation assays. The DNA encoding rpoS was amplified using primers HK1 5'-AGGCTCGGATCCATGAGTCAGAATACGCTGAAAGTTCAT-3' 2 and HK2 5'-3 TTCCGAAAGCTTTTACTCGCGGAACAGCGCTTCGATATT-3' and cloned into the BamHI and HindIII sites of pQE30 to yield pQE30rpoS. The nucleotide sequence of the rpoS 4 insert in pQE30*rpoS* was checked by DNA sequencing. S. Typhimurium His₆- σ^{s} produced 5 from pQE30*rpoS* was purified using Ni-affinity chromatography as described for His₆- σ^{S}_{T2} . 6 E. coli core (Epicentre) was used to reconstitute the σ^{s} -holoenzyme, which was incubated 7 with variant Crl proteins (wild-type or mutants) or buffer. 9 µl of the protein complexes were 8 then added to 3 µl of [³²P]-labelled *katN* fragment in buffer A (40 mM Hepes pH 8.0, 10 mM 9 MgCl₂, 100 mM K-glutamate, 2 mM DTT) containing 500 µg ml⁻¹ BSA, and incubated at 10 28°C for 20 min. The final concentrations of core, His₆- σ^{s} , Crl and *katN* fragments were 11 12 respectively 8 nM, 32 nM, 3 µM and 1 nM. After addition of 3 µl loading buffer (buffer A containing 50% sucrose, 0.025 % xylene cyanol blue and 150 µg ml⁻¹ of heparin) the mixture 13 14 was loaded onto a 6 % native polyacrylamide gel run in TG buffer (25 mM Tris, 192 mM 15 Glycine pH 8.5) at 8V/cm. The gel was dried before being autoradiographed and quantified 16 using a PhosphorImager (Molecular Dynamics).

17

Abortive initiation assays. A mixture of the 207-bp *lacUV5* fragment (10 nM), ApA (2 mM), and 80 μ M [α -³²P] UTP was added to an equal volume of the holoenzyme (60 nM, His₆- σ ^S: core ratio = 2.5 : 1) preincubated with buffer or Crl (wild-type or variants) at 6 μ M concentration at 28° C. Aliquots were removed at various times. The aliquots were spotted onto Whatman 3MM paper prespotted with 100 mM EDTA, and chromatograms were developed as previously described (40).

2

RESULTS

4 A Blast search in 1264 sequenced bacterial genomes (www.ncbi.nlm.nih.gov) revealed 5 143 genomes containing a *crl* gene. They are all γ -Proteobacteria and contain an *rpoS* gene 6 (Table 2). Crl was absent from many rpoS-containing bacteria, for example Pseudomonas, in 7 which RpoS plays an important role (45). Thus, *crl* is not as widely distributed as *rpoS* and it 8 is also less conserved at the sequence level (Table 2). Sequence comparison of the 143 Crl 9 proteins identified 60 Crl sequences that differed by at least one amino acid. Alignment of 10 these 60 Crl sequences by clustalW revealed 17 residues that are conserved in all members of 11 the Crl family (Figure 1).

12 Analysis of the 16 completed genomic sequences of Salmonella 13 (www.ncbi.nlm.nih.gov) indicated that the Crl sequence is highly conserved (Figure 1 and 14 Table 2). Interestingly, the crl gene in S. enterica serovar Paratyphi B SPB7 lacks a G at 15 position 73 in the open-reading frame, causing the appearance of a premature stop codon and 16 the likely production of a truncated, non-functional Crl protein.

17

The first 71 amino acids of σ^{s} are non-essential for Crl binding 18

19 Crl is less conserved than the full length RpoS protein (Table 2). However, the first 55 20 residues of RpoS are poorly conserved, whereas the remaining 275 residues of RpoS are 21 highly conserved (data not shown). For instance, the level of sequence identity between RpoS_{STM} and the RpoS proteins from the Enterobacteria Proteus mirabilis and Providencia 22 23 stuartii is 43% and 40% for aminoacids 1-55 compared to 87% and 86% for amino acids 56-330, respectively. These first 55 residues define a specific domain of σ^{s} that is not well 24 conserved in other sigma factors besides σ^{70} (corresponding roughly to region $\sigma_{1,1}$, see also 25 26 Discussion section and 32, 51), and might thus play a role in the specificity of recognition of 1 σ^{s} by Crl. In addition, $\sigma_{1.1}$ and Crl might play complementary roles in the formation of $E\sigma$. 2 Crl increases the affinity of σ^{s} for the core enzyme (12). $\sigma_{1.1}$ stabilizes the interaction 3 between σ^{70} and core RNAP (18) and might play a similar role in σ^{s} -core interaction. These 4 possibilities prompted us to determine whether this domain of σ^{s} is required for σ^{s} activity 5 and for Crl activation.

We first examined the sensitivity to Crl activation of a σ^{s} mutant lacking the first 39 6 residues of σ^{S} (σ^{S}_{T2} , Figures 2 and 3) and thus lacking most of region $\sigma_{1.1}$ (32, 51). *rpoS*_{T2} is a 7 natural rpoS allele from S. Typhi strain T2 that contains a deletion of eight nucleotides in the 8 9 5' end of the coding sequence resulting in the appearance of a premature stop codon (Figure 2A). Surprisingly, *rpoS*_{T2} was able to direct the synthesis of a 35 kDa σ^{s} protein (σ^{s}_{T2} , Figure 10 3A). N-terminal sequence analysis indicated that σ^{s}_{T2} has a N-terminal Met residue followed 11 by a sequence that is identical to that downstream of residue 39 in native σ^{s} (Figure 2B). We 12 inferred that σ^{s}_{T2} is encoded from the mutant gene in which translation initiates at a GTG start 13 codon preceding codon 40 of σ^{s} in $rpoS_{T2}$, and upstream of which is a putative ribosome 14 binding site (SD, Figure 2B). Translational reinitiation within the E. coli rpoS gene has also 15 been reported previously (15). To facilitate characterization of the $rpoS_{T2}$ allele in isogenic 16 backgrounds, the rpoS gene from S. Typhimurium ATCC14028 was replaced by rpoS_{T2} 17 yielding strain 2922KrpoS_{T2}. During exponential growth, σ^{S}_{T2} was produced in higher 18 amounts than the wild-type σ^{s} protein (OD600 0.1 to 1.4, Figure 3B), likely because the N-19 terminal 11 residues of σ^{s} involved in ClpX recognition and σ^{s} proteolysis by ClpP (13, 49) 20 were not present in σ_{T2}^{s} . σ^{s} is required for bacterial resistance to various stresses during 21 stationary phase (the so-called general stress resistance) (20,25). The σ^{s} -dependent gene *katE* 22 23 encodes a catalase, an enzyme that detoxifies hydrogen peroxide (H_2O_2) and contributes to the 24 resistance of Salmonella to oxidative stress in stationary phase (38). Despite the high amount 1 of σ^{s}_{T2} produced, strain 2922K*rpoS*_{T2} was slightly less resistant to H₂O₂ than the wild-type 2 strain 2922K and exhibited lower expression of the *katE-lacZ* fusion (Figure 2CD). This 3 suggested that σ^{s}_{T2} is less active than wild type σ^{s} . *In vitro* transcription experiments with 4 σ^{s}_{T2} confirmed this hypothesis (data not shown). Interestingly, the levels of H₂O₂ resistance 5 and the expression of the catalase gene *katE* were more dependent on Crl activation in 6 2922K*rpoS*_{T2} than in the wild-type strain (Figure 2CD). This indicated that the σ^{s}_{T2} protein is 7 able to interact with Crl and that its activity is highly dependent on Crl.

8 We previously used the bacterial two-hybrid system (BACTH system, 22) to detect interaction between Crl and σ^{s} (12). C-terminal fusions of S. Typhimurium σ^{s} to T25 (T25-9 σ^{S}) and N-terminal fusions of S. Typhimurium Crl to T18 (Crl-T18) were produced in the E. 10 *coli cya lac*⁺ strain BTH101 from pKTRpoS and pUTCrl, respectively, and the resulting β -11 galactosidase activities were measured (Table 3). As expected (12), T25- σ^{s} interacted with 12 Crl-T18 (Table 3). In this system, the first 38 residues of σ^{s} were dispensable for efficient 13 binding of Crl (compare T25- σ^{s} and T25- $\sigma^{s}_{\Lambda 1-38}$ in the presence of Crl-T18, Table 3). 14 Moreover, deletion of residues 1 to 38 from σ^{s} resulted in increased levels of the chimera and 15 β -galactosidase activity in the BACTH assay (Figure 3C lanes 2 and 5, Table 3). This is 16 probably again due to the absence of the ClpX recognition site in T25- $\sigma^{s}_{\Delta 1-38}$. Interestingly, 17 deletion of residues 1 to 71 in T25- $\sigma^{s}_{\Delta 1-71}$, and deletion of residues 12 to 71, which does not 18 remove the ClpX recognition site (13, 49) in T25- $\sigma^{s}_{\Delta 12-71}$, had a more pronounced effect on 19 σ^{s} levels and β -galactosidase activity than deletion of residues 1 to 38 (in T25- $\sigma^{s}_{\Lambda 1-38}$) (Figure 20 3C lanes 4-6, Table 3). This indicated that the first 71 residues in σ^{S} are non-essential for Crl 21 binding, and further suggested that the deletion of this region induces a conformational 22 change in σ^{s} that favors its stability. Altogether, these results suggested that Crl interacts with 23 the conserved amino acid region (72-330) of σ^{s} . 24

2

Identification of conserved residues important for Crl activity

3 Alanine substitution mutagenesis was performed to investigate the functional relevance of the conserved residues in Crl. The conserved residues are located in four regions 4 5 of Crl (one third of them are aromatic amino-acids, Figure 1). As a first step, one or two 6 residues in each conserved region were mutated. We also mutated the three cysteine residues 7 in Crl. pACcrl-1 expresses the crl gene from S. Typhimurium under the control of the vector 8 cat promoter, and is able to complement a Δcrl mutant of Salmonella (40). Site-directed 9 mutagenesis of pACcrl-1 yielded eleven derivatives expressing altered Crl proteins (Table 1 10 and Figure 4A). The ability of the mutated *crl* genes to complement the *Salmonella* Δcrl 11 mutant was assessed using three different tests (Figures 4 and 5). The development of the rdar 12 morphotype depends on Crl (40) and was used as a qualitative test (Figure 5). This colony 13 morphology is caused by production of curli and cellulose and is correlated with biofilm 14 formation and expression of the regulatory gene csgD (42). In ATCC14028, the expression 15 level of a *csgD-lacZ* fusion and the H₂O₂ resistance level at 28°C both depend on Crl (38, 40), 16 and were used as quantitative assays (Figure 4BC).

17 Derivatives of pACcrl-1 expressing the Crl proteins with F35A, C28A, C37A, G74A, G80A and F103A substitutions complemented Δcrl as well as did pACcrl-1 (Figures 4-5 and 18 19 data not shown). Two substitutions, W56A and F53A, located in a conserved motif of Crl 20 (residues 48 to 57, Figure 1), substantially diminished complementation in all three tests 21 (Figure 4 and Figure 5 spots 7-8). Two additional substitutions, Y22A and W82A, also 22 decreased the complementation levels but to a lesser extent (Figure 4 and Figure 5 spots 3 and 23 11). These four substitutions did not significantly decrease the cellular level of Crl (Figure 24 6A) and, thus, likely affected its activity.

1 The Salmonella Crl protein contains three cysteines, (C28, C37, C41), one of which 2 (C41) is conserved in all Crl members. Alanine substitution of cysteine residues (C28A, 3 C37A and C41A) did not significantly affect the levels of Crl protein or its activity (Figures 4, 5 and 6A). Only a slight effect of C41A on H_2O_2 resistance was observed (Figure 4B). 4 5 Interestingly, substitution of all three cysteine residues substantially decreased Crl levels (Figure 6A lanes 17 and 20). Examination of the Crl family members indicated that their all 6 contain at least 2 cysteine residues. Altogether, these results suggested that the presence of at 7 least one, and of probably two, cysteine residue(s) is required for Crl stability. Double 8 9 mutations in the cysteine residues in Crl (Crl_{C28-37}, Crl_{C28-41}, Crl_{C37-41}) were constructed to 10 determine what combination of cysteine substitution affected Crl stability. The combinations 11 of C28A and C37A substitutions and C28A and C41A substitutions did not significantly 12 affected the levels of Crl protein (Figure 6A lanes 14 and 15). In contrast, alanine substitution 13 of both C37 and C41 substantially decreased Crl levels (Figure 6A lane 16) suggesting that 14 these two cysteine residues are important for Crl stability. Levels of the triple Crl_{C28-37-41} 15 mutant protein were lower than that of Crl_{C37-41} (Figure 6A, lanes 19-20). This suggests that 16 the C28A substitution further decreased the stability of Crl_{C37-41}.

17

18 Screening of the *crl* mutations in the BACTH system and *in vitro* assays

19 We next determined whether the mutations selected above affected Crl- σ^{s} interaction. 20 DNA encoding proteins with Y22A, F53A, W56A or W82A substitutions or the neutral 21 G80A substitution were introduced into pUTCrl and the ability of the Crl-T18 derivatives to 22 interact with σ^{s} was assessed in the presence of T25- σ^{s} (Table 3). Interestingly, the Y22A, 23 F53A, W56A, W82A substitutions, but not the G80A substitution, all abolished the 24 interaction between Crl-T18 and T25- σ^{s} (Table 3). These substitutions did not affect levels of 25 Crl-T18 (Figure 6B). The altered Crl-T18 hybrid proteins were also unable to interact with 1 the T25- σ^{s} derivative lacking amino acids 12 to 71 (data not shown). In conclusion, the 2 Y22A, F53A, W56A and W82A substitutions all abolished the interaction between Crl and σ^{s} 3 either directly or indirectly through conformational changes in Crl.

4 The activities of the variant proteins were also probed in vitro by gel retardation and 5 abortive initiation assays, based on the ability of Crl to increase the amount of promotercomplexes formed by the σ^{s} -RNA polymerase (E σ^{s}) and activate transcription initiation (12, 6 40). As shown on the autoradiogram of the native polyacrylamide gel (Figure 7A lane 3), $E\sigma^{s}$ 7 binding to the [³² P]-labelled katN promoter was stimulated 2-fold by wild-type Crl. The 8 Y22A and W82A variants were still able to increase the formation of $E\sigma^{S}$ -katN complexes by 9 about 1.8 and 1.6 fold (Figure 7A lanes 4 and 6). In contrast, the F53A and W56A variants of 10 11 Crl (Figure 7A lanes 5 and 7) had hardly any effect, in agreement with their highly reduced 12 ability to stimulate *csgD-lacZ* expression or hydrogen peroxide resistance. The gel retardation 13 methodology is only semi-quantitative, however, and the addition of heparin to disrupt non-14 specific core-DNA binding before electrophoresis might have disturbed weak interactions between Crl variants and σ^{s} . Therefore, abortive initiation assays, which do not require 15 heparin addition and can be exploited for more quantitative studies, were conducted using the 16 well-characterised *lac*UV5 promoter, which can be transcribed *in vitro* by $E\sigma^{70}$ and $E\sigma^{S}$ and, 17 18 unlike katN, has the advantage of producing high amounts of abortive products (29). As expected, $E\sigma^{s}$ produced less ApApUpU tetranucleotide from *lac*UV5 in the presence of the 19 Crl variants than with the wild-type protein (Figure 7B). The results were more obvious than 20 in gel retardation assays since wild-type Crl stimulated $E\sigma^{s}$ -*lac*UV5 abortive transcription by 21 22 4-fold under our conditions. After 30 min incubation at 28°C, the observed decrease in ApApUpU synthesis was 35 % for Y22A, 40% for W82A and 90% for F53A. The addition of 23 24 W56A Crl resulted in an even lower activity than in the absence of Crl, suggesting a slight inhibitory effect of this variant on σ^{s} -dependent transcription. Altogether, these results 25

emphasise the role of the REF-GWW motif (Figure 1) in Crl activation of σ^{s} -dependent promoters. Consistent with this hypothesis, the R51A, W57A substitutions and to a lesser extent the E52A and G55A substitutions affected the interaction between Crl and σ^{s} in the BACTH assay (Table 3) but did not affect levels of Crl-T18 (Figure 6B).

5

6 Fur-Crl interactions in Salmonella

To determine whether the four substitutions studied above affected Crl- σ^{s} interaction 7 specifically, we assessed their effect on the interaction of Crl with another partner. In a 8 9 previous study using *E. coli* W3110, an interaction between Crl and the ferric uptake regulator Fur was detected by co-purification (28). To investigate the interaction between the 10 11 Salmonella Crl and Fur proteins in the BACTH assay, we constructed plasmids pUTFur 12 encoding a Fur-T18 protein and pKTFur encoding a T25-Fur protein. Significant interactions 13 between the Crl- and the Fur- hybrid proteins were not detected (Table 3). The Fur hybrid 14 proteins were stable, as indicated by the fact that the expected dimerization of Fur-T18 and 15 T25-Fur (8 and references therein) gave a positive signal (Table 3).

16 crl transcription in E. coli W3110 is repressed by Fur (by 100-fold) and by Crl (by 17 more than 10-fold) (28). To assess the possible relationship between Crl and Fur in 18 Salmonella ATCC14028, we examined the effects of Fur and Crl on crl expression. We first 19 compared Crl levels in the wild-type and *fur* strains by immunodetection of the Crl protein 20 (Figure 6C). Cells were grown to stationary phase in LB at 30°C and 37°C, and also on LB 21 agar at 30°C, a growth condition used to study the regulation of *crl* expression by Fur in *E*. 22 coli W3110 (28). The amount of Crl detected in the fur strain was the same as in the wild-type under all three conditions (Figure 6C, compare lanes 1 and 2, 4 and 5, 7 and 8). In addition, 23 24 the levels of *crl-lacZ*–encoded β -galactosidase were not significantly affected either by the *fur* mutation or by the production of Crl *in trans* from plasmid pAC*crl*-1 (Figure 8AB). These
 results suggested that *crl* transcription is not repressed by Fur and Crl in ATCC14028.

In the presence of iron, Fur negatively controls transcription of a small RNA, RyhB, which facilitates degradation of the *sdhCDBA* mRNA encoding succinate dehydrogenase (30). Consistent with this finding, *sdhA-lacZ* -encoded β -galactosidase in ATCC14028 was dependent on both iron and Fur (Figure 8C). This demonstrated that regulation by Fur is indeed functional in ATCC14028. Fur did not exert any effect on the *crl-lacZ* fusion (Figure 8C). Also, no effect of Fur on *crl-lacZ* expression was detected in LB supplemented with iron (data not shown).

10 E. coli W3110 has many genes that are differentially expressed in E. coli MG1655 (52), and many laboratory stocks of W3110 contain mutations in rpoS (21, 52 and 11 12 www.ncbi.nlm.nih.gov). These findings prompted us to evaluate the effect of the fur mutation 13 on *crl* expression in an *rpoS* background. Crl production and *crl-lacZ*-encoded β galactosidase levels were not affected by the *fur* mutation in the Salmonella rpoS mutant (Figure 6C lanes 14 15 12 and 14, and Figure 8D). As previously shown (38), Crl production was increased by the 16 rpoS mutation (Figure 6C lanes 10, 12, 15) by a mechanism that likely operates at the post-17 transcriptional level (Figure 8D).

18 Altogether, these results suggest that, contrary to *E. coli* W3110, Crl and Fur do not 19 interact in *Salmonella*.

DISCUSSION

2 Analysis of the protein sequence databases revealed the narrow distribution of Crl 3 homologues in bacteria. In contrast, RpoS homologues are found in many gram-negative bacteria of the γ -, δ - and β - subdivisions (<u>www.ncbi.nlm.nih.gov</u>). Several hypotheses might 4 explain the absence of Crl in bacteria containing a RpoS homologue. Crl compensates for a 5 low affinity of σ^{s} for the E, and might be dispensable in bacteria expressing a σ^{s} protein with 6 a high affinity for E. Moreover, the physiological impact of Crl is greatest at low 7 concentrations of σ^{s} (40, 41), and the environment can modulate the impact of Crl by 8 affecting σ^{s} level (38). Thus, depending on the bacterial lifestyles and environment 9 10 encountered, cellular physiology and the mechanisms of regulation of *rpoS* expression, Crl 11 might be dispensable. Alternatively, functional homologues of Crl might exist in some rpoScontaining bacteria or some species might use alternative strategies to favor σ^{s} interaction 12 13 with E. Interestingly, we did not find Crl homologues in bacteria that do not contain rpoS. This is consistent with previous work suggesting that the main function of Crl is to favor σ^{s} 14 15 interaction with RNA polymerase(41, 50).

16 Of the 16 Salmonella strains whose genome sequence is completed, one, S. Paratyphi 17 SB7, contains a frameshift mutation in *crl* that results in the appearance of a premature stop 18 codon. This observation is reminescent of previous findings that rpoS null mutants can be 19 found in natural isolates of S.Typhi (7, 37, 39), and is consistent with our previous 20 observation that a Δcrl mutation increases the competitive fitness of Salmonella by attenuating σ^{s} activity (41). The selection of *rpoS* mutants in bacterial populations likely 21 22 results from a growth advantage of *rpoS* mutants in the absence of environmental stress (24, 33, 34). However, in environments where bacteria encounter mild stress, $\Delta rpoS$ mutants are 23 24 outcompeted by the wild-type strain, whereas rpoS attenuated mutants exhibit increased 1 fitness (33, 41, 53). We can predict that the beneficial effect of a *crl* mutation is significant 2 when the survival strategy requires "enough but not too much" σ^{s} activity.

We showed that four conserved residues (Y22, F53, W56, W82) are important for Crl 3 activity and for Crl- σ^{s} interaction but not for Crl stability. In contrast, a Crl protein in each of 4 its three cysteine residues was substituted was unstable. The Y22A and W82A substitutions 5 6 had a more dramatic effect in the BACTH system than in the complementation tests and the in 7 vitro assays (Table 3 and Figures 4, 5 and 7). The IPTG-inducible lac promoter on pKT25 and pUT18 is induced by cAMP and, thus, a mutation that decreases the efficiency of the 8 interaction between T25- σ^{s} and Crl-T18 also lowers the expression level of the hybrid 9 proteins, thereby amplifying the effect of the mutation (24 and Daniel Ladant personal 10 11 communication).

12 The F53A and W56A substitutions are located in a conserved motif in the middle part 13 of the protein (REF-GWW Figure 1). Analysis of four additional substitutions (R51A, E52A, G55A W57A) in the BACTH assay confirmed the role of the REF-GWW motif in Crl- σ^{s} 14 interactions (Table 3, Figure 6B). The two other substitutions, Y22A and W82A, are outside 15 of this motif. At least some of these mutations likely affect Crl binding to σ^{s} indirectly 16 through conformational changes. The three-dimensional structure of Crl would provide new 17 18 insight into this issue. Unfortunately, our attempts to crystallize Crl have been unsuccessful. 19 In addition, deletion experiments in the BACTH system did not allow us to delineate a subregion of Crl involved in σ^{s} binding. Crl might spread along σ^{s} or, more likely, a specific 20 conformation of Crl might be required for efficient interaction with σ^{s} . 21

The primary determinants of σ^{S} involved in the binding of Crl are unknown, but the lack of interaction between Crl and σ^{70} (12) suggests that the interaction involves sequence determinants or structural features that are specific to σ^{S} . The vast majority of σ factors, including σ^{S} , belong to the so-called σ^{70} family, reflecting their relationship to the principal σ

factor of *E. coli*, σ^{70} . Sequence alignments of σ^{70} family members of groups 1 and 2, to which 1 σ^{S} belongs, reveal that they have four conserved regions (regions 1-4) (32, 35). Among these, 2 region 2 (subregions 2.1 to 2.4) and region 4 (subregions 4.1 and 4.2) contain DNA-binding 3 domains that mediate recognition of the conserved -10 and -35 elements of σ^{70} -dependent 4 promoters respectively. The linear division of σ^{70} factors into functionally distinct regions is 5 6 largely confirmed by structural data, which revealed that primary sigma factors have four 7 flexibly linked domains, $\sigma_{1.1}$, σ_2 , σ_3 and σ_4 , containing regions 1.1, 1.2-2.4, 3.0-3.1, and 4.1-8 4.2, respectively (5, 32). Free σ factors generally do not specifically bind promoter DNA, and the N-terminal $\sigma_{1,1}$ region is autoinhibitory (4, 9). $\sigma_{1,1}$ might act indirectly to inhibit promoter 9 10 binding by stabilizing a compact conformation of σ that is incompatible with promoter 11 recognition (47). Binding to core RNAP induces large movements of the σ domains (3), 12 converting σ into an active conformation in which the DNA binding determinants in σ_2 and 13 σ_4 are exposed (32).

 $\sigma_{1,1}$ shows little conservation between σ^{70} and σ^{S} and is not found in the alternative 14 sigma factors of the σ^{70} family (32, 35). Therefore, the specificity of recognition between Crl 15 and σ^{S} might involve $\sigma_{1,1}$. However, our findings that (1) σ^{S}_{T2} , a protein lacking residues 1 to 16 39, including most of region $\sigma_{1,1}$ in native σ^{s} (32, 51) is dependent on Crl activation (Figure 17 2), and (2) deletion of residues 1 to 71 of σ^{s} , which removes region σ_{11} and part of subregion 18 1.2 (32, 51) does not abolish Crl- σ^{s} interaction (Table 3), ruled out this possibility. 19 Experiments are in progress to determine which region of σ^{s} is involved in Crl binding. In 20 addition to its autoinhibitory role, $\sigma_{1,1}$ stabilizes the interaction between σ^{70} and core RNAP 21 (18). Our finding that σ^{s}_{T2} is -less active but -more dependent on Crl activation than wild-type 22 σ^{s} (Figure 2CD and data not shown) is consistent with a similar role for $\sigma_{1.1}$ in σ^{s} -core 23

1 interaction. Crl increases the affinity of σ^{S} for core RNAP (12) and might partially 2 compensate for the absence of a full region $\sigma_{1,1}$ in σ^{S}_{T2} .

3 Acknowledgments:

- 4 We are grateful to Anthony Pugsley for critical reading of the manuscript. We thank James
- 5 Slauch for providing the *Salmonella fur* strain. We are grateful to Gouzel Karimova and
- 6 Daniel Ladant for providing the *cya* strain and BACTH vectors and for helpful discussions.

| 1 | REFERENCES |
|----|---|
| 2 | |
| 3 | |
| 4 | 1. Bang, IS., J. G. Frye, M. McClelland, J. Velayudhan, and F.C. Fang. 2005. |
| 5 | Alternative sigma factor interactions in <i>Salmonella</i> : σ^{E} and σ^{H} promote antioxidant defences |
| 6 | by enhancing σ^{s} levels. Mol. Microbiol. 56 : 811-823. |
| 7 | |
| 8 | 2. Bougdour, A., C. Lelong, and J. Geiselmann. 2004. Crl, a low temperature-induced |
| 9 | protein in Escherichia coli that binds directly to the stationary phase sigma subunit of RNA |
| 10 | polymerase. J. Biol. Chem. 279:19540-50. |
| 11 | |
| 12 | 3. Callaci S., E. Heyduk, and T. Heyduk. 1999. Core RNA polymerase from <i>E. coli</i> induces |
| 13 | a major change in the domain arrangement of the sigma 70 subunit. Mol. Cell. 3 : 229-138. |
| 14 | |
| 15 | 4. Camarero, J. A., A. Shekhtman, E.A. Campbell, M. Chlenov, T. M. Gruber, D. A. |
| 16 | Bryant, S. A. Darst, D. Cowburn, and T. W. Muir. 2002. Autoregulation of a bacterial σ |
| 17 | factor explored by using segmental isotopic labeling and NMR. Proc. Natl. Acad. Sci USA |
| 18 | 99 : 8536-8541. |
| 19 | |
| 20 | 5. Campbell, E. A., L. F. Westblade, and S. A. Darst. 2008. Regulation of bacterial RNA |
| 21 | polymerase σ factor activity: a structural perspective. Current Opin. Microbiol. 11 : 1-7. |
| 22 | |
| 23 | 6. Chang, A. C., and S. N. Cohen. 1978. Construction and characterization of amplifiable |
| 24 | multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J. Bacteriol. |
| 25 | 134: 1141-56. |

| 2 | 7. Coynault, C., V. Robbe-Saule, and F. Norel. 1996. Virulence and vaccine potential of |
|---|---|
| 3 | Salmonella typhimurium mutants deficient in the expression of the RpoS (sigma S) regulon. |
| 4 | Mol. Microbiol. 22: 149-60. |
| 5 | |

- 8. D'Autréaux, B., L. Pecqueur, A. Gonzalez de Peredo, R. E. M. Diederix, C. CauxThang, L. Tabet, B. Bersch, E. Forest and I. Michaud-Soret. 2007. Reversible Redox- and
 Zinc-dependent dimerization of the *Escherichia coli* Fur protein. Biochemistry 46: 13291342.
- 10
- 9. Dombroski A. J., W. A. Walter, M. T. Jr Record, D. A. Siegele, and C. A. Gross. 1992.
 Polypeptides containing highly conserved regions of transcription initiation factor sigma 70
- 13 exhibit specificity of binding to promoter DNA. Cell. **70**: 501-512.
- 14

15 10. Dong, T., M. G. Kirchhof, and H. E. Schellhorn. 2008. RpoS regulation of gene
expression during exponential growth of *Escherichia coli* K12. Mol. Genet. Genomics 279:
267-77.

18

11. Ellermeier, C. D., A. Janakiraman, and J. M. Slauch. 2002. Construction of targeted
single copy *lac* fusions using lambda Red and FLP-mediated site-specific recombination in
bacteria. Gene 290:153-61.

^{12.} England, P., L. F. Westblade, G. Karimova, V. Robbe-Saule, F. Norel, and A. Kolb.
2008. Binding of the unorthodox transcription activator, Crl, to the components of the
transcription machinery. J Biol Chem. 283:33455-64.

| 1 | | | | |
|----|--|--|--|--|
| 2 | 13. Flynn J. M., S. B. Neher, Y-I. Kim, R. T. Sauer, and T. A. Baker. 2003. Proteomic | | | |
| 3 | discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX- | | | |
| 4 | recognition signals. Mol. Cell 11: 671-683. | | | |
| 5 | | | | |
| 6 | 14. Gaal, T., M. J. Mandel, T. J. Silhavy, and R. L. Gourse. 2006. Crl facilitates RNA | | | |
| 7 | polymerase holoenzyme formation. J. Bacteriol. 188 : 7966-70. | | | |
| 8 | | | | |
| 9 | 15. Gowrishankar, J., K. Yamamoto, P. R. Subbarayan, and A. Ishihama. 2003. In vitro | | | |
| 10 | properties of RpoS (σ^{S}) mutants of <i>Escherichia coli</i> with postulated N-terminal subregion 1.1 | | | |
| 11 | or C-terminal region 4 deleted. J. Bacteriol. 185: 2673-2679. | | | |
| 12 | | | | |
| 13 | 16. Grigorova, I. L., N. J. Phleger, V. K. Mutalik, and C. A. Gross. 2006. Insights into | | | |
| 14 | transcriptional regulation and σ competition from an equilibrum model of RNA polymerase | | | |
| 15 | binding to DNA. Proc. Natl. Acad. Sci. USA 103: 5332-5337. | | | |
| 16 | | | | |
| 17 | 17. Gruber, T. M., and C. A. Gross. 2003. Multiple sigma subunits and the partitioning of | | | |
| 18 | bacterial transcription space. Annu. Rev. Microbiol. 57: 441-466. | | | |
| 19 | | | | |
| 20 | 18. Hinton, D. M., S. Vuthoori, and R. Mulamba. 2006. The bacteriophage T4 inhibitor and | | | |
| 21 | coactivator AsiA inhibits Escherichia coli RNA polymerase more rapidly in the absence of | | | |
| 22 | σ^{70} region 1.1: Evidence that region 1.1 stabilizes the interaction between σ^{70} and core. J. | | | |

23 Bacteriol. **188**: 1279-1285.

| 1 | 19. Ikeda, J. S., A. Janakiraman, D. G. Kehres, M. E. Maguire, and J. M. Slauch. 2005. |
|----|--|
| 2 | Transcriptional regulation of sitABCD of Salmonella enterica serovar Typhimurium by MntR |
| 3 | and Fur. J. Bacteriol. 187: 912-922. |
| 4 | |
| 5 | 20. Ishihama, A. 2000. Functional modulation of Escherichia coli RNA polymerase. Annu. |
| 6 | Rev. Microbiol. 54: 499-518. |
| 7 | |
| 8 | 21. Jishage, M., and A. Ishihama. 1997. Variation in RNA polymerase sigma subunit |
| 9 | composition within different stocks of Escherichia coli W3110. J. Bacteriol. 179: 959-963. |
| 10 | |
| 11 | 22. Karimova G., J. Pidoux, A. Ullmann, and D. Ladant. 1998. A bacterial two-hybrid |
| 12 | system based on a reconstituted signal transduction pathway. Proc. Natl. Acad. Sci. USA. 95: |
| 13 | 5752-5756. |
| 14 | |
| 15 | 23. Karimova, G., C. Robichon, and D. Ladant. 2009. Characterization of YmgF, a 72- |
| 16 | residue inner membrane protein that associates with the Escherichia coli cell division |
| 17 | machinery. J. Bacteriol. 191: 333-346. |
| 18 | |
| 19 | 24. King, T., A. Ishihama, A. Kori, and T. Ferenci. 2004. A regulatory trade-off as a source |
| 20 | of strain variation in the species Escherichia coli. J. Bacteriol. 186: 5614-5620. |
| 21 | |
| 22 | 25. Klauck, E., A. Typas, and R. Hengge. 2007. The σ^{s} subunit of RNA polymerase as a |
| 23 | signal integrator and network master regulator in the general stress response in Escherichia |
| 24 | <i>coli</i> . Sci. Prog. 90 : 103-127. |
| | |

| 1 | 26. Kowarz, L., C. Coynault, V. Robbe-Saule, and F. Norel. 1994. The Salmonella |
|----|---|
| 2 | typhimurium katF (rpoS) gene: cloning, nucleotide sequence, and regulation of spvR and |
| 3 | spvABCD virulence plasmid genes. J. Bacteriol. 176: 6852-60. |
| 4 | |
| 5 | 27. Lelong, C., K. Aguiluz, S. Luche, L. Kuhn, J. Garin, T. Rabilloud, and J. |
| 6 | Geiselmann. 2007. The Crl-RpoS regulon of Escherichia coli. Mol. Cell. Proteomics 6: 648- |
| 7 | 659. |
| 8 | |
| 9 | 28. Lelong, C., M. Rolland, M. Louwagie, J. Garin, and J. Geiselmann. 2007. Mutual |
| 10 | regulation of Crl and Fur in Escherichia coli W3110. Mol. Cell. Proteomics 6: 660-668. |
| 11 | |
| 12 | 29. Malan, T. P., A. Kolb, H. Buc, and W. R. McClure. 1984. Mechanism of CRP-cAMP |
| 13 | activation of <i>lac</i> operon transcription initiation activation of the P1 promoter. J. Mol. Biol. |
| 14 | 180: 881–909. |
| 15 | |
| 16 | 30. Massé, E., and S. Gottesman. 2002. A small RNA regulates the expression of genes |
| 17 | involved in iron metabolism in Escherichia coli. Proc. Natl. Acad. Sci.USA 99: 4620-4625. |
| 18 | |
| 19 | 31. Miller, J. H. 1972. Experiments in Molecular Genetics, Cold Spring Harbour, N. Y. |
| 20 | |
| 21 | 32. Murakami K.S., and S. A. Darst. 2003. Bacterial RNA polymerase: the wholo story. |
| 22 | Current Opin. Struct. Biol. 13: 31-39. |
| | |

| 1 | 33. Notley-McRobb, L., T. King, and T. Ferenci. 2002. rpoS mutations and loss of general |
|----|---|
| 2 | stress resistance in Escherichia coli populations as a consequence of conflict between |
| 3 | competing stress responses. J. Bacteriol. 184: 806-811. |
| 4 | |
| 5 | 34. Nyström T. 2004. Growth versus maintenance: a trade-off dictated by RNA polymerase |
| 6 | availability and sigma factor competition? Mol Microbiol. 54: 855-62. |
| 7 | |
| 8 | 35. Paget , M. S. B., and J. Helmann. 2003. The σ^{70} family of sigma factors. Genome Biol. |
| 9 | 4 : 203 |
| 10 | |
| 11 | 36. Pratt, L. A., and T. J. Silhavy. 1998. Crl stimulates RpoS activity during stationary |
| 12 | phase. Mol. Microbiol. 29: 1225-36. |
| 13 | |
| 14 | 37. Robbe-Saule, V., G. Algorta, I. Rouilhac, and F. Norel. 2003. Characterization of the |
| 15 | RpoS status of clinical isolates of Salmonella enterica. Appl. Environ. Microbiol. 69: 4352- |
| 16 | 4358. |
| 17 | |
| 18 | 38. Robbe-Saule, V., I. Carreira, A. Kolb, and F. Norel. 2008. Effect of growth |
| 19 | temperature on Crl-dependent regulation of σ^{S} activity in Salmonella enterica serovar |
| 20 | Typhimurium. J. Bacteriol. 190: 4453-4459. |
| 21 | |
| 22 | 39. Robbe-Saule V., C. Coynault, and F. Norel 1995. The live oral typhoid vaccine Ty21a |
| 23 | is a <i>rpoS</i> mutant and is susceptible to various environmental stresses. FEMS Microbiol. Lett. |
| 24 | 126 : 171-176. |
| 25 | |

| 1 | 40. Robbe-Saule, V., V., Jaumouille, M. C., Prevost, S., Guadagnini, C. Talhouarne, H. |
|----|---|
| 2 | Mathout, A., Kolb A, and F. Norel. 2006. Crl activates transcription initiation of RpoS- |
| 3 | regulated genes involved in the multicellular behavior of Salmonella enterica serovar |
| 4 | Typhimurium. J Bacteriol. 188 : 3983-94. |
| 5 | |
| 6 | 41. Robbe-Saule, V., M. D., Lopes, A., Kolb, and F. Norel. 2007. Physiological effects of |
| 7 | Crl in <i>Salmonella</i> are modulated by σ^{s} level and promoter specificity. J. Bacteriol. 189 : 2976- |
| 8 | 2987. |
| 9 | |
| 10 | 42. Römling, U. 2005. Characterization of the rdar morphotype, a multicellular behaviour in |
| 11 | Enterobacteriaceae. Cell. Mol. Life Sci. 62:1234-46. |
| 12 | |
| 13 | 43. Sambrook, J., E. F. Fritsch and T. Maniatis. 1989. Molecular cloning: a laboratory |
| 14 | manual (2nd edn.), Cold Spring Harbor, N. Y. |
| 15 | |
| 16 | 44. Schmieger, H. 1972. Phage P22-mutants with increased or decreased transduction |
| 17 | abilities. Mol. Gen. Genet. 119: 75-88. |
| 18 | |
| 19 | 45. Schuster, M., A. C. Hawkins, C. S. Harwood, and E. P. Greenberg. 2004. The |
| 20 | Pseudomonas aeruginosa RpoS regulon and its relationship to quorum sensing. Mol. |
| 21 | Microbiol. 51 : 973-985. |
| 22 | |
| 23 | 46. Silhavy, T. J., M. L. Bernman and L. W. Ernqvist. 1984. Experiments with Gene |
| 24 | Fusions, Cold Spring Harbour, N. Y. |
| 25 | |
| | |

| 1 | 47. Sorenson M. K., and S.A. Darst. 2006. Disulfide cross-linking indicates that FlgM- |
|----|--|
| 2 | bound and free σ^{28} adopt similar conformations. Proc. Natl. Acad. Sci USA 103: 16722- |
| 3 | 16727. |
| 4 | |
| 5 | 48. Sternberg, N. L., and R. Maurer. 1991. Bacteriophage-mediated generalized |
| 6 | transduction in <i>Escherichia coli</i> and <i>Salmonella typhimurium</i> . Methods Enzymol. 204: 18-43. |
| 7 | |
| 8 | 49. Stüdemann, A., M. Noirclerc-Savoye, E. Klauck, G. Becker, D. Schneider, and R. |
| 9 | Hengge. 2003. Sequential recognition of two distinct sites in σ^{s} by the proteolytic targeting |
| 10 | factor RssB and ClpX. EMBO J. 22: 4111-4120. |
| 11 | |
| 12 | |
| 13 | 50. Typas, A. C., Barembruch, A., Possling, and R. Hengge. 2007. Stationary phase |
| 14 | reorganisation of the Escherichia coli transcription machinery by Crl protein, a fine-tuner of |
| 15 | σ^{S} activity and levels. The EMBO J. 26 : 1569-1578. |
| 16 | |
| 17 | 51. Vassylyev D. G., S-I. Sekine, O. Laptenko, J. Lee, M. N. Vassylyeva, S. Borukhov, |
| 18 | and S. Yokoyama. 2002. Crystal structure of a bacterial RNA polymerase holoenzyme at |
| 19 | 2.6A resolution. Nature 417 : 712-719. |
| 20 | |
| 21 | 52. Vijayendran, C., T. Polen, V. F. Wendisch, K. Friehs, K. Niehaus, and E. Flaschel. |
| 22 | 2007. The plasticity of global proteome and genome expression analysed in closely related |
| 23 | W3110 and MG1655 strains of a well-studied model organism, Escherichia coli-K12. J. |
| 24 | Biotechnol. 128: 747-761. |

86:181-184.

| 1 Table 1: Bacterial strains and plasmids used in this s | study |
|--|-------|
|--|-------|

| 3 | Strain or plasmid | Characteristics | Source or reference ^a | |
|----|--------------------------------|---|----------------------------------|--|
| 4 | Escherichia coli | | | |
| 5 | MC1061K | araD139 ∆(ara-leu)-767 ∆(lacIPOZY)X74 | rpsL 26 | |
| 6 | | galU galK rpoS::kan | | |
| 7 | JM109 | recA1 supE44 endA1 hsdR17 gyrA96 relA | 1 <i>thi</i> 43 | |
| 8 | | $\Delta(lac-proAB)$ F'(traD36 proAB ⁺ lacI ⁹ lacZ | ZAM15) | |
| 9 | BTH101 | F- cya-99 araD139 galE15 galK16 rpsL1 | (Str ^R) D. Ladant | |
| 10 | | hsdR2 mcrA1 mcrB1 | | |
| 11 | Salmonella serovar Typhi | | | |
| 12 | T2 | Wild-type | CIS ^b | |
| 13 | Salmonella serovar Typhimurium | | | |
| 14 | C52 | Wild-type | 26 | |
| 15 | C52K | C52 $\Delta rpoS::kan$ | 26 | |
| 16 | ATCC14028 | Wild-type | ATCC ^c | |
| 17 | ATCC <i>rpoS</i> | ATCC14028 ΔrpoS | 41 | |
| 18 | ATCCcrl | ATCC14028 Δ <i>crl</i> ::Cm | 40 | |
| 19 | JS402 | $\Delta fur-41::cat$ | 19 | |
| 20 | ATCCfur | АТСС14028 <i>Дfur</i> -41:: <i>cat</i> | | |
| 21 | ATCCrpoSfur | ATCC14028 ΔrpoS Δfur-41::cat | | |
| 22 | ATCCcrl csgD-lacZ | ATCC14028 $\Delta crl::cm \ csgD-lacZY$ | 40 | |
| 23 | ATCCsdhA-lacZ-k | ATCC14028 sdhA-lacZY-Km | 41 | |

| 1 | ATCCfur sdhA-lacZ-k | ATCC14028 Δfur-41::cat sdhA-lacZY-Km | |
|----|--|--|----|
| 2 | ATCCcrl-lacZ | ATCC14028 crl-lacZY-Km | |
| 3 | ATCCrpoS crl-lacZ | ATCC14028 Δ <i>rpoS</i> ::Cm <i>crl-lacZY</i> -Km | |
| 4 | ATCCfur crl-lacZ | ATCC14028 Δ <i>fur</i> -41:: <i>cat crl-lacZY</i> -Km | |
| 5 | ATCCrpoSfur crl-lacZ | ATCC14028 Δ <i>rpoS</i> Δ <i>fur-</i> 41:: <i>cat</i> crl-lacZY-Km | |
| 6 | 2922K | ATCC14028 ΔSTM2922::Km | 41 |
| 7 | 2922Kcrl | ATCC14028 ΔSTM2922::Km Δ <i>crl</i> ::Cm | 41 |
| 8 | 2922KrpoS | ATCC14028 ΔSTM2922::Km Δ <i>rpoS</i> ::Cm | 41 |
| 9 | 2922KrpoS _{T2} | ATCC14028 ΔSTM2922::Km <i>rpoS</i> _{T2} | |
| 10 | 2922KrpoS _{T2} crl | 2922K $rpoS_{T2} \Delta crl$::Cm | |
| 11 | 2922K katE-lacZ | 2922K Tn5B21-2.4 | 41 |
| 12 | 2922KrpoS katE-lacZ | 2922K <i>rpoS</i> Tn5B21 –2.4 | 41 |
| 13 | 2922Kcrl katE-lacZ | 2922Kcrl Tn5B21 –2.4 | 41 |
| 14 | 2922KrpoS _{T2} katE-lacZ | 2922K <i>rpoS</i> _{T2} Tn5B21-2.4 | |
| 15 | 2922KrpoS _{T2} crl katE-lacZ2922K rpoS _{T2} Δcrl ::Cm Tn5B21-2.4 | | |
| 16 | | | |
| 17 | Plasmids | | |
| 18 | pACYC184 | cloning vector, Cm ^R , Tet ^R | 6 |
| 19 | pACcrl-1 | pACYC184 with the promoterless | 40 |
| 20 | | crl gene cloned into the cat gene | |

| 20 | | crl gene cloned into the cat gene |
|----|--------------------------------|---|
| 21 | | $(crl \text{ is transcribed from the } cat \text{ promoter}), \text{Tet}^{R}$ |
| 22 | pAC <i>crl</i> _{Y22A} | pACcrl-1 with mutation Y22A in Crl |
| 23 | pAC <i>crl</i> _{F35A} | pACcrl-1 with mutation F35A in Crl |
| 24 | pAC <i>crl</i> _{C28A} | pACcrl-1 with mutation C28A in Crl |
| 25 | pACcrl _{C37A} | pACcrl-1 with mutation C37A in Crl |

| 1 | pAC <i>crl</i> _{C41A} | pACcrl-1 with mutation C41A in Crl | |
|----|-------------------------------------|--|-----------|
| 2 | pAC <i>crl</i> _{F53A} | pACcrl-1 with mutation F53A in Crl | |
| 3 | pACcrl _{W56A} | pACcrl-1 with mutation W56A in Crl | |
| 4 | pACcrl _{G74A} | pACcrl-1 with mutation G74A in Crl | |
| 5 | pACcrl _{G80A} | pACcrl-1 with mutation G80A in Crl | |
| 6 | pACcrl _{W82A} | pACcrl-1 with mutation W82A in Crl | |
| 7 | pACcrl _{F103A} | pACcrl-1 with mutation F103A in Crl | |
| 8 | pACcrl _{C28-37} | pACcrl-1 with mutation C28A, C37A in Crl | |
| 9 | pACcrl _{C28-41} | pACcrl-1 with mutation C28A, C41A in Crl | |
| 10 | pACcrl _{C37-41} | pACcrl-1 with mutation C37A, C41A in Crl | |
| 11 | pAC <i>crl</i> _{C28-37-41} | pACcrl-1 with mutation C28A, C37A, C41A in Crl | |
| 12 | pUC4K | source of Km resistance cartridge | Pharmacia |
| 13 | pSTF4 | <i>spvRAB-lacZ</i> fusion in pQF50, Cb ^R | 26 |
| 14 | pAC <i>rpoS</i> _{T2} | pACYC184 with a 6 kb BglII fragment | |
| 15 | | carrying $rpoS_{T2}$, Cm ^R | |
| 16 | pUC <i>rpoS</i> _{T2} | pUC19 with a 3.3 kb SphI-ScaI fragment | |
| 17 | | carrying $rpoS_{T2}$, Cb^R | |
| 18 | pUC <i>rpoS</i> _{T2} K | pUCrpoS _{T2} with STY3047::Km, Cb ^R Km ^R | |
| 19 | pQE60 | vector for expression of His-tagged proteins, Cb ^R | Qiagen |
| 20 | pQE30 | vector for expression of His-tagged proteins, Cb ^R | Qiagen |
| 21 | pQE <i>rpoS</i> _{T2} | pQE60:: $rpoS_{T2}$ expresses a σ^{S}_{T2} -His ₆ protein, Cb ^R | |
| 22 | pQE30rpoS | pQE30:: $rpoS$ expresses a His ₆ - σ^{S} protein, Cb ^R | |
| 23 | pQEcrl | pQE30:: <i>crl</i> expresses a His ₆ -Crl protein, Cb ^R | 40 |
| 24 | pQEcrl _{Y22A} | pQEcrl with mutation Y22A in Crl | |
| 25 | pQEcrl _{F53A} | pQEcrl with mutation F53A in Crl | |

| 1 | pQEcrl _{W56A} | pQEcrl with mutation W56A in Crl | |
|----|--------------------------|--|----|
| 2 | pQEcrl _{W82A} | pQEcrl with mutation W82A in Crl | |
| 3 | pKT25 | BACTH vector designed to express a given 23 | |
| 4 | | polypeptide fused in frame at its N-terminal end with | |
| 5 | | T25 fragment, p15 ori, Km ^R | |
| 6 | pKNT25 | BACTH vector designed to express a given | 23 |
| 7 | | polypeptide fused in frame at its C-terminal end with | |
| 8 | | T25 fragment, p15 ori, Km ^R | |
| 9 | pUT18 | BACTH vector designed to express a given | 23 |
| 10 | | polypeptide fused in frame at its C-terminal end with | |
| 11 | | T18 fragment, ColE1 ori, Ap ^R | |
| 12 | pUTCrl | pUT18 expressing Crl-T18 | |
| 13 | pUTCrl _{Y22A} | pUT18 expressing Crl-T18 with mutation Y22A in Crl | |
| 14 | pUTCrl _{R51A} | pUT18 expressing Crl-T18 with mutation R51A in Crl | |
| 15 | pUTCrl _{E52A} | pUT18 expressing Crl-T18 with mutation E52A in Crl | |
| 16 | pUTCrl _{F53A} | pUT18 expressing Crl-T18 with mutation F53A in Crl | |
| 17 | pUTCrl _{G55A} | pUT18 expressing Crl-T18 with mutation G55A in Crl | |
| 18 | pUTCrl _{W56A} | pUT18 expressing Crl-T18 with mutation W56A in Crl | |
| 19 | pUTCrl _{W57A} | pUT18 expressing Crl-T18 with mutation W57A in Crl | |
| 20 | pUTCrl _{G80A} | pUT18 expressing Crl-T18 with mutation G80A in Crl | |
| 21 | pUTCrl _{W82A} | pUT18 expressing Crl-T18 with mutation W82A in Crl | |
| 22 | pKNTCrl | pKNT25 expressing Crl-T25 | |
| 23 | pKTRpoS | pKT25 expressing T25-RpoS | |
| 24 | $pKTRpoS_{\Delta 12-71}$ | pKTRpoS with in frame deletion of residues 12-71 in RpoS | |
| 25 | $pKTRpoS_{\Delta 1-38}$ | pKTRpoS with deletion of residues 1-38 in RpoS | |

- 1 $pKTRpoS_{\Delta 1-71}$ pKTRpoS with deletion of residues 1-71 in RpoS
- 2 pKTFur pKT25 expressing T25-Fur
- 3 pUTFur pUT18 expressing Fur-T18
- 4
- 5 ^a This study, unless otherwise noted.
- ^b WHO Reference Center for *Salmonella* (Institut Pasteur, Paris)
- 7 ^c American Type Culture Collection

| 2 | | | | |
|----|--------------------------|----------------------------------|-------------------------------|--|
| 3 | Genus ^a | identity with Crl _{STM} | $\left(\% \right)^{b}$ ident | ity with $\operatorname{RpoS}_{\operatorname{STM}}(\%)^{\operatorname{C}}$ |
| 4 | | | | |
| 5 | Salmonella | 96-100 | 99-100 | |
| 6 | Citrobacter | 90 | 100 | |
| 7 | Escherichia | 83-84 | 98-99 | |
| 8 | Shigella | 83-84 | 99 | |
| 9 | Enterobacter | 74-77 | 97 | |
| 10 | Klebsiella | 75-76 | 98 | Enterobacteriaceae |
| 11 | Serratia | 69 | 94 | |
| 12 | Erwinia | 66 | 95 | |
| 13 | Yersinia | 63-66 | 84-97 | |
| 14 | Pectobacterium | 64 | 91 | |
| 15 | Proteus | 49 | 79 | |
| 16 | Providencia | 40-45 | 78 | |
| 17 | Aeromonas | 43-44 | 85-84 | |
| 18 | Vibrio | 38-50 | 72-82 | |
| 19 | Unclassified Vibrionales | 46 | 81 | |
| 20 | Photobacterium | 42-49 | 79-80 | |
| 21 | Moritella | 38 | 76 | |
| 22 | Psychromonas | 35-44 | 77 | |
| | | | | |

1 Table 2: Distribution and sequence conservation of Crl in bacterial sequenced genomes.

^a bacterial sequenced genomes containing a *crl* gene

^c Range of percentage of identity between the aminoacid sequence of the RpoS proteins in
strains belonging to the indicated genus and that of the *S*. Typhimurium RpoS protein
(RpoS_{STM}, 330 residues, X77752, 26).

| 2 | | |
|----|--|--|
| 3 | Co-expressed proteins | β -galactosidase activity ^a |
| 4 | | |
| 5 | T18 + T25 | 47 ± 6 |
| 6 | Crl-T18 + T25 | 47 ± 6 |
| 7 | $T18 + T25 - \sigma^{s}$ | 43 ± 3 |
| 8 | Crl-T18 + T25- σ^{s} | 263 ± 38 |
| 9 | Crl-T18 + T25- $\sigma^{s}_{\Delta 1-38}$ | 455 ± 33 |
| 10 | Crl-T18 + T25- $\sigma^{s}_{\Delta 1-71}$ | 1830 ± 186 |
| 11 | Crl-T18 + T25- $\sigma^{s}_{\Delta 12-71}$ | 2352 ± 146 |
| 12 | Crl_{Y22A} -T18 + T25- σ^{s} | 44 ± 5 |
| 13 | Crl_{R51A} -T18 + T25- σ^{s} | 59 ± 3 |
| 14 | Crl_{E52A} -T18 + T25- σ^{S} | 102 ± 12 |
| 15 | Crl_{F53A} -T18 + T25- σ^{S} | 48 ± 5 |
| 16 | Crl_{G55A} -T18 + T25- σ^{S} | 138 ± 19 |
| 17 | Crl_{W56A} -T18 + T25- σ^{S} | 44 ± 4 |
| 18 | Crl_{W57A} -T18 + T25- σ^{S} | 74 ± 4 |
| 19 | Crl_{G80A} -T18 + T25- σ^{S} | 205 ± 27 |
| 20 | Crl_{W82A} -T18 + T25- σ^{S} | 55 ± 6 |
| 21 | Crl-T18 + T25-Fur | 43 ± 1 |
| 22 | Fur-T18 + Crl-T25 | 58 ± 6 |
| 23 | Fur-T18 + T25-Fur | 243 ± 32 |
| 24 | Fur-T18 +T25 | 55 ± 3 |

1 Table 3: BACTH analysis of Crl interactions with σ^{S} and Fur.

| 1 | T18 + T25-Fur | 42 ± 2 |
|---|---------------|------------|
| 2 | T18 + Crl-T25 | 52 ± 5 |
| 3 | | |

^a The efficiencies of functional complementation between the indicated proteins were quantified by measuring β -galactosidase activities in *E. coli* BTH101 cells harboring the corresponding plasmids as described in Material and Methods. β -galactosidase activity was measured according to the method of Miller (31).

| 1 | |
|---|--|
| н | |
| T | |
| | |

3 FIGURE 1. Alignment of the proteins from the Crl family.

Alignment of the Crl protein sequences was performed with ClustalW. Residues that are 4 5 identical in all of the Crl proteins are shown in black boxes and position of these conserved 6 residues in the Crl sequence from S. enterica serovar Typhimurium (Crl_{STM}, 133 residues, 7 NP-459316, line 16) is shown below the alignment. The accession numbers for the Crl sequences are as follows: Salmonella: ZP-02346845 (14), ZP-02663633 (15)NP-459316 (16), 8 YP-215307 (17), YP-001571691 (18); Citrobacter: YP-001454497 (19); Escherichia: NP-9 10 285957 (22), NP-414775 (23), YP-001742400 (25), YP-002383844 (26), YP-002327819 (27), 11 NP-752324 (28), YP-539315 (29), ZP-02902139 (30); Shigella: YP-309300 (20), YP-402173 (21), NP-706240 (24); Enterobacter: YP-001175503 (31), ZP-03281271 (32), YP-001439177 12 (35); Klebsiella: YP-001333935 (33), YP-002240279 (34); Serratia: YP-001477200 (13); 13 14 Erwinia: YP-001908518 (36); Yersinia: YP-001007379 (6), ZP-00829910 (7), ZP-00821198 15 (8), ZP-00826726 (9), ZP-00833806 (10), YP-001402111 (11), NP-668295 (12); 16 Pectobacterium: YP-051554 (37); Proteus: YP-002150140 (38); Providencia: ZP-03313695 17 (39), ZP-03317513 (40), ZP-02958311 (41); Aeromonas: YP-857911 (42), YP-001140774 18 (43); Vibrio: ZP-01235672 (1), NP-231906 (44), ZP-01957597 (45), ZP-01980856 (46), NP-19 759329 (47), NP-933650 (48), ZP-01065753 (49), YP-002418001 (50), ZP-00993088 (51), ZP-01870715 (53), NP-797054 (54), ZP-01260295 (55), YP-001444378 (56), ZP-01987497 20 21 (57), ZP-02197346 (58); Unclassified Vibrionales: ZP-01813308 (52); Photobacterium: ZP-22 01159718 (2), YP-129053 (3), ZP-01218949 (4); *Moritella*: ZP-01898329 (5); 23 Psychromonas: YP-944255 (59), ZP-01216084 (60).

- 24
- 25
- 26

2 FIGURE 2. Characterization of the *rpoS* mutant allele *rpoS*_{T2}.

3 A) Relevant portion of the *rpoS* sequence in the wild-type *rpoS* allele and the $rpoS_{T2}$ mutant 4 allele. The deletion of eight nucleotides in $rpoS_{T2}$, compared to rpoS wild-type, is shown. This 5 mutation results in the appearance of a premature stop codon in rpoS_{T2}. B) Determination of the N-terminal sequence of σ^{s}_{T2} . Production of σ^{s}_{T2} likely results from translational 6 7 reinitiation in $rpoS_{T2}$ at the GTG codon that is preceded by a putative ribosome binding site 8 (Shine-Dalgarno sequence SD). C) Resistance to hydrogen peroxide of the S. Typhimurium 9 strains indicated. Cells were grown to stationary phase in LB at 37°C, washed, resuspended in 10 PBS to an OD₆₀₀ of 0.1 and H₂O₂ 15 mM was added. A representative experiment is shown. 11 Similar results were obtained in repeat experiments. D) Expression of a katE-lacZ gene fusion 12 in Salmonella carrying the wild-type rpoS and mutant rpoS_{T2} alleles. 1: 2922KkatE-lacZ, 2: 13 2922KrpoS katE-lacZ, 3: 2922Kcrl katE-lacZ, 4: 2922KrpoS_{T2} katE-lacZ, 5: 2922KrpoS_{T2}crl 14 katE-lacZ. β-galactosidase activity was measured in overnight LB cultures at 37°C according to the method of Miller (31). 15

16

17 FIGURE 3. Expression of σ^{S} wild-type and mutant proteins.

A) Detection of σ^{S}_{T2} produced from the *rpoS*_{T2} allele. Overnight LB cultures at 37°C were 18 analysed by western blotting with anti- σ^{s} antibodies. 5 µg of total protein was loaded in each 19 20 slot. 1: S. Typhimurium wild-type strain C52, 2: S. Typhimurium $\Delta rpoS$ mutant C52K, 3: S. Typhi strain T2, 4: *E.coli* MC1061K harboring plasmid pAC*rpoS*_{T2}. B) Expression of σ^{S} and 21 σ^{S}_{T2} as a function of bacterial cell growth. *Salmonella* strains 2922K and 2922K*rpoS*_{T2} were 22 23 grown in LB at 37° C. Exponential-phase cultures (optical density at 600 nm = 0.5) were 24 diluted into LB pre-warmed at 37°C to prolong the exponential phase. Aliquots were removed at various time intervals and analysed by western blotting with anti- σ^{s} antibodies. 10 µg of 25

1 total protein was loaded in each slot. The growth phase was determined by measuring culture turbidity at an optical density of 600 nm (OD600). C1: 2922KrpoS (OD600=4), C2: 2922K 2 (OD600=1.4). C) Expression of wild-type and truncated T25- σ^{s} hybrid proteins. The *E. coli* 3 4 cya strain BTH101 harboring pKT25 and its derivatives were grown to stationary phase 5 (OD600= 3.5) in LB in the presence of 2 mM cAMP and 0.5 mM IPTG to fully induce the *lac* 6 promoter on pKT25. 10 µg of total protein was loaded in each slot and analysed by western blotting with anti- σ^{s} antibodies. 1: no plasmid, 2: pKTRpoS, 3: pKT25, 4: pKTRpoS_{Δ 12-71}, 5: 7 pKTRpoS $_{\Delta 1-38}$, 6: pKTRpoS $_{\Delta 1-71}$. 8

9

FIGURE 4 Site-directed mutagenesis of the *Salmonella crl* gene and characterization of the *crl* mutant alleles.

Site-directed mutagenesis of pACcrl-1 was performed, yielding eleven derivatives expressing 12 13 Crl proteins in which residues Y22, F35, C28, C37, C41, F53, W56, G74, G80, W82, and 14 F103 are substituted by alanine. The ability of the mutated crl genes to complement the Salmonella Δcrl mutant was assessed using three different tests: the H₂O₂ resistance level at 15 28°C, the expression level of a *csgD-lacZ* fusion, and development of the rdar morphotype. 16 A) Summary of the results obtained: (+) complementation level similar to that obtained with 17 18 pACcrl-1 expressing the wild-type Crl protein, (-) complementation level similar to that 19 obtained with the negative control pACYC184, and (±) partial complementation level. 20 Complementation experiments for the rdar morphotype are shown in Figure 5. B) 21 Complementation of the Δcrl mutant for resistance to H₂O₂. ATCC*crl* derivatives containing 22 the indicated plasmids were grown to stationary phase in LB at 28°C, washed, resuspended in 23 PBS to an OD₆₀₀ of 1 and H₂O₂ 15 mM was added. A representative experiment is shown. Similar results were obtained in repeat experiments. C) Complementation of the Δcrl mutant 24 for expression of a csgD-lacZ fusion. ATCCcrl csgD-lacZ derivatives containing the 25

1 indicated plasmids were grown in LB0 (LB without NaCl) at 28°C. Exponential-phase 2 cultures (optical density at 600 nm = 0.5) of *Salmonella* were diluted into LB0 pre-warmed at 3 28°C to prolong the exponential phase. Aliquots were removed at various time intervals and 4 β -galactosidase activity measured (lines) according to the method of Miller (31). The growth 5 phase was determined by measuring the culture turbidity at an optical density of 600 nm 6 (dashed line, the growth curve was similar for all the strains). The measurements were 7 repeated twice, and a representative experiment is shown.

8

9 FIGURE 5. Rdar morphotypes of *Salmonella* ATCC*crl* harbouring different plasmids.

10 The Δcrl mutant harbouring the vector pACYC184, the wild-type crl gene on pACcrl-1 and 11 mutated pACcrl-1 derivatives (Table 1) were grown five days on CR plates at 28°C. 12 Plasmids are 1 : pACYC184, 2 : pACcrl-1, 3 : pAC crl_{Y22A} , 4 : pAC crl_{F35A} , 5 : pAC crl_{C37A} , 6 : 13 pAC crl_{C41A} , 7 : pAC crl_{F53A} , 8 : pAC crl_{W56A} , 9 : pAC crl_{G74A} , 10 : pAC crl_{G80A} , 11 : 14 pAC crl_{W82A} , 12 : pAC crl_{F103A} , 13 : pAC crl_{C28A} , 14 : pAC $crl_{C28-37-41}$. Complementation of the 15 crl mutation was observed with pACcrl-1 and its derivatives containing mutations F35A, 16 C28A, C37A, C41A, G74A, G80A, F103A.

17 **FIGURE 6. Expression of Crl wild-type and mutant proteins.**

A) Detection of Crl produced from pAC*crl*-1 and its mutant derivatives in *Salmonella* ATCC*crl*. ATCC*crl* strains, harboring the vector pACYC184 and the pAC*crl*-1 derivatives as indicated below, were grown to stationary phase (OD600=4) in LB at 37°C and analysed by western blotting with anti-Crl antibodies. 1 µg of total protein was loaded in slots 1 to17. 1: pAC*crl*-1, 2: pACYC184, 3: pAC*crl*_{Y22A}, 4: pAC*crl*_{F35A}, 5: pAC*crl*_{C37A}, 6: pAC*crl*_{C41A}, 7: pAC*crl*_{F53A}, 8: pAC*crl*_{W56A}, 9: pAC*crl*_{G74A}, 10: pAC*crl*_{G80A}, 11: pAC*crl*_{W82A}, 12: pAC*crl*_{F103A}, 13: pAC*crl*_{C28A}, 14: pAC*crl*_{C28-37}, 15: pAC*crl*_{C28-41}, 16: pAC*crl*_{C37-41}, 17: pAC*crl*_{C28-37-41}. 5 µg

1 of total protein was loaded in slots 18 to 20. 19: pACcrl_{C37-41}, 20: pACcrl_{C28-37-41}, 19: a 2 stationary phase culture of ATCC14028 in LB at 37°C was used as control. B) Expression of 3 wild-type and altered Crl-T18 hybrid proteins. The E. coli cya strain BTH101 harboring pUT18 and its derivatives were grown to stationary phase (OD600= 3) in LB in the presence 4 5 of 2 mM cAMP and 0.5 mM IPTG to fully induce the lac promoter on pUT18. 5 µg of total 6 protein was loaded in each slot and analysed by western blotting with anti-Crl antibodies. 1: no plasmid, 2: pUT18, 3: pUTCrl, 4: pUTCrl_{Y22A}, 5: pUTCrl_{F53A}, 6: pUTCrl_{W56A}, 7: 7 pUTCrl_{G80A}, 8: pUTCrl_{W82A}, 9: pUTCrl_{R51A}, 10: pUTCrl_{E52A}, 11: pUTCrl_{G55A}, 12: pUTCrl_{W57A}. 8 9 C) Detection of Crl in Salmonella ATCC14028 (1, 4, 7, 10, 15) and its mutant derivatives 10 ATCCfur (2, 5, 8, 13), ATCCcrl (3, 6, 9, 11), ATCCrpoS (12), and ATCCrpoSfur (14). 11 Cultures were grown in different conditions, as indicated, and analysed by western blotting 12 with anti-Crl antibodies. 1 µg of total protein was loaded in each slot. LB 37°C: overnight LB 13 cultures at 37°C. LB 30°C: overnight LB cultures at 30°C. LB agar 30°C: colonies grown on 14 LB agar at 30°C as described (28).

15

16 FIGURE 7. Effects of the Crl variants on $E\sigma^{S}$ promoter binding and abortive 17 transcription.

A) Band shift analysis of $E\sigma^{S}$ binding to the *katN* labelled fragment in the absence of Crl 18 19 (lanes 2 and 8) and in the presence of wild-type Crl (lane 3) or its variants Y22A (lane 4), 20 F53A (lane 5), W82A (lane 6) and W56A (lane 7), lane 1: no protein. A typical 21 autoradiogram is shown. The bands corresponding to free and bound DNAs are indicated by 22 arrows and the percentage of bound DNA indicated below each lane is the average of two experiments. B) Abortive initiation assays. $E\sigma^{s}$ was preincubated in the absence of Crl or in 23 the presence of wild-type Crl or its variants Y22A, F53A, W82A and W56A. After addition 24 of a mixture containing the *lac*UV5 fragment, the ApA dinucleotide and $[\alpha^{-32}P]$ -UTP, the 25

- 1 incorporation of labeled $[\alpha^{-32}P]$ UMP into abortive transcripts was monitored as a function of
- 2 time. The measurements were repeated twice, and a representative experiment is shown.
- 3

4 FIGURE 8. Expression of *crl-lacZ* and *sdhA-lacZ* gene fusions in *Salmonella*.

Expression of the gene fusions in the Salmonella strains indicated was determined in cultures 5 6 grown in different conditions. LB 37°C: overnight LB cultures at 37°C. LB 30°C: overnight LB cultures at 30°C. LB agar 30°C: colonies grown on LB agar at 30°C as described (28). 7 M9: Minimal M9 medium with 20 mM glucose. M9 + FeCl₂: M9 with 100 µM ferrous 8 9 chloride. M9 + DP: M9 with the iron chelator 2,2'-Dipyridyl (100 μ M). β -galactosidase activity was measured according to the method of Miller (31). A) Expression of the crl-lacZ 10 11 gene fusion in ATCC crl-lacZ strains harboring pACYC184 and pACcrl-1 that expresses the 12 crl gene under the control of the cat promoter of pACYC184 (40). B) Expression of the crllacZ gene fusion in ATCC crl-lacZ and ATCCfur crl-lacZ. C) Expression of the sdhA-lacZ 13 14 gene fusion in ATCC sdhA-lacZ-k and ATCCfur sdhA-lacZ-k and expression of the crl-lacZ 15 gene fusion in ATCC crl-lacZ and ATCCfur crl-lacZ. D) Expression of the crl-lacZ gene fusion in ATCC crl-lacZ, ATCCrpoS crl-lacZ, ATCCfur crl-lacZ and ATCCrpoSfur crl-lacZ. 16 17

| | FPPHGRLM1 | KLTALGI | YLREKKSKEGQF | FEDCLAS | SCVNADKE | EQREFWGWWLILTATDI | NGYEYCYDF <mark>G</mark> RFDI | LNGEWI | NSGKLPAKHKDAVMKTLSDFHGKLTPFIVDECQLTLSSSSELTQTPLTELA |
|--|--|---|---|--|--|---|--|---|---|
| 2 | MDGFAMTSTTLFPPHGRLMT | KLTAL GI | YLREKKSKEGQF | FFDCLAS | SCVNADKE P | EQREFWGWWLILTATDI | NGYEYCYDF <mark>G</mark> RFDI | LNGEWI | NNGKLPAKHKDAVMKTLSDFHGKLTPFIVDDCQLTLSSSSELTQTPLTELA |
| 3 | HTTDTIFPPHGRLMI | KLTALGI | YLRKQQSKEGEF: | FFDCLAS | SCISANKS | EEREFWGWWLVLTATD: | IGFEYRYDF <mark>G</mark> RYD | AKGNW: | IKGTLPAKHTEAVLKTLDDFYVKLTKFVKEDCQLDLQASAELEEAVLGSA- |
| 4 | HTTDTIFPPHGRLMT | KLTALGI | YLRKQQSKEGEF: | FFDCLAS | SCISANKS | EEREFWGWWLVLTATD: | IGFEYRYDF <mark>G</mark> RYD | AKGNW: | IKGSLPAKHAEAVLKTLDDFYVKLTKFVKEDCQLDIQASAELEEAVLGSA- |
| 5 | F AVIAT DPNLIYFRKRIF | ILNALG | YLREHNCQPTSF | YFDCFSV | CIDANIE: | DEREFYGWWLEMNLVDI | OTFEYQYQF <mark>G</mark> TYNI | KAGEW: | IITPIPKALQHDVTKSLNVFYEKLSVCLTEQLSFNLKPSSILAKTLILSAA |
| 6 | MTSTSAHPKSKLMK | RFAALGE | YLREGQCQNDRF | FFDCLAV | CONVELAT | EKREFWGWWIELEPAAI | OHFTYVYQLGLFNI | KDGDWI | KAEKIKDPEVQEKLETTLRGFHKRLAEMLTSIEMRLEPAQDFSEQPVKLSA |
| 7 | MTSTSAHPKSKLMK | RFAALGE | YLREGQCQNDRF | FFDCLAV | CONVELAT | EKREFWGWWIELEPAAI | OHFTYVYQLGLFNI | KDGNWI | KAEKIKDPEVQEKLETTLRGFHKRLADMLTSIEMRLEPAQDFSDQPVKLSA |
| 8 | MR | RFAALGI | YLREGQCQNDRF: | FFDCLAV | CVNVKLAP | EKREFWGWWIELEPAAI | OHFTYVYQFGLFNI | KEGSWI | KAENIKDPEVQEKLETTLRGFHKRLAEMLTSIEMRLVPAQDFSEQPVKLSA |
| 9 | MR | RFAALGI | YLREGQCQNDRF: | FFDCLAV | CVNVKLAP | EKREFWGWWIELEPAAI | OHFTYVYQFGLFNI | KEGSWI | KAENIKDPEVQEKLETTLRGFHKRLAEMLTSIEMKLVPAQDFSEQPVKLSA |
| 10 | MTSTSAHPKSKLMK | RFAALGI | YLREGQCQSDHF. | FEDCLAV | CVNVKLAI | EKREFWGWWIELVPAEI | DHFTYVYQFGLFNI | KEGNWI | KAEKIKDPEIQEKLETTLRGFHKRLAEMLTSIEMRLEPAQNFSEQPIKLSA |
| П | MTLTSAHPKSKLMK | RFAALGI | YLREGQCQNDHF. | FEDCLAY | CINVKLA | EKREFWGWWIEQEPSAC | GRFTYVYQLGLFNI | KEGNWI | NAEKISDPEVQDKLESTLRSFHLRLEEMLASIDMKLEPAADFNDQPVKLSA |
| 12 | MTLTSAHPKSKLMK | RFAALGI | MLREGQCQNDHF. | FFDCLA | CINVKLA | EKREFWGWWIELEPSAC | 3RFTY VYQLGLFNI | KEGNWI | NAEKISDPEVQDKLESTLRSFHLRLEEMLASIDMKLEPAADFNDQPVKLSA |
| 13 | MTLPSGHPKSRLMK | RFASLCI | YLREGQCENDRF: | FEDCLAV | VCVNVKPAL | ELREFWGWWIELHAED | FRFTYSYQFGLFD | KKGQW | TPEKIKDTEVKTKLETTLRDFHRRLGELLATMELTLEPAADFKEKLIKLSA |
| 14 | MTLPSGHPKSRLIK | KFTALGI | MIREGQCEDNRF. | FEDCLAV | VCVNVKPAL | EKREFWGWWMELEAQEI | KRFTYRYQFGLFDI | KEGNW | TAVPINETEVVERLEYTLREFHEKLRDLLISMELALEPSDDFNDEPVKLSA |
| 15 | MTLPSGHPKSRLIK | KFTALGI | MIREGQCEDNRF. | FFDCLA | CVNVKPA | EKREFWGWWMELEAQEI | KRFTYRYQFGLFDI | KEGNW: | TAVSINETEVVERLEYTLREFHEKLRDLLISMELALEPSDDFNDEPVKLSA |
| 10 | MTLPSGHPKSRLIK | KFTALGI | MIREGQCEDNRF. | FFDCLA | CVNVKPA | EKREFWGWWMELEAQEI | KRFTYRYQFGLFDI | KEGNW: | TVVPINETEVVERLEYTLREFHEKLRDLLISMELALEPSDDFNDEPVKLSA |
| 17 | MTLPSGHPKSRLIK | KFTALGI | MIREGQCEDNRF. | FFDCLA | CVNVKPA | EKREFWGWWMELEAQEI | KRFTYRYQFGLFDI | KEGNW: | TAVPINETEVVERLEYTLREFHEKLRDLLISMELTLEPSDDFNDEPVKLSA |
| 18 | MTLPSGHPKSRLIK | KFTALGI | MIREGQCEDNRF. | FFDCLA | CVNVKPA | EKREFWGWWMELEAQEI | KSFTYRYQFGLFDI | KEGNW: | TAVPIKETEVVERLEYTLREFHGKLRDLLISMGLALEPSDDFNDEPVKLSA |
| 19 | MTLPSGHPKSRLIK | KFTALGI | MIREGQCEDNRF. | FFDCLA | CVNVKPA | EKREFWGWWMELEAQEI | KRFTYSYQFGLFDI | KEGNW: | TAVPIKETEVIERLEYTLREFHEKLRALLTSLNLALEPADNFK-EPVKLTA |
| 20 | MTLPSGHPKSRLIK | KFTALGE | MIREGRCEDNRF. | FFDCLA | CVNVKPAL | EVREFWGWWMELEAQE | SRFTISIQFGLFD | KAGDWI | KSVPVKDTEVVERLEHTLREFHEKLRELLTTLNLKLEPADDFRDEPVKLTA |
| 21 | MTLPSGIPKSRLIK | KFTALGE | MIREGRCEDNRF. | FFDCLA | V C V N V K P A P | EVREFWGWWMELEAQES | SRFTISIQFGLFD | KAGDWI | KSVPVKDTEVVERLEHTLREFHEKLRELLTTLNLKLEPADDFRDEPVKLTA |
| 22 | MTLPSGHPKSRLIK | KFTVLGF | MIREGRCEDNRF. | FFDCLA | VCVNVKPAP | EVREFWGWWMELEAQES | SRFTISIQFGLFD | KAGDWI | KSVPVKDTEVVERLEHTLREFHEKLRELLTTLNLKLEPADDFRDEPVKLTA |
| 23 | MTLPSGHPKSRLIK | KFTALGE | MIREGRCKDNRF. | FFDCLA | VCVNVKPAP | EVREFWGWWMELEAQES | SRFTISIQFGLFD | KAGDWI | KSVPVKDTEVVERLEHTLREFHEKLRELLTTLNLKLEPADDFRDEPVKLTA |
| 24 | MTLPSGHPKSRLIK | KFTALGE | MIREGRCEDNRF. | FFDCLA | V C V N V K P A P | EVREFWGWWMELEAQES | SRFTISIQFGLFD | KAGDWI | KSVPVKDTEVVERLEHTLREFHEKLRELLTTLNLKLEPANDFRDEPVKLTA |
| 25 | MTLPSGHPKSRLIK | KFTALGE | MIREGRCEDNRF. | FFDCLA | V C V N V K P A P | EVREFWGWWMELEAQES | SRFTISIQFGLFD | KAGDWI | KNVPVKDAEVVERLEHTLREFHEKLRELLTTLNLKLEPADDFRDEPVKLTA |
| 20 | MTLPSGHPKSRLIK | KFTALGE | MIREGRCEDNRF. | FFDCLA | VCVNVKPAP | EVREFWGWWMELEAQES | SRFTISIQFGLFD | KAGDWI | KSVPVKDAEVVERLEHTLREFHEKLRELLSTLNLKLEPADDFRDEPVKLTA |
| 2/ | MTLPSGHPKSKLIK | KFTALGE | MIREGRCEDNRF. | FFDCLA | V C V N V K P A P | EVREFWGWWMELEAQES | SRFTISIQFGLFDI | KAGDW: | TSVPVKDTEVVERLEHTLREFHEKLRELLATLNLKLEPADDFRDEPVKLTA |
| 20 | | KF TALGE | MINEGRCEDNEF. | FFDCLA | V C V N V K P A P | EVREFWGWWMELEAOES | | KAGDW. | ISVKIKDAEVVERUEHIDEEHEKIDELIAHINIKUEPADDERUEVKUIA |
| 29 | | KF TALGE | MINEGRCEDNEF. | FFDCLA | V C V N V K P A P | EVREFWGWWMELEAOES | | KAGDW. | TO VOINDARUTEDI RUEI DERUEVI DELLOTI CIVI EDADDEDEDUVI TA |
| 30 | MTT DCCUDVCDT TV | VEMALGE | NTREGACEDNAF. | FEDCLAY | ZUNNUODA | EVREFWGWWMELEAQES | V P P T Y T Y U P OL P NI | KAGDW. | INVQVNDAEVIERDEIIDREFNEREREDGIESEREPADDFRDEFVERIA |
| 22 | MTI DCCUDVCDI TV | V PMAL OF | NTREEQCEENEF. | FFDCLA | ZUNNUEDA | ESREFWGWWMEMEAEEI | CRETYCYOFCI PNI | V FOND | QAITIKDKEVVEKDEQIDKOFHEKAKADDEIFEDKDEFADDFINEKDKUSA |
| 22 | MTT DCCUT VCVT TV | VETALO | NTREEOCUDNEE | FFDCLA | U CUNUK PA | EXPERMENTEMENT | CRETYSYOF CLEN | | OAUDTDDDEUCDDI EVET DEEUEDAAAI ADDAIDDIKDEFADDFSSERVKUIR |
| 34 | MTLPSCHLKSKLTK | KETALCE | VIDEEOCHDNRE | FEDCLAN | VOUNVEDAD | EKREEWGWWMEMEAOE | SPETYSVOEGLENI | KEGGW | QAVDIDDI HVODKHIKTIREFHERAAAIIIIIKKKKIKERADDES- FOVRI.RA |
| 35 | MASOKETAMTLPSCHPKSPMTK | KETALCE | VIDEEOCEGNEE' | FEDCLAS | VOUNVEDAD | EKREEWGWWMELEAHEI | DRETYTWOEGLEDI | KECDW | TAVDI KDDEVKEKI SOTI POEHERI PALI ASMDI KI EPVTEDI KI SA |
| 36 | MADQKHIAMIDI DGHI KOKMIN | REHST.CE | VIDESKCEGDEE | FEDCLAN | VCVNVKPO | EL REEWGWWMELEAHTI | HETYEVHECLEDI | KECNWI | K DTA VK GKDNNA PLEETI, PNEHVPLKA I, LA FLEI, FLEI, TOATDEADODVKI, TA |
| 30 | | KI HOHOI | TREDREAGDER. | I DCHA | | | | THE MAN | |
| 37 | | NETALC | VIDEAOCEDTAE | FILLICTAT | VNTKPA | EEEEWGWWINLEDTG | CARSYEYHY(HIFD) | KOCNWI | REEKIK(#KAVIK()VENAK()AFHVRIEKI, RSIEPA("TLVARP |
| 37 38 | MMLPSGHPKSRLMK | NFTALGE | YIREAQCEDTAF: YIREOOCOESOF | FEDCLAY | CVNIKPAP VCVNKKVTP | EEREFWGWWLNLEDTGI EKREEWGWWMELERNGI | KAFSYEYHYGLFDI | KQGNWI | REEKIKGKAVIEQVENAKQAFHVRLEKLLRSLEPACTLVARP |
| 37 38 39 | RMMLPSGHPKSRLMK | NFTALGE RFAQIGE KFAAIGE | YIREAQCEDTAF YIREQQCQESQF YLREKOSLKDCY | FFDCLAV FFDCLAV FFDSLVV | ZVNIKPAF ZVNKKVTF ZVNANIAF | EEREFWGWWLNLEDTGI EKREFWGWWMELERNGI EKREFWGWWLELVSTHI | KAFSYEYHYGLFD EQLIYYYQVGLFD EGFEFAYHLGMYDI | KQGNWJ KNGDW NOGNW | REEKIKGKAVIEQVENAKQAFHVRIEKLIRSLEPACTLVARP VNQVISKKDVIESIHETLIRFHDFLQAAVSELEMTLVPDEKMSNFPLPLKP DAKTLKNSETTKAVEKNLVSFHKSLSSORISELETITVPSPLMTELKLELSA |
| 37 38 39 40 | MMLPSGHPKSRLMK MTSSLHPTRGKLLK MLKRDLPKKEILNMALPTDYSNGKYLK MLKRDLPKKEILNMALPTDYTNGKFLK | NFTALGE RFAQIGE KFAAIGE KFAAMGE | YIREAQCEDTAF YIREQQCQESQF YLREKQSLKDCY YLREKQSVKNCY | FFDCLAN FFDCLAN FFDSLVN FFDSLVN | ZVNIKPAF CVNKKVTF CVNANIAF CVNANVYF | EEREFWGWWLNLEDTGI EKREFWGWWMELERNGI EKREFWGWWLELVSTHI EKREFWGWWLELASTDI | KAFSYEYHYGLFD 3QLIYYYQVGLFD 3GFEFAYHLGMYD 1GFEFAYNLGIYT(| KQGNWI KNGDW NQGNW GOGKW | KEEKIKGKAVIBQUENAKQAFHVKEEKIKSEEPACTIVAKP VNQVISKKDVIESIHETIIRFHPIQAAVSELEMTIVPDEKKSNFLPLKP QAKTLKNSETTKAVEKNLVSFHKSLSQRISELEITIYPSPLMTELKLEISA OTKTLKNGSTARAVEKNLHAFHORISOO.SALEKSIPSPLMTELKLEISA |
| 37 38 39 40 41 | | NFTALGE RFAQIGE KFAAIGE KFAAMGE KFSALGE | YIREAQCEDTAF YIREQQCQESQF YLREKQSLKDCY YLREQQSVKNCY YLREKOCLDGCY | FFDCLAV FFDCLAV FFDSLVV FFDSLVV FFDSLTV | /CVNIKPAF /CVNKKVTF /CVNANIAF /CVNANVVF /CVNVNSAF | EEREFWGWWLNLEDTGI EKREFWGWWMELERNGI EKREFWGWWLELVSTHI EKREFWGWWLELASTDI ERREFWGWWMLLSPNEI | KAFSYEYHYGLFD EQLIYYYQVGLFD EGFEFAYHLGMYD NGFEFAYNLGIYT OGFEYSYHLGLYN | KQGNWI KNGDW NQGNW GQGKW VHGTW | KEEK LKGRAVIEQVENARQAFH VKLEKLIKSLEPACTIVARP VNQVISKADVIESIHETIIR PHPIQAAVSELEMTIVP DERMSNF PIPLKP QAKTLKNSETTKAVEKNLVSFHKSLSQRLSELELTIVP SPLMTELKLELSA QTKTLKDGSTAEAVEKNLHAFHQRLSQQLSALELSLFP SSHMTELKLELSA DNRPLKNARTLDITENNISSKOLSSKOLSSKELSLSP SSAMAEFKLELSA |
| 37 38 39 40 41 42 | MMLPSGHPKSRLM | NFTALGE RFAQIGE KFAAIGE KFAAMGE KFSALGE OFAAIGE | YIREAQCEDTAF YIREQQCQESQF YLREKQSLKDCY YLREQQSVKNCY YLREKQCLDGCY YLREDLCEEGRY | FFDCLA\ FFDCLA\ FFDSLV\ FFDSLV\ FFDSLT\ RFDCLS\ | CVNIKPAF CVNKKVTF CVNANIAF CVNANVF CVNANVF CVNVNSAF | EEREFWGWWLNLEDTGI EKREFWGWWMELERNGJ EKREFWGWWLELVSTHI EKREFWGWWLELASTDI ERREFWGWWLLSPNEI DKREFWGWWLLVLEODD | KAFSYEYHYGLFD] 3QLIYYYQVGLFD] 3GFEFAYHLGMYD] NGFEFAYNLGIYT COGFEYSYHLGLYN HSFSYHYOVGFYNJ | KQGNWI KNGDW NQGNW GQGKW VHGTW AEGEW | KEEKIKGRAVIEQVENARQAFHVRLEKLIKSLEPACTIVARP VNQVISKRVVIESIHETLIRFHDFLQAAVSELEMTIVPDEKMSNFPLPLKP QAKTLKNSETTKAVEKNLVSFHKSLSQRLSELELTLYPSPLMTELKLELSA QYKTLKDGSTAEAVEKNLHAFHQRISQQLSALELSLFPSPLMTELKLELSA QKPLKDNATLDIIEQNLISFHQSLSKQLSELELSLFPSAKMAEFKLELSA LMKPLP-KHOAEVERTNHFYKLISDKLHOLGCOLAPACOVTNELTVTAA |
| 37 38 39 40 41 42 43 | MLPSGHPKSRLM MISSLHPTRGKLK MLKRDLPKKEILNMALPTDYSNGKVLK MLKRDLPKKEILNMALPTDYSNGKVLK MLKRDLPKKEILNMALPTDYSNGKVL MLPTDYSNGKVL MSERVTYKRLLF | NFTALGE RFAQIGE KFAAIGE KFAAMGE KFSALGE QFAAIGE OFAAIGE | YIREAQCEDTAF YIREQQCQESQF YLREKQSLKDCY YLREQQSVKNCY YLREKQCLDGCY YLREDLCEEGRY YLREDLCEEGRY | FFDCLA\ FFDCLA\ FFDSLV\ FFDSLV\ FFDSLT\ RFDCLS\ RFDCLS\ | VCVNIKPAF VCVNKKVTF VCVNANIAF VCVNANVVF VCVNVNSAF VCVSAKPAF VCVSAKPAF | EEREFWGWWLNLEDTGJ EKREFWGWWNELEVSTH EKREFWGWWLELASTDI ERREFWGWWLLEASTDI DKREFWGWWLVLEQQDI DKREFWGWWLVLEQODI | XAFSYEYHYGLFD) 3QLIYYYQVGLFD) 3GFEFAYHLGMYD] NGFEFAYNLGIYT 0GFEYSYHLGLYN 1SFSYHYQVGFYN 2SFSYHYQVGFYN | KQGNWI KNGDW NQGNWO GQGKWO VHGTWO AEGEWI AEGEW | KEEK LKGRAVIEQUENAKQAFHVKLEKLIKSLEPACTIVAKF NQVISKADVIESIHETIIRFHPIQAAVSELEMTIVPDEKKSNFLPLKP QAKTLKNSETTKAVEKNIVSFHKSLSQRISELETIYPSPLMTELKLELSA QTKTLKDGSTABAVEKNIHAFHQRISQQLSALELSIFPSPLMTELKLELSA DKKPLKDNATLDIIEQNIISFHQSLSKQLSELELSIFPSAKMAEFKLELSA LMKPLP-KKHQAEVERTINHFYKLISDKIHQLGCQLAPAEQVTNELTVTAA INKPLP-KHQAEVERTINHFYKLISDKIHQLGCQLAPAEQVTNELTVTAA |
| 37 38 39 40 41 42 43 44 | MLPSGHPKSRLM MISSLHPTRGKLL MLKRDLPKKEILNMALPTDYSNGKYL MLKRDLPKKEILNMALPTDYTNGKFL MSERVTYKRLL MSERVTYKRLL MSERVTYKRLL MSERVTYKRLL | NFTALGF RFAQIGF KFAAIGF KFAAMGF KFSALGF QFAAIGF QFAAIGF TLKAMGF | YIREAQCEDTAF YIREQQCQESQF YLREKQSLKDCY YLREKQCVKNCY YLREKQCLDGCY YLREDLCEEGRY YLREDLCEEGRY YLREDCEEGRY | FFDCLA\ FFDCLA\ FFDSLV\ FFDSLV\ FFDSLT\ RFDCLS\ RFDCLS\ LFDCLAS | VCVNIKPAF VCVNANIAF VCVNANIAF VCVNANVF VCVNVNSAF VCVSAKPAF VCVSAKPAF SCVNDKKSF | EEREFWGWWLILEDTG EKREFWGWWLLELVSTH EKREFWGWWLLELVSTH EKREFWGWWLLELSTD DKREFWGWWLLVLEQQD DKREFWGWWLVLEQQD DKREFWGWWLVLEQQD | XAFSYEYHYQLFDJ EQLIYYYQVGLFDJ EGFEFAYHLGMYDJ NGFEFAYNLGIYT OGFEYSYHLGLYN ISFSYHYQVGFYNJ RSFSYHYQVGFYNJ RSFSYHYQVGFYNJ DEMSACYHIGRYT | KQGNWI KNGDW GQGKW VHGTW AEGEWI AEGEWI LAGDW | KEEK LKGRAVIEQVENARQAPH VKLEKLIKSLEPACTIVARP VNQVISKADVIESIHETIIRPHPIQAAVSELEMTIVP DEKKSNFLPLKP QAKTLKNSETTKAVEKNLVSPHKSLSQRLSELELTLYPSPLMTELKLELSA QYKTLKNGSTAEAVEKNLHAFHQRLSQQLSALELSLFPSPLMTELKLELSA DKPLKDNATLDIIEQNIISPHQSISKQLSELESLFPSIMTELKLELSA LNKPLP-KKHQAEVERTINHFYKLISDKLTLLECQLAPAEQVTNELTVTAA NKPLP-KKHQAEVERTINHFYKLISDKLTLLECQLAPAEQVTNELAVTAA VAEATPESAQAEVHTTOAFFHKKLVKTLERFEISVYSTSSAPFA |
| 37 38 39 40 41 42 43 44 45 | MLPSGHPKSRLM MISSLHPTRGKLK MLKRDLPKKELLNMALPTDYSNGKVLK MLKRDLPKKELLNMALPTDYSNGKVLK MALPTDYSNGKVLK MALPTDYSNGKVLK MSERVTYKRLLF MSERVTKRLF MSERMTKTPTHYRLLS MSERMTKTPTHYRLLS | NFTALGF RFAQIGF KFAAIGF KFAAMGF KFSALGF QFAAIGF QFAAIGF TLKAMGF | YIREQOCEDTAF YIREQOCQESQF YIREKQSLKDCY YIREKQCLDGCY YIREDLCEEGRY YIREDLCEEGRY YIREDLCEEGRY YIREGOCSERFY YIREGOCSERFY | FFDCLAY FFDCLAY FFDSLVY FFDSLVY FFDSLTY RFDCLSY RFDCLSY LFDCLAS | VCVNIKPAF VCVNKKVTF VCVNANIAF VCVNANVVF VCVNVNSAF VCVSAKPAF SCVNDKKSF SCVNDKKSF | EEREFWGWWINLEDTG EKREFWGWWIELERNG EKREFWGWWIELENSTH EKREFWGWWIELSPST EREFWGWWILSENST DKREFWGWWIVLEQQD DKREFWGWWIELTQNE EKREFWGWWIELTQNE | XAFSYEYHYQLGLFD 2QLIYYYQUGLFD 3GFEFAYHLGNYD 1GFEFAYHLGIYT 1SFSYHYQVGFYN 2SFSYHYQVGFYN 2EMSACYHIGRYT 2EMSACYHIGRYT | KQGNW KNGDW NQGNW GQGKW VHGTW AEGEW AEGEW LAGDW LAGDW | KEEK LKGRAVIEQVENARQAFH VKLEKLIKSLEPACTIVAKP VNQVISKROVIESIHETIIRFHPIQAAVSELEMTIVPDEKKSNFLPLKP QAKTLKNSETTKAVEKNLVSFHKSLSQRLSELEHITYPSPLMTELKLELSA QTKTLKDGSTABAVEKNIHAFHQSISKQLSELELSLFPSPLMTELKLELSA QDKPLKDNATLDIIEQNIISFHQSISKQLSELELSLFPSAKMAEFKLEISA DKPLP-KKHQABVERTINHFYKLISDKHQLGCLAPAEQVTNELTVTAA INKPLP-KKHQABVERTINHFYKLISDKHTLECCLAPAEQVTNELTVTAA VAEAIPESAQAEVNHTQAEFHKKLVKTLRERFEISVTVSTESAPFA |
| 37 38 39 40 41 42 43 44 45 46 | MLPSGHPKSRLM MLSLPTRGKLK MLKRDLPKKEILMALPTDYSNGKYL MLKRDLPKKEILMALPTDYTNGKPL MLKRDLPKEILMALPTDYSNGKYL MSERVTYKRLK MSERVTYKRLK MSERVTKPTHYRLLS MSEMTKTPTHYRLS | NFTALGE RFAQIGE KFAAIGE KFSALGE QFAAIGE QFAAIGE TLKAMGE TLKAMGE | YIREQQCEDTAF YIREQQCQESQF YLREKQSLKDCY YLREKQCLDGCY YLREKQCLDGCY YLREDLCEEGRY YLREDLCEEGRY YLREGQCSERFY YLREGQCSERFY YLREGQCSERFY | FFDCLA\ FFDSLV\ FFDSLV\ FFDCLS\ LFDCLS\ LFDCLAS LFDCLAS | VCVNIKPA VCVNKKVIP VCVNANIA VCVNANVVP VCVNANVSA VCVSAKPAP VCVSAKPAP SCVNDKKSP SCVNDKKSP | EEREFWGWWINLEDTG; EKREFWGWWILELVSTH EKREFWGWWILLSTSTH EKREFWGWWILLSTSTD DKREFWGWWILLSPNB DKREFWGWWILLQOD EKREFWGWWISLTQNE EKREFWGWWISLTQNE EKREFWGWWISLTQNE | CAFSYEYHYGLFD BQLIYYQVGLFD GGFEFAYHLGMYD NGFEFAYHLGIYT DGFEYSYHLGIYT NGFSYHYQVGFYN RSFSYHYQVGFYN 2EMSACYHIGRYT 2EMSACYHIGRYT 2EMSACYHIGRYT | KQGNWI KNGDW NQGNWG GQGKWG VHGTWG AEGEW AEGEW LAGDW LAGDW LAGDW | KEEK L KGKAV I BQUENAKQAFH VKLEKLIKSLEPACTIVAKP VNQVI SKAVI B SI HETI I I FIPIP L QAAVSELEMTI VP DEKMSNF LIPLKP QAKTLKNS TTKAVEKNU VS FHKSL SQRISELEITI YP SPLMTELKLELSA QTKTLKDGSTABAVEKNU NS FHQSL SQQL SALELSL FP SPLMTELKLELSA DKPLKDNATLDI I EQNI I SFHQSL SKQL SELELSL FP SAKMAEFKLELSA LMKPLP-KHQAEVERTINHFYKLI SDK. HQLGCQLA PAEQVTNELTVTAA INKPLP-KHQAEVERTINHFYKLI SDK. HQLGCQLA PAEQVTNELTVTAA VAEA I PESAQAE VNHTQAEFHKKLVKTL RERFEI SVTVSTESAPFA VAEA I PESAQAE VNHTQAEFHKKLVKTL RERFEI SVTVSTESAPFA |
| 37 38 39 40 41 42 43 44 45 46 47 | MLPSGHPKSRLM MLKRDLPKKEILNMALPTDYSNGKYL MLKRDLPKKEILNMALPTDYSNGKYL MSERVTYKRLF MSERVTYKRLF MSERVTYKRLF MSEMTKTPTHYRLLS MSEMTKTPTHYRLLS MEIHMSEVTNNPTHRLA | NFTALGE RFAQIGE KFAAIGE KFSALGE QFAAIGE TLKAMGE TLKAMGE TLKAMGE | YIREQCCEDTAF YIREQOCQESQF YIREKQSLKDCY YIREKQSVKNCY YIREKCLDGCY YIRECLCEEGRY YIREQCSERFY YIREQCSERFY YIREQCSERFY YIREQCSERFY YIREQCSERFY | FFDCLAX FFDCLAX FFDSLVV FFDSLTV RFDCLSX LFDCLAS LFDCLAS LFDCLAS LFDCLAS | VCVNIKPAF VCVNKKVTF VCVNANIVF VCVNANVVF VCVSAKPAF VCVSAKPAF SCVNDKKSF SCVNDKKSF SCVNDKKSF VCIDDRKSF | EEREFWGWWINDLEDTG EKREFWGWWILELVSTH EKREFWGWWILLSVSTH EKREFWGWWILLSPNE DKREFWGWWIVLEQQD DKREFWGWWIVLEQQD DKREFWGWWIVLQQD EKREFWGWWMELTQNE EKREFWGWWMELTQNE EKREFWGWWMELTQNE | CAFSYEYHYGLFD 3GITYYVQVGLFD 3GFEFAYHLGMYD 3GFEFAYHLGIYT OCFEYSYHLGIYT OFFSYHYQVGFYN 3SFSYHYQVGFYN 2EMSACYHIGRYT 2EMSACYHIGRYT 2EMSACYHIGRYT 3GFTANYHIGKYD | KQGNWI KNGDW QQGNWQ GQGKWQ VHGTWQ AEGEW AEGEW LAGDW LAGDW LAGDW VEGNWI | KEEK LKGRAVIEQUENARQAPH VKLEKLIKSLEPACTIVARP VNQVISKROVIESIHETIIRFHPIQAAVSELEMTIVP DEKMSNFLPLKP QAKTLKNSETTKAVEKNLVSPHKSLSQRLSELEIITLYP SPLMTELKLELSA QTKTLKDGSTARAVEKNLUSPHKSLSQQLSALELSIFPSPLMTELKLELSA DNFLKDNATLDIIEQNIISFHQSLSKQLSELEISIFPSPLMTELKLELSA LNKPLP-KKHQAEVERTLNHFYKLISDKLTLLECQLAPAEQUTNELTVTAA NKPLP-KKHQAEVERTLNHFYKLISDKLTLLECQLAPAEQUTNELTVTAA VAEAIPESAQAEVHHTQAEFHKKLVKTLREFFEISVTVSTESAPFA VAEAIPESAQAEVHHTQAEFHKKLVKTLREFFEISVTVSTESAPFA VAEAIPESAQAEVNHTQAEFHKKLVKTLREFFEISVAVSTESAPFA VAEAIPESAQAEVNHTQAEFHKKLVKTLREFFEISVAVSTESAPFA VAEAIPESAQAEVNHTQAEFHKKLVKTLREFFEISVAVSTESAPFA |
| 37 38 39 40 41 42 43 44 45 46 47 48 | MLPSGHPKSRLM MTSSLHPTRGKLK MLKRDLPKKEILNMALPTDYSNGKVLK MLKRDLPKKEILNMALPTDYSNGKVLK MLKRDLPKKEILNMALPTDYSNGKVLK MSERVTYKRLLF MSERVTYKRLLF MSERMTKTPTHYRLLS MSEMTKTPTHYRLLS MEIMSEVTNNPTHNRLLA | NFTALG RFAQIG KFAAIG KFAAMGE KFSALG QFAAIG QFAAIG TLKAMGE TLKAMGE TLKAMGE KLRAMGE | YIREQCCEDTAF YIREQCQESQF YIREQCSVKNCY YIREKQCLDGCY YIREKQCLDGCY YIREDLCEEGRY YIREDLCEEGRY YIREQCSERFY YIREQCSERFY YIREQCSERFY YIREQCSERFY YIREQCSERFY YIREQQSKEGLY YIRDPQSKEGLY | FFDCLAX FFDSLVV FFDSLVV FFDSLVV RFDCLSV RFDCLSV LFDCLAS LFDCLAS LFDCLAS YFDCLSV | VCVNIKPAF VCVNANIAF VCVNANIAF VCVNANVVF VCVSAKPAF VCVSAKPAF SCVNDKKSF SCVNDKKSF SCVNDKKSF VCIDDRKSF VCIDDRKSF | EEREPWGWWEALEATS EKREPWGWWELLVSTH EKREPWGWWLLVSTH EKREPWGWWLLASTD DKREPWGWWLLEQQD EKREPWGWWELTQNE EKREPWGWWMELTQNE EKREPWGWWMELTQNE EKREPWGWWMELTQNE | CAFSYEYHYGLFD 3CJIYYYQVGLFD 3CFEFAYHLCMYD 4CFEFAYHLCIYM 4SFSYHYQVGFYN, 3SFSYHYQVGFYN, 3SFSYHYQVGFYN, 2EMSACYHICRYT 3CFTANYHICKYD 3CFTANYHICKYD | KQGNW KNGDW QQGKW QQGKW VHGTW AEGEW LAGDW LAGDW LAGDW LAGDW VEGNW | KEEK L KGKAV JI ESI HEQVENAKQAFH VKLEKLIKSLEPACTIVAK VNQVI SKKAV JI ESI HETI II R FHDI QAAVSELEMTI VY DEKMSNF LJPLKP QAKTLKNSETTKAVEKNI HAS FHKSL SQRISELETI Y PSPLMTELKLELSA QTKTLKDG STABAVEKNI HAS FHQSL SKQL SELELSI FPSPLMTELKLELSA QDKPLKDNATLDI I EQNI IS FHQSL SKQL SELELSI FPSAKMAEFKLELSA NKPLP-KKHQABVERTINHFYKLI SDK LTLECQLA PAEQVTNELTVTAA NKPLP-KKHQABVERTINHFYKLI SDK LTLECQLA PAEQVTNELTVTAA VABAI PESAQABVNHTQABFHKKLVKTLREFEI SVTVSTESAPFA VABAI PESAQTEVNHTQABFHKKLVKTLREFEI SVTVSTESAPFA VABAI PESAQAEVNHTQABFHKKLVKTLREFEI SVAVSTESAPFA LDLAI PKNALE VNKTQNCFHLKLVKTLENFFL SVAVSTESAPFA LDLAI PKNALE VNKTQNCFHLKLVKTLENFFQL SVAYHEGSVEFV |
| 37 38 39 40 41 42 43 44 45 46 47 48 49 | | NFTALGE RFAQIGE KFAANGE KFSALGE QFAAIGE QFAAIGE TLKAMGE TLKAMGE KLRAMGE KLRAMGE VLKAIGE | YIREQCESTA YIRECCESS YIRECSLKOCY YIRECCESC YIRECCESC YIRECCESE YIRECCESE YIRECCESE YIRECCESE YIRECCESE YIRECCESE YIRECESE YIRECESE YIRECESE YIRECESE YIRECESE YIRECESE YIRECESE YIRECS | FFDCLAX FFDSLVV FFDSLVV FFDSLVV RFDCLSV LFDCLSSV LFDCLAS LFDCLAS LFDCLSV YFDCLSV | VCVNIKPA VCVNANIA VCVNANIA VCVNANVP VCVSAKPA VCVSAKPA SCVNDKKS SCVNDKKS SCVNDKKS VCIDDRKS VCIDDRKS VCIDDRKS | EEREFWGWWILLEDTG EKREFWGWWILLVSTH EKREFWGWWILLSTSTH EKREFWGWWILLSTSTH DKREFWGWWILLSPNB DKREFWGWWILLQQD DKREFWGWWILLQQD EKREFWGWWILLQNE EKREFWGWWIELTQNE ELREFWGWWIELTQNE ELREFWGWWIELEATE ELREFWGWWIELEATE | CAFSYEYHYGLFD 3GIIYYYQ'GLFDJ 3GFEFAYHLGWYD 3GFEFAYHLGIYT 0GFEYSYHLGLYT 0FFSYHYQVGFYN 7SFSYHYQVGFYN 7SFSYHYQVGFYN 2EMSACYHIGRYT 3GFTANYHIGKYD 3GFTANYHIGKYD 3GFSAKYMTGKYD | KQGNWI KNGDW NQGNW GQGKW VHGTW AEGEW LAGDW LAGDW LAGDW VEGNWI VEGNWI IDGNW | KEEK L KGKAD I BEQUENAKQAPH VKLEKLIKSLEPACTIVAKP VNQVI SKKAD I BEJ HETI IT FIP PL QAAVESLEMTI VP DERMSNF LIP LKP QAKTLKNESTTKAVEKNUVS FHKSL SQRISELEITI YP SPLMTELKLEL SA QTKTLKDGSTABAVEKNUNS FHQSL SQQL SALELSL FP SAKMAEFKLELSA DKPLKDNATLDI I EQNI IS FHQSL SKQL SELELSL FP SAKMAEFKLELSA LMKPLP-KHQAEVERTINNF YKLI SDKL HQLGCQL AP AEQVTNELIVTAA INKPLP-KHQAEVERTINNF YKLI SDKL TILECQL AP AEQVTNELIVTAA VAEA I PESAQAEVNHTQAEFHKKLVKTL RERFEI SVTVSTESAPFA VAEA I PESAQAEVNHTQAEFHKKLVKTL RERFEI SVTVSTESAPFA LDLAI PKNALEEVNKTQNCFHLKLVKTL EENFQL SVAYHEQSVEFV LDLAI PKNALEEVNKTQNCFHLKLVKTLEENFQL SVAYHEQSVEFV |
| 37 38 39 40 41 42 43 44 45 46 47 48 49 50 | | NFTALGE RFAQIGE KFAAIGE KFSALGE QFAAIGE QFAAIGE TLKAMGE TLKAMGE KLRAMGE KLRAMGE VLKAIGE VLKAIGE | YIREOCCESSF YIREOCCESSF YIREOCSKNCY YIREOCSKNCY YIREOCSESF YIREOCSESF YIREOCSEF YIREOCSEF YIREOCSEF YIREOCSEF YIREOCSEF YIREOCSEF YIREOSEG | FFDCLAY FFDSLVY FFDSLVY FFDCLSY RFDCLSY LFDCLAS LFDCLAS LFDCLAS YFDCLSY IFDCLSY IFDCLSY | C VNK KVT C VNANVY C VNANVY C VNANVY C VSAKPAE S VNDK KSE S VNDK KSE C VNDK KSE C VNDK KSE C VDDK KSE C VDDK KSE C VNDK KSE | EEREPWGWWELLENS EKREPWGWWELEVSTH EKREPWGWWELIVSTH ERREPWGWWELSSPNE DKREPWGWWELSSPNE DKREPWGWWELLQQD EKREPWGWWELLQQD EKREPWGWWELTQNE EKREPWGWWELTQNE EKREPWGWWELEATE EREPWGWWELEATE | CAFSYEYHYGLFD 3CJIYYYQVGLFDJ 3CFEFAYHLGWYD NGFEFAYHLGIYN 3SFSYHYQVGFYN 2SFSYHYQVGFYN 2ENSACYHIGRYT 2ENSACYHIGRYT 3GFTANYHICKYD 3GFTANYHICKYD 3GFTANYHICKYD | KQGNW KNGDW QQGKW GQGKW VHGTW AEGEW AEGEW LAGDW LAGDW LAGDW LAGDW VEGNW VEGNW IDGNW | KEEK L KGKAD I BQUENAKQAPH VKLEKLIKSLEPACTLVAK VNQUI SKKAD I BSIHETII IR PHD L QAAVSELEMIL VY DERMSNF LPLKP QAKTLKNSETTKAVEKNL VS FHKSL SQRLSELEMIL VY DERMSNF LPLKL QIKTLKDGSTABAVEKN HAF APROSLSKQLSELELSL FP SPLMTELKLELSA QDKPLKDMATLDI I EQNLIS FHQSLSKQLSELELSL FP SAKMAEFKLELSA DIKPLP - KKHQABVERTINHFYKLISDKL NLGCLAPAEQVTNELAVTAA INKPLP - KKHQABVERTINHFYKLISDKL NLGCLAPAEQVTNELAVTAA VAEAI PESAQAEVNHTQAEFHKKLVKTLREFEISVTVSTESAPFA VAEAI PESAQAEVNHTQAEFHKKLVKTLREFEISVTVSTESAPFA VAEAI PESAQAEVNHTQAEFHKKLVKTLREFEISVTVSTESAPFA LDLAI PKNALESVNKTQNCFHLKLVKTLREFFISVAVSTESAPFA DLAI PKNALESVNKTQNCFHLKLVKTLEENFQLSVAYHEQSVEFV DFLPLPKKAIAEVSRTQEAFHQKLVDELQKKFEIGVQLDEESVEFV DFLPLPKKAIAEVSRTQEAFHQKLVDELQKKFEIGVQLDEESVEFV |
| 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 | MLPSGHPKSRLM MTSSLHPTRGKLL MLKRDLPKKEILNMALPTDYSNGKVL MLKRDLPKKEILNMALPTDYSNGKVL MLKRDLPKKEILNMALPTDYSNGKVL MALPTDYSNGKVL MSERVTYRKLL MSERVTYRKLL MSERVTKPTHYRLL MSEMTKTPTHYRLL MSEMTKTPTHYRLL MSEVTNNPTHNRLLA MSEVTNNPTHNRLLA MSEVTNQPTHYRLLN MSEVTQQPTHYRLLN MSEVTQQPTHYRLLN | NFTALGF RFAQIGF KFAANGF KFSALGF QFAAIGF QFAAIGF TLKAMGF TLKAMGF KLRAMGF KLRAMGF VLKAIGF VLKAIGF | VIREACCESTA VIREACCESS VIREXOSIMO VIREXOSIMO VIREXOSIMO VIRENCEESE VIRENCEESE VIRENCESS VIRENCES | FFDCLAY FFDSLVV FFDSLVV FFDSLVV FFDCLSV FFDCLSV LFDCLSV LFDCLSV YFDCLSV YFDCLSV IFDCLSV IFDCLSV | YC VNK K VTF YC VNAN VTF YC VNAN VTF YC VNAN VTF YC VAN SAF YC VSAK FAF SC VND K SF SC VND K SF YC IDD R SF YC IDD R SF YC VND K SF YC VND K SF YC VND K SF | EEREPWGWWILLETNG EKREPWGWWILLVSTH EKREPWGWWILLVSTH EKREPWGWWILLSPNB DKREPWGWWILLSPNB DKREPWGWWILLQOD EKREPWGWWILLQOD EKREPWGWWMELTQNE EKREPWGWWMELTQNE EKREPWGWWMELTQNE ELREPWGWWMELTQNE EREPWGWWILLTQNE EREPWGWWILLDXSE EREPWGWWILLDXSE | CAFSYEYHYGLFD 3CFEFAYHLCMYD 3CFEFAYHLCYT 0CFEYSYHLCYT 0CFEYSYHLCYT 2CFSYHYQVCFYN 3SFSYHYQVCFYN 2EMSACYHICRYT 2CFSACYHICRYT 3CFTANYHICKYD 3CFTANYHICKYD 3CFTANYHICKYD 3CFSAKYNTCKYN 3CFSAKYNTCKYN | KQGNWI KNGDW NQGKWU VHGTW AEGEW AEGEW LAGDW LAGDW LAGDW ULAGDW ULAGDW IDGNWI IDGNWI IDGNWI | KEEK L KGRAV I BQVENAKQAFH VKLEKLIKSLEPACTIVAK VNQVI SKKDVI E SIHETIIT FHPI QAAVSELEMTI VP DERMSNF LPLKP QAKTLKNETTKAVEKNLKAFHQEN SQLSALELSI FP SPLMTELKLELSA QTKTLKDGSTABAVEKNLKAFHQEN SQLSALELSI FP SPLMTELKLELSA DKPLKDNATLDII E QNI IS FHQSLSKQLSELELSI FP SAKMAEFKLELSA LNKPLP-KKHQAEVERTINNFYKLISDK HQLGCQLA PAEQVTNELTVTAA NNFLP-KKHQAEVERTINNFYKLISDK HILGLGCQLA PAEQVTNELTVTAA VAEAI PESAQAEVNHTQAEFHKKLVKTI RERFEISVTVSTESAPFA VAEAI PESAQAEVNHTQAEFHKKLVKTI RERFEISVTVSTESAPFA ULAI PKNALEEVNKTQNCFHLKLVKTLEENFFLSVAVSTESAPFA DLLAI PKNALEEVNKTQNCFHLKLVKTLEENFPLSVAVSTESAPFA DPLPLPKKAIAEVSRTQEAFHQKLVDELQKKFEIGVQLDEESVEFV |
| 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 | | NFTALGE RFAQIGE KFAAIGE KFSALGE QFAAIGE TLKAMGE TLKAMGE KLRAMGE KLRAMGE KLRAMGE VLKAIGE VLKAIGE VLKAIGE | YIREQCESSF YIRECCESSF YIRECCESSF YIRECCESSF YIRECCESSF YIRECCESSF YIRECCESSF YIRECCESSF YIRECCESSF YIRECCESSF YIRECESSF YIRECESSF YIREPESSE YIREPE | FFDCLAY FFDSLVY FFDSLVY FFDSLVY FFDCLSY RFDCLSY LFDCLSY LFDCLSY YFDCLSY IFDCLSY IFDCLSY IFDCLSY IFDCLSY | YC VNK KVTE YC VNANIAF YC VNANIAF YC VNANVA SC VNANYAF SC VNAKSE SC VNAKSE SC VNAKSE YC IDDRKSE YC VNAKSE YC VNAKSE YC VNAKSE YC VNAKSE | EEREFWGWWILLEDTG EKREFWGWWILELVSTH EKREFWGWWILELVSTH EKREFWGWWILELSTD DKREFWGWWILLSPNE DKREFWGWWILLQQD DKREFWGWWILLQQD EKREFWGWWILTQNE EKREFWGWWILTQNE ELREFWGWWILLQNE ELREFWGWWILLATE ELREFWGWWILLATE EREFWGWULELDKSE EREFWGWUILDKSE | CAFSYEYHYGLFD 3GIIYYYQ'GLFDJ 3GFEFAYHLGWYD 3GFEFAYHLGIYT 0GFEYSYHLGIYT 0FFSYHYQVGFYN 7SFSYHYQVGFYN 7SFSYHYQVGFYN 7DFAC 2EMSACYHIGRYT 3GFTANYHIGKYD 3GFTANYHIGKYD 3GFSAKYNTGKYN 3GFSAKYNTGKYN | KQGNW KNGDW NQGNW QQCKW VHGTW AEGEW LAGDW LAGDW LAGDW VEGNW IDGNW IDGNW IDGNW IDGNW IIDGNW | KEEK L KGKAV I BQVENAKQAPH VKLEKLIKSLEPACTIVAK VNQVI SKNDVI BSI HETII IR PHPL QAAVSELEMTI VP DERMSNF LIPLKP QAKTLKNSETTKAVEKNU VS FHSL SQRISELEII I VP SPLMTELKLELSA QTKTLKDGSTABAVEKNU VS FHSL SQQL SALELSL FP SPLMTELKLELSA DKPLKDNATLDI I EQNI IS FHQSL SKQL SELELSL FP SAKMAEFKLELSA LNKPLP-KHQAEVERTINNFYKLI SDKLTILECQL AP AEQVTNELIVTAA INKPLP-KHQAEVERTINNFYKLI SDKLTILECQL AP AEQVTNELIVTAA VAEA I PESAQAEVNHTQAEFHKKLVKTL RERFEI SVTVSTESAPFA VAEA I PESAQAEVNHTQAEFHKKLVKTL RERFEI SVTVSTESAPFA LDLAI PKNALE SVNKTQNCFHLKLVKTLEENFQL SVAYHEQSVEFV LDLAI PKNALE SVNKTQNCFHLKLVKTLEENFQL SVAYHEQSVEFV DFLPLPKKAI AE VSRTQEAFHQKLVDELQKKFEI GVQLDEESVEFV DFLPLPKKAVAE VSRTQEAFHQKLVDELQKKFEI GVQLDEESVEFV |
| 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 | | NFTALGE RFAQIGE KFAAIGE KFAAIGE QFAAIGE QFAAIGE TLKAMGE TLKAMGE KLRAMGE KLRAMGE VLKAIGE VLKAIGE VLKAIGE VLKAIGE | YIR 0 CEDTAP YIR 0 CLGSOF YIR COLOCSOF YIR CLGCSOF YIR CLGCSOF YIR CLGEGR YIR CLEGGR YIR CLEGGR YIR CCEGGR YIR CCSERFY YIR COCSERFY YIR COCSERFY YIR COSERFY YIR POSEGOF YIR POSEGOF YIR POSEGOF YIR POSEGOF YIR POSEGOF YIR POSEGOF YIR POSEGOF | FFDCLAX FFDSLVV FFDSLVV FFDSLTV RFDCLSV RFDCLSV LFDCLSV LFDCLSV YFDCLSV IFDCLSV IFDCLSV IFDCLSV IFDCLSV | VCVNKKVTE VCVNKNVTE VCVNKNSAE VCVSAKPAE VCVSAKPAE VCVSAKPAE VCVSAKPAE VCVSAKPAE VCVDKKSE VCUDKKSE VCVDKKSE VCVNDKKSE VCVNDKKSE VCVNDKKSE | | CAFSYEYHYGLFD 3CFEFAYHLCMYD 3CFEFAYHLCMYD 3CFEFAYHLCYM 3CFEFAYHLCYM 3CFSYHYQVGFYN. 3CFSYHYQVGFYN. 3CFSYHYQVGFYN. 2EMSACYHICRYT 3CFTANYHICKYD 3CFTANYHICKYD 3CFTANYHICKYD 3CFSAKYNTCKYN 3CFSAKYNTCKYN 3CFSAKYNTCKYN | KQGNW KNGQW NQGNW VHGTW AEGEW AEGEW LAGDW LAGDW LAGDW LAGDW IDGNW IDGNW IDGNW IDGNW IDGNW IDGNW IDGNW IDGNW QVGEW | KEEK L KGRAV I BQVENAKQAPH VKLEKLIKSLEPACTIVAK VNQVI SKNDVI E SIHETIIT FHPI QAAVSELEMTI VY DEKMSNF LPLKP QAKTLKNSETTKAVEKNIHA FHCI SQLISELEITI Y SPLMTELKLELSA QTKTLKOSTABAVEKNIHA FHCI SQLISALEISIF PSPLMTELKLELSA QDKPLKDNATLDI I EQNI ISFHQSI SKQLSELELSIF PSPLMTELKLELSA UNKPLP-KKHQABVERTINHFYKIISDK HUQLGCQIA PAEQVTNELTVTAA NABAI PESAQAEVNHTQABFHKKLVKTIREFEISVTVSTESAPFA VABAI PESAQAEVNHTQABFHKKLVKTIREFEISVTVSTESAPFA VABAI PESAQAEVNHTQABFHKKLVKTIREFEISVTVSTESAPFA VABAI PESAQAEVNHTQABFHKKLVKTIREFEISVAVSTESAPFA DLDLAI PKNALEEVNKTQNCFHLKLVKTLEENFQLSVAYHEQSVEFV DLLAI PKNALEEVNKTQNCFHLKLVKTLEENFQLSVAYHEQSVEFV DPLPLPKKAIAEVSRTQEAFHQKLVDELQKKFEIGVQLDEESVEFV DPLPLPKKAIAEVSRTQEAFHQKLVDELQKKFEIGVQLDEESVEFV DSLPLPKKAVAEVSRTQEAFHQKLVDELQKKFEIGVQLDEESVEFV DSLPLPKKAVAEVTRTQEAFHQKLVDELQKKFEIGVQLDEESVEFV |
| 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 | | NFTALGP RFAAIGP RFAAIGP KFAAMGP QFAAIGP TLKAMGP TLKAMGP TLKAMGP KLRAMGP VLKAIGP VLKAIGP VLKAIGP ALKAIGP KLKAIGP | YIREQCESTA YIREQCESS YIREXCESS YIREXCESS YIREXCESS YIRESCESS YIRESCESS YIRECESS YIRECESS YIRECESS YIRECESS YIRECESS YIREQCESS | FFDCLAX FFDSLVV FFDSLVV FFDSLTV FFDCLSX LFDCLAS LFDCLAS LFDCLAS YFDCLSV IFDCLSV IFDCLSV IFDCLSX IFDCLSX IFDCLSX IFDCLSX XFDCLSX XFDCLSX XFDCLSX XFDCLSX | YC VNK K VTE YC VNK N TE YC VNANIAF YC VNANIAF YC VSAKPAE YC VSAKPAE SC VNDKKSE YC IDDRKSE YC IDDRKSE YC VNDKKSE YC VNDKKSE YC VNDKKSE YC VNDKKSE | EEREFWGWWILLEDTGJ EKREFWGWWILLISTH EKREFWGWWILLISTH EKREFWGWWILLISTH DKREFWGWWILLSPNE DKREFWGWWILLEQQD EKREFWGWWILLQNE EKREFWGWWILLQNE ELREFWGWWILLIQNE ELREFWGWWILLATE EREFWGWWILLDKSE EREFWGWWILLDKSE EREFWGWWILLDKSE EREFWGWWILLDKSE | CAFSYEYHYGLFD 3CJIYYYQVCLFDJ 3CFEFAYHLCIYT 0CFEYSYHLCIYT 0CFEYSYHLCIYT 0CFEYSYHUQVCFYN 3SFSYHYQVCFYN 2EMSACYHICRYT 1CFTANYHICKYD 3CFTANYHICKYD 3CFTANYHICKYD 3CFSAKYNTCKYN 3CFSAKYNTCKYN 3CFSAKYNTCKYN 3CFSAKYNTCKYN 3CFSAKYNICKYN | KQGNW KNGDW NQGKW VHGTW AEGEW LACDW LACDW LACDW LACDW LACDW ULACDW IDGNW IDGNW IDGNW IDGNW IDGNW IDGNW KEGKW | KEEK L KGRAV I BQVENAKQAPH VKLEKLIKSLEPACTIVAK VNQVI SKNDVI E SIHETIIT FIP PL QAAVSELEMTI VP DERMSNF LIP LKP QAKTLKNSETTKAVEKNI MAPHQRI SQQLSALELSI FP SPLMTELKLEI SA QTKTLKDGSTABAVEKNI MAPHQRI SQQLSALELSI FP SPLMTELKLEI SA DKPLKDNATLDI I EQNI I SFHQSI SKQLSELELSI FP SAKMAEFKLEI SA INKPLP-KHQABVERTINHFYKLI SDK HQLGCQLAPABQVTNELTVTAA INKPLP-KHQABVERTINHFYKLI SDK HIQLGCQLAPABQVTNELTVTAA VABAI PESAQAEVNHTQAEFHKKLVKTI RERFEI SVTVSTESAPFA VAEAI PESAQAEVNHTQAEFHKKLVKTI RERFEI SVTVSTESAPFA ULAI PESAQAEVNHTQAEFHKKLVKTI RERFEI SVTVSTESAPFA DLLAI PKNALEEVNKTQNCFHLKLVKTI EENFQLSVAYHEGSVEFV DLLAI PKNALEEVNKTQNCFHLKLVKTI EENFQLSVAYHEGSVEFV DPLPLPKKAIAEVSRTQEAFHQKLVDELQKKFEI GVQLDEESVEFV DFLPLPKKAVAEVSTQEAFHQKLVDELQKKFEI GVQLDEESVEFV |
| 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 | | NFTALGE RFAANGE RFAANGE KFFAANGE QFAAIGE QFAAIGE QFAAIGE TLKAMGE TLKAMGE TLKAMGE VLKAIGE VLKAIGE VLKAIGE VLKAIGE KLKAIGE KLKAIGE | YIR 0,0C8057 YIR 0,0C9505 YIR 0,0C9505 YIR 0,0C9505 YIR 0,0C5057 YIR 0,0C5057 YIR 0,0C5057 YIR 0,0C5057 YIR 0,05057 YIR 0,050577 YIR 0,050577 YIR 0,050577 YIR 0,05057777 YIR 0,0505777777777777777777777777777777777 | FFDCLAV FFDSLVV FFDSLVV FFDSLVT RFDCLSV LFDCLS LFDCLS LFDCLS LFDCLS VFDCLS IFDCLS IFDCLS IFDCLS IFDCLS VFDCLS VFDCLS VFDCLS VFDCLS | YC VNK K VTP YC VN& VTP YC VNANIA YC VNANIA YC VNANY S VNNSA S VNDKS S VNDKS S VNDKS YC IDDRKS YC VNDKS YC VNDKS YC VNDKS YC VNDKS YC VNDKS YC VNDKS YC VNDKS YC VNDKS YC VNDKS | EEREPWGWWILLEDTSJ EKREPWGWWILLVSTH EKREPWGWWILLVSTH EKREPWGWWILLSPNEJ DKREPWGWWILLSPNEJ DKREPWGWWILLQQD EKREPWGWWILLQQD EKREPWGWWILLQNE EKREPWGWWILLEATE EREFWGWWILLATE EREFWGWWILLATE EREFWGWWILLDKSE EREFWGWWILLDKSE EREFWGWWILLDKSE EREFWGWWILLDKSE EREFWGWWILLDKSE | CAFSYEYHYGLFD CGEFAYHLCMYD CGFEFAYHLCIYN CGFEFAYHLCIYN CGFEFAYHLCIYN CGFEAYHUCYGFYN CGFAYUGFYN CGFX | KQGNW KNGDW NQGNW GQGKW VHGTW AEGEW LAGDW LAGDW LAGDW LAGDW VEGNW IDGNW IDGNW IDGNW IDGNW IDGNW IDGNW KEGKW KEGW | KEEK L KGKAV I BQVENAKQAPH VKLEKLIKSLEPACTLVAK VNQVI SKKAV I BSI HETI II R PHD L QAAVSELEMI LVP DEKMSNF LPLKP QAKTLKNSETTKAVEKN. HAF PHC SQ SQ LSELEMI LVP DEKMSNF LPLKLELSA QTKTLKOSTABAVEKN. HAF PHQ SJ SKQ LSELELS LFP SPLMTELKLELSA QDKPLKDNATLDI I EQNI IS PHQ SJ SKQ LSELESL SFP SAKMAEPKLELSA QDKPLKDNATLDI I EQNI SPHQ SJ SKQ LSELESL FP SAKMAEPKLELSA QDKPLKDNATLDI I EQNI SPHQ SJ SKQ LSELESL FP SAKMAEPKLELSA VAEA I PESAQA EV ENTINNF YKLI SDKL NLLECQL AP AEQUTNELAVTAA NABAI PESAQA EVNHTQAEPHKKLVKTL RERFEI SVTVSTESAPFA VAEA I PESAQA EVNHTQAEPHKKLVKTL RERFEI SVTVSTESAPFA VAEA I PESAQA EVNHTQAEPHKKLVKTL RERFEI SVAVSTESAPFA LDLA I PKNALE EVNKTQNCPHLKLVKTL EENFQL SVAYHEQSVEFV DFLPLPKKAI AE VSRTQEAPHQKLVDEL QKKFEI GVQLDEESVEFV DFLPLPKKAI AE VSRTQEAPHQKLVDEL QKKFEI GVQLDEESVEFV DFLPLPKKAI AE VSRTQEAPHQKLVDEL QKKFEI GVQLDEESVEFV DSLPLPKKAVAE VSRTQEAPHQKLVDEL QKKFEI GVQLDEESVEFV VLESAPESATQ BI TRTQEVFHEKLVSAL QEKFESLDVAI HDDSVEFV VLESAPESATQ BI TRTQEVFHEKLVSAL QEKFESLDVAI HDDSVEFV VSEALPQKVVE VYLTQSTFHQKLIDTLAEHFKLEVE YHTESFDFA |
| 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 | MLPSGHPKSRLM MTSSLHTRGKLK MLKRDLPKKEILNMALPTDYSNGKVLK MLKRDLPKKEILNMALPTDYSNGKVLK MLKRDLPKKEILNMALPTDYSNGKVLK MSENTYTKRLLF MSENTYTKRLLF MSENTKTPTHYRLLS MSENTKTPTHYRLLS MSENTKTPTHYRLLS MSEVTQPTHYRLLN MSEVTQPTHYRLLN MSEVTQPTHYRLLN MSEVTQPTHYRLLN MSEVTQPTHYRLLN MSEVTQPTHYRLLN MSEVTQPTHYRLLN MSEVTQPTHYRLN MSETTQPTHYRLN MSETTQPTHFRLMS | NFTALGE RFAALGE KFAALGE KFAALGE VFAALGE QFAALGE QFAALGE TLKAMGE TLKAMGE KLRAMGE KLRAMGE VLKALGE VLKALGE VLKALGE KLKALGE KLKALGE KLKALGE KLKALGE | YIREQCCESSF YIREQCCESSF YIREQCCESSF YIREXCSLKCCY YIREXCCESSF YIREXCESSF YIREDLCESSF YIREDLCESSF YIREDLCESSF YIREQCSESF YIREQCSESF YIREPOSES YIREPOSES YIREPOSES YIREPOSES YIREPOSES YIREPOSES YIREPOSES YIREPOSES YIREPOSESF YIREPOSESF YIREPOSESF YIREPOSESF | FPDCLAV FFDSLVV FFDSLVV FFDSLVV FFDSLVV FFDSLVV FFDCLSV LFDCLSV LFDCLSV LFDCLSV LFDCLSV IFDCLSV IFDCLSV HFDCLSV HFDCLSV YFDCLSV YFDCLSV | YC VNK K VTF YC VNK K VTF YC VNANI AF YC VNANI AF YC VSAK FAF SC VND KK SF SC VND KK SF YC IDD KK SF YC VND KK SF YC VND KK SF YC VND KK SF YC VND KK SF YC VDD KK SF YC VDD KK SF YC VDD KK SF YC VDD KK SF | EEREPWGWWILLENG EKREPWGWWILLVSTH EKREPWGWWILLVSTH EKREPWGWWILLSSPN DKREPWGWWILLSPN DKREPWGWWILLQQD EKREPWGWWILLQQD EKREPWGWWILLQNE EKREPWGWWILLQNE EKREPWGWWILLDNE EREFWGWWILLDKSE EREFWGWWILLDKSE EREFWGWWILLDKSE EREFWGWWILLDKSE EREFWGWWILLDKSE EREFWGWWILLDKSE | CAFSYEYHYGLFD 3CFEFAYHLCMYD 3CFEFAYHLCYW 1CFFAYHLCYW 3CFEFAYHLCYW 3CFEYYUVCFFYN 3CFXHYVVCFYN 2EMSACYHICRYT 3CFTANYHICKYD 3CFTANYHICKYD 3CFTANYHICKYN 3CFSAKYNTCKYN 3CFSAKYNTCKYN 3CFSAKYNTCKYN 3CFSAKYNTCKYN 3CFSAKYNTCKYN 3CFSAKYNTCKYN 3CFSAKYNTCKYN 3CFSAKYNTCKYN 3CFSAKYNTCKYN 3CFSAKYNTCKYN 3CFSAKYNTCKYN | KQGNW KNGDW NQGNW GQQKW VHGTW AEGEW AEGEW LAGDW LLAGDW LLAGDW UVEGNW LLAGDW UVEGNW LLAGDW ILAGDW ILAGDW ILAGDW VEGNW ILGNW ILDGNW ILDGNW ILGNW KEGQW KEGQW KEGQW | KEEK L KGKAV I BQVENAKQAPH VKLEKLIKSLEPACTIVAK VNQVI SKKDVI E SIHETIIT FHPI QAAVSELEMTI VP DEKMSNF LIPLKP QAKTLKNSETTKAVEKNILAS FHKSL SQRLSELEMI VP DEKMSNF LIPLKLELSA QTKTLKDGSTABAVEKNILAS FHKSL SQRLSELELSL FP SPLMTELKLELSA QDKPLKDNATLDI I EQNI IS FHQSL SKQLSELELSL FP SPLMTELKLELSA QDKPLKDNATLDI I EQNI IS FHQSL SKQLSELELSL FP SPLMTELKLELSA XDKPLP-KKHQAEVERTINNFYKLISDK ITLECQLA PAEQVTNELTVTAA NKPLP-KKHQAEVERTINNFYKLISDK ITLECQLA PAEQVTNELTVTAA VABA I PESAQAEVNHTQAEFHKKLVKTLREFEISVTVSTESAPFA VAEA I PESAQAEVNHTQAEFHKKLVKTLREFEISVTVSTESAPFA ULLA I PKNALE VNKTQNCFHLKLVKTLEENFQLSVAVHEQSVEFV DLLA I PKNALE VNKTQNCFHLKLVKTLEENFQLSVAVHEQSVEFV DPLPLPKKAIAEVSRTQEAFHQKLVDELQKKFEIGVQLDEESVEFV DSLPLPKKAVAEVSRTQEAFHQKLVDELQKKFEIGVQLDEESVEFV DSLPLPKKAVAEVSRTQEAFHQKLVDELQKKFEISULDESSVEFV SSELPCKVEVSTQEYFHQKLVDLQKKFEISULDESSVEFV |
| 37 38 39 40 41 42 43 44 50 51 52 53 54 55 56 57 | MLPSGHPKSRLM - MTSSLHPTRGKLL MLKRDLPKKEILNMALPTDYSNGKYL MLKRDLPKKEILNMALPTDYSNGKYL - MSENTYSKRLL - MSENTYSKRLL - MSENTKTPTHYRLL - MSENTKTPTHYRLL - MSENTKTPTHYRLL - MSEVTNPTHNRLL - MSEVTNPTHNRLL - MSEVTQPTHYRLL - MSEVTQPTHYRLL - MSEVTQPTHYRLL - MSEVTQPTHYRLL - MSEVTQPTHYRLL - MSEVTQPTHYRLL - MSEVTQPTHYRLL - MSEVTQPTHYRLL - MSETTQGFTHFRLM - MSETTQGFTHFRLM - MSETTQGFTHFRLM | NFTALGE KFAALGE KFAALGE KFSALGE QFAALGE QFAALGE TLKAMGE TLKAMGE KLRAMGE KLRAMGE VLKALGE VLKALGE VLKALGE KLKALGE KLKALGE KLKALGE | $\begin{array}{c} \mathbf{YIR}_{\mathbf{E},0}(\mathbf{CEDTAF})\\ \mathbf{YIR}_{\mathbf{E},0}(\mathbf{CQESOF})\\ \mathbf{YIR}_{\mathbf{E},0}(\mathbf{QSVRNCY})\\ \mathbf{YIR}_{\mathbf{E},0}(\mathbf{QSVRNCY})\\ \mathbf{YIR}_{\mathbf{E},0}(\mathbf{CSEGRF})\\ \mathbf{YIR}_{\mathbf{E},0}(\mathbf{CSEFF})\\ \mathbf{YIR}_{\mathbf{E},0}(\mathbf{CSEFF})\\ \mathbf{YIR}_{\mathbf{E},0}(\mathbf{QSEFF})\\ \mathbf{YIR}_{\mathbf{E},0}(Q$ | F FD CLA \ F FD SL V\ F FD CLS \ L FD CLA S L FD CLA S L FD CLA S L FD CLA S L FD CLS \ L FD CLS \ L FD CLS \ Y FD CLS \ | YC VNK K VT B YC VNK K VT B YC VNANIAF YC VNANV B YC VSAK PA SC VND K S SC VND K S SC VND K S YC IDD R S YC VND K S YC VND K S YC VND K S YC VND K S YC VDD K S YC VD K S YC V S YC YC S YC V S YC YC YC YC YC S YC YC YC YC S YC YC Y | EEREFWGWWILLEDTG EKREFWGWWILLSTSTH EKREFWGWWILLSTSTH EKREFWGWWILLSPNB DKREFWGWWILLSPNB DKREFWGWWILLQDD EKREFWGWWILLQNE EKREFWGWWILLQNE ELREFWGWWILLATE ELREFWGWWILLATE EREFWGWWILLATE EREFWGWWILLDKSE EREFWGWWILLDKSE EREFWGWWILLDKSE EREFWGWWILLSIE EREFWGWWILSIESIE EKREFWGWWIDLSSIE | CAPSYEYHYGLFD COLIYYYQ'GLFDD COFEFAYHLGWYD COFEFAYNLGIYT COFEFSYHQVGFYNN SFSYHYQVGFYNN SFSYHYQVGFYNN DEMSACYHIGRYT DEMSACYHIGRYT COFTANYHICKYD COFSAKYNTGKYN COFSAKYNTGKYN COFSAKYNTGKYN COFSAKYNTGKYN COFTANYHICKYD COFTANYHICKYN COFTANYHICKYN COFTANYHICKYN | KQGNW KNGDW NQGNW GQQKW VHGTW AEGEW AEGEW LAGDW UEGNW UEGNW IDGNW IDGNW IDGNW IDGNW IDGNW IDGNW KEGQW KEGQW KEGDW KEGDW | KEEK L KGKAV I BOUENAKQAPH VKLEKLIKSLEPACTUVAK VNQVI SKKAVI B SI HETI II R PHDL QAAVSELEMI UV DERKSNF LIPLKP QAKTLKNSE TTKAVEKN US FHKSL SQRLSELEMI UV DERKSNF LIPLKLELSA UTKTLKDST TAAVEKN HAA PHQSL SQLLSELELSI FP SPLMTELKLELSA DDKPLKDNATLDI I EQNLIS FHQSL SKQLSELELSI FP SPLMTELKLELSA LNKPLP - KKHQABVERTINRFYKL I SDKL TLLECCLAPAEQVTNELAVTAA VAEAI PESAQAEVNHTQAEFHKKLVKTLREFFEISVTVSTESAPFA VAEAI PESAQAEVNHTQAEFHKKLVKTLREFFEISVTVSTESAPFA VAEAI PESAQAEVNHTQAEFHKKLVKTLREFFEISVTVSTESAPFA ULDLAI FKNALE EVNKTQNCFHLKLVKTLEENFQLSVAHEQSVEFV DDLPLPKKAIAEVSRTQEAFHQKLVDELQKKFEIGVQLDEESVEFV DPLPLPKKAIAEVSRTQEAFHQKLVDELQKKFEIGVQLDEESVEFV DSLPLPKKAVA SVETQEAFHQKLVDELQKKFEIGVQLDEESVEFV DSLPLPKKAVAEVSTQEAFHQKLVDELQKKFEIGVQLDEESVEFV VSEALPCSAVESVHTQEAFHQKLVDELQKKFEIGVQLDEESVEFV VSEALPCSAVE SVETQEAFHQKLVDELQKKFEIGVQLDEESVEFV VSEALPCSAVE SVETQEAFHQKLVDLELGKFEIGULDESSVEFV SSEALPKKVEEVYLTQSTFHQKLIDTLAHFKLEVEYHTESFDFA TSEPLPAKVTEEVYLTQSTFHQKLIDTLAHFKLEVEYHTESFDFA |
| 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 55 56 57 58 | MLPSGHPKSRLM MLSSLHPTRGKLL MLKRDLPKKEILNMALPTDYSNGKYLM MLKRDLPKKEILNMALPTDYSNGKYLM MSRVTYSNGKYLM MSRVTYKRLLF MSERVTYKRLLF MSERVTYRLLF MSEMTKTPTHYRLLS MSENTKTPTHYRLLS MSEVTQPTHYRLN MSEVTQPTHYRLN MSEVTQPTHYRLN MSEVTQPTHYRLN MSEVTQPTHYRLN MSETTQGFTHFRLMS MSETTQGPTHFRLMS MSETTQGPTHFRLMS | NFTALGE KFAALGE KFAALGE KFAALGE QFAALGE QFAALGE TLKAMGE TLKAMGE TLKAMGE TLKAMGE VLKALGE VLKALGE VLKALGE KLKALGE KLKALGE KLKALGE KLKALGE | YIR 0,0 CEDTAP YIR 0,0 CQESOF YIR 0,0 CQESOF YIR 0,0 CQESOF YIR 0,0 CSENT YIR 0,0 SEGHY YIR 0,0 SEGHY | FPDCLAV FFDCLAV FFDSLVV FFDSLVV FFDSLVV FFDSLVV FFDCLSV FFDCLSV FFDCLSV YFDCLSV YFDCLSV YFDCLSV YFDCLSV YFDCLSV YFDCLSV YFDCLSV YFDCLSV YFDCLSV YFDCLSV YFDCLSV YFDCLSV | YC VNK K VTP YC VNK K VTP YC VNANIA YC VNANIA YC VNANV YC VNANY YC VSAKPA SC VNDK SS SC VNDK SS YC VDDK SS YC VNDK SS YC VNDK SS YC VNDK SS YC VNDK SS YC VDDK SS YC VDC SS YC VC SS YC YC SS YC VC SS YC VC SS YC YC YC YC SS YC YC YC SS YC YC YC YC YC SS YC YC Y | | CAFSYEYHYGLFD CAFSYEYHYGVGLFD CFEFAYHLCMYD CFEFAYHLCYM CFFAYHLCYM CFFAYHCYVFYN SFSYHYQVGFYN SFSYHYQVGFYN DEMSACYHICRYT DEMSACYHICRYT DEMSACYHICRYT CFFANYHICKYD CFFANYHICKYD CFFANYHICKYD CFFAXYHICKYN CGFAXYHICKYN CGFAXYHICKYN CGFTAXYHICKYN CGFTAXYHICKYN | | KEEK L KGRAV I BQVENAKQAPH VKLEKLIKSLEPACTIVAK NQVI SKKDVI E SIHETIIT R PHD L QAAVSELEMTI VY DEKKSNF LIPLKP QAKTLKNSETTKAVEKNI HAFHOL SQRLSELEMTI VY DEKKSNF LIPLKLELSA QTKTLKDGSTABAVEKNI HAFHOL SQLSALEISLFP SPLMTELKLELSA QDKPLKDNATLDI I EQNI ISFHQSI SKQLSELELSLFP SAKMAEFKLEI SA DKPLP-KKHQABVERTINHFYKLI SDK HQLGCQIA PAGVTNELTVTAA INKPLP-KKHQABVERTINHFYKLI SDK HILLCCLA PAGVTNELTVTAA VAEAI PESAQAEVNHTQABFHKKLVKTLRERFEI SVTVSTESAPFA VAEAI PESAQAEVNHTQAFFHKKLVKTLRERFEI SVTVSTESAPFA VAEAI PESAQAEVNHTQAFFHKKLVKTLERFFI SVAVSTESAPFA VAEAI PESAQAEVNHTQAFFHKKLVKTLERFFI SVAVSTESAPFA DLDLAI FKNALEE VNKTQNCFHLKLVKTLERFFI SVAVSTESAPFA DLLDLAI PKNALE SVKTQEAFHQKLVDELQKKFEI GVQLDESSVEFV DFLPLPKKAI AEVSRTQEAFHQKLVDELQKKFEI GVQLDESSVEFV DSLPLPKKAI AEVSRTQEAFHQKLVDELQKKFEI GVQLDESSVEFV DSLPLPKKAVE VSTTGEAFHQKLVDELQKKFEI SVI LESSVEFV DSLPLPKANAEVSRTQEAFHQKLVDELQKKFEI SVI DESSVEFV DSLPLPKKAVAEVSRTQEAFHQKLVDELQKKFEI SVI DESSVEFV SSELPGKVE SVI TRTGEVFHEKLVSALGKFS LIVAIHDDSVEFV SSELPGKVE SVI TRTGSTHQKLIDTI LAEHFKLEVEYHTESFDFA TSEPLPAKVTE SVI LQSTFHQKLIDTI LAEHFKLEVEYHTESFDFA TSEPLPAKVTE SVI LGSTFHQKLIDTI VAEHFKLEVEYHTESFDFA |
| 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 253 55 56 57 58 59 | MLPSGHPKSRLM MLRDLPKKELNMALPTDYSNGKVL MLKRDLPKKELNMALPTDYSNGKVL MLKRDLPKKELNMALPTDYSNGKVL MLKRDLPKKELNMALPTDYSNGKVL MSENTVTYKRLL MSENTVTYKRLL MSENTKTPTHYRLL MSENTKTPTHYRLL MSENTKTPTHYRLL MSEVTQPTHYRLL MSEVTQPTHYRLL MSEVTQPTHYRLL MSEVTQPTHYRLL MSETTQGPTHFRLMS MSETTQGPTHFRLMS MSETTQGPTHFRLMS MSETTQGPTHFRLMS MSETTQGPTHFRLMS MSETTQGPTHFRLMS MSETTQGPTHFRLMS MSETTQGPTHFRLMS MSETTQGPTHFRLMS MSETTQGPTHFRLMS MSETTQGPTHFRLMS MSETTQGPTHFRLMS MSETTQGPTHFRLMS MSETTQGPTHFRLMS MSETTQGPTHFRLMS MSETTQGPTHFRLMS | NFTALGE KFAALGE KFAALGE QFAALGE QFAALGE QFAALGE QFAALGE TLKAMGE TLKAMGE TLKAMGE VLKALGE VLKALGE VLKALGE KLKALGE KLKALGE KLKALGE KLKALGE KLKALGE KLKALGE | YIREQCCSCPTAF YIREQCCSSOF YIREQCCSSOF YIREXCSLKDCY YIREDCESCR YIREXCSLKDCY YIREDCESCR YIREQCCSERFY YIREDCESCR YIREQCSERFY YIREOSESCH YIREOSESCH YIREOSESCH YIREOSESCH YIREOSESCH YIREOSESCH YIREOSESCH YIREOSOG YI | | YC VNK K VTF YC VNK K VTF YC VNANIAF YC VNANIAF YC VSAKPAF SC VNDK SF SC VNDK SF YC VDDK SF YC VDDK SF YC VNDK SF YC VNDK SF YC VNDK SF YC VDDK SF YC VDC SF YC VC YC YC SF YC VC YC SF YC VC SF YC VC YC SF YC YC YC YC YC SF YC YC Y | EEREPWGWWILLEDTG EKREPWGWWILLVSTH EKREPWGWWILLSTSTH EKREPWGWWILLSTSTH DKREPWGWWILLSPNE DKREPWGWWILLQOD EKREPWGWWILLQOD EKREPWGWWILLQOD EKREPWGWWILLQOD EKREPWGWWILLQOE ELREPWGWWILLATE EEREPWGWWILLATE EEREPWGWWILLATE EEREPWGWWILLDKSE EEREPWGWWILLSIE EKREPWGWWILSIESIE EKREPWGWWILLSIE EKREPWGWWILLSIE EKREPWGWWILLSIE | CAFSYEYHYGLFD 3CJIYYYQVGLFD 3CFEFAYHLGWYD 4GFEFAYHLGIYT 0GFEYSYHLGIYT 0GFEYSYHLGIYT 2EMSACYHIGRYT 2EMSACYHIGRYT 2EMSACYHIGRYT 3GFTANYHICKYD 3GFSAKYNTGKYN 3GFSAKYNTGKYN 3GFSAKYNTGKYN 3GFTAKYHIGKYN 3GFTAKYHIGKYN 3GFTAKYHIGKYN 3GFTAKYHIGKYN 3GFTAKYHIGKYN 3GFTAKYHIGKYN | | KEEK L KGRAV I BQVENAKQAPH VKLEKLIKSLEPACTIVAK VNQVI SKNDVI E SI HETI II R PHD L QAAVSELEMI VV DERKSNF LPLKP QAKTLKNETTKAVEKNIKAPH KSI SQRISLEMI VV DERKSNF LPLKE LSA QTKTLKDGSTABAVEKNIKAPHQRI SQLSALELSI FP SPLMTELKLEI SA QTKTLKDGSTABAVEKNIKAPHQEI SQLSALELSI FP SPLMTELKLEI SA UNKPLP-KKHQABVERTINNFYKLISDK HQLGCQI APABQVTNELTVTAA INKPLP-KKHQABVERTINNFYKLISDK ITLECQLA PABQVTNELTVTAA VABAI PESAQAEVNHTQABFHKKLVKTI RERFEI SVTVSTESAPFA VABAI PESAQAEVNHTQABFHKKLVKTI RERFEI SVTVSTESAPFA ULAI PKNALE EVNKTQNCFHLKLVKTI RERFEI SVTVSTESAPFA DLLAI PKNALE EVNKTQNCFHLKLVKTI EENFQLSVAYHEGSVEFV DPLPLPKKAIAEVSRTQEAFHQKLVDELQKKFEI GVQLDESVEFV DPLPLPKKAIAEVSRTQEAFHQKLVDELQKKFEI GVQLDESVEFV DSLPLPKKAVAEVTTQEAFHQKLVDELQKKFEI SQLDESVEFV DSLPLPKKAVAEVTTQEAFHQKLVDELQKKFEI SQLDESVEFV VIESAPESATQ EITRQEVFHEKLVSALEKFSLDVAI HHDSVEFV DSLPLPKKAVAEVTTQEAFHQKLVDELQKKFEI SQLDESVEFV DSLPLPKKAVAEVTTQEAFHQKLVDELQKKFEI SQLDESVEFV VIESAPESATQ EITRQEVHEKLVSALEKFSLDVAI HHDSVEFV TSEPLPAKVTE VILGSTFHQKLIDTLAEHFKLEVEYHTESFDFA TSEPLPAKVTE VILGSTFHQKLIDTLEHFKLEVEYHTESFDFA TSEPLPAKVTE VILGSTFHQKLIDTLAEHFKLEVEYHTESFDFA |
| 37 38 40 41 42 43 44 45 46 47 48 49 51 52 53 54 55 56 57 58 59 60 | MLPSGHPKSRLM MLYSGHPKSLMALPTDYSNGKYLM MLKRDLPKKEILNMALPTDYSNGKYLM MLKRDLPKKEILNMALPTDYSNGKYLM MLKRDLPKKEILNMALPTDYSNGKYLM MSERVTYKRLLF MSERVTYKRLLF MSERVTYKRLLF MSERVTYRLLF MSEKTKTPTHYRLLC MSEVTNNPTHNRLLA MSEVTNNPTHNRLLA MSEVTQPTHYRLN MSEVTQPTHYRLN MSEVTQPTHYRLLN MSEVTQPTHYRLL MSEVTQPTHYRLL MSETTQFTHRLMS MSETTQGFTHRLMS MTMDVKSSPGR | NFTALGE KFAALGE KFAALGE KFAALGE QFAALGE QFAALGE TLKAMGE TLKAMGE TLKAMGE VLKALGE VLKALGE VLKALGE KKKALGE KKKALGE KKKALGE KKKALGE KKKALGE KKKALGE KKKALGE KKKALG | YIRE 0 CCBSPF YIRE 0 CCBSPF YIRE 0 CCBSPF YIRE 0 CCBSPF YIRE 0 CSEFFY YIRE 0 CSEFFY YIRE 0 CSEFFY YIRE 0 CSEFFY YIRE 0 SEEGHY YIRE 0 SEEGHY | F PD CLA V F FD SL VV F FD SL VV F FD SL VV F FD SL TV F FD SL TV FD CLS V FD CLS V | YC VNK K VTP YC VNK K VTP YC VNANIA YC VNANIA YC VSAKPA SC VNDKS SC VNDKS SC VNDKS SC VNDKS YC IDDRKS YC VNDKS YC VNDKS YC VNDKS YC VNDKS YC VNDKS YC VNDKS YC VDDKS YC VDDA YC VDDA YC VDDA YC VDDA YC VDA YC YC Y | EEREPWGWWILLEDTG EKREPWGWWILLVSTH EKREPWGWWILLVSTH EKREPWGWWILLSPNE DKREPWGWWILLSPNE DKREPWGWWILLQQD EKREPWGWWILLQQD EKREPWGWWILLQQD EKREPWGWWILLQNE EKREPWGWWILLATE EREPWGWWILLATE EREPWGWWILLATE EREPWGWWILLATE EREPWGWWILLATE EREPWGWWILLSIE EREPWGWWILLSIE EREPWGWWILLSIE EKREPWGWWILLSIE EKREPWGWWILLSIE EKREPWGWWILLSIE EKREPWGWWILLSIE | XAFSYEYHYGLFD XGFSYEYHYGLFD 2GFEFAYHLGWYD 2GFEYNLGYT 2GFEYSYHLGLYN 2GFEYSYHLGLYN 2GFYN 2GFYN 2GFYN 2GFYN 2GFYN 2GFYN 2GFYN 2GFANYHICKYN 2GFANYHICKYN 3GFYANYHICKYN 3GFYANYHICKYN 3GFYANYHICKYN 3GFYANYHICKYN 3GFFAXYHICKYN 3GFTAXYHICKYN 3GFYN 3GFTAXYHICKYN 3GFTAXYHICKYN 3GFTAXYHICKYN 3GFTAXYHICKYN 3GFTAXYHICKYN 3GFTAXYHICKYN 3GFTAXYHICKYN 3GFYN | K QGNW KNGDW KNGCKW VHGTW AEGEW LAGDW VEGNW LLAGDW VEGNW LLAGDW VEGNW IDGNW IDGNW IDGNW IDGNW IDGNW KEGQW KEGDW KEGDW KEGDW KCGDW | KEEK L KGKA VI E SI HEQVENAKQAPH VKLEKLIKSLEPACTLVAK VNQVI SKKD VI E SI HETI II R PHD L QAAVSELEMI LV DERMSNF LIPLKP QAKTLKNSE TTKAVEKNL VS FHKSL SQRLSELEMI LV DERMSNF LIPLKLELSA QIKTLKDGSTABAVEKN HAA FHQSL SQLLSELEMI LV SPLMTELKLELSA QIKTLKDGSTABAVEKN HAAF HQSL SQLLSELESL FP SPLMTELKLELSA QIKTLKDASTLD II E QNI IS FHQSL SKQLSELESL FP SAKMAEFKLELSA QIKTLKDASTLD II E QNI IS FHQSL SKQLSELESL FP SAKMAEFKLELSA QIKTLY VISTABAVEKTINNF YKLI SDKL TLLECQL AP AEQVINELAVTAA INKPLP - KKHQAEVERTINNF YKLI SDKL TLLECQL AP AEQVINELAVTAA VABA I PESAQA EVNHTQAEFHKKLVKTL RERFEI SVTVSTESAPFA VABA I PESAQA EVNHTQAEFHKKLVKTL RERFEI SVTVSTESAPFA VABA I PESAQA EVNHTQAEFHKKLVKTL RERFEI SVAVSTESAPFA DLDLA I FKNALE EVNKTQNCFHLKLVKTL BENFQL SVAYHEQSVEFV DFLPLPKKAI AE VSRTQEAFHQKLVDEL CKKFEI GVQLDESSVEFV DFLPLPKKAI AE VSRTQEAFHQKLVDEL CKKFEI GVQLDESSVEFV DFLPLPKKAI AE VSRTQEAFHQKLVDEL CKKFEI GVQLDESSVEFV DSLPLPKKAVAE VSRTQEAFHQKLVDEL CKKFEI GVQLDESSVEFV SSELPQKKVE VYTQESTFHQKLUDELQKKFEI SVA HIDSSVEFV VLESAPESATQE ITRTQEVFHEKLVSAL QEKFEI DLA I DDSVEFV SSELPQKKVE VYTQSTFHQKLVDTL E LFFKLEVEYHTESFDFA VSEALPKKVVE YYTQSTFHQKLVDTL E SFFI SVA VSTRSFFF TSEPLPAKVTE EVITQESFFN CULDEL SKFFI SVA VSTRSFFF |
| 37 38 40 41 42 43 44 45 46 47 48 49 51 52 53 54 55 55 56 57 58 59 60 | MLPSGHPKSRLM MLYSGHPKSLMAALPTDYSNGKIL MLKRDLPKKEILNMALPTDYSNGKYL MLKRDLPKKEILNMALPTDYSNGKYL MLKRDLPKKEILNMALPTDYSNGKYL MSENTYKRLLF MSENTYKRLLF MSENTKTPTHYRLLS MSENTKTPTHYRLLS MSENTKTPTHYRLLS MSEVTQQPTHYRLLN MSEVTQQPTHYRLLN MSEVTQQPTHYRLLN MSEVTQQPTHYRLLN MSEVTQQPTHYRLLN MSEVTQQPTHYRLLN MSETTQGPTHFRLMS | NFTALGE KFAALGE KFAALGE QFAALGE QFAALGE QFAALGE QFAALGE TLKAMGE TLKAMGE TLKAMGE VLKALGE VLKALGE VLKALGE KLKALG | YIR 0.0 CEDTAP YIR 0.0 CEDTAP YIR 0.0 CSURCY YIR 0.0 SYLKOCY YIR 0.0 CEG 0.0 Y YIR 0.0 SEG 0.0 Y Y YIR 0.0 SEG 0.0 Y Y YIR 0.0 SEG 0.0 Y Y Y YIR 0.0 SEG 0.0 Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y | F PD CLA V F FD SL VV F FD SL VV F FD SL TV F FD SL TV | YCVNKKVTF YCVNANIAT YCVNANIAT YCVNANIAT YCVSAKPAF SCVNDKKSF SCVNDKKSF YCTDDKKSF YCVNDKKSF YCVNDKKSF YCVNDKKSF YCVNDKKSF YCVDDKKSF YCVDDKSF YCVDDKSF YCVDDKSF YCVDDKSF YCVDDKSF YCVDDKSF YCVDDKSF | EEREPWGWWILLENTG EKREPWGWWILLVSTH EKREPWGWWILLVSTH EKREPWGWWILLSPNE DKREPWGWWILLSPNE DKREPWGWWILLQVE EKREPWGWWILTQNE EKREPWGWWMELTQNE EKREPWGWWMELTQNE EREPWGWWMELATE EREPWGWWILLDKSE EREPWGWWILLDKSE EREPWGWWILLDKSE EREPWGWWILLSIE EKREPWGWWILSIE EKREPWGWWILSIE EKREPWGWWILSIE EKREPWGWWILSIE | CAFSYEYHYGLFD 3CJIYYYQVGLFDJ 3CFEFAYHLGWYD 4GFEFAYHLGIYM 4GFEFAYHLGIYM 4SFSYHYQVGFYN. 3SFSYHYQVGFYN. 2EMSACYHIGRYT 2EMSACYHIGRYT 3GFTANYHIGKYD 3GFTANYHIGKYD 3GFSAKYNTGKYN. 3GFSAKYNTGKYN. 3GFSAKYNTGKYN. 3GFSAKYNIGKYN. 3GFTAKYHIGKYN. 3GFTAKYHIGKYN. 3GFTAKYHIGKYN. 3GFTAKYHIGKYN. 3GFTAKYHIGKYN. 3GFTAKYHIGKYN. 3GFTAKYHIGKYN. 3GFTAKYHIGKYN. 3GFTAKYHIGKYN. 3GFTAKYHIGKYN. 3GFTAKYHIGKYN. 3GFTAKYHIGKYN. 3GFTAKYHIGKYN. 3GFEFERFDGLFD | K QGNW KNGQNW KNGQNW KNGQCKW VAEGEW AEGEW LAGDW VEGNW VEGNW ULAGDW VVEGNW IDGNW IDGNW IDGNW IDGNW IDGNW LAGDW KEGQW KEGQW KEGQW LPGNW | KEEK L KGKAV I EGVENAKQAPH VKLEKLIKSLEPACTIVAK NQVI SKKAV I EGVENAKQAPH VKLEKLIKSLEPACTIVAK ZAKTLKNSETTKAVEKNI US FHKSL SQRLSELEMI VP DEKMSNF LPLKD QAKTLKNSETTKAVEKNI HAFNGL SQRLSELEMI VP DEKMSNF LPLKLELSA QTKTLKOG STABAVEKNI HAFNGL SQLSALEIG SLEPSAKMAEFKLELSA QDKPLKDNATLDI I EQNI IS FHQSL SKQL SELELSI FP SPLMTELKLELSA QTKTLKOG STABAVEKNI HAFNKLI SDK HQLGCQLA PAEQUTNELTVTAA INKPLP-KKHQAEVERTINNFYKLI SDK TILECQLA PAEQUTNELTVTAA VAEAI PESAQAEVNHTQAEFHKKLVKTLREFEISVTVSTESAPFA VAEAI PESAQAEVNHTQAEFHKKLVKTLREFEISVTVSTESAPFA VAEAI PESAQAEVNHTQAEFHKKLVKTLREFEISVTVSTESAPFA DLDLAI FKNALE VNKTQNCFHLKLVKTLEENFQLSVAYHEGSVEFV DPLPLPKKAIAEVSRTQEAFHQKLVDELQKKFEIGVQLDEESVEFV DPLPLPKKAIAEVSRTQEAFHQKLVDELQKKFEIGVQLDEESVEFV DSLPLPKKAVAEVTROEAFHQKLVDELQKKFEIGVQLDEESVEFV DSLPLPKKAVAEVTROEAFHQKLVDELQKKFEIGVQLDESVEFV DSLPLPKKAVAEVTROEAFHQKLVDELGKFFIJ ULDESVEFV VSEALPKVVEEVYLTQSTFHQKLIDTLAEHFKLEV YHTESFDFA TSEPLPAKVTE VYLTQSTFHQKLIDTLAEHFKLEV YHTESFDFA TSEPLPAKVTE VYLTQSTFHQKLTDTVAEHFKLEV YHTESFDFA |
| 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 8 59 60 | MLPSGHPKSRLM MLPSGHPKSRLMALPTDYSNGKIL MLKRDLPKKELLNALPTDYSNGKYL MLKRDLPKKELNALPTDYSNGKYL MLKRDLPKKELNALPTDYNGKPL MSERVTYKRLL MSERVTYKRLL MSERVTYKRLL MSERTKPTHYRLL MSEKTKPTHYRLL MSEVTNPTHNRLL MSEVTNPTHNRLL MSEVTNPTHNRLL MSEVTQPTHYRLN MSEVTQPTHYRLN MSEAAKNTHYRLI MSEAAKNTHYRLL MSETQGPTHFRLM MSETTGGPTHFRLM MSETTGGPTHFRLM MSETTGGPTHFRLM MSETTGGPTHFRLM MSETTGGPTHFRLM MSETTGGPTHFRLM MSETTGGPTHFRLM MSETTGGPTHFRLM MSETTGGPTHFRLM | NFTALGEN KFAALGEN KFAALGEN KFAALGEN KFAALGEN KLFAALGEN KLFAALGEN KLFAALGEN KLFAALGEN KLFAALGEN KLFAALGEN KLFALG | YIREQCCESSF YIREQCCESSF YIREQCCESSF YIREXCLGCY YIREDLCESSF YIREQCCSERFY YIREQCCSERFY YIREQCSERFY YIREQCSERFY YIREQCSERFY YIREQCSERFY YIREQCSEGHY YIREPCSEGHY YIREPCSEGHY YIREPCSEGHY YIREPCSEGHY YIREPCSQECHY YIREPCSQECHY YIREPCSQECHY YIREPCSQECHY YIREPCSQECHY YIREPCSQECHY YFRINZSY | | CVNIKPA CVNKKVTI CVNANIA CVNANVU CVNANVU CVSAKPA CVSAKPA SCVNDKKS CVNDKKS CVNDKKS CVNDKKS CVNDKKS CVNDKKS CVNDKKS CVNDKKS CVNDKKS CVNDKKS CVDDKKS CVDDKKS CVDDKKS CVDDKKS CVDDKKS CVDDKKS CVDDKKS CVDDKKS | EEREPWGWWILLEDTG EKREPWGWWILLISPNE EKREWGWWILLISPNE EREPWGWWILLISPNE DKREPWGWWILLISPNE DKREPWGWWILLISPNE DKREPWGWWILLOPNE EKREPWGWWILLIQNE EKREPWGWWILLIQNE ELREPWGWWILLIATSE EREPWGWWILLIDKSE EREPWGWWILLIDKSE EREPWGWWILLIDKSE EREPWGWWILLIDKSE EREPWGWWILLISIE EKREPWGWWILLISIE EKREPWGWWILLISIE EKREPWGWWILLISIE EKREPWGWWILLISIE EKREPWGWWILLISIE EKREPWGWWILLISIE EKREPWGWWILLISIE | CAFSYEYHYQCLFD 3CJIYYYQ'CLFD 3CFEFAYHLCWYD 3CFEFAYHLCIYN 3SFSYHYQVCFYN 3SFSYHYQVCFYN 3SFSYHYQVCFYN 3CFXYQVCFYN 3CFXYQVCFYN 3CFXYQVCFYN 3CFXYQVCFYN 3CFXXYUCKYD 3CFXXYUCKYD 3CFXXYUCKYD 3CFXXYUCKYD 3CFXXYUCKYD 3CFXXYUCKYD 3CFXXYUCKYD 3CFXXYUCKYD 3CFXXYUCKYD 3CFXXYUCKYD 3CFXXYUCKYD 3CFXXYUCKYD 3CFXXYUCKYD 3CFXXYUCKYD 3CFXXYUCKYD 3CFXXYUCKYD 3CFXXYUCKYD | K COCONO COCONO KNOCKKIN COCONO KNOCKKIN COCONO KNOCKKIN COCONO C | KEEKI KGKAD VI BSI HEQVENAKQAPH VKLEKLINKSLEPACTIVAK VNQVI SKKAD VI BSI HETI IIR FNDI QAAVSELEMI VY DERMSNF LPLKP QAKTLKNSETTKAVEKNI VSFHKSL SQRLSELEMI VY DERMSNF LPLKLELSA QTKTLKDSTABAVEKNI HAFNGI SQLSALEKSI FP SPLMTELKLELSA QTKTLKDSTABAVEKNI HAFNGI SQLSALEKSI FP SPLMTELKLELSA QDKPLKDNATLDI I EQNIISFHQSI SQLSELEKI FP SPLMTELKLELSA DIKPLP-KKHQAEVERTINKFYKLISDK ITLECQI APAEQVTNELAVTAA INKPLP-KKHQAEVERTINKFYKLISDK ITLECQI APAEQVTNELAVTAA VABAI PESAQAEVNHTQAEFHKKLVKTIRERFEISVTVSTESAPFA VABAI PESAQAEVNHTQAEFHKKLVKTIRERFEISVTVSTESAPFA ULDAI PKNALE EVNKTQNCFHLKLVKTLERFFISVAVSTESAPFA DILAI PKNALE EVNKTQNCFHLKLVKTLERFQISVAYHEQSVEFV DILPLPKKAIAEVSRTQEAFHQKLVDELQKKFEIGVQLDESSVEFV DILPLPKKAIAEVSRTQEAFHQKLVDELQKKFEIGVQLDESSVEFV DILPLPKKAVAEVSTQEAFHQKLVDELQKKFEIGVQLDESSVEFV DSLPLPKKAVAEVSTQEAFHQKLVDELQKKFEIGVQLDESSVEFV SSEALPGKVVEEVYTQSTFHQKLIDTLAHFKLEVYHTESFDFA TSEPLPAKVTE VILQSTFHQKLIDTLEHFKLEVEYHTESFDFA TSEPLPAKVTEVYLQSTFHQKLIDTVAEHFKLEVEYHTESFDFA TSEPLPAKVTEVYLQSTFHQKLIDTVAEHFKLEVEYHTESFDFA |
| 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 | MLPSGHPKSRLM MLSSLHPTRGKLL MLKRDLPKKEILNMALPTDYSNGKYLM MLKRDLPKKEILNMALPTDYSNGKYLM MSRVTYSNGKYLM MSRVTYKRLLF MSENTKTPTHYRLLS MSENTKTPTHYRLLS MSENTKTPTHYRLLS MSEVTQPTHYRLLS MSEVTQPTHYRLM MSEVTQPTHYRLM MSEVTQPTHYRLM MSETTQGPTHFRLMS MSETTGGPTHFRLMS | NFTALGE KFAAIG KFAAIG GU VIKAIG GTAAIG CAA VIKAIG VIKAIG VIKAIG VIKAIG VIKAIG K | YIREOCCESSF YIREOCCESSF YIREOCCESSF YIREOCSINCY YIREOCSEST YIREOCSEF YIREOCSEF YIREOCSEF YIREOCSEF YIREOCSEF YIREOSES YI | FPDCLAW FFDSLVV FFDSLVV FDSLTV FDSLTV FDSLTV FDSLTV FDCLSV LFDCLAS LFDCLAS LFDCLAS LFDCLAS LFDCLAS LFDCLSV LFDCLSV LFDCLSV LFDCLSV FFDCLSV FFDCLS FFDCLEI | C VNK KVTF C VNK KVTF C VNANIA C VNANIA C VNANVSA C VNVNSA C VNOKSA S VNDKKS S VNDKKS S VNDKKS C VNDKKS C VNDKKS C VNDKKS C VNDKKS C VNDKKS C VNDKKS C VDDKKS C VDDKS C VDC C V | EEREPWGWWILLENTG EKREPWGWWILLVSTH EKREPWGWWILLSPNE DKREPWGWWILLSPNE DKREPWGWWILLQQDD EKREPWGWWILLQQDD EKREPWGWWILTQNE EKREPWGWWILLQNE ELREPGWWIMELTQNE EEREPWGWWIMELATE EEREPWGWWIMELATE EEREPWGWWIMELATE EEREPWGWWIMELSTE EEREPWGWWIMELSIE EKREPWGWWIMELSIE EKREPWGWWIMELSIE EKREPWGWWIMELSIE EKREPWGWWIMELSIE EKREPWGWWIMELSIE EKREPWGWWIMELSIE EKREPWGWWIMELSIE EKREPWGWWIMELSIE | CAPSYEYHYGLFDJ SQLIYYYQVGLFDJ SGPEYAYHLGWYDJ GGPEYSYHLGIYM SGPEYSYHLGIYM SGPEYSYHQVGFYN. SGPSYHQVGFYN. SGPSYHQVGFYN. SGPSAYHIGKYD SGPTANYHIGKYD SGPSAKYNTGKYM. SGPSAKYNTGKYM. SGPSAKYNTGKYM. SGPTAKYHIGKYM SGPTAKYHIGKYM. SGPTAKYHIGKYM. SGPTAKYHIGKYM. SGPTAKYHIGKYM. | K COCONSTANTS OF CONSTANTS OF C | KEEKI KUKAKVI EUQUENAKQAPH VKLEKLIKSLEPACTIVAK VNQVI SKKAVI ESIHETIIIR PHDIQAAVSELEMTI VYDEKKNSNF LPLKP 2AKTLKNSETTKAVEKNI WSFHKSL SQRLSELEMI VYDEKKNSNF LPLKP 2AKTLKNSETTKAVEKNI HAFHGI SQLSALEISLFP SPLMTELKLELSA 2TKTLKDGSTABAVEKNI HAFHGI SQLSALEISLFP SPLMTELKLELSA 2DKPLKDNATLDI I EQNI ISFHQSI SKQLSELEISLFP SPLMTELKLELSA 2DKPLP-KKHQABVERTINNFYKLI SDK HILGCGIA PAGVTNELTVTAA 1NKPLP-KKHQABVERTINNFYKLI SDK HILGCGIA PAGVTNELTVTAA 2000 AND |
| 37 38 39 41 42 43 44 45 50 51 52 53 54 55 56 57 58 59 60 | MLPSGIPKSRLM MISSLHPTRGKLL MLKRDLPKKEILNMALPTDYSNGKVL MLKRDLPKKEILNMALPTDYSNGKVL MLKRDLPKKEILNMALPTDYSNGKVL MSENTYTKRLL MSENTYTKRLL MSENTKTPTHYRLL MSENTKTPTHYRLL MSENTKTPTHYRLL MSENTKTPTHYRLL MSEVTQPTHYRLL MSEVTQPTHYRLL MSEVTQPTHYRLL MSEVTQPTHYRLL MSEVTQPTHYRLL MSEVTQPTHYRLL MSETTQGPTHFRLM MSETTQGPT | NFTALGE KFAALGE KFAALGE QFAALGE QFAALGE QFAALGE QFAALGE VLKALGE KLRAMGE KLRAMGE VLKALGE VLKALGE KLKALG | VIRE OCCREST VIRE OCCREST VIRE OCCREST VIRE COLOCY VIRE COLOCY VIRE COLOCY VIRE OCCREST VIRE OCC | FPDCLAV FFDSLVV FFDSLVV FFDSLVV FFDSLVV FFDSLVV FFDSLV FFDCLSV LFDCLAS LFDCLAS LFDCLAS LFDCLSV UFDCLSV IFDCLSV YFDCLSV YFDCLSV YFDCLSV YFDCLSV YFDCLSV YFDCLSV YFDCLSV YFDCLSV YFDCLSV | CVNKKVTF CVNKKVTF CVNANIA CVNANVUF CVSAKPAF SCVNDKSF SCVNDKSF CIDDRKSF CUDDKSF CVNDKSF CVNDKSF CVNDKSF CVDCKSF CVDDKSF CVDDKSF CVDDKSF CVDDKSF CVDCKSF CVCKS | | CAPSYEYHYGLFD 3CJIYYYQVGLFDJ 3CFEFAYHLGWYD 3CFEFAYHLGIYM 1SFSYHYQVGFYN 3SFSYHYQVGFYN 2EMSACYHIGRYT 2EMSACYHIGRYT 3CFTANYHIGKYD 3CFTANYHIGKYD 3CFTANYHIGKYN 3CFSAKYNTGKYN 3CFSAKYNTGKYN 3CFSAKYNTGKYN 3CFTAKYHIGKYN 3CFTAKYHIGKYN 3CFTAKYHIGKYN 3CFTAKYHIGKYN 3CFTAKYHIGKYN 3CFTAKYHIGKYN 3CFTAKYHIGKYN 3CFTAKYHIGKYN 3CFTAKYHIGKYN 3CFTAKYHIGKYN 3CFTAKYHIGKYN 3CFTAKYHIGKYN 3CFTAKYHIGKYN 3CFTAKYHIGKYN 3CFTAKYHIGKYN 3CFTAKYHIGKYN 3CFTAKYHIGKYN 3CFTAKYHIGKYN 3CFTAKYHICKYN 3CFTAKYHICKYN 3CFTAKYHICKYN 3CFTAKYHICKYN 3CFTAKYHICKYN 3CFTAKYHICKYN 3CFTAKYHICKYN | KENGCHEREN AN AGEN AN | KEKE K KGKAV I BQVENAKQAPH VKLEKLIKSLEPACTIVAK VNQVI SKKDVI E SIHETIIT FHPI QAAVSELEMTI VP DEKMSNF LPLKP QAKTLKNSETTKAVEKNIUS FHKSL SQRISELEMI VP DEKMSNF LPLKLE QAKTLKNSETTKAVEKNIUS FHKSL SQRISELEMI VP DEKMSNF LPLKLE QAKTLKDG STABAVEKNIUS FHKSL SQRISELEMI VP DEKMSNF LPLKLE QAKTLKDG STABAVEKNIUS FHKSL SQRISELEMI VP DEKMSNF LFLKLELSA QTKTLKDG STABAVEKNIUS FHKSL SQRISELEN LAVIA QDKPLKNATLDI I EQNI IS FHQSL SKQL SELEL SL FP SILMTELKLELSA QTKPLP-KKHQAEVERTINNFYKLISDK HQLGCQLA PAEQUTNELTVTAA NNFPL-KKHQAEVERTINNFYKLISDK ITLECQLA PAEQUTNELTVTAA VABAI PESAQAEVNHTQAEFHKKLVKTLRERFEISVTVSTESAPFA VABAI PESAQAEVNHTQAEFHKKLVKTLRERFEISVTVSTESAPFA ULAI PKNALEEVNKTQNCFHLKLVKTLEENFQLSVAYHEQSVEFV DDLAI PKNALEEVNKTQNCFHLKLVKTLEENFQLSVAYHEQSVEFV DPLPLPKKAIAEVSRTQEAFHQKLVDELQKKFEIGVQLDEESVEFV DDLPLPKKAIAEVSRTQEAFHQKLVDELQKKFEIGVQLDEESVEFV DSLPLPKKAVAEVTRTQEAFHQKLVDELQKKFEISULDESSVEFV DSLPLPKKAVAEVTRTQEAFHQKLVDELQKKFEISULDESSVEFV VSEALPKKVVEEVYTQSTFHQKLIDTLAEHFKLEVEYHTESFDFA TSEPLPAKVTEEVYLTQSTFHQKLIDTVAEHFKLEVEYHTESFDFA TSEPLPAKVTEEVYLTQSTFHQKLIDTVAEHFKLEVEYHTESFDFA TSEPLPAKVTEEVYLTQSTFHQKLTDTVAEHFKLEVEYHTESFDFA TSEPLPAKVTEEVYLTQSTFHQKLTDTVAEHFKLEVEYHTESFDFA TSEPLPAKVTEEVYLTQSTFHQKLTDTVAEHFKLEVEYHTESFDFA |
| 37 38 39 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 C(| MLPSGIPKSRLM MLSSLHPTRGKLL MLKRDLPKKEILNMALPTDYNGKPL MLKRDLPKKEILNMALPTDYNGKPL MLKRDLPKKEILNMALPTDYNGKPL MSENTYKRLLP MSENTYKRLLP MSENTYKRLLP MSENTKPTHYRLLS MSENTKPTHYRLLS MSENTKPTHYRLLS MSENTKPTHYRLLS MSEVTQPTHYRLLN MSEVTQPTHYRLLN MSEVTQPTHYRLLN MSETTQPTHFRLMS MSETTQFTHFLMS MSETTQFTHF | NFTALGE RFAQIES KFAAMGE KFAAMGE KFAAMGE KLRAMGE KLRAMGE VLKAIGE VLKAIGE VLKAIGE KLRAMGE KLRAMGE KLRAMGE KLRAMGE KLRAIGE KLR | VIREOCCUSEOF VIREOCCUSEOF VIREOCSUNCY VIREOCSUNCY VIREOCSUSEOF VIREOCSERFY VIREOCSERFY VIREOCSERFY VIREOCSERFY VIREOCSERFY VIREOSECHY | FDCLAX FFDSLVF FFDSLVF FFDSLVG FFDSLTV FFDSLTX RFDCLSX LFDCLAS LFDCLAS LFDCLAS LFDCLAS LFDCLSX IFDCLSY IFDCLSY YFDCLSY YFDCLSY YFDCLSY YFDCLSY YFDCLSY YFDCLSY FFDCLSY | | | CAPSYEYHYQCLFDJ 20CIIYYYQCLFDJ 20CFFAYHLCIYT 20GFEAYHLCIYT 20GFEAYHLCIYT 20GFEAYHQVGFYN 28FSYHYQVGFYN 28FSYHYQVGFYN 20FSACYHICRYT 20FSACYHICRYT 20FSACYHICRYT 30GFTANYHICKYD 30GFTANYHICKYD 30GFTANYHICKYN 30GFTAKYHICKYN 30GFTAKYHICKYN 30GFTAKYHICKYN 30GFTAKYHICKYN 30GFTAKYHICKYN 30GFTAKYHICKYN 30GFTAKYHICKYN 30GFTAKYHICKYN 30GFTAKYHICKYN 30GFTAKYHICKYN 30GFTAKYHICKYN | KENGUGTEN KANGANANANANANANANANANANANANANANANANANAN | KEEKI KUKAN UI BOUENAKQAPH VKLEKLIKSLEPACTUVAK VNQVI SKKAVI BOUENAKQAPH VKLEKLIKSLEPACTUVAK ZITTLKLOSTABAVEKNILYSFHKSLSQRISELEMI VYDEKMSNFLPLKLE AUKPLEPACKENIHAFHKSLSQRISELEMI VYDEKMSNFFLPLKLE AUKPLEPACKENIHAFHKSSQLISELESI FPSPLMETELKLELSA DVKPLKDNATLDI I EQNIISFHQSI SQLISELESI FPSPLMETELKLELSA DVKPLKDNATLDI I EQNIISFHQSI SQLISELESI FPSPLMETELKLELSA DVKPLKDNATLDI I EQNIISFHQSI SQLISELESI FPSPLMETELKLELSA DVKPLKDNATLDI I EQNIISFHQSI SQLISELESI FPSPLMETELKLELSA VARAI PESAQAEVNHTQAFFHKKLVKTIRERFEISVTVSTESAPFA VARAI PESAQAEVNHTQAFFHKKLVKTIRERFEISVTVSTESAPFA VARAI PESAQAEVNHTQAFFHKKLVKTIRERFEISVAVSTESAPFA LDLAI PKNALEEVNKTQNCFHLKLVKTIENFPGISVAVHEGSVEFV DPLPLPKKAIAEVSRTQEAFHQKLVDELGKKFEIGVQLDESVEFV DDLPLPKKAIAEVSRTQEAFHQKLVDELGKKFEIGVQLDESVEFV DDLPLPKKAIAEVSRTQEAFHQKLVDELGKKFEIGVQLDESVEFV VLESAPESATQEITRTQEVHEKLVSALGEKFEISDVAIHDSVEFV VSEALPQKVVE EVYTTQSTFHQKLIDTIAEHFKLEVEYHTESFDFA TSEPLPAKVTEVYLTQSTFHQKLIDTVAEHFKLEVEYHTESFDFA TSEPLPAKVTEVYLTQSTFHQKLIDTVAEHFKLEVEYHTESFDFA TSEPLPAKVTEVYLTQSTFHQKLIDTVAEHFKLEVEYHTESFDFA TSEPLPAKVTEVYLTQSTFHQKLIDTVAEHFKLEVEYHTESFDFA |
| 37 38 39 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 | MLPSGIPKSRIM MLSSLIPTRGKLL MLKRDLPKKEILMAAPTDYSNGKVL MLKRDLPKKEILNAAPTDYSNGKVL MLKRDLPKKEILNAAPTDYSNGKVL MSENTYTYRLL MSENTYTYRLL MSENTYTYRLL MSENTYTYRLL MSENTKPTHYRLL MSENTKPTHYRLL MSEVTQPTHYRLL MSEVTQPTHYRLL MSEVTQPTHYRLL MSEVTQPTHYRLL MSETTQGPTHFRLM | NFTALGE KFAALGE KFAALGE KFAALGE QFAALGE QFAALGE QFAALGE QFAALGE VILALG | YIREOCCUSSOF YIREOCCUSSOF YIREOCUSSOF YIREOCUSSOF YIREOCUSEGOR YIREOCCEGOR YIREOCCEGOR YIREOCCEGOR YIREOCCEGOR YIREOCCEGOR YIREOCCEGOR YIREOCCEGOR YIREOCSEGOR YIR | FPDCLAY FFDSLVY FFDSLVY FDSLTY FFDSLTY FFDSLTY FFDSLTY FFDCLSY FFDCLSY FFDCLSY FFDCLSY FFDCLSY FFDCLSY FFDCLSY FFDCLSY FFDCLSY FFDCLSY FFDCLSY FFDCLSY FFDCLSY FFDCLSY FFDCLSY | CVNKKVTF CVNKKVTF CVNANUA CVNANUA CVNVNSAF CVSAKPAF SCVNDKKSF SCVNDKKSF CIDDRKSF CUDDKKSF CVNDKKSF CVNDKKSF CVNDKKSF CVNDKKSF CVDDKKF CVDDKKF CVDDKKF CVDDKKF CVDDKKF CVDDKKF CVDDKKF CVDDKKF CVDDKKF CVDDKKF CVDDKKF CVDDKKF CVDDKKF CVDDKKF CVDDKKF CVDDKF CVD | EEREPWGWWELLDKSE EKREPWGWWELLZVSTH EKREPWGWWELLVSTH EKREPWGWWELLVSTH EKREPWGWWELLSPNE DKREPWGWWELLEQOD EKREPWGWWELLQOD EKREPWGWWELTQNE EKREPWGWWMELTQNE EEREPWGWWMELTQNE EEREPWGWWMELTQNE EEREPWGWWMELTQNE EEREPWGWWMELEATE EEREPWGWWMELEATE EEREPWGWWMELEATE EEREPWGWWMELSIE EEREPWGWWMELSIE EKREPWGWWMELSIE EKREPWGWWMELSIE EKREPWGWWMELSIE EKREPWGWWMELSIE EKREPWGWWMELSIE EKREPWGWWMELSIE | CAPSYEYHYGLFDJ CAPSYEYHYGLFDJ CGUIYYYGYGLFDJ CGPEYXHLCHYM CGPEYYHLCHYM CGPEYYHLCHYM CGPEYYHLCHYM CGPEYYHLCHYM CGPEYYHLCHYM CGPEYYHLCHYM CGPEYYHCYGYM CGPEYHCHCWY CGPAXYHICKYM CGPAXYHICKYM CGFAXYHICKYM CGFAXYHICKYM CGFAXYHICKYM CGFAXYHICKYM CGFAXYHICKYM CGFEFERFDCLFD CG H CG CG CG CG CG CG CG CG CG CG | | KEKKIKGKAVIEQUENAKQAPH VKLEKLIKSLEPACTIVAK VNQVISKKDVIESIHETIIRFHPIQAAVSELEMTIVPDEKKSNFLPLKP 2AKTLKNSETTKAVEKNIKAFHQKISQLISELEMTIVPDEKKSNFLPLKLELSA 2TKTLKDGSTABAVEKNIKAFHQKISQLISALELSIFPSPLMTELKLELSA 2DKPLKDNATLDIIEQNIISFHQSISKQLSELELSIFPSPLMTELKLELSA 2DKPLKDNATLDIIEQNIISFHQSISKQLSELELSIFPSPLMTELKLELSA 2DKPLP-KKHQAEVERTINHFYKLISDKITLECCLAPARQVTNELTVTAA 1NKPLP-KKHQAEVERTINHFYKLISDKITLECCLAPARQVTNELTVTAA 2000 AND |





















