The Rv3799–Rv3807 gene cluster in *Mycobacterium tuberculosis* genome corresponds to the 'Ancient Conserved Region' in CMN mycolyltransferases

Hemalatha G. Ramulu, Adindla Swathi and Lalitha Guruprasad

School of Chemistry, University of Hyderabad, Hyderabad 500 046, India.

Abstract: We have identified based on gene cluster analysis that the genes between Rv3799–Rv3807 in *M. tuberculosis* have orthologs in Corynebacteria, Mycobacteria and Nocardia (CMN) genomes. Therefore, this gene cluster possibly corresponds to the 'Ancient Conserved Region' of CMN mycolyltransferases. The evolutionary trace analysis suggests that twelve amino acid residues; Leu39, Trp51, Pro71, Trp82, Trp97, Phe100, Gly124, Ser126, Asp192, Glu230, Gly260 and Trp264 are 'absolutely conserved'. These amino acid residues constitute the active site and conserved hydrophobic tunnel in CMN mycolyltransferases.

Keywords: Corynebacteria; Mycobacteria; Nocardia; Mycolyltransferases; Gene cluster, Evolutionary trace analysis.

Introduction

The organisms belonging to the Corynebacterium, Mycobacterium and Nocardia (or 'CMN') genera have been grouped together based on factors that distinguish them individually. These include; complex cell wall components, presence and type of mycolic acids, adjuvant activity, presence of cord factor, sulfolipids, iron-chelating compounds, polyphosphate, and serological cross-reactivity. The cell wall of the CMN group organisms consist interconnected peptidoglycan and polysaccharide-mycolate complex and are characterized by the presence of mycolic acid on their surface (Cocito and Delville 1985). Mycolic acids are long chain fatty acids that form a part of the unique cell envelope responsible for the pathogenesis and the survival of the organism inside the host. These mycolic acids are identified by different names depending upon the genus and comprise different carbon chain lengths; corynomycolic acids from the genus Corynebacterium has 22–36 long carbon chain, mycolic acids from the genus Mycobacterium has nearly 60–90 long carbon chain and nocardomycolic acids from the genus Nocardia has 40–60 long carbon chain (Collins et al 1982; Minnikin 1982; Daffé M and Draper 1992).

In *M. tuberculosis*, the antigen 85 complex enzymes constitute three secreted proteins (Wiker and Harboe 1992); Ag85A (gene identifier: Rv3804), Ag85B (Rv1886) and Ag85C (Rv0129) that comprise a signal peptide at the N-terminus followed by a carboxylesterase domain. These are known to catalyse the transfer of mycolic acids to the α, α' - trehalosemonomycolate (TMM) and arabionogalactan. It has been demonstrated that Ag85 complex enzymes catalyse the transfer of mycolyl residue from one molecule of TMM to another TMM leading to the formation of α, α' - trehalosedimycolate (TDM) and hence these enzymes are termed mycolyltransferases (Belisle et al 1997). Also, in Corynebacterium and Nocardia, orthologous proteins synthesize trehalosedicorynomycolate (TDCM) and trehalosedino-cardiomycolate (TDNM), respectively. Mycolyltransferases are also termed fibronectin binding proteins, since they are involved in binding fibronectin that aids entry of the organism into host cells (Abou-Zeid et al 1988; Ratliff et al 1988). The compound TDM commonly known as "cord factor" extracted from *M. tuberculosis* has been shown to be toxic in mouse model (Kato 1968). Hence a study related to the analysis of the structure, function and evolution of proteins responsible for the synthesis of TDM in these species is important.

The three dimensional structures of Ag85A (PDB code: 1SFR; Ronning et al 2004), Ag85B (1F0P; Anderson et al 2001) and Ag85C (1DQZ, 1VA5; Ronning et al 2000) are known for both native and substrate bound forms and comprises a α/β hydrolase fold. The catalytic triad constituting S126, G230 and H262 (numbering is according to PDB code: 1F0P) is responsible for the mycolyltransferase activity. The structural comparison

Correspondence: Lalitha Guruprasad, School of Chemistry, University of Hyderabad, Hyderabad 500 046, India. Tel: + 91 40 2313 4820; Fax: + 91 40 2301 2460; Email: lgpsc@uohyd.ernet.in

of these mycolyltransferases revealed that the active sites are virtually identical indicating that they share the same substrate. However, in contrast to the high level conservation within the substrate-binding and active site, it was observed that surface residues disparate from the active site are guite variable indicating that all three Ag85 enzymes are needed to evade the host immune system (Ronning et al 2004). The multiple sequence alignment (see Appendix-A in supplementary data) suggests that the three sequences corresponding to Ag85A, Ag85B and Ag85C share more than 69% sequence identity. In our previous work (Adindla et al 2004a), we identified mycolyltransferases in the C. glutamicum and C. efficiens genome and analyzed the threedimensional computer models that were constructed based on comparative modeling methods. The mycolyltransferases are restricted to the CMN genera and the complete genome sequences of M. tuberculosis (Cole et al 1998), C. glutamicum (Kalinowski et al 2003), C. efficiens (Kawarabayasi et al 2002), C. diphtheria (Cerdeno-Tarraga et al 2003) and Nocardia farcinica (Ishikawa et al 2004) are now available. Therefore, we intended to identify and analyse all the mycolyltransferases from various species in order to get an insight into their substrate specificity. Since there are several isoforms in each genome, we intended to understand the origin of their evolution and therefore carried out the evolutionary trace analysis. Further, synteny or colinearity of gene order is observed when a group of genes are present in the same order in two or more genomes as a cluster. Two species that have recently diverged from a common ancestor might be expected to share a similar set of genes present in the same order. During evolution the sequence of each pair of genes is accompanied by changes, such as, gene duplication and gene loss. Genetic analyses reveal that genes with related function are frequently clustered at one chromosomal location in evolution. Despite M. tuberculosis being an ancient species compared to Corynebacterium and Nocardia, evolution has maintained the conservation of proteins involved in the synthesis of the cell envelope. Also, it is interesting to note that there are varying numbers of mycolyltransferases in these different species. We therefore intended to identify the evolutionary origin of CMN mycolyltransferases based on the computational analysis of 'gene neighborhood' and to identify the role of conserved amino acid residues using the evolutionary trace analysis.

Materials and methods

The amino acid sequences corresponding to the mycolyltransferases; Ag85A, Ag85B and Ag85C were obtained from the website at www.srs.ebi.ac.uk/. The homologous proteins from C. glutamicum, C. efficiens, C. diptheria and N. farcinica completed genome database were identified using the BLASTP and PSI-BLAST (Altschul et al 1990; 1997) programs available at the website www.ncbi.nlm.nih.gov/ BLAST/ and by using the sequence corresponding to Ag85B as query. The blosum62 matrices were used and the results were sorted based on p-value. The analysis of gene clusters was carried out by performing BLAST searches using mycolyltransferases and their neighbouring proteins as query on all the finished and unfinished genomes. The evolutionary trace (ET) analysis was carried out using TraceSuite II server (Innis et al., 2000) available at the website http://www.cryst.bioc.cam.ac.uk/~jiye/evoltrace/ evoltrace.html by submitting the sequences corresponding to the carboxylesterase domain of CMN mycolyltransferases and the crystal structure of Ag85B (PDB code: 1F0P). A trace is generated by comparing the consensus sequences for groups of proteins that originate from a common node in a phylogenetic tree and is characterized by a common evolutionary time cut-off (ETC).

Results and discussion

Identification of CMN mycolyltransferases

We identified 32 mycolyltransferases in the genera of CMN group; 4 proteins in M. tuberculosis, 4 in C. diphtheria, 6 in C. glutamicum, 5 in C. efficiens, and 13 in N. farcinica. The four mycolyltransferases corresponding to each of the mycobacterial species; *M. tuberculosis*, *M. leprae* and *M. bovis* are highly similar, therefore only the mycolyltransferases from *M. tuberculosis* is referred to in the discussion. There are between 350-480 amino acid residues in these proteins. However, proteins corresponding to gene identifiers; Ncgl2777, Ce2709 and Dip2193 (in Corvnebacterium) and Nfa1840 (in N. farcinica) are associated with an additional ~300 amino acid residue domain towards the C-terminus that is not a part of the carboxylesterase domain. The carboxylesterase domain that is responsible for the mycolyltransferase activity, corresponds to approximately 280 amino acid residues. The multiple sequence alignment is attached as



Figure 1. Schematic representation of genes corresponding to the 'Ancient Conserved Regions' in five completed genomes based on gene neighbourhood analysis. '/' indicates insertion of gene. Nfa1810-30 indicates three genes Nfa1810, Nfa1820 and Nfa1830.

supplementary data along with this manuscript in Appendix-A. The *N. farcinica* proteins; Nfa1820 and Nfa1810, comprise a 'long insertion' sequence of 22 and 27 amino acid residues that is rich in glycine and serine respectively relative to the other CMN proteins. This region is located between the 'absolutely conserved' W82 and W97 residues. The glycine/serine rich sequences are often known to be associated with cell-surface proteins. Another insertion region in some corynomycolyltransferases and nocardiomycolyl transferases relative to mycolyltransferases is located between the 'absolutely conserved' D192 and E230 amino acid residues. We predict this loop to be close to the substrate binding site based on a comparison with the crystal structure of Ag85B.

Gene cluster analysis

The analysis of all mycolyltransferases and their neighbouring proteins revealed that genes between Rv3799–Rv3807 in the *M. tuberculosis* genome has corresponding orthologs in Corynebacterium and Nocardia genera as shown in Figure 1. The ten protein orthologs shown in Figure 1 share high sequence similarity in the five different species analyzed. In addition to mycolyltransferase (Rv3804) and its precursor protein (Rv3803) this cluster also comprises propionyl CoA carboxylase (Rv3799), polyketide synthase (Rv3800), acyl CoA synthase (Rv3801), membrane proteins (Rv3806, Rv3807), and hypothetical proteins (Rv3802, Rv3805). We

observed that the Nocardia proteins are arranged in the reverse order relative to the other species. We report that this set of genes represents the only mycolyltransferase comprising gene cluster during divergence of a common ancestoral organism into individual genera, such as, Corynebacterium, Mycobacterium and Nocardia (CMN group). Therefore, we propose that this gene cluster corresponds to the "Ancient Conserved Regions -ACR's" among the mycolyltransferases across the CMN genera. It was reported that Rv3800 (*pks13*) is involved in the final condensation step in mycolic acid synthesis (Damien et al 2004). It was also reported that the genes; Rv3799, Rv3800 and Rv3801 (accD4-pks13-fadD32) play an essential role in the biosynthesis of mycolic acids (Gande et al 2004). This indicates that the proteins in this cluster are important for the mycolic acid synthesis and its transfer to trehalose. Since, functionally related genes are often clustered, we suggest that the other "uncharacterized" proteins belonging to the ACR gene cluster may also have a role in associated functions. Further, we observed that the mycolyltransferase gene neighbours; Rv0129 and Rv1886 are conserved among *M. tuberculosis* and *M. bovis* suggesting that gene duplication events have occurred before speciation.

Evolutionary Trace Analysis

The TraceSuite II server generates a phylogenetic tree split into 10 evenly distributed partitions (P01–P10)



Figure 2(a). TraceSuite II analysis representing partition based on evolutionary time cut off. (b) 'absolutely conserved' residues corresponding to P01 partition.

in the order of increasing evolutionary time cut-off (ETC) as shown in Figure 2a. The conserved amino acid residues associated with each partition is shown in Figure 2b. Analysis of amino acid residues corresponding to P01 partition (see Figure 2b) revealed that 12 amino acid residues are "absolutely conserved". By examining the equivalent residues in the crystal structure of the protein (PDB code: 1F0P), we infer that the residues; L39, P71, W82, W97 and F100 constitute the 'hydrophobic tunnel' as shown in Figure 3a. This figure also indicates the amino acid residues involved in the catalytic triad. The residues in the 'hydrophobic tunnel' are needed in order to accommodate the alkyl chain of mycolic acid indicating a functional conservation in these proteins. The invariant S126 and G260 are close to the catalytic active site comprising the amino acid residue E230. The indole side chains of the W51 and W264 are perpendicular to each other and are in proximity to G124 associated with the β 5 strand. The amino acid residue D192 is away from the active site





110

01203405070910

110

0122345567

08 09 10

1100

012345557891

HOp

023455678910

110

110

0123455678910

110

012034056768910

indicating that the conservation extends beyond the catalytic site in CMN mycolyltransferases. According to Figure 2a, the 14 proteins indicated in the lower half, from Corynebacterium, Mycobacterium, and Nocardia represent the 'Ancient Conserved Region'. The 18 proteins in the upper half, comprise only the Nocardia and Corynebacterium. From the multiple sequence



Figure 3(a). Stereo-view showing three-dimensional model corresponding to the protein with gene identity Nfa1840. The amino acid residues comprising the catalytic triad (pink), hydrophobic tunnel (blue) and trehalose (red) are also indicated.

alignment, we observed that the proteins in the upper half of Figure 2a are associated with an insertion loop of variable length between 4 to 20 amino acid residues and this loop is close to the active site. The positions of these insertion loops are shown in Figure 3b. Further, the amino acid residues comprising the specificity pockets defined by interactions with trehalose substrate in the protein with PDB code: 1F0P are mutated in these proteins. Primarily, the mutations associated with the substrate binding sites in some Corynebacterium (Adindla et al., 2004a) and Nocardia proteins accompanied by the presence of 'insertion loops' close to the active site suggest that these may interfere with trehalose binding. These Corynebacterium and Nocardia proteins are possibly a result of divergent evolution accompanied by gene duplication and mutation events in order to accommodate different substrates in the binding site. This suggests that the ancient proteins form a distinct cluster and are different from proteins that evolved later.

We previously reported that the corynemy colyltransferase Ncgl2777 gene in *C. glutamicum* (protein with ~300 amino acid residue C-terminal extension) is associated with a 55 amino acid residue 'LGFP' tandem repeat that is likely to be associated with maintaining cell-wall integrity (Adindla et al 2004b). Our hypothesis was based on the work of Brand et al., 2003 who have demonstrated that the deletion of Ncgl2777 gene in *C. glutamicum* resulted in a 10-fold increase in cell volume of the organism thereby suggesting its involvement in cell

shape formation. In this work, we observed that the 'LGFP' tandem repeats are also present in the C-terminal region of Nocardia (NfaA1840) and *C. diphtheria* (Dip2193) proteins and accordingly may be involved in maintaining cell wall integrity.



Figure 3(b). Structural overlay corresponding to the proteins with PDB code: 1F0P and gene identities: Nfa1810, NcgI0336 and NcgI0987 indicating the position of the two loops; loop 1 (D192 – E230 loop) and loop 2 (W82-W97 loop).

Conclusions

The comparative analysis of mycolyltransferase proteins from different genomes suggested that the gene cluster corresponding to the ten gene families located between Rv3799 and Rv3807 in *Mycobacterium tuberculosis* genome represents the 'Ancient Conserved Region' in CMN genera. According to the evolutionary trace analysis twelve amino acid residues are 'absolutely conserved' in all CMN proteins analyzed. These CMN proteins fall into two distinct clusters in the phylogeny that correlates with the presence or absence of insertion loop close to the active site. Some Corynebacterium and Nocardia proteins with extra C-terminal 300 amino acid residues are associated with the LGFP tandem repeats.

Acknowledgments

HGR thanks UGC, New Delhi for a JRF fellowship. SA thanks CSIR New Delhi for a SRF fellowship. LGP thanks DBT, New Delhi for research funding.

References

- Abou-Zeid C, Ratliff TL, Wiker HG, et al. 1988. Characterisation of fibronectin-binding antigens released by *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG. *Infect Immun*, 56:3046–3051.
- Adindla S, Guruprasad K, Guruprasad L. 2004a. Three-dimensional models and structure analysis of corynemycolyltransferases in *Corynebacterium glutamicum* and *Corynebacterium efficiens*. Int J Biol Macromol, 34:181–189.
- Adindla S, Inampudi KK, Guruprasad K, et al. 2004b. Identification and analysis of novel repeats in the cell surface proteins of archaeal and bacterial genomes using computational tools. *Comp Funct Genom*, 5:2–16.
- Altschul SF, Gish W, Miller W, et al. 1990. Basic local alignment search tool. J Mol Biol, 215:403–410.
- Altschul SF, Madden TL, Schäffer AA, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25:3389–3402.
- Anderson DH, Harth G, Horwtiz MA, et al. 2001. An interfacial mechanism and a class of inhibitors inferred from two crystal structures of the *Mycobacterium tuberculosis* 30 kDa Major secretory protein (antigen 85B), a mycolyl transferase. J. Mol. Biol, 307:671– 681.
- Belisle JT, Vissa VD, Sievert T, et al. 1997. Role of the major antigen of Mycobacterium tuberculosis in the cell wall biogenesis. Science, 276:1420–1422.

- Brand S, Niehaus K, Puhler A, et al. 2003. Identification and functional analysis of six mycolyltransferase genes of Corynebacterium glutamicum ATCC 13032: the genes cop1, cmt1, and cmt2 can replace each other in the synthesis of trehalose dicorynomycolate, a component of the mycolic acid layer of the cell envelope. *Arch. Microbiol*, 180:33–44.
- Cocito A, Delville J. 1985. Biological, chemical, immunological and staining properties of bacteria isolated from tissues of leprosy patients. *Eur J. Epidemiol*, 1:202–231.
- Cole ST, Brosch R, Parkhill J, et al. 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*, 393(6685):537–544.
- Cerdeno-Tarraga AM, Efstratiou A, Dover LG, et al. 2003. The complete genome sequence and analysis of Corynebacterium diphtheriae NCTC13129. *Nucleic Acids Res*, 31(22):6516–6523.
- Collins MD, Goodfellow M, Minnikin DE. 1982. A survey of the structures of mycolic acids in Corynebacterium and related taxa. *J Gen. Microbiol*, 128:129–149.
- Daffé M, Draper P. 1998. The envelope layers of mycobacteria with reference to their pathogenicity. *Adv Microb Phys*, 39:131–203.
- Damien P, Célia de Sousa-D'Auria, Christine H, et al. 2004. A polyketide synthase catalyzes the last condensation step of mycolic acid biosynthesis in mycobacteria and related organisms. *Proc Natl Acad Sci* USA, 101:314–319.
- Gande R, Gibson KJC, Brown AK, et al. 2004. Acyl-CoA Carboxylases (accD2 and accD3), Together with a Unique Polyketide Synthase (Cg-pks), Are Key to Mycolic Acid Biosynthesis in Corynebacterianeae Such as Corynebacterium glutamicum and Mycobacterium tuberculosis. J. Biol. Chem, 279:44847–44857.
- Innis CA, Shi J, Blundell TL. 2000. Evolutionary trace analysis of TGF- β and related growth factors: implications for site-directed mutagenesis. *Prot. Engin.* 13:839–847.
- Ishikawa J, Yamashita A, Mikami Y, et al. 2004. The complete genomic sequence of *Nocardia farcinica* IFM 10152. *Proc Natl Acad Sci* USA, 101:14925–14930.
- Kalinowski J, Bathe B, Bartels D, et al. 2003. The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *J Biotechnol*, 104(1-3):5-25.
- Kawarabayasi Y, Yamazaki J, Hino Y, et al. 2002. The entire genomic sequence of *Corynebacterium efficiens* YS-314. Unpublished.
- Minnikin DE. 1982. The Biology of the Mycobacteria. In: Ratledge C, Stanford JL, eds. London: Academic Press, p 95–184.
- Ratliff TL, McGarr JA, Abou-Zeid C, et al. 1988. Attachment of mycobacteria to fibronectin-coated surfaces. J. Gen. Microbiol,134:1307-1313.
- Ronning DR, Klabunde T, Besra GS, et al. 2000. Crystal structure of the secreted form of antigen 85C reveals potential targets mycobacterial drugs and vaccines. *Nat. Struc. Biol*, 7(2):141–146.
- Ronning DR, Vissa V, Gurdyal B, et al. 2004. *Mycobacterium tuberculosis* antigen 85A and 85C structures confirm binding orientation and conserved substrate specificity. *J. Biol. Chem*, 279:36771–36777.
- Wiker HG, Harboe M. 1992. The antigen 85 complex: a major secretion product of *Mycobacterium tuberculosis*. *Microbiol. Rev*, 56:648–661.

Appendix-A

Multiple sequence alignment corresponding to carboxylesterase domain in CMN mycolyltransferases.

cmytD	LNGLRLEKWSVASP	SMQRNVDVQI	MKSAEADSPAI	PMLYMLDGIGGN	KNSSGWINGG	EGPKVFADENV	TVV
Ce1488	MDGLRLERWTVASP	SMQRNVDVQI	MRSVDAGAPAI	PMLYMLDGIGGN	RNSSGWINHG	QGPKVFGDENV	TVV
CmytE	VDGDRIRQINAYSP	SMGRTIPLVW	VVPEDNTVPG	PTVYALGGGDGG	QGGQNWVTRT	DLDELTSENNI	NLI
cmytF	VDGDRIRQINAYSP	SMGRTIPLVW	VVPEDNTVPGI	PTVYALGGGDGG	QGGQNWVTRT	DLEELTSDNNI	NLI
Ce0984	VDGERIRQINAYSP	SMERWIPLVW	IVPEDTSEPRI	PTLYALGGGDGG	QGSANWITKT	DMPELMSSNNV	HVI
cmytC	AADERVKEMWAYSP	SMDRNVPLVV	ITADESAGPRI	PVIYLLNGGDGG	EGAANWVMQT	DVLDFYLEKNV	NVV
Ce0356	ASGERVKEMWAYSP	SMDRDVPLVV	ITADESAGPRI	PVIYLLNGGDGG	EGNANWIMQT	DVIDFYLEKNV	NVV
Dip0365	ATGDRVVEMWAHSP	SMNRNVPLVV	LKAANPGRI	PTIYLLNGGDGG	EGSANWVMQT	KALDFYRDKDV	NVV
Nfa38260	VVDARTVRLRVYSA	AMGRVIDIDV	ORPADTGAPRI	PTLYLLAGAGGG	EDSASWAKOT	SVLEFLADKNV	NVV
Nfa7210	IKDDRNLRLYVYSA	AMDENVIIDV	~ ORPADASVPRI	PTLYLLNGAGGG	EDDASWVAKS	DALKFLSDKNV	NVI
Nfa32420	AKEGRTWHLTVYSA	AMDTEIAVDV	~ ORPADDSVPAI	PNLYMLNGLDGG	EGTASWAAAT	HALDWLADKPV	NVI
Nfa23770	GTPARLVDLAVYSP	AMORSTAVKV	~ LRPADTTRPAI	PTLYLINGAGGG	EDAANWEGOT	DAVEFFADKHV	NVV
Nfa30260	PRSDREVEVIVHSA	AMAAEIPIRL	LRAADPDRPAI	PTLYLLNGITGG	GDGGNWFDRT	GVAAFFAGEOV	NVA
Nfa43800	PENDRIJDIETHSP	AMDSTTRVI	TRAPDPDRPA	PTLYLLNGASGH	VDG-SWHDRT	DYORFFADKOV	NVV
Nfa45560	PLGGROLEVVVHSA	AMNRPITLWM	SHPGPGAI	PALYLLNAVDGG	EDGGPWMNRT	DVAAFFADKNV	NVI
Nfa25110	PLAPRVDOVOVYSP	SMDAVVSSTV	TRADGPAI	PTIYLLAGAGGG	TDGISWWHHT	DVROFFADKNV	NVV
Nfa5610	ELSPTRSAVEVDSP	AMGRVTOVOV	LHP-AGGAARI	PSYYLLDGLDPG	VGOSTWTNAT	DAEAFFRGKNV	NVV
Dip2339	DEREDVDRIFTESP	AMERIVOVOV	OHPKDRTTPAI	PMI.YI.I.DGVTAP	-SOSGWLRKG	DVOGAMANEHV	TVT
CmvtA	TLTTOSAAMPER	PIKVOLLPR	DWYSSPNREFI	PETWALDGURAT	EEOSGWTIET	NTEOYYADKNA	TVV
Ce2709	VLSTOSAAMPER	PIKVOLLPR	DWYSSPDRDFI	PETWALDGLRAT	EKOSGWTIET	NTEOFFADKNA	TVV
Din2193			WIGGIBICH	SVWALDGLEATI	VENGWTIGT	NTEOFYSDKNV	NIVT
Nfa1840			NAKDEVDEI NVKDEVDEI		DDFSCWTKDA	CVEEEEVDKWA UIFÖLISDUUA	TV V I
Co2710	CVCVWVOPCDVV	Q V Q LL LA KDW	OTOPAORCENI	ACI VI I DOMRAT		NAAAT VADUNT	
Cez / IU			QIQFAQAGGNZ	AGLILLDGMAAI		NAAALIAFINI NAADIVADNNI	
Dip2104			QIQFAGRGGNZ	AGLILLDGMAAI	EISNAWLVDI ETNNAWLVDI	NAARD I EVDUNT	
DIP2194		SPAMGRNIIV	UIUPAUKGGNA	AALILLUGARAN		UND Y T Y OMNIT	
NIA1820	AF DPAAF DF WVD	SGMGPIKSRI.	LRA-ADGNINI	RVVIVLDGMRAP	ETLNGWEIET	DVPALLASWNI	IN V V
NIALOIU		SUMGPIKSKI	FRA-ADGNIN	RVVIALDGMRAR	NDLSGWEIDT	EVARELIKWNI	
NIA1830	APAGGYEELMVP	SVMGPIKVQV	QWA-SRG-GDA	AALYLLDGLRAR	DDRNAWSFET	NAMEQFKNDNI	TLV
RV3804C	RPGLPVEYLQVP	SPSMGRDIKV	QFQ-SGGANSI	PALYLLDGLRAQ	DDFSGWDINT	PAFEWIDQSGL	SVV
AGUIL	RPGLPVEYLQVP	SPSMGRDIKV	() E () = S(-(-N)NS)	PAVYLLDGLRAO	(1) (1) (2) (1) (1) (1) (1) (1) (1) (1) (1) (1) (1	PAFEWYYOSGL	/
5 0100					DDINGWDINI		SIV
Rv0129c	RPGLPVEYLQVP	SASMGRDIKV	QFQ-GGGPI	AVYLLDGLRAQ	DDYNGWDINT	PAFEEYYQSGL	SIV
Rv0129c Rv3803c	RPGLPVEYLQVP AKAAPYENLMVP	SASMGRDIKV SPSMGRDIPV	QFQ-GGGPI AFL-AGGPI	HAVYLLDGLRAQ HAVYLLDAFNAG	DDINGWDINI DDYNGWDINT PDVSNWVTAG	PAFEEYYQSGL NAMNTLAGKGI	SIV SVI SVV
Rv0129c Rv3803c	RPGLPVEYLQVP AKAAPYENLMVP	SASMGRDIKV SPSMGRDIPV	QFQ-GGGPI AFL-AGGPI	AVYLLDGLRAQ HAVYLLDAFNAG *	DDINGWDINI DDYNGWDINT PDVSNWVTAG *	PAFEEYYQSGL NAMNTLAGKGI	SIV SVI SVV
Rv0129c Rv3803c	RPGLPVEYLQVP AKAAPYENLMVP	SASMGRDIKV SPSMGRDIPV.	QFQ-GGGPI AFL-AGGPI	AVYLLDGLRAQ HAVYLLDAFNAG * 	DDINGWDINI DDYNGWDINI PDVSNWVTAG *	PAFEEYYQSGL NAMNTLAGKGI	SVI SVV
Rv0129c Rv3803c	RPGLPVEYLQVP AKAAPYENLMVP 10	PSASMGRDIKV PSPSMGRDIPV 20	QFQ-GGGPI AFL-AGGPI 30	AVYLLDGLRAQ HAVYLLDAFNAG * 40	DDINGWDINI DDYNGWDINT PDVSNWVTAG * 50	PAFEEYYQSGL NAMNTLAGKGI 60	SVI SVV
Rv0129c Rv3803c	RPGLPVEYLQVP AKAAPYENLMVP 10	PSASMGRDIKV PSPSMGRDIPV 20	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG 40	DDINGWDINT DDYNGWDINT PDVSNWVTAG * 50	PAFEEYYQSGL NAMNTLAGKGI 60	SVI SVV
Rv0129c Rv3803c	RPGLPVEYLQVP AKAAPYENLMVP 10	PSASMGRDIKV PSPSMGRDIPV 20	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG * 40	DDYNGWDINT PDVSNWVTAG * 50	PAFEEYYQSGL NAMNTLAGKGI 60	SVI SVV
Rv0129c Rv3803c	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL	SASMGRDIKV SPSMGRDIPV 20 JEEDP	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG * 40	DDYNGWDINT DDYNGWDINT PDVSNWVTAG * 50 ALGRIKWE	PAFEEYYQSGL NAMNTLAGKGI 60 TTFIVEELA-PL	SVI SVV
Rv0129c Rv3803c cmytD Ce1488	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAAASMYSDWL	SASMGRDIKV SPSMGRDIPV 20 EEDP	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG * 40	DDYNGWDINT DDYNGWDINT PDVSNWVTAG 50 ALGRIKWE ALGRIKWE	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL	SVI SVV SVV
Rv0129c Rv3803c cmytD Ce1488 cmytE	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAAASMYSDWV MPMLGSFSFYADWA	SASMGRDIKV SPSMGRDIPV 20 EEDP GESES	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG * 40	DDYNGWDINT PDVSNWVTAG + 50 ALGRIKWE ALGRIKWE MGGAQQWE	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL TFLWHELP	SVI SVV SVV LEA LEA
Rv0129c Rv3803c cmytD Ce1488 cmytE cmytF	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAAASMYSDWV MPMLGSFSFYADWA MPMLGSFSFYADWA	SASMGRDIKV SPSMGRDIPV 20 EEDP GESES GESES	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG * 40	DDYNGWDINT PDVSNWVTAG + 50 ALGRIKWE ALGRIKWE MGGAQQWE MGGAQQWE	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL TFLMHELP TFLMHELP	SVI SVV SVV LEA LEA EPL EPL
Rv0129c Rv3803c cmytD Ce1488 cmytE cmytF Ce0984	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAAASMYSDWV MPMLGSFSFYADWA MPMLGSFSFYADWA MPMLGSHSFYADWV	SASMGRDIKV SPSMGRDIPV 20 EEDP GESES GESES FEENDS	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG / 40	ALGRIKWE ALGRIKWE ALGRIKWE MGGAQQWE MGGAQQWE LGGKQQWE	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL TFLMHELP TFLMHELP TFLTHELP	LEA LEA LEA EPL EPL
Rv0129c Rv3803c cmytD Ce1488 cmytE cmytF Ce0984 cmytC	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAASMYSDWV MPMLGSFSFYADWA MPMLGSFSFYADWA MPMLGSHSFYADWV IPMEGKFSYYTDWV	SASMGRDIKV SPSMGRDIPV 20 EEDP GESES GESES FEENDS FEENAS	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG / 40	ALGRIKWE ALGRIKWE ALGRIKWE MGGAQQWE MGGAQQWE LGGKQQWE LGGKQMWE	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL TFLMHELP TFLMHELP TFLTHELP TFLTHELP	SIV SVI SVV SVV LEA LEA EPL EPL GPL
Rv0129c Rv3803c cmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAASMYSDWV MPMLGSFSFYADWA MPMLGSFSFYADWA MPMLGSHSFYADWV IPMEGKFSYYTDWV IPMEGKFSYYTDWV	SASMGRDIKV SPSMGRDIPV 20 EEDP GESES GESES EENDS ZEENAS QENAA	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG / 40	ALGRIKWE ALGRIKWE ALGRIKWE MGGAQQWE MGGAQQWE LGGKQQWE LGGKQMWE LGGKQMWE	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL TFLMHELP TFLMHELP TFLTHELP TFLVKELP TFLVKELP	LEA SVV SVV LEA EPL EPL GPL GPL
Rv0129c Rv3803c cmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAASMYSDWV MPMLGSFSFYADWA MPMLGSFSFYADWA MPMLGSHSFYADWV IPMEGKFSYYTDWV IPMEGKFSYYTDWV	SASMGRDIKV SPSMGRDIPV 20 EEDP GESES GESES EENDS ZEENAS ZEENAS ZEENAS	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG / 40	ALGRIKWE ALGRIKWE ALGRIKWE MGGAQQWE MGGAQQWE LGGKQQWE LGGKQMWE LGGKQMWE LGGKQMWE LGGKQNWE	PAFEEYYQSGL NAMNTLAGKGI 60 TTFIVEELA-PL TTFIVEELA-PL TTFLMHELP TTFLMHELP TTFLMHELP TTFLVKELP TTFLVKELP	LEA LEA LEA EPL EPL GPL GPL GPI
Rv0129c Rv3803c cmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAASMYSDWV MPMLGSFSFYADWA MPMLGSFSFYADWA MPMLGSHSFYADWV IPMEGKFSYYTDWV IPMEGKFSYYTDWV IPMAGKFSYYTDWV QPIGGAWTYYTDWR	SASMGRDIKV SPSMGRDIPV 20 EEDP GESES GESES EENAS ZEENAS ZEENAS SEAPS	2FQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG * 40	ALGRIKWE ALGRIKWE ALGRIKWE MGGAQQWE MGGAQQWE LGGKQWE LGGKQMWE LGGKQMWE LGGKQNWE LGGKQNWE ALGVNKWK	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL TFLMHELP TFLMHELP TFLTHELP TFLVKELP TFLVKELP TFLVKELP TFLTKELP	LEA LEA LEA EPL EPL GPL GPL GPI PVI
Rv0129c Rv3803c CmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260 Nfa7210	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAASSMYSDWV MPMLGSFSFYADWA MPMLGSFSFYADWA MPMLGSFSFYADWA IPMEGKFSYYTDWV IPMEGKFSYYTDWV QPIGGAWTYTDWR QPIGGKWSYYTDWI	SASMGRDIKV SPSMGRDIPV 20 JEEDP GESES GESES ZENDS ZENAS ZENAS ZENAS SEAPS KDDP	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG * 40 	ALGRIKWE ALGRIKWE ALGRIKWE MGGAQQWE MGGAQQWE LGGKQWE LGGKQWE LGGKQWWE LGGKQWWE LGGKQNWE ALGVNKWF	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL TFLMHELP TFLMHELP TFLTHELP TFLVKELP TFLVKELP TFLVKELP TFLTKELP TFLTEELP	LLEA SVV LLEA LLEA EPL EPL GPL GPL GPI PVI PVI
Rv0129c Rv3803c cmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260 Nfa32210 Nfa322420	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAASSMYSDWV MPMLGSFSFYADWA MPMLGSFSFYADWA MPMLGSFSFYADWV IPMEGKFSYYTDWV IPMEGKFSYYTDWV QPIGGAWTYTDWR QPIGGKWSYYTDWI OPIGGRGSYTDWI	SASMGRDIKV SPSMGRDIPV 20 JEEDP GESES GESES ZEENDS ZEENDS ZEENAS ZEAPS KDPP KDDP	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG * 40	ALGRIKWE ALGRIKWE ALGRIKWE MGGAQQWE MGGAQQWE LGGKQWE LGGKQWE LGGKQWE LGGKQWE LGGKQWE LGGKQWE LGGKQWE LGRNKWE	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL TFLMHELP TFLMHELP TFLTKELP TFLVKELP TFLTKELP TFLTEELP TFFTEELP	LEA SVV LLEA EPL EPL GPL GPL GPL PVI PLV PLL
Rv0129c Rv3803c cmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260 Nfa7210 Nfa22420 Nfa22420 Nfa23770	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAASSMYSDWV MPMLGSFSFYADWA MPMLGSFSFYADWA IPMEGKFSYYTDWV IPMEGKFSYYTDWV QPIGGAWTYYTDWR QPIGGRWSYYTDWI IPMEGAFSYYTDWI IPMEGAFSYYTDWE	SASMGRDIKV SPSMGRDIPV 20 LEEDP GESES GESES ZENAS ZENAS ZENAS ZENAS ZENAS ZENAS RAPP RRDP RRDP	2FQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG * 40	<pre>>DJYNGWDINT PDVSNWVTAG * 50 ALGRIKWE ALGRIMWE MGGAQQWE LGGKQWWE LGGKQWWE LGGKQMWE LGGKQNWE ALGVNKWF TLGRNKWF ELGMNKWF ELGMNKWF ELGMNKWF ELGMNKWF CNGRNWW</pre>	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL TFLMHELP TFLMHELP TFLTHELP TFLVKELP TFLVKELP TFLTEELP TFFTEELP TFFTEELP TFFTEELP	LLEA SVV LLEA EPL EPL GPL GPL GPL PVI PLV PLL
Rv0129c Rv3803c CmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260 Nfa32420 Nfa22770 Nfa30260	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAASMYSDWV MPMLGSFSFYADWA MPMLGSFSFYADWA MPMLGSFSFYADWV IPMEGKFSYYTDWV IPMEGKFSYYTDWV QPIGGAWTYYTDWR QPIGGKWSYYTDWI IPMEGAFSYYTDWI IPMEGAFSYYTDWE IPMEGAFSYYTDWE	SASMGRDIKV SPSMGRDIPV 20 EEDP GESES GESES GESES GENAS ZENAS ZENAS ZEAPS SADEP RADEP RADEGLAETL	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG * 40	<pre> ALGRIKWE ALGRIKWE ALGRIKWE ALGRIMWE MGGAQQWE LGGKQQWE LGGKQWWE LGGKQNWE LGGKQNWE LGGKQNWE LGGKQNWE GNNGRNKWF GNNGRNMW</pre>	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL TFLMHELP TFLMHELP TFLKELP TFLVKELP TFLVKELP TFLTKELP TFFTEELP TFFTEELP TFFTEELP TFFTEELP TFFTEELP	LLEA SVV LLEA LLEA LLEA EPL GPL GPL GPL PVI PVI PVI PVI PVI
Rv0129c Rv3803c CmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260 Nfa7210 Nfa32420 Nfa23770 Nfa30260	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAASMYSDWV MPMLGSFSFYADWA MPMLGSFSFYADWA MPMLGSFSFYADWV IPMEGKFSYYTDWV IPMEGKFSYYTDWV QPIGGAWTYYTDWR QPIGGRWSYYTDWI QPIGGRGSYYTDWI IPMEGAFSYYTDWE IPMEGAFSYYTDWE MPIGGAGSYFADWR	SASMGRDIKV SPSMGRDIPV 20 EEDP GESES GESES GESES GENAS ZENAS ZENAS ZENAS SAPDP RADEP RADEGLAETL ARDP	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG * 40	<pre>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>></pre>	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL TFLMHELP TFLMHELP TFLTHELP TFLTKELP TFLTKELP TFLTEELP TFFTEELP TFFTEELP TFFTEELP TFFTEELP TFFTEELP	LLEA SVV SVV LLEA LLEA EPL EPL GPL GPL PVI PVI PVI PVI PVI PVI
Rv0129c Rv3803c CmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260 Nfa7210 Nfa32420 Nfa23770 Nfa30260 Nfa43800 Nfa43800	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAAASMYSDWV MPMLGSFSFYADWA MPMLGSFSFYADWA MPMLGSHSFYADWV IPMEGKFSYYTDWV IPMEGKFSYYTDWV QPIGGAWTYYTDWR QPIGGAWTYYTDWR QPIGGRGSYYTDWL IPMEGAFSYYTDWE MPIGGAGSYFADWR IPLGGAGSYFTDWR	SASMGRDIKV SPSMGRDIPV 20 EEDP GESES GESES EENDS EENAS CENAS CENAS SCAPP SRADEGLAETL ARDP AEDP	2FQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG * 40	<pre>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>></pre>	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL TFLMHELP TFLMHELP TFLTHELP TFLTKELP TFLTKELP TFLTEELP TFFTEELP TFFTEELP TFFTEELP TFFLTEELP TFFLTEELP	LEA SVI SVV LLEA EPL EPL GPL EPL GPL PVI PLV PLL PLL
Rv0129c Rv3803c CmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260 Nfa7210 Nfa32420 Nfa23770 Nfa30260 Nfa43800 Nfa43800 Nfa45560	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAASMYSDWV MPMLGSFSFYADWA MPMLGSFSFYADWA MPMLGSHSFYADWV IPMEGKFSYYTDWV IPMEGKFSYYTDWV IPMEGKFSYYTDWV QPIGGAWTYYTDWR QPIGGRGSYYTDWI IPMEGAFSYYTDWE MPIGGAGSYFADWR IPMEGASYYTDWV	SASMGRDIKV SPSMGRDIPV 20 EEDP GESES GESES EENAS EENAS SEAPS SEAPS SEAPS SRADEGLAETL ARDP SRADEGLAETL ARDP ADDP	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG * 40	<pre>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>></pre>	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL TFLMHELP TFLMHELP TFLTHELP TFLTKELP TFLTKELP TFLTEELP TFFTEELP TFFTEELP TFFTEELP TFFLTEELP TFLTEELP TFLTEELP TFLTEELP	LEA SVV SVV LEA EPL EPL GPL GPL PLV PLL PLL PLL
Rv0129c Rv3803c CmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260 Nfa7210 Nfa32420 Nfa23770 Nfa30260 Nfa43800 Nfa45560 Nfa25110	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAASMYSDWV MPMLGSFSFYADWA MPMLGSFSFYADWA MPMLGSHSFYADWV IPMEGKFSYYTDWV IPMEGKFSYYTDWV QPIGGAWTYTDWR QPIGGRGSYYTDWI QPIGGRGSYYTDWI IPMEGAFSYYTDWE MPIGGAGSYFADWR IPLGGAGSYFTDWR VPMGGRASYYTDWV MPIGGRFSLYTDWV	SASMGRDIKV SPSMGRDIPV 20 EEDP GESES GESES GESES GENAS	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG * 40	<pre> ALGRIKWE ALGRIKWE ALGRIKWE ALGRIMWE MGGAQQWE LGGKQQWE LGGKQWWE LGGKQNWE LGGKQNWE LGGKQNWE LGGKQNWE LGGKQNWE LGGKQNWE CUGGRNKWG CUGRNKWG VLGRNKWS VLGRNKWS VLGRNKWS VLGRNKWS</pre>	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL TFLMHELP TFLMHELP TFLTHELP TFLTKELP TFLTKELP TFFTEELP TFFTEELP TFFTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP	LLEA SVV SVV LLEA LLEA EPL EPL EPL EPL GPL PVI PLL PLL PLL PLL AAM
Rv0129c Rv3803c CmytD Ce1488 CmytE CmytF Ce0984 CmytC Ce0356 Dip0365 Nfa38260 Nfa7210 Nfa32420 Nfa32420 Nfa32420 Nfa30260 Nfa30260 Nfa45560 Nfa25110 Nfa5610	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAASMYSDWV MPMLGSFSFYADWA MPMLGSFSFYADWA MPMLGSFSFYADWA IPMEGKFSYYTDWV IPMEGKFSYYTDWV IPMEGKFSYYTDWI QPIGGAWTYYTDWR QPIGGRGSYYTDWI IPMEGAFSYYTDWE MPIGGAGSYFADWR IPLGGAGSYFADWR IPLGGAGSYFTDWR VPMGGRASYYTDWQ LPVGGQASYYTDWQ	SASMGRDIKVO SPSMGRDIPV. 20 EEDP GESES GESES GESES GESES GESES GENAS	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG * 40	ALGRIKWE ALGRIKWE ALGRIKWE ALGRIMWE MGGAQQWE MGGAQQWE LGGKQWWE LGGKQMWE LGGKQMWE LGGKQMWE LGGKQNWE LGGKQNWE LGGKQNWE LGGKQNWE ULGRNKWE VLGRNKWE VLGRNKWE VLGRNKWE VLGRNKWE VLGRNKWE VLGRNKWE	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL TFLMHELP TFLMHELP TFLTHELP TFLVKELP TFLVKELP TFLTEELP TFFTEELP TFFTEELP TFFTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP	LLEA LLEA LLEA EPL EPL EPL EPL PLV PLL PLL PLL PLL AAM PII
Rv0129c Rv3803c CmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260 Nfa32420 Nfa32420 Nfa32420 Nfa32420 Nfa30260 Nfa30260 Nfa30260 Nfa43800 Nfa45560 Nfa25110 Nfa5610 Dip2339	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAASSMYSDWV MPMLGSFSFYADWA MPMLGSFSFYADWA MPMLGSFSFYADWA IPMEGKFSYYTDWV IPMEGKFSYYTDWV QPIGGAWTYTDWR QPIGGRGSYYTDWI IPMEGAFSYYTDWI IPMEGAFSYYTDWE MPIGGAGSYFADWR IPLGGAGSYFADWR VPMGGRASYYTDWV MPIGGRFSLYTDWQ LPVGGQASYYTDWN	SASMGRDIKV SPSMGRDIPV 20 JEEDP GESES GESES ZENAS ZENAS ZENAS ZENAS ZENAS SRAPEP RADEP RADEGLAETL ARDP AEDP JADDP JADDP JETDP	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG * 40	<pre>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>></pre>	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL TFLWELP TFLMHELP TFLTHELP TFLVKELP TFLVKELP TFLTEELP TFFTEELP TFFTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP	LLEA LLEA LLEA LLEA EPL GPL GPL EPL PLV PLL PLL PLL AAM PII VQP
Rv0129c Rv3803c CmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260 Nfa32420 Nfa32420 Nfa30260 Nfa30260 Nfa30260 Nfa43800 Nfa45560 Nfa45560 Nfa55110 Nfa5610 Dip2339 cmytA	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAASSMYSDWV MPMLGSFSFYADWA MPMLGSFSFYADWA MPMLGSFSFYADWA IPMEGKFSYYTDWV IPMEGKFSYYTDWV QPIGGAWTYYTDWR QPIGGRGSYYTDWI QPIGGRGSYYTDWI IPMEGAFSYYTDWE MPIGGAGSYFADWR IPLGGAGSYFADWR VPMGGRASYYTDWQ MPIGGRFSLYTDWQ LPVGGQASYYTDWN LPVGGESSFYSDWE	SASMGRDIKV SPSMGRDIPV 20 JEEDP GESES GESES GESES ZENDS ZENDS ZENDS ZENDS ZENDS SADDP RADEGLAETL ARDP ADDP ADDP SADDP SEDP SEDP SEDP SEDP SEDP SEDP SEDP SEDP SEDP SEDP	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG * 40	<pre>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>></pre>	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL TFIVEELA-PL TFLMHELP TFLMHELP TFLTKELP TFLTKELP TFLTKELP TFFTEELP TFFTEELP TFFTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTRELP TFLTRELP	LLEA LLEA LLEA LLEA EPL GPL GPL EPL PLV PLL PLL PLL AAM PIL VQP PLL PLL PLL PLL PLL PLL PLL PLL PLL P
Rv0129c Rv3803c CmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260 Nfa32420 Nfa32420 Nfa32420 Nfa30260 Nfa30260 Nfa30260 Nfa43800 Nfa45560 Nfa45560 Nfa25110 Nfa5610 Dip2339 cmytA Ce2709	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAASSMYSDWV MPMLGSFSFYADWA MPMLGSFSFYADWA MPMLGSFSFYADWA IPMEGKFSYYTDWV IPMEGKFSYYTDWV QPIGGAWTYYTDWR QPIGGGWSYYTDWI QPIGGGGSYYTDWI IPMEGAFSYYTDWE MPIGGAGSYFADWR IPLGGAGSYFADWR VPMGGRASYYTDWQ LPVGGQASYYTDWQ MPIEAGGTNYTDWN LPVGGESSFYSDWE LPVGGESSFYTDWN	SASMGRDIKV SPSMGRDIPV 20 LEEDP GESES GESES GESES ZEENDS ZEENDS ZEENDS ZEENS ZEENS SAPDP RADEGLAETL LARDP RADEGLAETL LARDP RADEGLAETL LARDP SADDP SEPNNGK	2FQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG * 40	<pre>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>></pre>	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL TFIVEELA-PL TFIMHELP TFLTHELP TFLTKELP TFLTKELP TFLTEELP TFFTEELP TFFTEELP TFFTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP TFLTEELP	LEA LEA LEA LEA EPL GPL GPL GPL PLL PLL PLL AAM PIL VQP TLD TLD
Rv0129c Rv3803c CmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260 Nfa7210 Nfa32420 Nfa2420 Nfa2420 Nfa2420 Nfa2420 Nfa2510 Nfa45560 Nfa45560 Nfa45560 Nfa5610 Dip2339 cmytA Ce2709 Dip2193	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAASSMYSDWV MPMLGSFSFYADWA MPMLGSFSFYADWA MPMLGSFSFYADWV IPMEGKFSYYTDWV IPMEGKFSYYTDWV QPIGGAWTYTDWR QPIGGRGSYTDWI QPIGGRGSYTDWI IPMEGAFSYYTDWV MPIGGAGSYFADWR IPLGGAGSYFTDWR VPMGGRASYYTDWV MPIGGRFSLYTDWQ LPVGGQSSFYSDWE LPVGGSSFYSDWE	SASMGRDIKV SPSMGRDIPV 20 LEEDP GESES GESES GESES CENAS ZENAS ZENAS ZENAS ZENAS SAPDP RADEP RADEP ARDP CAPDP CAPDP CAPDP CAPDP CAPDP CAPDP CAPDP CAPDP CAPDP CAPDP CAPDP CAPDP CAPDP CAPDP CAPDP CAPDP CAPDP CAPDP	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG * 40	<pre>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>></pre>	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL TFIVEELA-PL TFLMHELP TFLMHELP TFLTKELP TFLTKELP TFLTEELP TFFTEELP TFFTEELP TFLTEELP TFLTEELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP	LEA LLEA LLEA EPL EPL GPL GPL GPL GPL PVI PLL PVI PLL AAM PIL VQP VLK
Rv0129c Rv3803c CmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260 Nfa7210 Nfa32420 Nfa2420 Nfa2420 Nfa2420 Nfa2420 Nfa2560 Nfa43800 Nfa45560 Nfa45560 Nfa5610 Dip2339 cmytA Ce2709 Dip2193 Nfa1840	RPGLPVEYLQVP AKAAPYENLMVP 10 MPLGAASSMYSDWL MPLGAASSMYSDWV MPMLGSFSFYADWA MPMLGSFSFYADWA MPMLGSFSFYADWV IPMEGKFSYYTDWV IPMEGKFSYYTDWV QPIGGAWTYYTDWR QPIGGRWSYYTDWV QPIGGRGSYTDWL IPMEGAFSYYTDWE MPIGGAGSYFADWR IPLGGAGSYFTDWR VPMGGRASYYTDWV MPIGGRFSLYTDWQ LPVGGQSSFYSDWE LPVGGQSSFYSDWQ LPVGQQSSFYADWM	SASMGRDIKV SPSMGRDIPV 2 20 2 20 2 20 2 2 2 2 2 2 2 2 2 2 2 2	QFQ-GGGPI AFL-AGGPI 30	HAVYLLDGLRAQ HAVYLLDAFNAG * 40	<pre>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>></pre>	PAFEEYYQSGL NAMNTLAGKGI 60 TFIVEELA-PL TFIVEELA-PL TFIVEELA-PL TFLMHELP TFLTHELP TFLTKELP TFLTKELP TFLTEELP TFFTEELP TFFTEELP TFLTEELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP TFLTRELP	LEA SVI SVV LEA EPL EPL EPL EPL EPL PVI PLV PVI PLL PLL PLL VUP VLK CLLE

cmytB	MPVGGAGSFYADWNS(QASLSSS -			DPVIYM	WETFLTQELP	AYLE	
Dip2194	MPVGGAGSFYTDWVG	PAGPQN			AIYR	WETFLTQELP	GYLA	
Nfa1820	MPVGGMSSFYADWNAH	PSEFFGIP	AGS GSSS	GSGALNAFTO	GPGKSYRYQ	WETFLTNELR	WALR	
Nfa1810	MPVGGMSSFYADWNAH	STILGIG	GGSSGSASGSSS	GSGALQMFAG	GPGKSTRYT	WETFLTNNLR	WALR	
Nfa1830	MPVGGQSSFYTDWYA	SNTN			GQKTTYK	WETFLTQELP	NFLA	
Rv3804c	MPVGGQSSFYSDWYQF	PACGK			AGCQTYK	WETFLTSELP	GWLQ	
1f0pA	MPVGGQSSFYSDWYSF	PACGK			AGCQTYK	WETFLTSELP	QWLS	
Rv0129c	MPVGGQSSFYTDWYQ	SQSN			GQNYTYK	- GQNYTYKWETFLTREMPAWLQ		
Rv3803c	APAGGAYSMYTNWEQI	DG			SKQ	WDTFLSAELP	DWLA	
	* *					* *		
					I			
	70 80				90	100	110	
cmytD	EEELNFNGHRGIGGL	SMGATGAV	HLANSNPDLFDG	GVIGISGCYST	LDPIGQTTV	SLIVNSR-GG	NVENMW	
Ce1488	EEELNFNGHRGIGGLS	SMGATGAV	HLANANPDFFDA	VIGISGCYST	LDPIGQATV	SLIVKSR-GG	DVENMW	
cmytE	EAAIGADGQRSIVGMS	SMSGGSVLI	NFATHDPNFYSS	SVGSFSGCAEI	NSWMGRRGI	AATAYNG-NV	VPEQIF	
cmytF	EAAIGADGQRSIVGMS	SMSGGSVLI	NFATHDPNFYSS	SVGSFSGCAEI	NSWMGRRGI	AATAYNG-NV	VPEQIF	
Ce0984	EAAIGGDGQRSIIGMS	SMSGGSVVI	NIASHQPNFYSS	SVASLSGCAET	NSWMGRRGV.	AATVYSG-NA	FPTQIF	
cmytC	EEKLNTDGQRAIAGMS	3MSATTSL	LFPQHFPGFYDA	AASFSGCAAI	SSLLPWEYL	KLTLDRG-NA	TPEQMW	
Ce0356	EEELNADGQRAIAGMS	3MSATTSL	LFPQHYPGFYDA	AASFSGCASI	SQPLPWEYI	RLTLDRG-NA	TPEQMW	
Dip0365	ERHLGASNKRAIAGLS	SMSATSAL	VLAEHAQGFYDA	AGSFSGCAAI	SSPLTYHFL	RLTLERG-GA	ΓΡΕQMW	
Nfa38260	DAALGTNGVNALAGLS	SMSGTSAL	QLPIAAPGLYRA	VAAYSGCAQI	SDPVGHHFV.	ATVVAAG-HG	DVVNMY	
Nfa7210	DGALGTNGINAIAGLS	STSGTTVL	ALPIAKPGLYKA	AAAYSGCAQI	SDPVGSEFV	KLTVETWGGG	DTENMW	
Nfa32420	DATLRSTGRNALTGLS	STSGTSVL	QLAEAKPGLWRS	SVAAYSGCAQI	ADPTGRQFV	KLAVETWAGG	DTENMY	
Nfa23770	DATFGATGANALAGIS	SMAGSSVL	DLTIQAPTRYRS	SVAAYSGCAMI	SDPLGRMFV	TVVISLG-GG	DPENMW	
Nfa30260	DNAFRGTGANAVIGVS	SMAGTSVF	QLALHAPGVYRA	IGSFSGCVP	SDARGRAVV	NTVVRAY-GG	DPVNLW	
Nfa43800	DEHFHGSGANAVAGVS	SMSGTSVF	QLALAAPGLYRA	IGSFSGCVRI	SDPQGQVMV	NAVVASH-RGI	NPVNMW	
Nfa45560	EQRFGMTGRNAVAGLS	SMSATSAL	NLALDAPGRYQA	VGAYSGCARI	SDPAGRALI	YAELAVF-GA	NATNMW	
Nfa25110	TPWLGATGRDAIAGVS	SMSAASAI	DLAIQAGDRYRA	VAAYSGCPWF	ADPPGMTLV.	AAQVLRG-GG	NPVNMW	
Nfa5610	DAQFAGNGVNGIGGLS	SMGGNAAY	ILAARNPHLYTA	VAGYSACPDI	GLATG AVM	FSIANRGG	NPLNMW	
Dip2339	ETKIAYNGKSYIGGLS	SMGGSAAVI	RLANLYPEKFVO	GTFGVSGCYSI	VNTSGRELF	NLAARVI-GGI	NPDLMW	
cmytA	KGFRSN-TDRAITGIS	SMGGTAAVI	NIATHHPDMFKE	VGSFSGYLDI	TSAGMPIAI	SAALADAGGY	DANAMW	
Ce2709	KGFRSN-GERAITGIS	3MGGTAAVI	NIATHNPEMFNE	'VGSFSGYLD'	TSNGMPAAI	GAALADAGGYI	NVNAMW	
Dip2193	NGFRTN-DDRAVVGLS	3MGG'I'AA I I	NLAERRPDLFKF	'VGSF'SGYLD'I	"TSIGMPAAI.	RAAQKDAGGY	DS'I'AMW	
Nial840	SQWRA'I'-DVRGMQGLS	3MGG'I'AAM.	E'LAGRNPGE'VRY	AASYSGFL'I'I	"I'ILGMPQAI	QF'AMRDAGGF'	DSAAMW	
Ce2/10	QHFGVARNNNSVAGLS	SMGGTAALI	NLAAKHPGQFRQ	2AMSYSGYL'I''	"TAPGMQTML	RLAMLD'I'GGF'	NVNAMY	
CMYTB Di 0104	QNFGVARNNNSIGGL	SMGGTAAL	NLAAKHPDQFRQ	2AMSWSGYLNI	TAPGMQTLL.	RLAMLD'I'GGF'	NVNAMY	
D1p2194	ANFGVSPTNNSIAGL	SMGATAAM	NLAALHPDQFRQ	2VLSISGILSM	ISVPGTYLMM	TLALQEVGGEI	NINNMY	
NIAL820	DRLGENPNRNGVEGL	SMGGSAAL	TLAAIHPDQFSF	AGSESGILNI	SAPGMREAL	RVAMLDAGGI	NVDAMA	
NIALOIU NEo1920		CMCCCAAL	I LAAI NPDQF 5 I	AGSISGILN	SAPGMREAL	RVAMIDAGGI		
NIA1030	G-IGVSKINNAVAGL	CMAACCAT	TTATTUDOOFW	AASISGILNI	SAFGMALAI.	CIAMEDAGE		
1f0n7	ANGAVKDTCSAVVGL	SMARSSAL SMACGGAM	TIAIINFQQFVI	AGAMSGLLDI	SQAMGETLT	GLAMGDAGGII		
PT/01290	ANKCUSPTCNAAUCI	SMAGDDAM SMGCCGAT	TIAATIII QQF I I	AGSISALLDI	SCGMUDTI T	GLAMNDSCOVI		
RV0129C	ANIGUSI IGNAAVGL	MOCCVCAM	AI AA FUDDDECE	AASUSGEUNI	SEGWWIILL			
1000000	*	1000101111		*	51111110/111		DINORIN	
	I	I	1	I.	1	I		
	120	130	1/0	1	1			
			140	150	160	170	180	
			140	150	160	170	180	
			THO	150	160	170	180	
cmytD	GPTGSETWKAHDVTSM	IPEGLRD	-MAVYLSAANGV	150 VDDIDLAD -	160	170 SEKEPFYNLLJ	 180 Agvvle	
cmytD Ce1488	GPTGSETWKAHDVTS1 GPVGSRTWQEHDVVS1	1PEGLRD 1PEGLRN	-MAVYLSAANGV -MAVYLSAANGV	150 VVDDIDLAD - VVDEIDREE -	160	170 SEKEPFYNLL YADEPFYNLL	 180 AGVVLE AGTVLE	
cmytD Ce1488 cmytE	GPTGSETWKAHDVTSN GPVGSRTWQEHDVVSN GEVDSDYSRYNDPLLN	1PEGLRD 1PEGLRN 1AAKLEEQ-	-MAVYLSAANGV -MAVYLSAANGV -DNLYIFAGSGV	150 VVDDIDLAD - VVDEIDREE - VFSELDVI	160 GDNAPI	170 SEKEPFYNLL YADEPFYNLL DEDAFKNRVL	 180 AGVVLE AGTVLE VGFEIE	
cmytD Ce1488 cmytE cmytF	GPTGSETWKAHDVTSN GPVGSRTWQEHDVVSN GEVDSDYSRYNDPLLN GEVDSDYSRYNDPLLN	VPEGLRD VPEGLRN VAAKLEEQ VAAKLEEQ-	-MAVYLSAANGV -MAVYLSAANGV -DNLYIFAGSGV -DNLYIFAGSGV	150 VVDDIDLAD VDEIDREE YFSELDVI YFSELDVI	160 GDNAPI GDNAPI	170 SEKEPFYNLL YADEPFYNLL DEDAFKNRVL DEDAFKNRVL	 180 AGVVLE AGTVLE VGFEIE VGFEIE	
cmytD Ce1488 cmytE cmytF Ce0984	GPTGSETWKAHDVTSN GPVGSRTWQEHDVVSN GEVDSDYSRYNDPLLN GEVDSDYSRYNDPLLN GEVDSDYARYNDPVIN	VPEGLRD VPEGLRN VAAKLEEQ- VAAKLEEQ- VAAKLEEQ-	-MAVYLSAANGV -MAVYLSAANGV -DNLYIFAGSGV -DNLYIFAGSGV -DNLYVFAASGV	150 VDDIDLAD - VDEIDREE - VFSELDVI VFSELDVI WSEVDVE	160 GDNAPI GDNAPI GENAPE	170 SEKEPFYNLL YADEPFYNLL DEDAFKNRVL DEDAFKNRVL DEKGLKNRIT	 180 AGVVLE AGTVLE VGFEIE VGFEIE VGFRIE	
cmytD Ce1488 cmytE cmytF Ce0984 cmytC	GPTGSETWKAHDVTSM GPVGSRTWQEHDVVSM GEVDSDYSRYNDPLLM GEVDSDYSRYNDPLLM GEVDSDYARYNDPVIM GPRGGEYNIYNDALIM	VPEGLRD VPEGLRN VAAKLEEQ- VAAKLEEQ- VAHRLAKQ VSDKLRG	-MAVYLSAANGV -MAVYLSAANGV -DNLYIFAGSGV -DNLYIFAGSGV -DNLYVFAASGV -TELYVSNASGI	150 VDDIDLAD - VDEIDREE - YFSELDVI - VFSELDVI - WSEVDVE - AGEWESVDSE	160 GDNAPI GDNAPI GENAPE RFEGLNQQVO	170 SEKEPFYNLL YADEPFYNLL DEDAFKNRVL DEDAFKNRVL DEKGLKNRIT QSIAMAETVV	I 180 AGVVLE AGTVLE VGFEIE VGFEIE VGFRIE IGGIIE	
cmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356	GPTGSETWKAHDVTS1 GPVGSRTWQEHDVVS1 GEVDSDYSRYNDPLL1 GEVDSDYSRYNDPLL1 GEVDSDYARYNDPVI1 GPRGGEYNIYNDAL11 GPRGGEVNIYNDAL11	IPEGLRD- IPEGLRN- VAAKLEEQ VAAKLEEQ VAAKLEEQ VAHRLAKQ VSDKLRG- VSDKLRG-	-MAVYLSAANGV -MAVYLSAANGV -DNLYIFAGSGV -DNLYIFAGSGV -DNLYVFAASGV -TELYVSNASGI -TELYVSNASGI	150 VDDIDLAD - VDEIDREE - FSELDVI WSEVDVE AGEWESVDSE AGHWESANSE	160 GDNAPI GDNAPI GENAPE RFEGLNQQVG RFNGLDQAY	170 SEKEPFYNLL YADEPFYNLL DEDAFKNRVL DEKGLKNRIT QSIAMAETVV LSLAMTETIV	AGVVLE AGTVLE VGFEIE VGFEIE VGFRIE FGGIIE FGGIIE	
cmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365	GPTGSETWKAHDVTSM GPVGSRTWQEHDVVSM GEVDSDYSRYNDPLLM GEVDSDYSRYNDPLLM GEVDSDYARYNDPVIM GPRGGEYNIYNDALIM GPRGGEVNIYNDALIM GPQGSEVNRYNDALIM	NPEGLRD- NPEGLRN- VAAKLEEQ VAAKLEEQ VAHRLAKQ VSDKLRG- VSDKLRG- VAERLRG-	-MAVYLSAANGV -MAVYLSAANGV -DNLYIFAGSGV -DNLYIFAGSGV -DNLYVFAASGV -TELYVSNASGI -TDLYISNASGI -TELYVSNNSGA	150 VDDIDLAD VDEIDREE FSELDVI WSEVDVE AGEWESVDSE AGHWESANSE VGKYDLPSS	160 GDNAPI GDNAPI RFEGLNQQV RFNGLDQAY RLAGKDPVT	170 SEKEPFYNLL YADEPFYNLL DEDAFKNRVL DEKGLKNRIT 2SIAMAETVV LSLAMTETIV IFATNLITA	AGVVLE AGVVLE VGFEIE VGFEIE VGFRIE FGGIIE FGGLIE EGGIIE	
cmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260	GPTGSETWKAHDVTSM GPVGSRTWQEHDVVSM GEVDSDYSRYNDPLLM GEVDSDYSRYNDPLLM GEVDSDYARYNDPVIM GPRGGEYNIYNDALIM GPRGGEVNIYNDALIM GPPGGSEVNRYNDALIM GPPDDPMWAANDPYVQ	NPEGLRD- NPEGLRN- VAAKLEEQ VAAKLEEQ VAHRLAKQ VSDKLRG- VSDKLRG- VAERLRG- VAERLRG-	-MAVYLSAANGV -MAVYLSAANGV -DNLYIFAGSGV -DNLYIFAGSGV -DNLYVFAASGV -TELYVSNASGI -TELYVSNASGI -TEVYVSNNSGA -LELFLSTGTGI	150 VDDIDLAD VDEIDREE FSELDVI WSEVDVE AGEWESVDSE AGHWESANSE VGKYDLPSSE PGKWDTLN PGCVDTL	160 GDNAPI GDNAPI RFEGLNQQVG RFNGLDQAY: RLAGKDPVT GPHAPP	170 SEKEPFYNLL YADEPFYNLL DEDAFKNRVL DEKGLKNRIT QSIAMAETVV LSLAMTETIV IFATNLITATI GSDGLTNQU	AGVVLE AGVVLE AGTVLE VGFEIE VGFEIE VGFRIE FGGIIE EGGIIE LGGILE	
cmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260 Nfa7210	GPTGSETWKAHDVTSM GPVGSRTWQEHDVVSM GEVDSDYSRYNDPLLM GEVDSDYSRYNDPLLM GEVDSDYARYNDPVIM GPRGGEYNIYNDALIM GPRGGEVNIYNDALIM GPPDGSEVNRYNDALIM GPPDDPMWAANDPYVM GPPGSEEWVKNDPYVM	NPEGLRD- VPEGLRN- VAAKLEEQ VAAKLEEQ VAHRLAKQ VSDKLRG- VSDKLRG- VAERLRG- VAERLRG- VAERLRG- VAERLRG-	-MAVYLSAANGV -MAVYLSAANGV -DNLYIFAGSGV -DNLYIFAGSGV -DNLYVFAASGV -TELYVSNASGI -TELYVSNASGI -TEVYVSNNSGA -LELFLSTGTGI -LELFLSTGTGI	150 VDDIDLAD VDEIDREE FSELDVI WSEVDVE AGEWESVDSE AGHWESANSE VGKYDLPSSE PGFYDTLN PGFYDTLN	160 GDNAPI GDNAPI RFEGLNQQVG RFNGLDQAY RLAGKDPVT GPHAMP GPYALP	170 SEKEPFYNLL YADEPFYNLL DEDAFKNRVL DEKGLKNRIT QSIAMAETVV LSLAMTETIV IFATNLITATI GSDGLTNQLV GSYGLANQUL	AGVVLE AGVVLE VGFEIE VGFEIE VGFRIE IGGIIE EGGIIE LGGILE LGGVIE	
cmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260 Nfa7210 Nfa32420	GPTGSETWKAHDVTSI GPVGSRTWQEHDVVSI GEVDSDYSRYNDPLLM GEVDSDYSRYNDPLIM GPVDSDYARYNDPVIM GPRGGEYNIYNDALIM GPRGGEVNIYNDALIM GPPDDPMWAANDPYVY GPDDSPLWRENDPVVM GPDDSPLWRENDPVVM	NPEGLRD- NPEGLRN- VAAKLEEQ VAAKLEEQ VAHRLAKQ- VSDKLRG- VSDKLRG- VAERLRG- VAERLRG- VAERLRG- VAERLRG- VAEKLRG-	-MAVYLSAANGV -MAVYLSAANGV -DNLYIFAGSGV -DNLYIFAGSGV -DNLYVFAASGV -TELYVSNASGI -TELYISNASGI -TEVYVSNNSGA -LELFLSTGTGI -LELYISTGNGI -TQLYISTGSGI	150 VDDIDLAD VDEIDREE FSELDVI FSELDVI WSEVDVE AGEWESVDSE AGHWESANSE VGKYDLPSSE PGKWDTLN PGPYDTLN PVLEDVQY	160 GDNAPI GDNAPI GENAPE RFEGLNQQVG RFNGLDQAY RLAGKDPVT GPHAMP GPYALP YLNAAP	170 SEKEPFYNLL YADEPFYNLL DEDAFKNRVL DEKGLKNRIT QSIAMAETVV LSLAMTETIV IFATNLITATI GSDGLTNQLV GSYGLANQIL GPMGAVN-LC	AGVVLE AGVVLE AGTVLE VGFEIE VGFEIE VGFRIE IGGLIE EGGIIE LGGILE LGVIE LGVIE	
cmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260 Nfa7210 Nfa23770 Nfa30260	GPTGSETWKAHDVTSI GPVGSRTWQEHDVVSI GEVDSDYSRYNDPLLM GEVDSDYSRYNDPVIN GPRGGEYNIYNDALIN GPRGGEVNIYNDALIN GPPGGEVNRYNDALIN GPPDDPMWAANDPYVQ GPPDSEEWVKNDPYVY GPDDSPLWRENDPVVY GPTGGDGWREHDPVLQ	NPEGLRD- NPEGLRN- VAAKLEEQ VAAKLEEQ VAHRLAKQ- VSDKLRG- VSDKLRG- VAERLRG- VAERLRG- VAERLRG- VAEKLRG- VAEKLRG- 2AERLPP-	-MAVYLSAANGV -MAVYLSAANGV -DNLYIFAGSGV -DNLYIFAGSGV -DNLYVFAASGV -TELYVSNASGI -TELYVSNASGI -TEVYVSNNSGA -LELFLSTGTGI -LELFLSTGTGI -TQLYISTGSGI -TPMYISSGSGI	150 VDDIDLAD VDEIDREE FSELDVI FSELDVI WSEVDVE AGEWESVDSE AGHWESANSE NGKYDLPSSE PGFYDTLN PGFYDTLN PGFYDTLA PGFHDTLA	160 GDNAPI GDNAPI RFEGLNQQV RFNGLDQAY RLAGKDPVT GPHAMP GPYALP NPRLHN NPRLHN	170 SEKEPFYNLL YADEPFYNLL DEDAFKNRVL DEKGLKNRIT QSIAMAETVV LSLAMTETIV IFATNLITATI GSDGLTNQLV GSYGLANQIL GPMGAVN-QTL DRLALADOL	AGVVLE AGVVLE AGTVLE VGFEIE VGFEIE VGFRIE IGGIIE IGGLE IGGVIE LGGVIE VGGAIE	
cmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260 Nfa32420 Nfa32420 Nfa30260 Nfa30260 Nfa30260	GPTGSETWKAHDVTSM GPVGSRTWQEHDVVSM GEVDSDYSRYNDPLLM GEVDSDYSRYNDPLLM GPVGSEVNIYNDALIM GPRGGEVNIYNDALIM GPPGGEVNIYNDALIM GPPDDPMWAANDPYVQ GPPGSEEWVKNDPYVM GPDDSPLWRENDPVVM GPTGGDGWREHDPVLQ GPPEDPAWAANDPSLM GPDDTWRANDPVI	NPEGLRD- NPEGLRN- VAAKLEEQ VAAKLEEQ VAHRLAKQ- VSDKLRG- VSDKLRG- VAERLRG- VAERLRG- VAEKLRG- VAEKLRG- VAEKLRG- VAEKLRD- VAAELRD-	-MAVYLSAANGV -MAVYLSAANGV -DNLYIFAGSGV -DNLYIFAGSGV -DNLYVFAASGV -TELYVSNASGI -TELYVSNASGI -TEVYVSNNSGA -LELFLSTGTGI -LELFLSTGTGI -TQLYISTGSGI -TAVYVTAGTGF -TAVYVTSGSGI	150 VDDIDLAD VDEIDREE FSELDVI WSEVDVE AGEWESVDSE AGHWESANSE NGKWDTLN PGPYDTLN PGPYDTLN PGPHDTLA PGPHDTLA PGPLDND	160 GDNAPI GDNAPI RFEGLNQQV RFNGLDQAY RLAGKDPVT GPHAMP GPYALP VLNAAP NPRLHN GPGIDA	170 SEKEPFYNLL YADEPFYNLL DEDAFKNRVL DEKGLKNRIT QSIAMAETVV LSLAMTETIV IFATNLITATI GSDGLTNQLV GSYGLANQIL GPMGAVN-LGI DDRQLLNQTI	AGVVLE AGTVLE VGFEIE VGFEIE VGFRIE IGGLIE IGGLIE IGGVIE LGVIIE VGGAIE IGGALE FGAPIE	
cmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260 Nfa38260 Nfa32420 Nfa32420 Nfa30260 Nfa43800 Nfa45560	GPTGSETWKAHDVTSI GPVGSRTWQEHDVVSI GEVDSDYSRYNDPLLM GEVDSDYSRYNDPLLM GPRGGEYNIYNDALIN GPRGGEVNIYNDALIN GPPGSEVNRYNDALIN GPPDDPMWAANDPYVQ GPPGSEEWVKNDPYVY GPDDSPLWRENDPVVN GPTGGDGWREHDPYLQ GPPEDPAWAANDPSLH GPPTDPTWRANDPYLI GGPDSPLWAAHDPVI	NPEGLRD- NPEGLRN- VAAKLEEQ VAAKLEEQ VAHRLAKQ- SDKLRG- VSDKLRG- VAERLRG- VAERLRG- VAEKLRG- VAEKLRG- VAEKLRG- VAEKLRG- VAEKLRG-	-MAVYLSAANGV -MAVYLSAANGV -DNLYIFAGSGV -DNLYIFAGSGV -DNLYVFAASGV -TELYVSNASGI -TELYVSNASGI -TEVYVSNNSGA -LELFLSTGTGI -LELYISTGSGI -TQYISTGSGI -TAVYVTAGTGF -TAIYISSGSGI -TAIYISSGSGI -LAIYVSAGDGR	150 VDDIDLAD VDEIDREE FSELDVI WSEVDVE WSEVDVE AGEWESVDSE AGHWESANSE VGKYDLPSSE VGKYDLPSSE PGFWDTLN PGPHDTLA PGPHDTLA PGRHETLT	160 GDNAPI GDNAPI RFEGLNQQV RFNGLDQAY RLAGKDPVT GPHAMP GPYALP YLNAAP NPRLHN PGIDA AAVGG AAVGG	170 SEKEPFYNLL YADEPFYNLL DEDAFKNRVL DEKGLKNRIT QSIAMAETVV LSLAMTETIV IFATNLITATI GSDGLTNQLV GSYGLANQIL GPMGAVN-LG DDRQLLNQTL DPLALADQLL DPLALADQLL NPLDIVERTV	AGVVLE AGVVLE AGTVLE VGFEIE VGFEIE IGGIIE IGGLE IGGVIE LGVIIE IGGALE IGGALE FGAPLE VGGIME	
cmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260 Nfa7210 Nfa32420 Nfa32420 Nfa30260 Nfa43800 Nfa45560 Nfa25110	GPTGSETWKAHDVTSI GPVGSRTWQEHDVVSI GEVDSDYSRYNDPLLM GEVDSDYSRYNDPLLM GPRGGEYNIYNDALIN GPRGGEVNIYNDALIN GPPGSEVNRYNDALIN GPPDDPMWAANDPYVQ GPPGSEEWVKNDPYVY GPDDSPLWRENDPVVN GPTGGDGWREHDPYLQ GPPEDPAWAANDPSLH GPPTDPTWRANDPYLH GPPGDSPLWAAHDPVLH GPPGDPGWOSHDAFFN	NPEGLRD- NPEGLRN- VAAKLEEQ VAAKLEEQ VARLAKQ- VSDKLRG- VSDKLRG- VAERLRG- VAERLRG- VAEKLRG- VAEKLRG- VAEKLRG- VARELRD- VAELRG- VAGALAG-	-MAVYLSAANGV -MAVYLSAANGV -DNLYIFAGSGV -DNLYIFAGSGV -DNLYVFAASGV -TELYVSNASGI -TELYVSNASGI -TEVYVSNNSGA -LELFLSTGTGI -LELYISTGSGI -TQYYISSGSGI -TAVYVTAGTGF -TAIYISSGSGI -LAIYVSAGDGF -KTVYLSAASGT	150 VDDIDLAD VDEIDREE FSELDVI WSEVDVE WSEVDVE AGEWESVDSE AGHWESANSE VGKYDLPSSE VGKYDLPSSE PGFWDTLN PGPYDTLN PGPLDNP PGPLDNG PGPIDRG	160 GDNAPI GDNAPI RFEGLNQQV RFNGLDQAY RLAGKDPVT GPHAMP GPHAMP YLNAAP YLNAAP NPRLHN GPGIDA AAVGG APGIDG	170 SEKEPFYNLL YADEPFYNLL DEDAFKNRVL DEKGLKNRIT QSIAMAETVV LSLAMTETIV IFATNLITATI GSDGLTNQLV GSYGLANQIL GPMGAVN-LG DDRQLLNQTL DPLALADQLL DPMQLGYQLL NPLDLVERTV	AGVVLE AGTVLE VGFEIE VGFEIE VGFRIE IGGLIE IGGLIE IGGVIE LGVIIE IGGALE FGAPLE VGGLME PAPTLE	
cmytD Ce1488 cmytE cmytF Ce0984 cmytC Ce0356 Dip0365 Nfa38260 Nfa32420 Nfa32420 Nfa30260 Nfa30260 Nfa43800 Nfa45560 Nfa25110 Nfa5610	GPTGSETWKAHDVTSI GPVGSRTWQEHDVVSI GEVDSDYSRYNDPLLN GEVDSDYSRYNDPLLN GPVGSEYNIYNDALIN GPRGGEYNIYNDALIN GPRGGEVNIYNDALIN GPPGSEEWVKNDPYVY GPDDSPLWRENDPVVY GPDDSPLWRENDPVVY GPPGGDGWREHDPYLI GPPDPPWRANDPSLH GPPDSPLWAAHDPVLE GPPGDPGWQSHDAFRN GPPGDPGWQSHDAFRN	NPEGLRD- NPEGLRN- NAAKLEEQ VAAKLEEQ VAHRLAKQ- SDKLRG- VAERLRG- VAERLRG- VAERLRG- VAEKLRG- VAEKLRG- AAELRD- IADRLRG- VAELRG- VAGALAG- VAGALAG- LAGNLRG-	-MAVYLSAANGV -MAVYLSAANGV -DNLYIFAGSGV -DNLYIFAGSGV -DNLYVFAASGV -TELYVSNASGI -TELYVSNASGI -TELYISTGNGI -LELFLSTGTGI -TQYYISTGSGI -TAVYVTAGTGF -TAIYISSGSGI -LAIYVSAGDGF -KTVYLSAASGI -KTVYLSAASGI	150 VDDIDLAD VDEIDREE FSELDVI WSEVDVE WSEVDVE AGEWESVDSH AGHWESANSH VGKYDLPSSH PGKWDTLN PGPYDTLN PGPHDTLA PGPLDNP PGPLDNP PGPHEAEL	160 GDNAPI GDNAPI GENAPE RFEGLNQQV RFNGLDQAY RLAGKDPVT GPHAMP GPHAMP YLNAAP YLNAAP NPRLHN GPGIDA AAVGG APGIDG	170 SEKEPFYNLL YADEPFYNLL DEDAFKNRVL DEKGLKNRIT QSIAMAETVV LSLAMTETIV IFATNLITATI GSDGLTNQLV GSYGLANQIL GPMGAVN-LG DDRQLLNQTL DPLALADQLL DPMQLGYQLL NPLDLVERTV GL QLAENIF	AGVVLE AGTVLE VGFEIE VGFEIE VGFRIE IGGLIE IGGLIE IGGVIE LGGVIE IGGALE FGAPLE VGGLME PAPTLE LGGPVE	

cmytA	GPVGSERV	VQENDPKSI	IVDKLKGK'	TIYVSSGNGAD	DFGKEGSVAI -		GPANAAGV	GLE
Ce2709	GPAGSERV	VLENDPKRN	IVDQLRGK	QVYVSAGSGAD	DYGQDGSVAT -		GPANAAGV	GLE
Dip2193	GPDGSQDV	VIDHDPKLO	GVEALRGI	TTYVSAGSGRD	DFGEPGSVAN -		KKGSYAGI	GLE
Nfa1840	GPPTSPEV	VEAHDPYLI	LADKLRGV	SLYISSGSGTT	GPFDQASGIPG		VSTNYAGT	GLE
Ce2710	GSVINPRF	RFENDPFWN	MGGLRG-	-KDVYVSAASG	LWGPQDNGTR -		VDHRINGS	VLE
cmytB	GSIINPRF	RFENDPFWN	MGGLAN-	-TDVYISAASG	LWSPQDDGVR -		VDHRLTGS	VLE
Dip2194	GSFFGLRF	RFQLDPLVN	JAAGLAG-	-KDVYVSAASG	IWGGPDYSYA -		VNDRINGS	ILE
Nfa1820	PPWG-PQV	VLRMDPFVE	APNLIRN	GTRLWIAAASG	LPTSTDPPS		FNTLNGM	GLE
Nfa1810	PPWG-PQV	VLRMDPFVH	FAPRLKAN	NTRLWISAGSG	LPGPADGFN		FGTVNAM	GLE
Nfa1830	APWS-PQV	VLRMDPFVH	FAPQLRG-	-LPMYISAASG	LPGQHDRPNSP	VG	VFNTGNAM	ALE
Rv3804c	GPKEDPAV	VQRNDPLLN	VGKLIAN	NTRVWVYCGNG	KPSDLGGNN		LPAK	FLE
1f0pA	GPSSDPAV	VERNDPTQÇ)IPKLVAN	NTRLWVYCGNG	TPNELGGAN		IPAE	FLE
Rv0129c	GPSSDPAV	VKRNDPMVÇ)IPRLVAN	NTRIWVYCGNG	TPSDLGGDN		IPAK	FLE
Rv3803c	GAPQLGRV	VKWHDPWVH	HASLLAQN	NTRVWVWSPTN	-PGASDPAA		MIGQ	AAE
		*						*
				1	1			
		190	200	210	220			230
	cmvtD	RGSLSC	TEALDESM	ISRAGMNHON	VDYKDSGTHNW	RNFNPOLOPO	WDATKHAL	
	Ce1488	RGALSC	TEALDDAM	IODAGMTHON	VDYKGAGAHNW	RNFNEOLOPO	WDAVKDAL	
	CmvtE	AMSNTC	THNLKA	ATDOMGIDNIN	YDFRPTGTHAW	DYWNEALHRE	FPLMMOGE	
	CmytF	AMSNTC	THNLKA		IVDERPTGTHAW	DYWNEALHRE	FPLMMOGF	
	Coller	AT SNUC					FDIMMOCE	
	CEUJ04	ALONIC	- גא ומשוות דוווידועש	VIDGACID-AI			WTTENDAE	
	Colore	AAINKC	THDLNA	KIDUAGIE-AL	WINTED DECENICM	GWWQDDLKGS	WDEADCE	
	CE0336	AAINKC	INDLWI	KLDHAGIP-AL	WNLRPIGIRSW	GWWQDDLKG3	WDIFARSF	
	D1p0365	AGTNMC	THDLKV	KMDSLNIP-A	TNFRNTGTHSW	GIWEEDMVAS	SWELFNMAF	
	NIA38260	AGADHC	TRNMRD	RLTQLG-IPAT	'IDFQPRGTHSW	GIWEDALKLS	SWPVLAKGL	
	Nia/210	AGTNYC	'I'NNLK'I'	RLDELG-IPA1	YNFRPNGTHSW	GYWNEEFPKS	SWPVLAKGL	
	Nfa32420	AAVNQC	'I'ANLKN	RLDSLG-IPA'	'YEFTPVGTHYW	PYWEQALHDS	SWPMLAEGM	
	Nfa23770	SVTNLC	TTRLAQ	RTAELGRTDIT	'YNIRRPG'I'HSW	GYWQDDLRDS	SWPMIARSI	
	Nfa30260	AVAADC	TSELGA	RLRAAG-IPAT	VEVRPDGTHSW	GYWEQDLRRO	CWPLFAAAL	
	Nfa43800	AVTGMC	TRQLRD	RLQELR-IPAT	VDLRPTGTHAW	GYWQEDLHKA	AWPMFEAAL	
	Nfa45560	TVIGAC	TRPLVD	RLTSLA-VPA1	'LALRP-GTHSW	PYWQDDLHDS	SWPMFAAAI	
	Nfa25110	AIARTC	TAAFAD	RLAELG-IAAV	'HVDRPLGAHTW	GQFETDLHES	SWPHLAAAL	
	Nfa5610	VGVNIC	TVAFEQ	RLRGLG-IPAF	IDYSPVGTHSW	SYWQDTLHAS	SWSTIGRAL	
	Dip2339	KMSYRC	TQELEASV	REKIADPSRIT	FDYHDGGVHSW	PYYRQQLPVA	AWANVSKGQ	
	cmytA	VISRMT	SQTFVD	RASQAG-VEVV	ASFRPSGVHSW	EYWQFEMTQA	AFPHIANAL	
	Ce2709	LISRMT	SQTFVD	AANGAG-VNVI	ANFRPSGVHAW	PYWQFEMTQA	AWPYMADSL	
	Dip2193	VISRMT	TETFVA	HARRAG-VEVÇ	AFFRPSGVHDW	PYWQFEMTQA	AWPYMANAL	
	Nfa1840	ILSRLT	SQNFVT	KLGELQ-IPAT	VNYRASGTHSW	PYWDFEMRQS	SWPQAAAAL	
	Ce2710	AVSLAT	TRAWEA	KARAEG-LNVI	ADYPNTGIHSW	AQFSSQLHKT	TRDRVLNVM	
	cmytB	FVAMTS	TRIWEA	KARLQG-LNPI	ADYPMYGIHGW	AQFNSQLERT	TQGRVLDVM	
	Dip2194	IASRVS	TRIWEA	QARAIG-LNLT	TNYPLLGVHNW	VQWRYQIEQS	SKPRILDVM	
	Nfa1820	ALALAN	TRAFQV	RMATLGGGNAV	YSFPPFGIHAW	NNWRDEAVRM	MPDLSANI	
	Nfa1810	VLALAN	TRAFQV	RMATLGANNVI	YDFPAVGVHNW	RYWETEVYRM	MIPDLSANI	
	Nfa1830	ALSLVN	TRAFQV	RLKSLG-IPAC	FDFPATGTHSW	KYWEGQLWNS	GRQGILDAL	
	Rv3804c	GFVRTS	~ NIKFOD	AYNAGGGHNG	FDFPDSGTHSW	EYWGAOLNAN	IKPDLORAL	
	1f0pA	NFVRSS	~ NLKFOD	AYNAAGGHNAV	FNFPPNGTHSW	EYWGAOLNAN	~ MKGDLOSSL	
	Rv0129c	GLTIRT	NOTFRD	TYAADGGRNG	FNFPPNGTHSW	PYWNEOLVAN	IKADIOHVI	
	Rv3803c	AMGN	SRMFYN	OYRSVGGHNGH	FDFPASGDNGW	GSWAPOLGAN	ISGDIVGAT	
				£_1.0.0000000	* *	22		
			1	I	I	I	I	
			240	250	260	270	280	
				200				

Note: The codes are according to the gene identity for individual genomes. The conserved residues are indicated by '*' and the amino acid residue numbering is according to the PDB code:1F0P.