

UNIVERSIDADE DE SÃO PAULO
Instituto de Química

CARACTERIZAÇÃO DE CLONES QUE CODIFICAM MEMBROS DA FAMÍLIA
MULTIGÊNICA TC-85 DE *TRYPANOSOMA CRUZI*

Débora Rose de Oliveira
Tese de Doutorado

Maria Júlia Manso Alves
(orientadora)

São Paulo
22 de julho de 2004

Ficha Catalográfica
Elaborada pela Divisão de Biblioteca e
Documentação do Conjunto das Químicas da USP.

O48^a Oliveira, Débora Rose de
d Caracterização de clones que codificam membros da família
multigênica Tc-85 de *Trypanosoma cruzi* / Débora Rose de
Oliveira. -- São Paulo, 2004.
87p.

Tese (doutorado) - Instituto de Química da Universidade
de São Paulo. Departamento de Bioquímica.
Orientador : Alves, Maria Júlia Manso

I. Biologia molecular 2. Expressão gênica I. T. II.
Alves, Maria Júlia Manso, orientador.

574.88 CDD

**“Caracterização de clones que codificam
membros da família multigênica Tc-85 de
Trypanosoma cruzi”**

DÉBORA ROSE DE OLIVEIRA

Tese de Doutorado submetida ao Instituto de Química da Universidade de São Paulo como parte dos requisitos necessários à obtenção do grau de Doutor em Ciências - Área: Bioquímica.

Aprovado(a) por:

Profa. Dra. MARIA JÚLIA MANSO ALVES
IQ – USP
(Orientadora e Presidente)

Profa. Dra. SUELY LOPES GOMES
IQ – USP

Prof. Dr. RICARDO DE MARCO
IQ – USP

Profa. Dra. ELISABETH ANGÉLICA LEME MARTINS
INSTITUTO BUTANTÁ

Prof. Dr. MARCO AURÉLIO KRIEGER
FIOCRUZ – Paraná

SÃO PAULO
28 DE SETEMBRO 2004

Dedico aos meus pais, Moisés e Wolmar, meus irmãos, Eli e Filé,
ao Mauro, meu grande amigo e companheiro e a todos os animais
que me doaram ou que ainda me doam seu amor incondicional.

“ A ciência, uma fada
num canto de louco.
A luz é lavada-
Como o que nós vemos.
É nítido e pouco!”
 (“Uns versos quaisquer”- Fernando Pessoa)

“ Enquanto os homens não souberem respeitar os animais,
não saberão respeitar a si próprios.” (Michelangelo, século XVI)

AGRADECIMENTOS

À Profa. Dra. Maria Júlia Manso Alves, minha orientadora, pelo incentivo e apoio nos momentos árduos do laboratório.

Ao Prof. Dr. Walter Colli, pelo apoio e pelo bom humor que sempre eleva nosso espírito.

A todos os amigos do Laboratório: Célia, Roberto, Marinei, Marinez, Luiz, Prof. Ivan, Renata, Myriam, Paulo, Melissa, Paula pelos momentos divertidos e pela cumplicidade.

Aos amigos dos outros laboratórios do Instituto pela ajuda e pelas dicas.

Ao Prof. Sérgio Verjovski de Almeida pela concessão do uso do ABI 377.

A FAPESP, pela bolsa concedida.

ÍNDICE

Índice	I
Lista de figuras, tabelas e anexos	III
Abreviaturas	IV
Código de aminoácidos	V
Resumo	VI
Abstract	VII
1. Introdução	1
1.1. Relação parasita-célula hospedeira	3
1.2. Família das trans-sialidases	4
1.3. Família trans-sialidase símile	6
1.4. A interiorização do parasita e as moléculas envolvidas neste processo	9
1.5. Matriz Extracelular	10
1.6. Aspectos genéticos dos tripanosomatídeos	14
2. Objetivos	17
3. Material e Métodos	17
<i>3.a. Obtenção de parasitas</i>	17
<i>3.b. Extração de RNA</i>	17
3.b.1. Extração de RNA total	17
3.b.2. Extração de mRNA	18
<i>3.c. Obtenção da Biblioteca de cDNA</i>	18
3.c.1. Primers utilizados na PCR	18
3.c.2. Preparação do cDNA	19
3.c.3. PCR para amplificação do cDNA de Tc-85	20
3.c.4. Preparação de bactérias competentes com cloreto de cálcio	20
3.c.5. Construção da biblioteca de cDNA de Tc-85	21
3.c.6. Reação de ligação e transformação	21
<i>3.d. Análise dos clones</i>	22
3.d.1. Mini-preparações de DNA plasmidial	22
3.d.2. PCR dos clones (preparação para a análise de restrição)	22
3.d.3. Análise de restrição dos clones sequenciados	22
3.d.4. Purificação dos produtos de PCR	22
3.d.5. Determinação da concentração de DNA e RNA	22
3.d.6. Eletroforese em gel de agarose	23
3.d.7. Transformação de bactérias competentes (DH5 α e BI21(DE3)plysS)	23
<i>3.e. Sequenciamento</i>	23
3.e.1. Etapas do sequenciamento e primers utilizados	23
3.e.2. Reações de seqüenciamento (PCR)	24
3.e.3. Lavagem dos Terminadores	24
3.e.4. Gel de seqüenciamento	24
3.e.5. Tratamento das amostras de seqüenciamento	24
3.e.6. Aplicação das amostras no gel	25

3.e.7. Análise de seqüências	25
<i>3.f. Análise das proteínas</i>	26
3.f.1.Obtenção de proteína recombinante	26
3.f.2. Purificação de proteínas recombinantes	26
3.f.3. Dosagem de proteínas	27
3.f.4. Eletroforese em gel de poliacrilamida (SDS-PAGE)	27
3.f.5. Western blot	28
3.f.6. Dot Blot	28
4. Resultados e Discussão	30
4.1. Biblioteca enriquecida de cDNA de Tc-85	30
4.2. Sequenciamento e análise das seqüências	31
4.3. Expressão de proteínas recombinantes	52
4.4. Análise de ligação de proteínas de fusão a elementos da matriz extracelular Dot Blot	53
5. Considerações Finais	56
6. Conclusões	59
7. Bibliografia	60
8. Anexos	70
<i>Curriculum Vitae</i>	87

LISTA DE FIGURAS, TABELAS E ANEXOS

FIGURAS	Pg	TABELAS	Pg
Figura 1. Ciclo do Trypanosoma cruzi	2	Tabela 1. Membros das famílias das trans-sialidases e das trans-sialidase simile.	8
Figura 2. Estrutura dos membros das famílias das trans-sialidase e trans-sialidase simile.	7	Tabela 2. Primers utilizados nas etapas do seqüenciamento.	33
Figura 3. Primers desenhados.	19	Tabela 3. Identidades entre seqüências nucleotídicas	38
Figura 4. Vetor de clonagem e expressão.	21	Tabela 4. pI e massa molecular (Da) teóricos dos membros clonados.	39
Figura 5. cDNA de Tc-85 amplificado	30	Tabela 5. Análise comparativa das seqüências do grupo I	40
Figura 6. Insetos de cDNA digeridos com Bam HI e Eco RI.	31	Tabela 6. Análise comparativa das seqüências do grupo II	41
Figura 7. Cladograma obtido no algoritmo Neighbor joining	43	Tabela 7. Análise da distribuição de seqüências características nos grupos I e II	42
Figura 8. Cladograma obtido no programa Paup 4.0	44	Tabela 8. Alinhamento de nucleotídeos que resultaram em grande deleções	48
Figura 9. Comparação entre nucleotídeos dos clones 52 e 54	45		
Figura 10. Comparação entre aminoácidos dos clones 52 e 54	46		
Figura 11. Comparação entre nucleotídeos dos clones 43 e 48	47		
Figura 12. Comparação entre aminoácidos dos clones 43 e 48	48		
Figura 13. Domínios relevantes no clone Tc85-45	49		
Figura 14. Comparação entre as seqüências da superfamília das gp85/trans-sialidases e a Tc85-45.	50	ANEXOS	
Figura 15. Sítios de N-glicosilação potenciais da Tc85-45	51	Anexo 1. Alinhamento global das seqüências nucleotídicas.	70
Figura 16. Análise da expressão das recombinantes.	52	Anexo 2. Cromatogramas obtidos.	80
		Anexo 3. Comparação entre as seqüências de aminoácidos de Tc-85.	83
Figura 17. Reatividade da proteína de fusão Tc85-45 com laminina	53		
Figura 18. Reatividade da proteína de fusão Tc85-45 com laminina.	54		
Figura 19. Reatividade da proteína de fusão Tc85-45 com fibronectina.	54		
Figura 20. Reatividade da proteína de fusão Tc85-34 com laminina e fibronectina.	55		

ABREVIATURAS

- A: absorvância
- BSA: albumina de soro bovino
- cDNA: DNA complementar
- dNTP: desoxiribonucleotídeos
- D.O.: densidade óptica
- EDTA: ácido etilenodiaminotetracético
- GPI: glicosilfosfatidil-inositol
- IPTG: isopropil tio β -galactosídeo
- kDa: kilodalton
- LB: meio Luria-Bertani
- LLC-MK2: célula epitelial de rim de macaco Rhesus
- Mab: anticorpo monoclonal
- mRNA: RNA mensageiro
- ORF: “open reading frame”
- Pb: pares de base
- PBS: solução salina tamponada por fosfato (“phosphate buffer saline”)
- PCR: reação de polimerização em cadeia
- pH: potencial hidrogeniônico
- pI: ponto isoelétrico
- rpm: rotações por minuto
- SDS: lauril-sulfato de sódio (“sodium dodecyl sulfate”)
- SDS-PAGE: eletroforese em gel de poliacrilamida
- SOB: meio de cultura bacteriana
- SOC: meio de cultura bacteriana idêntico ao SOB acrescido de 20 mM de glicose.
- TAE: tampão Tris-acetato-EDTA
- TBE: tampão Tris-borato-EDTA
- TBS: tampão Tris-borato-sódio
- TCMM: tampão Tris-cálcio-manganês-magnésio
- TE: tampão Tris-EDTA
- TEMED: N,N,N',N'-tetrametiletlenodiamina
- Tris: Tris-(hidroximetilaminoetano)
- Triton X-114: “octylphenoxypolyethoxyethanol”

Tween 20: “polyoxyethylene-sorbitan-monolaurate”

tRNA: RNA transportador

CÓDIGO DE AMINOÁCIDOS

Alanina	Ala	A
Cisteína	Cys	C
Ác. Aspártico	Asp	D
Ác. Glutâmico	Glu	E
Fenilalanina	Phe	F
Glicina	Gly	G
Histidina	His	H
Isoleucina	Ile	I
Lisina	Lys	K
Leucina	Leu	L
Metionina	Met	M
Asparagina	Asn	N
Serina	Ser	S
Treonina	Thr	T
Valina	Val	V
Triptofano	Trp	W
Tirosina	Tyr	Y
Aminoácido qualquer	-	X

RESUMO

As formas tripomastigotas do *Trypanosoma cruzi* expressam as Tc-85, glicoproteínas de superfície. A Tc85-11, um dos membros da família da Tc-85, foi caracterizado como uma proteína de adesão, com os sítios de ligação a laminina e citoqueratina 18 localizados, respectivamente, nos domínios amino e carboxi-terminal. Utilizando-se primers consensos, um produto de cerca de 2000 pb foi amplificado do cDNA de formas tripomastigotas de *T. cruzi* da cepa Y. Uma biblioteca enriquecida em membros da família da Tc-85 foi construída utilizando-se esses primers e o plasmídeo pCR T7/NT Topo Cloning (Invitrogen). O sequenciamento de cerca de 60 clones resultou em 30 ORFs completas. O sequenciamento de DNA foi realizado com o método dideoxi de terminação de cadeia e as seqüências foram analisadas. Por análise comparativa, identidades de 40-90% entres os clones sequenciados e 30-70% com membros da superfamília de proteínas das gp85/trans-sialidases foram encontradas. Os dados foram compilados nas seguintes ferramentas: ORF finder (NCBI), Cap3, Smith-Waterman, ClustalW, BioEdit e BLASTX/BLASTN (NCBI). Para a análise filogenética, foi utilizado o programa Paup empregando máxima parsimônia (MP), “neighbor joining” (NJ) e máxima probabilidade (MP). A árvore possuía dois grupos e foi submetida a análise de bootstrapping, com busca heurística completa e com 100 e 500 repetições. Os mesmos dois grupos apareceram quando a comparação de seqüências e análise filogenética foram correlacionadas. Em experimentos utilizando-se elementos de matriz extracelular – laminina e fibronectina- foi testada a ligação de proteínas recombinantes purificadas. A proteína codificada pelo inserto de cDNA Tc85-45 foi capaz de se ligar de maneira saturável a laminina e a fibronectina, mas não a albumina e gelatina. Os dados sugerem fortemente que, apesar da família da Tc-85- e por consequência a superfamília das gp85/trans-sialidases- codificarem proteínas com graus variáveis de seqüências similares, as seqüências específicas de peptídeos para a ligação a diferentes moléculas de matriz e receptores celulares variam entre membros da família, constituindo, assim, uma família multiadesiva de glicoproteínas que capacita o parasita a ultrapassar as barreiras impostas pelas membranas celulares, matrizes extracelulares e lâminas basais.

ABSTRACT

Trypomastigote forms of *Trypanosoma cruzi* express Tc-85 surface glycoproteins. Tc85-11, one member of the Tc-85 family, was characterized as an adhesion protein, with laminin and cytokeratin-18 binding sites localized, respectively, on the amino and carboxy terminal domains. Using consensus primers, a 2000 bp product was amplified from cDNA of trypomastigote forms of *T. cruzi* of the Y strain. A library enriched in Tc-85 cDNA was constructed using these primers and pCR T7/NT Topo Cloning (Invitrogen) plasmid. Sequencing of about 60 clones gave 30 complete ORFs. DNA sequencing was performed by the dideoxy-chain termination method and the sequences have been analyzed. By comparative analysis, identity of 40-90% was found among the sequenced clones and 30-70%, with members of the gp85/trans-sialidase superfamily proteins. The data were compiled in the following tools: ORF finder (NCBI), Cap3, Smith-Waterman algorithm, ClustalW, BioEdit and BLASTX/BLASTN (NCBI). For phylogenetic analysis, the Paup program was employed using maximum parsimony (MP), neighbor joining (NJ), and maximum likelihood (ML). The inferred tree had two groups and was submitted to full heuristic bootstrapping with 100 and 500 repetitions. The same two groups appeared when comparative sequence and phylogenetic analyses were correlated. In experiments in which extracellular matrix elements – laminin and fibronectin– were tested for binding to purified recombinant proteins, the protein coded by the cDNA insert Tc85-45 was able to bind in a saturable manner to laminin and fibronectin, but not to albumin and gelatin. The data strongly suggest that although the Tc-85 family, and by extension the gp85/glycoprotein superfamily, encode glycoproteins with variable degrees of similar sequences, the peptide sequences specific for the binding to different matrix molecules and cell receptors varies among the family members, thus, constituting a multi-adhesion family of glycoproteins that enables the parasite to overcome the barriers imposed by cell membranes, extracellular matrices and basal laminae.

1. INTRODUÇÃO

O *Trypanosoma cruzi*, descoberto por Carlos Chagas (Chagas, 1909) é o protozoário flagelado agente etiológico da doença de Chagas que possui um ciclo de vida que alterna entre invertebrados e vertebrados. Emmanuel Dias (1934) apresentou uma descrição do ciclo evolutivo do parasita que, com algumas modificações, permanece até hoje (figura 1). Trata-se de um ciclo complexo, pois possui três formas de desenvolvimento: epimastigota, amastigota e tripomastigota. Os insetos vetores da família *Reduviidae* ao se alimentarem de sangue dos vertebrados parasitados, adquirem a forma sanguínea do hospedeiro. No intestino do inseto, o parasita diferencia-se em epimastigota, que é uma forma não infectiva e capaz de se dividir. Na porção superior do trato digestivo, os epimastigotas diferenciam-se em tripomastigotas metacíclicos (eliminados nas fezes) capazes de infectar o hospedeiro vertebrado. No hospedeiro vertebrado, os tripomastigotas metacíclicos invadem células do organismo ou podem ser fagocitados por células mononucleares e macrofágicas. Em ambos os casos, a população de parasitas se desenvolve. Após o rompimento do vacúolo parasitóforo, permanecem livres no citoplasma onde se diferenciam em amastigotas. Os amastigotas são as formas replicativas intracelulares do hospedeiro vertebrado, que posteriormente diferenciam-se em tripomastigotas infectivos e que são liberados no meio por rompimento das células. O processo pode ser reiniciado pela invasão de novas células ou pela captura de tripomastigotas pelo inseto vetor quando este se alimenta de sangue. Além deste ciclo básico, estágios intermediários entre as formas características do parasita têm sido descritas. Entre elas, as epimastigotas intracelulares, formas de transição entre os amastigotas intracelulares e os tripomastigotas (Almeida-de-Faria et al., 1999 e Tonelli, 2004), as formas esferomastigotas no inseto (De Souza, 1984). Da mesma forma, a descrição da capacidade invasiva dos amastigotas (Ley et al., 1988; Mortara, 1991), amplia o número de formas infectantes do *T. cruzi*.

Os vetores potenciais do *T. cruzi* abrangem mais de 130 espécies de insetos triatomíneos da família Reduviidae, mas cinco deles apresentam significância epidemiológica: *Triatoma infestans*, *Triatoma brasiliensis*, *Triatoma dimidiata*,

Rhodnius prolixus e *Panstrongylus megistus*. Há uma grande variedade de mamíferos que podem ser infectados e agir eventualmente como reservatório de parasitas. A distribuição geográfica diferencial de hospedeiros vertebrados e triatomíneos, de acordo com a preferência dos insetos por sua fonte de sangue, define dois ciclos de transmissão: o doméstico ou peridoméstico e o silvestre (WHO, 2002).

Fatores econômicos e de saúde pública, particularmente em alguns países latino-americanos, fazem com que a doença de Chagas continue sendo um problema de destaque. A doença ocorre desde o México, estendendo-se pela América Central e América do Sul, causando sérios danos à saúde em muitos países da região. De acordo com a OMS (WHO, 2002), a doença infecta cerca de 18 milhões de pessoas no México e Américas do Sul e Central e 25 milhões estão sob o risco de infecção. Na última década, os programas de controle em vários países endêmicos tiveram sucesso evidente. O programa de controle domiciliar brasileiro mediante o uso de inseticidas, na tentativa de interromper o ciclo de transmissão, tem permitido que 90% dos municípios infectados estejam livres da doença. Todavia, infecções congênitas e transfusão sanguínea são ainda importantes fontes de transmissão da doença em algumas regiões (Dias, 1992).

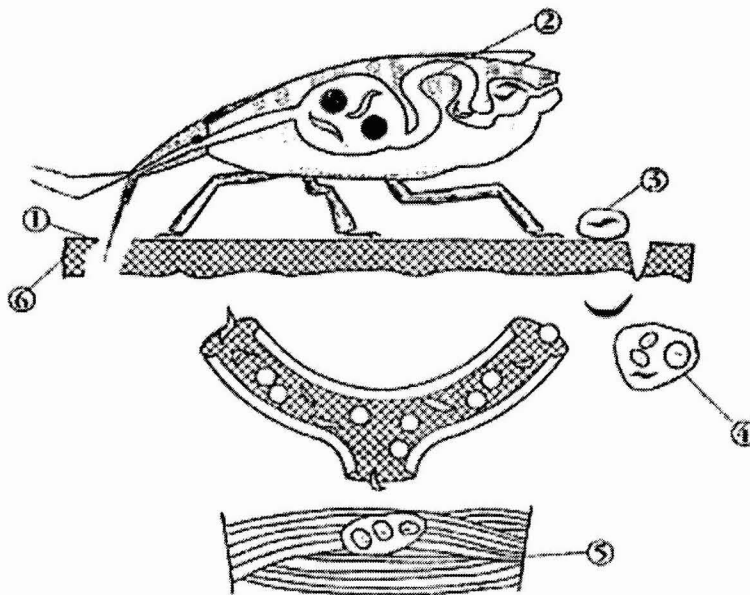


Figura 1. Ciclo do *Trypanosoma cruzi* mostrando os estágios de desenvolvimento. Primeiro desenho esquemático proposto por Carlos Chagas. 1. Picada do barbeiro; 2. Reprodução no intestino; 3. Fezes contaminadas; 4. Parasitas invadem primeiro células da pele e depois a circulação sanguínea; 5. *T. cruzi* nas fibras musculares, fase assintomática; 6. o ciclo recomeça com uma nova picada.

O aumento do número de casos da doença de Chagas tem sido também agravado pela existência de um ciclo silvestre do parasita, pela invasão das matas pelo homem e pela descontinuidade dos programas de controle domiciliar, mesmo em países onde a transmissão parecia controlada, como no Brasil (Silveira, 2000). Tem sido relatado o aumento da doença na região amazônica envolvendo uma variedade de animais silvestres e um mínimo de 10 espécies do inseto triatomíneo. Há que se considerar também a diversidade genética do parasita com alguns autores sugerindo pelo menos três subespécies de *T. cruzi* (Coura et al., 2002).

1.1. Relação parasita-célula hospedeira

Apesar de diversos grupos de pesquisa se dedicarem a este tema, o mecanismo de invasão não está completamente elucidado, com várias questões ainda a serem respondidas. Sem dúvida, adesão e invasão constituem as duas etapas da infecção.

O parasita é capaz de invadir uma ampla variedade de linhagens celulares, com exceção de hemácias e células hematopoiéticas (Dvorak et al., 1976; Colli, 1993). *In vivo*, observa-se parasitemia preferencial em células musculares, células inervadas pelo sistema nervoso autônomo e macrófagos, que parece ser dependente não apenas do tipo celular, mas também da cepa de *T. cruzi* (Andrade, 1974; Melo et al., 1978). Tripomastigotas, amastigotas e epimastigotas são engolfados por macrófagos com formação do vacúolo parasitóforo, mas somente as formas tripomastigota e amastigota são capazes de sobreviver e dar continuidade à infecção (Burleigh & Andrews, 1995; Burleigh & Woolsey, 2002).

Schenkman^b et al., 1991, demonstraram que o processo de adesão de tripomastigotas é dependente da energia metabólica do parasita e que a entrada de tripomastigotas ocorre preferencialmente pela região basolateral, por um mecanismo diferente da fagocitose. Evidências recentes sugerem que o reconhecimento da célula hospedeira é efetuado por diversas famílias de glicoproteínas de superfície do parasita que interagem com diferentes receptores da célula hospedeira e com moléculas da matriz extracelular (Ouassiss et al., 1984; Alves et al., 1986; Abuin et al., 1989; Giordano et al., 1994; Giordano et al., 1999; Ortega-Barria & Pereira,

1991) A superfamília das gp85/trans-sialidases contém seqüências fibronectina-like (Pereira et al., 1991) e domínio ligante de laminina (Giordano et al., 1994), sugerindo a participação das integrinas no processo invasivo. Um ligante específico da região conservada FLY da Tc-85 foi identificado como citoqueratina 18 (Magdesian et al., 2001).

Estão sendo desvendados pontos relevantes sobre o mecanismo de sinalização durante o processo de invasão às células do hospedeiro pelo parasita. Tardieux et al., 1992, verificaram que a migração dos lisossomos ocorre precocemente, sendo visualizada mesmo quando o parasita ainda está fora da célula, aderido à membrana celular e que a membrana dos lisossomos contribui para a formação do vacúolo parasitóforo. Os autores também observaram que tratamentos que induzem a movimentação de lisossomos da região perinuclear até próximo à membrana plasmática aumentam a infecção de células NRK (normal rat kidney cells) por tripomastigotas. O tratamento que inibe a migração também inibe a infecção. Nesse fenômeno foram detectados pulsos intracelulares de cálcio na célula hospedeira, levantando a hipótese de que este íon seja o segundo mensageiro utilizado (Tardieux et al., 1994, Burleigh & Andrews, 1998). A despolimerização dos filamentos de actina da célula hospedeira aumenta o número de lisossomos regulados por Ca^{2+} , mediando a entrada do parasita (Rodrigues et al., 1997). Recentemente, no entanto, dois grupos independentes mostraram que a invasão dos tripomastigotas se dá principalmente sem o recrutamento de lisossomos (Wilkowsky et al., 2002; Woolsey et al., 2003).

1.2. Família das trans-sialidases

Sialidases são enzimas hidrolíticas presentes de vírus a eucariotos superiores (Buschiazzo et al., 1997). A trans-sialidase do tripanosoma é a única que realiza a reação de transferência de um ácido siálico a partir de glicoconjugados pré-formados do hospedeiro ou do meio em que se encontram esses parasitas, em vez da reação típica de hidrólise encontrada na maioria das sialidases (Smith & Eichinger, 1997; Franchin et al., 1997). O ácido siálico, ligado por ligação α 2,3 nos glicoconjugados, é transferido para aceptores da superfície celular do parasita que

contêm resíduos de β -galactosil terminais (Colli, 1993, Cross & Takle., 1993; Schenkman et al., 1994; Ribeirão et al., 1997).

A família das trans-sialidases (família TS) (figura 2) possui no mínimo 140 membros (Frasch, 2000) que podem ser classificados em grupos, de acordo com a estrutura e função das proteínas. As proteínas dessa família têm como característica a presença de pelo menos um motivo conservado dos aminoácidos SxDxGxTW (Asp box), presente em neuraminidases bacterianas e uma seqüência de aminoácidos (VTVxNVxLYNR), altamente conservada, próxima à região C-terminal denominada seqüência FLY (Colli, 1993; Cross et al., 1993; Schenkman et al., 1994). O motivo FRIP (Phe-Arg-Ile-Pro), típico das sialidases, é encontrado próximo ao amino terminal das moléculas. Adotaremos aqui a classificação proposta por Frasc, 2000 (tabela 1). Dois grupos (TS e TS I) são expressos pelos tripomastigotas e são ancorados por glicosilfosfatidilinositol (GPI) na membrana do parasita. Estes grupos possuem duas regiões principais: N-terminal (catalítica) e C-terminal, com repetições dos aminoácidos DSSAHSTPSTPA (“SAPA repeats”). O grupo TS tem atividade de trans-sialidase, porém o grupo TS I não possui tal característica, devido à substituição da Tyr342 por histidina.

Os membros de terceiro grupo (TS-e) são expressos por epimastigotas, a região repetitiva SAPA é ausente e possuem atividade de trans-sialidase. Embora a presença da região repetitiva SAPA não esteja relacionada com a atividade enzimática, a presença das repetições modula a produção de anticorpos contra a região catalítica. Essas repetições são imunodominantes e se propõe que devem prevenir a formação precoce de anticorpos contra a região catalítica, com inibição da atividade enzimática. É sugerido que esta deve ser a maneira de escape usada pelo parasita no período agudo da infecção, possibilitando portanto a transferência de ácido siálico pela trans-sialidase para moléculas da superfície do parasita. Posteriormente, imunoglobulinas são produzidas também contra o sítio catalítico, inibindo a atividade enzimática (Buscaglia et al., 1998), diminuindo a invasão maciça das células pelo parasita.

1.3. Família trans-sialidase símile

Recentemente, alguns genes já identificados similares àqueles codificados pela família das trans-sialidasas, mas que não têm atividade enzimática foram incluídos nesta família trans-sialidase símile (Frasch, 2000) (figura 2). Estes genes foram agrupados desta forma, porque apresentam identidade de 30-40% com os genes das trans-sialidasas e também pelas seqüências deduzidas de proteínas da maioria dos membros terem 1-4 cópias do motivo SXDXGXTW e o motivo VTVXNVFLYNR (seqüência FLY), característico de todos os membros da família das trans-sialidasas e da família trans-sialidase símile.

Membros da família trans-sialidase símile podem ser classificados em três grupos de acordo com a identidade de seqüência, massa molecular e outros. Algumas características da família trans-sialidase símile chamam a atenção. Observa-se que a maioria dos membros são encontrados na membrana celular de tripomastigotas metacíclicos, tripomastigotas sangüíneos e amastigotas intracelulares, ou seja, são as formas encontradas no hospedeiro vertebrado. Os membros desta família apresentam funções diferentes, como proteínas regulatórias que restringem a ativação das vias clássica e alternativa do complemento, ligando-se aos componentes C3b e C4b (Tomlinson & Raper, 1998) e proteínas que se ligam a laminina e superfície celular (Giordano et al., 1999; Ramirez et al., 1993). São também um dos principais antígenos do parasita.

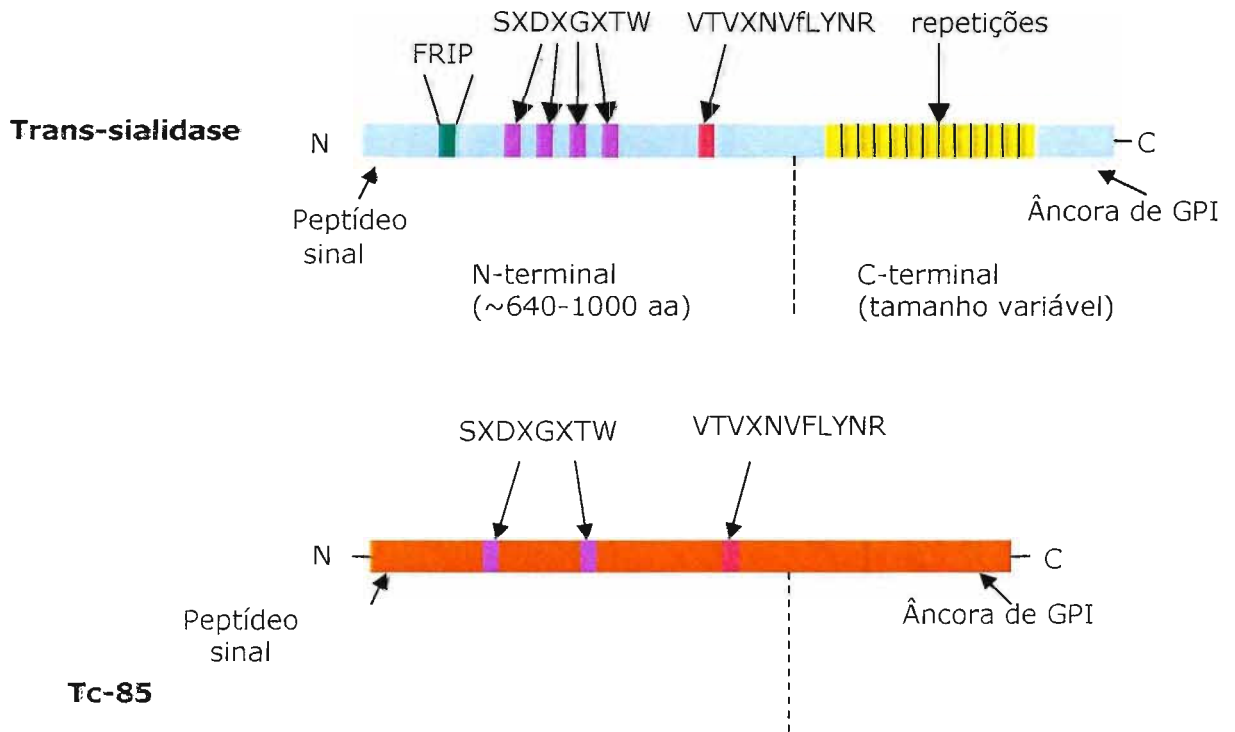


Figura 2: Estrutura geral dos membros das famílias das trans-sialidase e trans-sialidase símile. As funções propostas da região N-terminal em diferentes membros são: atividade de trans-sialidase, proteína regulatória do complemento, adesão celular, adesão a β -galactose e adesão a laminina. As repetições na região C-terminal podem estar envolvidas em: multimerização da proteína, estabilização da proteína no sangue e modulação da resposta imune contra a região N-terminal. As repetições do C-terminal estão presentes em alguns membros da família.

Tabela 1: Membros das famílias das trans-sialidases e das trans-sialidase simile

Nome ^a	Tamanho da proteína (kDa)	Características estruturais ^b	Estágio do parasita ^c	Nº de genes	Atividade/função ^d
Família das trans-sialidases					
TS	120-200	Região N-terminal catalítica (638 aa) Tyr342 Repetições SAPA no C-terminal (tamanho variável)	bT mT	70	Trans-sialidase/sialidase
TS-I	120-200	Região N-terminal catalítica (638 aa) His342 ^e Repetições SAPA no C-terminal (tamanho variável)	bT mT	70	Ligação a β -galactose
TS-e	85	Região catalítica (cerca de 640 aa) Tyr342	E	?	Trans-sialidase/sialidase
Família das trans-sialidases símile					
Tc-85 ^f	85	Aproximadamente 750-850 aa Alguns membros têm C-terminal com repetições de aminoácidos.	bT mT A	?	Adesão celular, ligação a laminina
FI/CEA/CRP	160	Aproximadamente 1000 aa	bT mT	700	Regulação do complemento
Tc13	85	Domínio N-terminal não repetitivo Domínio repetitivo C-terminal com 5 aa	bT mT	dúzias	?

^a Tc13, antígeno 13 de *T. cruzi*; TS, trans-sialidase; TS I, TS inativa; TS-e TS de epimastigota; FI, proteína de flagelo associada a superfície; CEA, exoantígeno da fase crônica; CRP, proteína regulatória do complemento.

^b Alguns membros têm cópias completas ou degeneradas dos motivos SxDxGxTW, FRIP e/ou uma cópia do motivo VTVxNVfLYNR. A maioria possui âncora de GPI.

^c bT: tripomastigotas sanguíneos (bloodstream trypomastigotes) mT: tripomastigotas metacíclicas (metacyclic trypomastigotes) E: epimastigotas

^d A função/atividade foi indicada para um ou alguns membros de cada família.

^e Outras diferenças de aminoácidos estão presentes no sítio catalítico de membros inativos da família de trans-sialidase, mas essas diferenças na posição Tyr 342 define membros naturalmente inativos

^f Nomes diferentes são dados às proteínas de 85kDa (TSA I, SA85, Tc34, ASP, etc). Trabalhos posteriores deverão determinar se estes membros serão agrupados em subfamílias

(modificado de Frasch et al., 2000)

1.4 interiorização do parasita e as moléculas da superfamília das gp85/trans-sialidases envolvidas neste processo

As famílias das trans-sialidases, das trans-sialidase símile e também a família das mucinas constituem os grupos de proteínas mais abundantes da superfície do *Trypanosoma cruzi*. A correlação de várias glicoproteínas de superfície com o processo invasivo do *T. cruzi*, principalmente com a etapa de adesão, tem sido apontada por diversos grupos de pesquisa, envolvendo proteínas de 80-90kDa pertencentes à superfamília das gp85/trans-sialidases. Como já ressaltado, foi proposto que a sialilação ocorrida na superfície do parasita está relacionada com a invasão da célula hospedeira (Schenkman^a et al., 1991). Outra função poderia ser a proteção à lise, por um caminho alternativo do complemento durante a circulação do parasita na fase aguda da doença ou seu escape do vacúolo parasitóforo (Colli, 1993).

Os tripomastigotas metacíclicos expressam uma glicoproteína de superfície de 82kDa que foi relacionada com a invasão celular (Ramirez et al., 1993). Análises da seqüência de aminoácidos sugerem a existência de várias modificações pós-transcricionais, como N-glicosilação e adição de âncora de glicosilfosfatitilinositol. De forma incomum, a gp82 não contém a seqüência amino-terminal correspondente ao peptídeo sinal, não sendo eficientemente processada em células VERO, possivelmente pela ausência de reconhecimento de sinais para a translocação até o retículo endoplasmático da célula de mamífero (Ramirez et al., 1999). Peptídeos derivados da gp82 (P4 e P8), previamente implicados na invasão de célula hospedeira, foram expressados como proteínas de fusão com LamB de *Escherichia coli* expondo o peptídeo na superfície celular. Mostrou-se que estas bactérias aderem às células HeLa, e as proteínas de fusão purificadas inibiram a infecção por *T. cruzi* (Pereira et al., 1999).

Uma das primeiras glicoproteínas de superfície 85 kDa (Tc-85) (Katzin & Colli, 1983), específica de tripomastigotas, que se liga a lectina WGA (wheat-germ agglutinin) foi descrita em nosso laboratório. A invasão de células de cultura pelo parasita é inibida (50-90%) por um anticorpo monoclonal (Mab H1A10) gerado contra esta glicoproteína (Alves et al., 1986). Hoje nós sabemos que a Tc-85 é uma família de glicoproteínas com o número de membros ainda indeterminado. Parte da

diversidade de proteínas da família pôde ser verificada quando se analisou a Tc-85 em eletroforese bidimensional, revelando um amplo polimorfismo (pIs 5,5-9,0) (Abuin et al., 1989). Anticorpos anti-laminina foram capazes de inibir a invasão de células de cultura por tripomastigotas (75-62%), sugerindo que a laminina poderia estar envolvida na adesão dos parasitas no hospedeiro. Por cromatografia de afinidade, foi isolada uma glicoproteína de 85-kDa (laminin binding glycoprotein, LBG) a partir de lisados de tripomastigotas, mas não de lisados de epimastigotas (Giordano et al., 1994). A clonagem de um membro ácido da família da Tc-85 (Tc85-11) mostrou que a proteína recombinante liga-se à laminina e tem identidade com todos os membros das gp85/trans-sialidases. O epítipo do anticorpo monoclonal (Mab H1A10) foi mapeado (Giordano et al., 1999) e localiza-se próximo à região FLY, ligante de citoqueratina 18 (Magdesian et al., 2001). A Tc-85 está ligada à membrana por uma âncora de GPI-glicosilfosfatidilinositol (Couto et al., 1993) e carboidratos N e O-ligados estão presentes na Tc-85. Uma única estrutura de oligossacarídeos N-ligado à proteína foi isolada e caracterizada e manoses, possivelmente como dissacarídeos, foram encontradas em O-ligações (Couto et al., 1990). É importante considerar que a galectina-3, uma lectina específica para β -galactosídeos, foi descrita como uma molécula moduladora, aumentando a ligação de *T. cruzi* à laminina, envolvendo moléculas de 45, 32 e 30 kDa da superfície do parasita (Moody et al., 2000).

Há ainda evidências de que outras moléculas, tanto do parasita quanto do hospedeiro, estão envolvidas na adesão e invasão, como a fibronectina, colágeno, heparan-sulfato, heparina e receptor de TGF β , entre outras (Ouaisi et al., 1984; Ouaisi et al., 1992; Ortega-Barria & Pereira, 1991; Ming et al., 1995). Foi relatado recentemente que a invasão do *Trypanosoma cruzi* a um grupo de células de mamíferos aumenta com a ativação dos receptores B₂ de bradicinina, juntamente com a elevação dos níveis de Ca²⁺ (Scharfstein et al., 2000).

1.5. Matriz Extracelular

As células de tecidos de organismos multicelulares estão embebidas em uma matriz extracelular constituída de vários polissacarídeos e proteínas que são

secretados localmente e estão organizados em forma de rede (Alberts et al., 2002). Esta matriz preenche os espaços intercelulares e mantém as células do tecido unidas. Ao simples papel de sustentação física da estrutura dos tecidos da matriz extracelular, foram acrescentadas diversas funções como papel na resposta inflamatória, cicatrização de feridas, adesão do espermatozóide ao óvulo, desenvolvimento embrionário e um complexo papel na regulação do comportamento das células que estão em contato influenciando seu desenvolvimento, migração, proliferação, forma e funções metabólicas. Há várias classes de proteínas que compõem a matriz extracelular: colágeno, fibronectina, laminina e proteoglicanas (glicosaminoglicanas) (Cooper, 1996). Dependendo do tipo de tecido, as camadas variam sua composição de forma específica. Um dos tipos de matriz extracelular são as membranas basais ou lâmina basal encontrada em camadas inferiores de células epiteliais e endoteliais. Todas as membranas basais possuem laminina, entactina-1/nidogênio-1, colágeno tipo IV e proteoglicanas. Pesquisas recentes estão demonstrando uma grande diversidade e uma alta complexidade nestas classes de proteínas, como a expansão da família entactina/nidogênio com a descrição de uma nova isoforma, E/N-2/osteonidogênio, a caracterização do colágeno tipo XVIII e a agrina reclassificados como proteoglicanas heparan-sulfato (HSPGs) ou a expansão da família da laminina (Erickson & Couchman, 2000).

As lamininas são glicoproteínas encontradas em todas as membranas basais, com diversos papéis funcionais no desenvolvimento e estabilização das estruturas. São moléculas heterotriméricas, compostas por uma cadeia α , uma cadeia β e uma cadeia γ . A estrutura primária completa de todas as nove cadeias ($\alpha 1$, $\alpha 2$, $\alpha 3$, $\alpha 4$, $\beta 1$, $\beta 2$, $\beta 3$, $\gamma 1$ e $\gamma 2$) já está descrita. Foi identificado, recentemente, um novo tipo de laminina que possui cadeias α e β conhecidas e a nova cadeia $\gamma 3$. A cadeia $\gamma 3$ é amplamente distribuída na superfície apical de células ciliadas do pulmão e trato reprodutivo, com a sugestão de que estas lamininas apicais são importantes na morfogênese e estabilidade estrutural destas células (Koch et al., 1999). A laminina-1 ($\alpha 1\beta 1\gamma 1$), a primeira a ser descrita, é expressa simultaneamente em epitélio e endotélio e purificada a partir de tumor murino EHS (Engelbreath-Holm-Swarm) (Timpl et al., 1979). Outro exemplo, é a laminina-5 ($\alpha 3\beta 3\gamma 2$) que já foi caracterizada em epitélios especializados em funções secretórias e protetoras, como a pele e a membrana mucosa (Ryan & Christiano, 1996). A laminina é fortemente associada a

outra proteína de adesão, a entactina ou nidogênio. Também associa-se ao colágeno tipo IV e ao perlecan, formando ligações cruzadas na lâmina basal. Um peptídeo derivado da cadeia $\alpha 1$ da laminina estabelece a interação com a integrina $\alpha 3\beta 1$ (Tashiro et al., 1999). $\alpha 3\beta 1$ é também um receptor para laminina-5 (Goldfinger et al., 1999), mas a exata seqüência da cadeia $\alpha 3$ da laminina que faz esta ligação não está definida (Belkin & Stepp, 2000). Interações de células com as cadeias α das lamininas são críticas nas associações célula-matriz. Pelo menos 7 integrinas distintas ($\alpha 1\beta 1$, $\alpha 2\beta 2$, $\alpha 3\beta 1$, $\alpha 7\beta 1$, $\alpha 9\beta 1$, $\alpha 6\beta 4$ e $\alpha 5\beta 3$) se ligam na cadeia α da laminina. Vale ressaltar, no entanto, que nem todas as lamininas estão restritas às membranas basais: a cadeia $\gamma 3$ é abundantemente expressa nas células endoteliais de arteríolas e capilares. Tem ampla distribuição entre os tecidos, sendo encontrada na pele, coração, pulmão, trato digestivo (Koch et al., 1999).

O colágeno, que constitui uma grande família de proteínas, é o mais abundante e a principal proteína estrutural da matriz extracelular. É formado por uma hélice tripla de cadeias polipeptídicas estabilizada por pontes de hidrogênio entre aminoácidos modificados, como hidroxiprolinas e hidroxilisinas. Um mecanismo de resposta ao estresse mecânico em tecidos conectivos foi descrito e tem o padrão alterado com a expressão diferencial de proteínas. Há evidências de regulação durante a transcrição gênica. Neste caso, a produção de tenascina e colágeno XII, proteínas típicas de tendões e ligamentos, é alta em fibroblastos acoplados a colágeno de matriz distendida e é suprimida em células de matriz relaxada. Este processo é acompanhado pela mudança no nível de mRNA destas proteínas (Chiquet, 1999). A interação entre colágeno e plaquetas ocorre na homeostase mas também está envolvida na trombose. Tratam-se de processos de adesão e ativação envolvendo reconhecimento seqüencial de diferentes receptores (Barnes et al., 1998). Hodivala-Dilke et al., 1998, demonstraram que fibroblastos depletados da integrina $\alpha 3\beta 1$ aumentavam sua adesão à fibronectina e colágeno IV, mostrando a dinâmica do processo.

As proteínas fibrosas estruturais da matriz extracelular estão embebidas em géis de polissacarídeos, os glicosaminoglicanos (GAGs), que são moléculas com repetições de dissacarídeos, ligados a uma proteína central formando as proteoglicanas. As espécies principais são: hialuronan, dermatan-sulfato, condroitin-

sulfato, queratan-sulfato e heparan-sulfato. Um dos açúcares desse dissacarídeo sempre é N-acetilglicosamina ou N-acetilgalactosamina e o segundo açúcar é geralmente ácido (ácido glicurônico ou idurônico). Como modificação característica, há a adição de grupos sulfato ao esqueleto de açúcares (Bosman & Stamenkovic, 2003). Heparan-sulfato (HS) é um dos membros com estrutura mais variável devido ao alto conteúdo de seqüências sulfatadas e que são responsáveis pela ligação a inúmeras proteínas. As cadeias de heparan-sulfato são ligadas por um resíduo de serina a um esqueleto protéico, localizado na superfície das células e na matriz. As cadeias de HS têm interações específicas com determinadas proteínas e essa interação resulta na regulação da atividade das proteínas, como fatores de crescimento, enzimas, moléculas de matriz extracelular ou de patógenos (Rostand & Esko, 1997). Foi descrita uma nova interação entre um domínio de lectina de lecticans e glicolipídeos da superfície celular que promove a adesão celular. Isso sugere que os lecticans da matriz extracelular sirvam como substrato para adesão e migração de células que expressam estes glicolipídeos *in vivo* (Míura et al., 1999).

As proteínas de adesão são responsáveis pela ligação de componentes da matriz um ao outro ou à superfície das células. A principal proteína de adesão dos tecidos conjuntivos é a fibronectina. As integrinas $\beta 1$, que se ligam à seqüência RGD da fibronectina e a molécula CD44 estão envolvidas no processo de invasão de células tumorais a matriz extracelular. Um trabalho recente reforçou o papel da fibronectina como ponto crítico neste processo (Brenner et al., 2000). A fibronectina é uma glicoproteína conhecida por mediar a adesão *in vitro* do *T. cruzi* a célula hospedeira (Wirth & Kierszenbaum, 1984; Ouaiissi et al., 1985). Em camundongos infectados, baixos níveis de IgM contra o sítio de adesão RGD da fibronectina foram detectados antes e durante as primeiras 3 semanas de infecção, no entanto, esses níveis eram altos nos animais sobreviventes (Truyens et al., 1995).

A trombospondina é uma glicoproteína que participa na agregação de plaquetas e também é o produto de secreção da maioria das células mesenquimais e epiteliais. A natureza e as conseqüências da interação da trombospondina com as células são complexas e controversas. Esta glicoproteína não é geralmente um fator de adesão celular. Embora seja secretada e retida localmente sendo incorporada na matriz extracelular, pode funcionar de uma forma autócrina ou parácrina, não tendo

função de sustentação dos tecidos do organismo adulto. Aparentemente, a trombospondina, juntamente com outras proteínas, tem um papel dinâmico na embriogênese e morfogênese (Sage & Bornstein, 1991). A trombospondina-1 tem propriedades de adesão e também modula o movimento e proliferação celular, a expansão de neuritos e a angiogênese (Adams, 1997).

1.6. Aspectos genéticos dos tripanosomatídeos

A infecção por *T. cruzi* apresenta manifestações clínicas variáveis em humanos. A base dessa variabilidade relaciona-se, em parte, com a complexidade da população de parasitas, que é constituída por múltiplos clones exibindo atividade biológicas distintas. O maior desafio atual em relação a esse tema é correlacionar a variabilidade genética do parasita com sua patogênese. Dois grupos filogenéticos principais estão bem estabelecidos em *T. cruzi*, presentes tanto no ciclo doméstico quanto no ciclo silvestre. Esses dois grupos possuem características ecológicas, epidemiológicas e biológicas diferentes, atribuindo-se aos parasitas do grupo II os sintomas fisio-patológicos associados à doença. Em alguns países, o grupo I é raramente encontrado em pacientes que vivem em áreas endêmicas e não parece estar associado aos sintomas patológicos. Existe relutância, no entanto, em estabelecer duas espécies para o taxon *T. cruzi* (Zingales et al., 1988; Briones et al., 1999; Buscaglia & Di Noia, 2003). O grupo II foi recentemente subdividido em IIa e IIe (Brisse et al., 2000).

Trabalhos recentes (Machado & Ayala, 2001; Gaunt et al., 2003) mostraram que o *T. cruzi* possui uma capacidade ampla de realizar trocas genéticas através da produção de clones híbridos. Dois clones biológicos transfectados com dois marcadores distintos de resistência de drogas (pTEX com marcador para higromicina e neomicina, respectivamente) perfizeram o ciclo biológico completo. A partir do ciclo no mamífero, obtiveram-se seis clones duplamente resistentes às drogas, demonstrando-se fusão com os genótipos parentais (Gaunt et al., 2003). O mecanismo observado é diferente daquele descrito no tripanosoma africano, *T. brucei* (Bingle et al., 2001). Neste caso, utilizando um sistema de expressão induzido por tetraciclina foi possível visualizar a produção de híbridos de *T. brucei* na mosca

tsetsé. Um clone de *T. brucei* foi transfectado com o gene da GFP (“Green Fluorescent Protein”) sob o controle do repressor Tet *in trans*. A essa construção também foram introduzidos genes de resistência a higromicina e fleomicina. Um cruzamento experimental com o segundo clone parental carregando o gene para resistência a geneticina produziu híbridos fluorescentes resistentes para geneticina e higromicina. Isso é consistente com segregação meiótica. Os híbridos fluorescentes apareciam apenas nas glândulas salivares da mosca, confirmando que a troca genética ocorre no estágio do ciclo de vida do parasita presente nas glândulas salivares.

Os mecanismos de controle da expressão gênica em tripanosomatídeos são encontrados em diversos graus de regulação, com a maioria das vias atuando na etapa de pós-transcrição. Em consequência, os parasitas podem sofrer rápidas mudanças nas transições entre inseto vetor e hospedeiro mamífero. Atualmente é mais evidente que as estratégias regulatórias escolhidas por diferentes espécies de tripanosomatídeos não são as mesmas. Algumas ferramentas disponíveis para manipulação genética como, sistemas indutivos de expressão de genes e inibição específica da expressão gênica por RNA de interferência, foram desenvolvidas ao longo dos estudos com os tripanosomatídeos.

Algumas características incomuns são encontradas nos tripanosomatídeos: transcrição policistrônica, processamento de pré-mRNA através de trans-splicing, editing mitocondrial e transcrição de genes codificadores para proteínas pela RNA polimerase I. A reação de clivagem do pré-mRNA ocorre no núcleo para que se formem mRNAs monocistrônicos. A reação de clivagem do transcrito primário ocorre juntamente com a adição de uma seqüência de 39 nucleotídeos denominada seqüência líder (SL) a 5’ (“cap” metilado) e uma cauda poli-A a 3’ de cada mRNA. A adição da SL é dada por uma reação de transesterificação (trans-splicing) que necessita de uma seqüência conservada AG como sítio de adição da SL. Nos tripanosomatídeos, nenhuma seqüência sinal (consenso) para poliadenilação foi identificada até o momento. Diversos estudos mostram fortes evidências que a adição de SL e a poliadenilação ocorrem concomitantemente ou imediatamente após a transcrição. A seleção de poli-A é governada pela localização do sítio de adição da SL na posição “downstream” do transcrito policistrônico primário. Desta maneira,

como nenhum sinal consenso para poliadenilação foi encontrado, o modelo de que haja uma distância conservada entre o sítio acceptor de “splice” e o sítio de poliadenilação, tornou-se aceito como a solução para este problema (Da Silveira, 2000).

Foi desenvolvida ao longo dos anos a noção da inexistência de introns nos tripanosomatídeos e, em consequência, a falta do aparato para cis-splicing. Estes fatos foram baseados em observações da falta de introns ou um análogo ao snRNA U1. No entanto, foi relatado recentemente a descoberta de seqüências intragênicas interrompendo os genes da poli (A) polimerase em *T. brucei* e *T. cruzi*, estabelecendo a existência de cis-splicing nos tripanosomatídeos (Mair et al., 2000).

A maioria do conhecimento sobre mecanismos de expressão gênica em tripanosomatídeos foi gerado dos protocolos de transfecção. O maior ganho que permitiu melhor controle da manipulação genética nestes parasitas foi o desenvolvimento de um sistema indutivo de expressão de produtos gênicos com repressor de tetraciclina. Trata-se de um sistema operador/repressor que responde a tetraciclina e se mostra como ferramenta excelente na descoberta de funções de genes essenciais e para a expressão de produtos tóxicos (Teixeira & DaRocha, 2003).

A metodologia do RNA de interferência tem sido muito conveniente no estudo de genes de tripanosomatídeos já que técnicas como RNA anti-sense e knockout gênico foram infrutíferas. O RNAi é um mecanismo de silenciamento específico de genes guiado por um RNA dupla fita (dsRNA) que contém seqüências derivadas do gene alvo. Este fenômeno foi relatado pela primeira vez em *C. elegans* (Fire et al., 1998) e também já descrito em *T. brucei* (Ullu et al., 2002). Resumidamente, dsRNAs homólogos a seqüência do gene alvo sintetizados exogenamente ou expressos internamente são processados em RNAs de 20-24 nucleotídeos que trabalham ativamente como guias para degradação de mRNA. O mecanismo exato ainda é desconhecido, mas hoje sabe-se que o processamento pode ocorrer através de um complexo de múltiplas nucleases e se especula que dsRNA ocorrem naturalmente no mecanismo de silenciamento de transposons.

2. OBJETIVOS

Os objetivos desta tese foram:

- a) Sequenciar clones selecionados da biblioteca parcial de cDNA de Tc-85 e analisar os dados de seqüência.
- b) Expressar e purificar algumas proteínas referentes aos clones selecionados.
- c) Realizar ensaios de ligação das proteínas de fusão obtidas desses clones a células de mamífero e a moléculas de matriz extracelular conhecidas (laminina, fibronectina).

3. MATERIAL E MÉTODOS

3.a. OBTENÇÃO DE PARASITAS.

Formas tripomastigotas de *T. cruzi* cepa Y foram obtidas do sobrenadante de cultura de células LLC-MK₂ (50 garrafas de 75 ml) previamente infectadas com 30 a 50 milhões de parasitas. Os parasitas presentes no sobrenadante foram centrifugados a 4.000 g por 15 minutos a 4°C, sendo o precipitado lavado 2 vezes com meio DME. Ao final foram obtidos aproximadamente 6×10^9 parasitas. As células de cultura de tecido foram cultivadas em meio DME suplementado com 10% de soro fetal bovino e infectados no mesmo dia com 2% de soro fetal bovino (Andrews & Colli, 1982).

3.b. EXTRAÇÃO DE RNA

3.b.1. Extração de RNA total.

Foram utilizados 15 ml do reagente Trizol™ (GibcoBRL) para 6×10^9 parasitas. Após a lise e incubação das células com o reagente por 5 minutos à temperatura ambiente, foram adicionados 200µl de clorofórmio para cada 1ml de

Trizol™, o extrato foi agitado gentilmente e incubado por 3 minutos à temperatura ambiente. As frações aquosa e orgânica foram separadas por centrifugação (12000g, 15 min., 4⁰C). O RNA da fase aquosa foi precipitado com isopropanol (0,5ml para cada 1 ml de Trizol™) e após 10 minutos à temperatura ambiente, foi submetido a centrifugação (12000g, 10 minutos, 4⁰C). O precipitado de RNA foi lavado com etanol 75%, com centrifugação de 7500g por 5 min. a 4⁰C. Após a secagem à temperatura ambiente (5-10 minutos), o precipitado foi dissolvido com 200µl de água livre de RNAase e incubado por 10 minutos a 56⁰C. Após a medida de absorbância a 260/280nm, foi feita a correção da concentração do RNA para 0,55 mg/ml, com ajuste da concentração de sal para 0,5M (NaCl 5M), para proceder o isolamento do mRNA.

3.b.2. Extração de mRNA. Foi utilizado o kit MessageMaker™ (GibcoBRL), de acordo com as instruções do fabricante. O RNA total foi aquecido por 5 minutos a 65⁰C e posto em gelo. Adicionou-se o volume de oligo d(T) indicado e incubou-se a mistura de RNA+oligo d(T) por 10 minutos a 37⁰C, transferiu-se a mistura para a seringa-filtro e lavou-se duas vezes o material, como descrito pelo fabricante. A eluição do mRNA foi feita com água aquecida a 65⁰C. A concentração do mRNA obtida foi de aproximadamente 68 ng/µl.

3.c. OBTENÇÃO DA BIBLIOTECA DE CDNA

3.c.1. Iniciadores utilizados na PCR. Estes iniciadores foram utilizados para a obtenção da biblioteca de cDNA enriquecida em membros de Tc-85 (BEM/Tc-85). O alinhamento de alguns genes codificantes para proteínas da família gp85/ transialidases {SA85-1 (n^o no GenBank: M62735), TC85-11 (AF085686), GP85-1 (M64836), GP85-2 (Q03877), SILVIO (L13844), TSA-1 (M58466) e TSA-E3 (U02615)}, mostra regiões altamente conservadas. Com base nestas regiões, foram sintetizados os seguintes iniciadores

(figura 3):

I (5') ATT TTT GTT CCG CAG AAG ACG CAG GTG

II (3') ATT CAG TGG GCG GTT GTA CAG AAA GAC

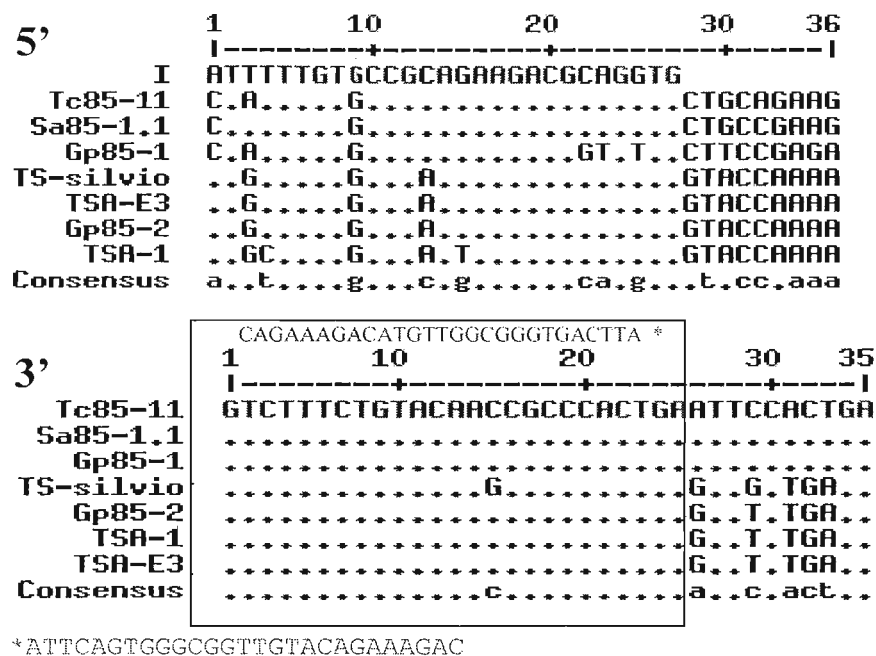


Figura 3: Alinhamento de diversos genes da família das gp85/trans-sialidase e a seqüência dos dois iniciadores desenhados para a construção da biblioteca enriquecida em genes codificantes da família.

3.c.2. Preparação do cDNA. Foi utilizado o kit First-strand cDNA Synthesis (Amersham Pharmacia) para obtenção da primeira fita de cDNA, com a seguinte mistura de reação em 33µl: Mix 11µl; iniciador Not (1:25) 1µl; DTT 1µl; mRNA (136 ng) 2µl. A reação foi realizada por 45 minutos a 37⁰C, 15 minutos a 42⁰C, inativada por 5 minutos a 90⁰C e o tubo de reação foi então mantido em gelo. Este kit é dotado de uma transcriptase reversa de vírus de leucemia murina e o iniciador Not possui uma cauda de 18 timinas. Posteriormente, a primeira fita foi selecionada e amplificada por PCR. À mistura anterior (33µl) adicionaram-se: 40pmol dos oligonucleotídeos I e II (I (5') ATTTTGTGCCGCAGAAGACGCAGGTG/ II (3') ATTCAGTGGGCGGTTGTACAGAAAGAC); 0,5µl de Taq DNA polimerase em volume final de reação de 100µl. O ciclo de PCR utilizado foi: 1min./94⁰C; 45 seg./55⁰C; 2min./72⁰C.

3.c.3. PCR para amplificação do cDNA de Tc-85. A reação de amplificação foi feita com a AmpliTaqTM Gold (Perkin Elmer) e cDNA de Tc-85, usando-se oligonucleotídeos iniciadores (5') ATT TTT GTT CCG CAG AAG ACG CAG GTG / (3') ATT CAG TGG GCG GTT GTA CAG AAA GAC, dNTP, tampão para PCR (Tris-HCl 10 mM, pH 8,3, KCl 50 mM) e MgCl₂ (25mM). A amplificação foi feita após estabelecer experimentalmente as condições de temperatura de reassociação, de concentração de MgCl₂ (4mM) e da quantidade de DNA adequadas à reação de polimerase em cadeia. O ciclo utilizado foi: 1min./94⁰C; 45 seg./55⁰C; 2min./72⁰C.

3.c.4. Preparação de bactérias competentes com cloreto de cálcio. Esta preparação foi feita segundo o método de Hanahan, 1983. Um tubo de bactéria da cepa desejada foi descongelado, mantendo-o por 15 minutos no gelo. As bactérias foram aplicadas em uma placa de LB contendo 10mM de MgSO₄ e incubada a 37°C durante a noite. Uma colônia isolada foi colocada em 2 ml de meio SOB pré-aquecido a 37°C e mantido por aproximadamente 2 horas a 37°C e 200 rpm até D.O._{600 nm} = 0,5. O pré-inóculo foi transferido para 50 ml de meio SOB aquecido e incubado a 37°C e 200 rpm até D.O._{600 nm} = 0,6. O frasco com a cultura de bactérias (tubo de centrífuga de 50 ml) foi mantido em gelo por 15 minutos e a seguir foi adicionado 0.5 ml de MgCl₂ 1M e incubado por mais 15 minutos. A cultura foi centrifugada por 15 minutos a 5000 rpm a 4°C, o sobrenadante descartado e o precipitado ressuspensionado em 10 ml de tampão RF I (vide abaixo) gelado e deixado em repouso no gelo por 15 minutos. Foi realizada a centrifugação por 10 minutos a 5000 rpm a 4°C, o precipitado foi ressuspensionado em 2 ml de RF II (vide abaixo) gelado, sempre se mantendo o material em gelo. As bactérias foram aliqüotadas em 100 µl, congeladas em banho de gelo seco/ acetona e armazenadas por até 2 meses a -80°C.

RF I : KCl 100mM; MnCl₂ . 2 H₂O 50 mM; acetato de potássio pH 7,0 30 mM; Ca Cl₂ . 2 H₂O 10 mM; glicerol 15%. O pH do tampão foi ajustado para 5,8 com ácido acético e esterilizado por filtração.

RF II : KCl 75 mM; Ca Cl₂ . 2 H₂O 75 mM; MOPS pH 7,0 10 mM; MgCl₂ 10 mM; glicerol 15%. O pH do tampão foi ajustado para 6,8 com ácido acético e esterilizado por filtração.

3.c.5. Construção da biblioteca de cDNA de Tc-85. A biblioteca foi construída com a utilização do pCR T7/NT TOPO ® Cloning (Invitrogen). Este vetor está linearizado pela topoisomerase e possui nas extremidades 3' uma timina (T) (figura 4). A ligação do vetor com o produto de PCR contendo adeninas na extremidade 3' é muito eficiente e ocorre em 5 minutos à temperatura ambiente. A topoisomerase liga-se à fita dupla do DNA e cliva o esqueleto fosfodiéster em uma fita no sítio específico 5' C/T (CCTT), podendo religar covalentemente a fita de DNA a que está presa ou pode religar um acceptor heterólogo criando uma molécula recombinante em uma reação de transesterificação (Cheng & Shuman, 2000).

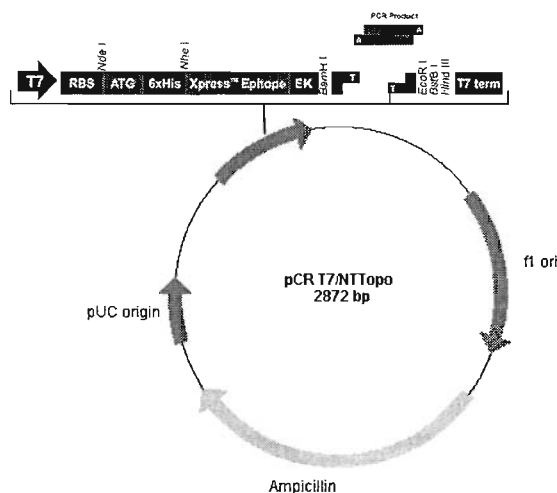


Figura 4. Vetor de clonagem e expressão utilizado na BEM/Tc-85.

3.c.6. Reação de ligação e transformação. A reação de ligação realizada foi a seguinte: 1µl do produto de PCR (56 ng de cDNA); solução salina (200 mM NaCl, 10 mM MgCl₂); 1µl do vetor pCR T7 (10 ng) em um volume final de 5 µl. A mistura foi mantida à temperatura ambiente por 10 minutos. As células competentes (*E. coli* TOP10F'), previamente mantidas em gelo, foram transformadas então com 2µl do produto da reação anterior.

3.d. ANÁLISE DOS CLONES

3.d.1. Mini-preparações de DNA plasmidial.

A cepa bacteriana de *E. coli* DH5 α foi utilizada para propagação e manutenção dos clones selecionados. Uma colônia de bactérias transformada contendo o plasmídeo foi inoculada em meio LB contendo ampicilina (100 μ g/ml) e incubada a 37°C durante a noite a 200rpm. As mini-preparações de DNA plasmidial foram feitas utilizando-se o sistema de extração e purificação de DNA de plasmídeo Concert (Gibco), como indicado no manual.

3.d.2. PCR dos clones (preparação para a análise de restrição). O procedimento executado foi o descrito no manual do fabricante da AmpliTaqTM Gold (Perkin Elmer), utilizando-se 30 ciclos de 94°C/1', 54°C/1', 72°C/2'. O produto da reação de PCR foi purificado com fenol/clorofórmio, precipitado com etanol e ressuspendido em 30 μ l de tampão TE (10 mM Tris, 1 mM EDTA, pH 8,0).

3.d.3. Análise de restrição dos clones sequenciados. Os clones da biblioteca purificados como descrito anteriormente, foram digeridos com Bam H1 e Eco RI (volume da reação de 20 μ l em tampão 50 mM NaCl, 100mM Tris-Cl, 10 mM MgCl₂, 0,025% Triton X-100, 200 μ g/ml BSA, pH 7,5, mantida a 37°C por 2 horas) e o produto da digestão foi aplicado em gel de agarose. O padrão de massa molecular utilizado foi o Lambda Hind III (Gibco).

3.d.4. Purificação dos produtos de PCR. A purificação dos produtos obtidos foi feita utilizando-se o sistema GFXTM Band Gel Purification (Amersham Pharmacia), segundo as instruções do fabricante.

3.d.5. Determinação da concentração de DNA e RNA. Aliquotas de 1 μ l e 5 μ l foram diluídas em 500 μ l de água deionizada. A medida de absorbância foi realizada a 260 nm (pico de absorção de DNA e RNA) e a 280 nm (pico de absorção de proteínas). Considerando que $A_{260nm} = 1$ corresponde a 50 ng/ μ l de DNA dupla fita e

40 ng/ μ l de RNA fita simples, foram determinados os valores de concentração. Através do quociente 260/280 nm foi avaliado o grau de pureza. O resultado perto de 2 revela um alto grau de pureza (Sambrook et al., 1989).

3.d.6. Eletroforese em gel de agarose. Foram utilizados géis de agarose 1% e tampão TAE (50x: Tris 2M, 57,1 ml de ácido acético glacial, EDTA 0,1 M, pH 8,0) na confecção do gel e na eletroforese. Ao gel foram adicionados 20 μ l de brometo de etídio 10 mg/ml (Sambrook et al., 1989). O padrão de peso molecular utilizado foi o DNA do fago lambda digerido com Hind III. As eletroforeses foram feitas a 80 V.

3.d.7. Transformação de bactérias competentes (DH5 α e BI21(DE3)plyS). As transformações realizadas na etapa de obtenção da BEM/Tc-85 seguiram este protocolo. Primeiramente, as células preparadas com cloreto de cálcio foram mantidas em gelo por 10 minutos. Foi adicionado 1 μ l de DNA às bactérias. A mistura foi deixada no gelo por 30 minutos, seguindo-se um choque térmico de 2 minutos a 42°C e posta novamente no gelo por 2 minutos. Foi adicionado 1 ml de meio SOC e as bactérias foram incubadas por 1 hora a 37°C e 200 rpm. As bactérias foram semeadas em placas de ágar LB contendo ampicilina (100 μ g/ml) e incubadas durante a noite a 37°C.

3.e. SEQUENCIAMENTO

3.e.1. Etapas do sequenciamento e iniciadores utilizados. O sequenciamento da biblioteca de cDNA foi efetuado através do método de terminação de cadeia (Sanger *et al*, 1977) pelo sequenciamento automático utilizando-se o kit DNA Sequencing-BigDye™ Terminator (Perkin-Elmer) e equipamento ABI 377 (Perkin-Elmer). Os iniciadores utilizados na primeira etapa associam-se ao vetor utilizado (pCR T7/NT-Invitrogen): T7 Forward TAATACGACTCACTATAGGG e pRSET Reverse TAGTTATTGCTCAGCGGTGG. Os iniciadores utilizados na segunda etapa foram: Tc300-5A TTT TAC TAC CCC ACA ACC AC; Tc300-5B ATT GTA GGA AGC CAT GAA GA, Tc300-3A ACA TGA CGC ACA TTA TTA ACC C, Tc300-3B ACT TAA AGC CAT TTT TGA CC. Os iniciadores utilizados na terceira etapa

foram: cDNA-5A AGGTGTTCTGATGGAGGATG; cDNA-5B
GAGGGGAGGAATAACAATGG; cDNA-3A GTATGGACGACTTGGAGAAG;
cDNA-3B TGGGTATAGAAAAGCTGGAG (tabela 2).

3.e.2. Reações de seqüenciamento (PCR). A reação de PCR, em volume final de 10 µl, foi composta por 100 ng de DNA molde, 10 pmol de um iniciador, 4,0 µl de tampão 2,5x (200 mM Tris-Cl pH 9,0; 5 mM MgCl₂), BigDye mix 2,0 µl e água até 10 µl. A mistura foi submetida a 35 ciclos de 95°C/30'', 56°C/30'' e 60°C/3'.

3.e.3. Lavagem dos Terminadores. Após cada reação de PCR (item 3.e.2.) foram adicionados aos poços da placa 25 µl (2,5x o volume da reação) de etanol absoluto, 1,0 µl de acetato de sódio (3 M, pH 5,2) e 1,0 µl de glicogênio (1g/l). Após a adição do etanol, acetato e glicogênio, a mistura foi bem homogeneizada e incubada por 15 minutos no gelo. Foi realizada uma centrifugação de 30 minutos, 6000 rpm a 7 °C e o sobrenadante desprezado e o líquido restante foi eliminado com uma centrifugação de 30 segundos, não superior a 700 g. Foram realizadas duas lavagens seguidas, adicionando-se 50 µl de etanol 70% por reação, homogeneizando por alguns segundos, seguida de centrifugação por 8 minutos, 6000 rpm a 7 °C. O líquido da placa de PCR foi retirado e o líquido restante foi eliminado com uma centrifugação de 30 segundos e não superior a 700 g.

3.e.4. Gel de seqüenciamento. Em um volume final de 50 ml, foram misturados para a constituição do gel 18g de uréia, 25 ml de água milli-Q, 5,3 ml de acrilamida comercial 50% (final 5,3 %), 5,0 ml de TBE 10x e, finalmente, para a polimerização do gel, 250 µl de persulfato de amônio 10% e 35 µl de Temed.

3.e.5. Tratamento das amostras de seqüenciamento. Foram aplicados em cada poço da placa 1,4 µl de "loading buffer" (6x diluído em formamida deionizada). O DNA foi denaturado a 95 °C por 4 minutos e mantido em gelo até a aplicação da amostra.

3.e.6. Aplicação das amostras no gel. Foram colocados TBE 0,2x no compartimento superior (660 ml) e 1x no inferior (700 ml) do suporte do gel. Foram aplicados 1,3 µl das amostras nos poços ímpares e realizada uma pré-corrída por 4 minutos para que as amostras migrassem pelo gel e não se misturassem com as amostras a serem aplicadas nos poços pares. Com uma pausa no aparelho, os poços pares foram lavados com uma seringa de 1,0 ml para a aplicação das amostras pares. Foram removidos 49,5 ml do tampão superior e adicionado o mesmo volume de TBE 10x para acerto da concentração do tampão. A pré-corrída foi cancelada e iniciou-se a corrída de 7 horas. Após a corrída, o gel foi analisado. Os erros de “tracking” do aparelho foram corrigidos manualmente para que houvesse um melhor aproveitamento da qualidade do gel. Em seguida, com base neste “tracking” os cromatogramas foram obtidos com a extração das fileiras.

3.e.7. Análise de seqüências. Para a montagem dos diversos fragmentos produzidos pelas reações de sequenciamento foram seguidas as seguintes etapas:

- a- obtenção dos arquivos no formato FASTA a partir dos cromatogramas;
- b- montagem dos fragmentos seqüenciados (CAP3: - Huang & Madan, 1999- <http://bio.ifomfirc.it/ASSEMBLY/assemble.html>) com os parâmetros de sobreposição para comprimento maior do que 30 nucleotídeos e, para porcentagem de identidade do “overlaps”, maior do que 75%;
- c- determinação do passo de leitura aberto (ORF) dos fragmentos (ORF finder, <http://www.ncbi.nlm.nih.gov> e algoritmo Smith-Waterman, <http://www.dna.affrc.go.jp/swsrch/swx.pl>). Montagem das ORFs geradas e comparação com os dados do GenBank: Blastp e Blastx (compara as seqüências traduzidas com as residentes no GenBank).
- d- ClustalW (<http://www.ebi.ac.uk/clustalw/>): realiza alinhamentos múltiplos.
- e- BioEdit: é um programa multitarefa. Faz a edição dos arquivos gerados pelo ClustalW para que o alinhamento tenha uma visualização facilitada, dentre outras funções.

3.f. ANÁLISE DE PROTEÍNAS.

3.f.1. Obtenção de proteína recombinante. Esta etapa foi realizada utilizando-se a bactéria BL21(DE3)plyS, própria para expressão. Foi adicionado 1 μ l (cerca de 50 ng) do DNA já purificado pela minipreparação às células descongeladas. A mistura foi incubada por 30 minutos em gelo, seguido de choque térmico (42°C por 30 segundos), retornando em seguida o tubo ao gelo. Foram adicionados às bactérias 250 μ l do meio SOC seguindo-se uma incubação de 30 minutos a 37°C e 200 rpm. As bactérias foram adicionadas a 10 ml de LB contendo 100 μ g/ml de ampicilina e 34 μ g/ml de cloranfenicol. A cultura foi incubada durante a noite a 37°C e 200 rpm. Para a expressão da proteína recombinante, foram inoculados 500 μ l dessa cultura incubada durante a noite em 10 ml de LB contendo 100 μ g/ml de ampicilina e 34 μ g/ml de cloranfenicol. O inóculo foi mantido por aproximadamente 2 horas a 37°C e 200 rpm até $D.O_{600\text{ nm}}=0,6$. Neste ponto foram retirados 1 ml da cultura (não induzido) e adicionado 1mM de IPTG. Alíquotas de 1 ml foram retiradas em 1 e 4 horas de incubação e tratadas com tampão de amostra para análise em SDS-PAGE.

3.f.2. Purificação de proteínas recombinantes. As proteínas recombinantes com cauda de histidina foram purificadas por cromatografia de afinidade em “Sephacel Chelating Fast Flow” (Amersham Pharmacia Biotech) ligada a Ni^{2+} . O precipitado bacteriano foi ressuscitado em tampão 50mM NaH_2PO_4 pH8,0, 300mM NaCl, 10 mM de imidazol, e lisado por sonicação (65 Hz e 5 pulsos de 15 segundos durante 10 minutos). O lisado foi centrifugado a 5.000 rpm por 15 minutos, o sobrenadante (fração solúvel) foi armazenado a -20°C e uma alíquota foi separada para análise em gel de SDS-PAGE. O precipitado (fração insolúvel) foi solubilizado em tampão de solubilização (8M de uréia, 0,5M NaCl, 50 mM de Tris-Cl pH 8,0, 5mM de Beta-mercaptoetanol em um volume de 10 ml/litro de cultura). Uma alíquota foi armazenada para posterior análise em gel de SDS-PAGE e o restante foi incubado por 4 a 18 horas, à temperatura ambiente. Após análise por SDS-PAGE das frações, a fração contendo a proteína de fusão foi centrifugada (15.000 rpm/30 minutos) para retirada de material insolúvel e submetida a cromatografia na coluna de níquel, previamente equilibrada com o mesmo tampão da amostra. A coluna foi lavada com tampão de lavagem (40 mM de imidazol, 0,5M NaCl, 50 mM, Tris-Cl pH 6,0) e

eluída com tampão de eluição (1M de imidazol, 0,1M NaCl, 50 mM Tris-Cl pH 6,0). O fluxo na coluna foi mantido a 0,25 ml/min e o material não retido foi recolhido em tubos gelados. Foram então passados pela coluna, em um fluxo de 1ml/min, 30 ml de cada um dos seguintes tampões (processo de “refolding” da proteína):

- a- 8M de uréia, 0,5M NaCl, 50 mM de Tris-Cl Ph 8,0, 5mM de Beta-mercaptoetanol;
- b- 4M de uréia, 0,25M NaCl, 50 mM de Tris-Cl pH 6,0, 2,5mM de Beta-mercaptoetanol;
- c- 2M de uréia, 0,125M NaCl, 50 mM de Tris-Cl pH 6,0, 1,25mM de Beta-mercaptoetanol;
- d- 40 mM de imidazol, 0,5M NaCl, 50 mM de Tris-Cl pH 6,0.

A proteína foi eluída a 0,25 ml/min com o tampão de eluição (1M de imidazol, 0,1M NaCl, 50 mM Tris-Cl pH 6,0) coletando-se frações de 1 ml e dialisadas amostras contra PBS.

3.f.3. Dosagem de proteínas. As proteínas purificadas foram dosadas pelo método de Bradford, 1976. Foi medida a absorvância no comprimento de onda de 595 nm. A concentração de proteínas foi calculada a partir de uma curva padrão de albumina sérica (BSA).

3.f.4. Eletroforese em gel de poliacrilamida (SDS-PAGE). O método utilizado de eletroforese vertical de gel de poliacrilamida contendo dodecil sulfato de sódio (SDS-PAGE) foi descrito por Laemmli, 1970. A separação foi feita em gel na concentração contínua de 9% de poliacrilamida e o gel de empilhamento foi preparado na concentração de 3% de poliacrilamida. As alíquotas a serem analisadas (extrato total de bactéria) foram coletadas como descrito e solubilizadas em 50µl de tampão de amostra (Tris-HCl pH 6,8, glicerol 38%, EDTA 0,2 M, SDS 20%, Beta-mercaptoetanol 4% e azul de bromofenol), aquecidas a 100°C por 5 minutos, centrifugadas em microcentrífuga por 2 minutos, sendo então o sobrenadante aplicado nas canaletas do gel de empilhamento. A eletroforese foi feita à amperagem constante de 20 mA até que o corante de acompanhamento (azul de bromofenol) atingisse a borda inferior do gel. Em seguida, o gel foi corado com uma solução de 0,2% de azul de Coomassie em metanol/ácido acético/água (proporção de 45:10:45),

por 30 minutos e depois descorado em ácido acético a 7% ou então submetido a Western blot. O padrão de peso molecular usado foi o SDS-6H (Sigma).

3.f.5. Western blot. As preparações foram submetidas a eletroforese em SDS-PAGE como descrito e foram transferidas para membrana de nitrocelulose (Bio Rad) a 20 V por 16-18 horas a 4°C em tampão de transferência (Towbin *et al.*, 1979). Após a transferência, a membrana de nitrocelulose foi corada com Ponceau S 0,2% por 10 minutos e lavada com água. O bloqueio da nitrocelulose foi feito em tampão TBS 50 mM e leite em pó desnatado 3%, pH 8,0 por 60 minutos à temperatura ambiente, com lavagens subsequentes em TBS 50mM-Tween 20 0,03%, pH 8,0. Em seguida, adicionou-se o anticorpo monoclonal anti-His tag em TBS 50mM-Tween 20 0,03% e Triton X-100 0,05%, pH 8,0, mantendo-se a 4°C por 16-18 horas. A membrana foi então lavada seqüencialmente com: TBS 50 mM pH 8,0, TBS 50mM-Tween 20 0,03% , pH 8,0 e TBS 50 mM, pH 8,0 por 5 a 10 minutos, sendo cada lavagem feita à temperatura ambiente. Repetiu-se o bloqueio como descrito anteriormente. Acrescentou-se como segundo anticorpo, anti-IgG de camundongo conjugado com peroxidase (Sigma), diluído em tampão de bloqueio, deixando-se à temperatura ambiente por 2 horas. Lavou-se 3 vezes com TBS 50mM-Tween 20 0,03%, pH 8,0 por 45 minutos e adicionou-se a solução de revelação ECL (Amersham Pharmacia).

3.f.6. Dot Blot. As membranas de nitrocelulose contendo quantidades crescentes de proteínas de matriz (laminina e fibronectina) na ordem de 0,5 µg, 1,0 µg, 2,0 µg, 3,0 µg e 5,0 µg, foram, antes de utilizadas, incubadas com Ponceau S 0,2 %. As membranas foram a seguir bloqueadas com leite em pó desnatado 3%, em TBS 50mM-Tween 20 0,03%, ou em TCMM (25 mM Tris, pH 7,5; 50 mM NaCl; 5 mM CaCl₂; 5mM MgCl₂; 1 mM MnCl₂) por 60 minutos à temperatura ambiente, com lavagens subsequentes em TBS 50mM-Tween 20 0,03%, pH 8,0 ou em TCMM. Adicionou-se proteína recombinante em diferentes quantidades (10 µg e 30 µg) em tampão apropriado (de acordo com a proteína de matriz) com 3% de leite em pó desnatado e incubou-se por 16 horas à temperatura ambiente. Em seguida, adicionou-se o anticorpo monoclonal anti-His tag em tampão apropriado. A membrana foi lavada seqüencialmente com o tampão utilizado por 5 a 10 minutos, sendo cada lavagem feita à temperatura ambiente. Acrescentou-se como segundo anticorpo, anti-

IgG de camundongo conjugado com peroxidase (Sigma), diluído em tampão de bloqueio, deixando-se à temperatura ambiente por 2 horas. Lavou-se 3 vezes tampão apropriado por 45 minutos e adicionou-se a solução de revelação ECL (Amersham Pharmacia).

4. RESULTADOS E DISCUSSÃO

4.1. Biblioteca enriquecida em membros da família da Tc-85 (BEM/Tc-85)

A biblioteca foi obtida utilizando-se os primers consensos I e II (I (5') ATTTTTGTTCCGCAGAAGACGCAGGTG/ II (3') ATTCAGTGGGCGGTTGTACAGAAAGAC) (figura 3) desenhados para flanquear as extremidades 5' e 3' dos genes da superfamília das gp85/transialidases {SA85-1.1 (AAA30245), TC85-11 (AF085686), GP85-1 (AAA30150), GP85-2 (A45622), SILVIO (AAA10827), TSA-1 (AAA30259) e TSA-E3 (AAD10620)}. Foi obtida uma única banda dos insertos com um tamanho de cerca de 2kb (figura 5). De 352 colônias bacterianas obtidas foram selecionadas 60 colônias para os estudos descritos neste trabalho. Todos os clones obtidos desta biblioteca foram analisados em gel de agarose 1%. A análise mostrou que 52 de um total de 60 clones isolados (86%) apresentam insertos com o tamanho esperado (cerca de 2 kb) (figura 6). Embora, nem todos os membros clonados devam pertencer à família da Tc-85, por facilidade, os novos genes e suas proteínas de fusão serão denominados de Tc85, seguido do número do clone, até que seja feita a classificação final dos membros da superfamília a partir do sequenciamento do genoma do *T. cruzi*.

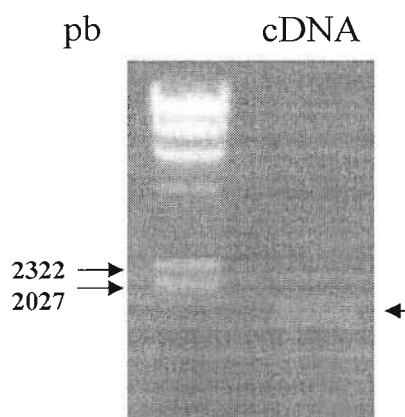


Figura 5. Gel de agarose 1% corado com brometo de etídio. O cDNA de Tc-85 (seta) foi amplificado e clonado no vetor pCR T7/NT TOPO (Invitrogen) para a construção da BEM/Tc-85.

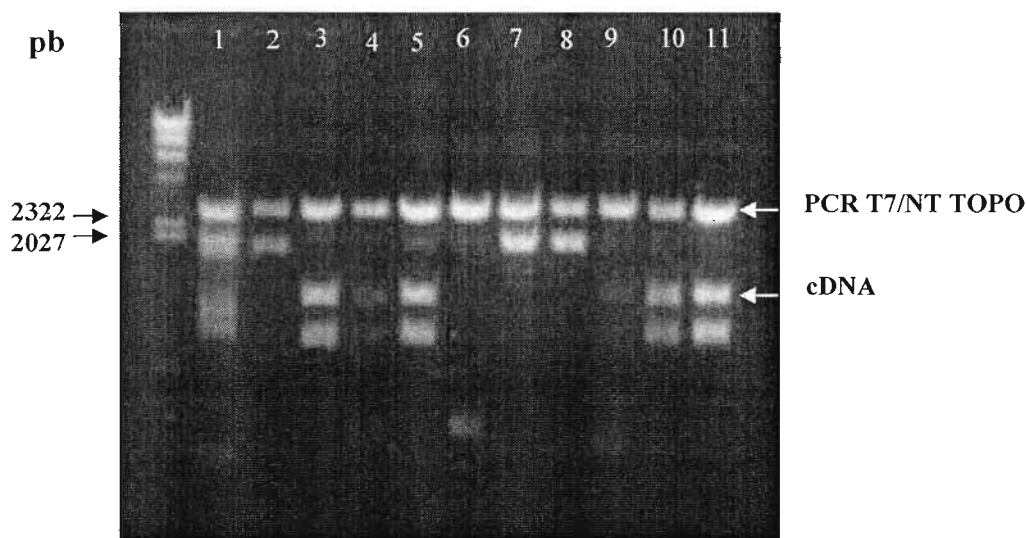


Figura 6. Gel de agarose 1% corado com brometo de etídio. Após purificação de DNA plasmidial de diferentes colônias de bactérias DH5 α transformadas com a BEM/Tc-85 em PCR T7/NT TOPO, os 60 clones foram digeridos com Bam HI e Eco RI. Nesta foto estão indicadas as digestões dos clones 1 ao 11 indicados pelos números na parte superior. As setas indicam as bandas correspondentes ao vetor e aos insertos de Tc-85 presentes na maioria dos clones selecionados. Os insertos possuem fragmentos de cDNA de cerca de 2kb.

4.2. Sequenciamento e análise de seqüências

Para uma ampla caracterização da BEM/Tc-85, foi necessário também o sequenciamento de um número razoável de genes e a obtenção das proteínas de fusão correspondentes para estudos de ligação às moléculas de matriz (citadas no item "Objetivos") e à superfície celular. Foi realizado o sequenciamento completo dos clones selecionados da BEM/Tc-85. O sequenciamento foi efetuado através do método de terminação de cadeia (Sanger *et al*, 1977) pelo sequenciamento automático utilizando-se o kit DNA Sequencing-BigDye™ Terminator (Perkin-Elmer) e equipamento ABI 377 (Perkin-Elmer).

O sequenciamento foi realizado em três etapas, sendo utilizados primers específicos em cada uma delas. Os primers utilizados na primeira etapa associam-se ao vetor utilizado (pCR T7/NT-Invitrogen): T7 Forward TAATACGACTCACTATAGGG e pRSET Reverse TAGTTATTGCTCAGCGGTGG. Os primers (associando-se na altura do nucleotídeo número 300 do inserto clonado) utilizados na segunda etapa foram:

Tc300-5A TTTTACTACCCCACAACCAC; Tc300-5B
ATTGTAGGAAGCCATGA AGA, Tc300-3A
ACATGACGCACATTATTAACCC, Tc300-3B
ACTTAAAGCCATTTTTGACC. Os primers (associando-se na altura do
nucleotídeo número 500) utilizados na terceira etapa foram: cDNA-5A
AGGTGTTCTGATGGAGGATG; cDNA-5B GAGGGGAGGAATAACAATGG;
cDNA-3A GTATGGACGACTTGGAGAAG; cDNA-3B
TGGGTATAGAAAAGCTGGAG (tabela 2).

O sequenciamento de 60 clones resultou em 30 ORFs completas, as quais foram analisadas. Por análises comparativas, identidades de 40-90% foram encontradas entre as seqüências nucleotídicas e de 30-65% com membros da superfamília de proteínas das gp85/trans-sialidases (BLASTX). Verificou-se que, dos 60 clones sequenciados, 34 apresentavam-se invertidos. Este fato já era esperado, uma vez que, pela própria natureza do vetor, os insertos poderiam ser ligados nos dois sentidos. A comparação entre as seqüências nucleotídicas obtidas e alguns cromatogramas estão, respectivamente, nos anexos 1 e 2.

Os genes identificados foram inseridos no GenBank (<http://www.ncbi.nlm.nih.gov/>) e estão descritos a seguir com números de acesso ainda não liberados. Foram clonadas as seguintes seqüências:

Tc85-2, Tc85-12, Tc85-16, Tc85-19, Tc85-20, Tc85-25, Tc85-27, Tc85-29,
Tc85-32, Tc85-35, Tc85-36, Tc85-40, Tc85-44, Tc85-45, Tc85-52, Tc85-54,
Tc85-10, Tc85-13, Tc85-14, Tc85-18, Tc85-22, Tc85-23, Tc85-26, Tc85-33,
Tc85-34, Tc85-41, Tc85-43, Tc85-48, Tc85-50, Tc85-57.

Tabela 2. A tabela contém as seqüências dos primers específicos utilizados nas etapas do seqüenciamento, derivadas do alinhamento das seqüências nucleotídicas obtidas e presentes em regiões conservadas. Clones que não geraram reações de seqüenciamento em algumas das condições indicadas, foram omitidos da tabela. Os clones não indicados são 4, 6, 9, 15, 21 (cDNA-5A), 15, 60, 12, 6, 31, 47, 4 (cDNA-3A), 37, 21, 42, 28, 56 (cDNA-3B), 1, 4, 6, 9, 12, 15, 21, 27, 28, 39, 42, 46, 56 (Tc300-5A e 5B), 1, 4, 6, 9, 12, 15, 21, 24, 28, 42, 47 (Tc300-3A e 3B).

Primers	Clones Tc-85
T7 <i>Forward</i> TAA TAC GAC TCA CTA TAG GG	Todos
pRSET <i>Reverse</i> TAG TTA TTG CTC AGC GGT GG	Todos
<i>Forward</i> cDNA-5A AGG TGT TCT GAT GGA GGA TG	7, 8, 19, 27, 36, 44, 20, 45, 37, 51, 2, 16, 32, 35, 40, 52, 54, 25, 11, 21
cDNA-5B GAG GGG AGG AAT AAC AAT GG	5, 3, 28, 47, 13, 18, 24, 50, 53, 60, 14, 17, 22, 38, 41, 48, 23, 26, 30, 34, 43, 57, 58, 10, 33, 55, 59, 49, 31
<i>Reverse</i> cDNA-3A GTA TGG ACG ACT TGG AGA AG	39, 51, 13, 59, 3, 5, 30, 34, 43, 17, 18, 22, 23, 26, 33, 41, 46, 48, 10, 58, 14, 38, 50, 53, 57, 55, 49
cDNA-3B TGG GTA TAG AAA AGC TGG AG	2, 25, 11, 40, 35, 32, 54, 52, 16, 27, 29, 20, 44, 45, 7, 8, 19, 36
<i>Forward</i> Tc300-5A TTT TAC TAC CCC ACA ACC AC	2, 20, 19, 37, 32, 52, 25, 7, 36, 45, 51, 40, 35, 31, 8, 44, 27, 16, 54, 11, 47
Tc300-5B ATT GTA GGA AGC CAT GAA GA	3, 14, 18, 59, 53, 48, 38, 33, 26, 5, 57, 58, 22, 30, 17, 24, 34, 49, 55, 41, 10, 13, 43, 60, 50, 23
<i>Reverse</i> Tc300-3A ACA TGA CGC ACA TTA TTA ACC C	2, 27, 54, 52, 44, 45, 37, 51, 7, 16, 35, 19, 29, 49, 60, 39, 8, 32, 40, 20, 36, 55, 31
Tc300-3B ACT TAA AGC CAT TTT TGA CC	3, 22, 58, 38, 26, 17, 34, 41, 5, 10, 48, 33, 23, 14, 50, 46, 53, 59, 43, 30, 18, 13, 57, 11

As 30 seqüências obtidas foram submetidas ao algoritmo Smith-Waterman de busca de “frame” (Smith & Waterman, 1981) para a determinação do passo de leitura aberto correto ou pelo menos o mais próximo do real. Com as seqüências geradas, foi feita a comparação uma a uma daquelas com identidade maior de 90% para verificar se havia alguma idêntica entre si. Por fim, restaram 24 seqüências seguramente distintas. As seqüências restantes podem ser, na verdade, idênticas ou possuir mutações pontuais ao longo dos genes. Encontram-se nesses possíveis casos os clones 44/45/36, 50/57 e 35/52/32. Na tabela 3 estão mostradas todas as identidades entre as seqüências nucleotídicas.

Os pesos moleculares esperados para os polipeptídeos, a partir das seqüências obtidas, é de cerca de 67.000 Da (a clonagem foi realizada de maneira a excluir o peptídeo sinal e a região de inserção da âncora de GPI), com uma grande variação de pontos isoelétricos (pI de 5,0 a 9,0) (tabela 4). A análise de BLASTX mostrou que todos os produtos tiveram um grau alto de identidade com seqüências da superfamília das gp85/trans-sialidases. Foi estabelecida uma divisão dos clones em dois grupos (I e II). O grupo T, com maior número de membros, possui identidade $\geq 50\%$ com glicoproteínas de superfície de tripomastigotas (tabela 5) e o grupo A possui identidade $\geq 50\%$ com glicoproteínas de superfície de amastigotas (tabela 6). Todos os clones foram analisados quanto à presença de seqüências características da família, como a FLY e o domínio Asp box (SxDxGxTW) (tabela 7). Todos os membros apresentam pelo menos um domínio Asp, enquanto que a maioria possui os dois domínios. Um terceiro domínio ASP degenerado- STDxSxxDTW está presente na maioria dos membros. A seqüência FLY (VTVXNVFLYNR), característica dos membros da superfamília, apresentou variantes na seqüência dos 3 primeiros aminoácidos: VkV (9), Vtm (13) e Vpm (1), mais representados no total do que a seqüência VTV. Magdesian et al., 2001, demonstram que a seqüência mínima ligante de citoqueratina 18 é VTxVFLYxR.

Outras seqüências conhecidas na família da Tc-85 foram analisadas. O epítipo do anticorpo monoclonal H1A10 está presente em 16 dos 30 clones, sendo encontrado em 100% dos membros do grupo I e ausente nos membros do grupo II. Este dado é interessante, uma vez que, o anticorpo monoclonal H1A10 inibe a invasão de células de cultura por tripomastigotas, mas também reconhece

especificamente as formas tripomastigotas e foi utilizado para a definição de Tc-85 (Alves et al., 1986; Abuin et al., 1989; Abuin et al., 1996). Estes dados sugerem que o grupo I possa ser específico das formas tripomastigotas. No entanto, na população de parasitas, somente cerca de 50% dos tripomastigotas de cultura expressam o epítipo reconhecido pelo Mab H1A10 (Abuin et al., 1989). Estes resultados poderiam ser explicados, por exemplo, por: (1) se houver uma substituição de membros do grupo II pelo grupo I durante o amadurecimento dos tripomastigotas após o rompimento da célula hospedeira, admitindo-se que só o tripomastigota infectante (maduro) expressaria o grupo I; (2) se os tripomastigotas infectantes expressam individualmente só membros do grupo I ou do grupo II (ou de outros grupos não selecionados pelo método de clonagem e também importantes na invasão).

Ainda dentro dos clones obtidos, foram analisadas seqüências envolvidas em adesão. A primeira delas foi a seqüência RGD que está presente em todos os clones do grupo I, excetuando o clone 12 (tabela 7). Proteínas que contêm RGD como sítio de adesão, juntamente com seus receptores, as integrinas, constituem um dos maiores sistemas de reconhecimento para adesão celular (Ruoslahti, 1996). Dentre as proteínas que contêm RGD, podem ser citadas a fibronectina, vitronectina, trombospondina e laminina. RGD está também presente nas desintegrinas, proteínas de veneno de cobras, capazes de se ligar a integrinas, interferindo na sua função (Silva et al., 2003). No entanto, um grande número de moléculas contém a seqüência RGD e não tem qualquer papel em adesão celular. O possível papel de RGD no grupo I necessita ser verificado. É interessante notar, no entanto, que a seqüência RGD_{FL/S} é conservada no grupo I e substituída por RGYT/A no grupo II (anexo 3). Proteínas contendo RGD e relacionadas com adesão foram descritas em *Bordetella pertussis* (Leininger et al., 1991), *Yersinia pseudotuberculosis* (Van Nhieu & Isberg, 1991) e em diversos vírus (Ensoli et al., 1990; Roivainen et al., 1994; Huang et al., 1995; Lea et al., 1995).

A segunda análise foi feita para verificar a presença de 3 peptídeos implicados na ligação de TC85-11 a laminina e localizados dentro do domínio amino-terminal da molécula: peptídeos N-17, N-20 e N-21 (Quelopana, 2003). Como se pode verificar (tabela 7), esses peptídeos são encontrados na grande

maioria dos membros do grupo I, mas não no grupo II. É interessante notar ainda que os peptídeos N-20 e N-21 estão presentes em 13 clones, ao passo que o N-17 está presente somente em 7 deles. Se essa variação é relevante para a possível ligação das proteínas a laminina é um ponto a ser verificado.

A terceira análise foi avaliar a presença de peptídeos da gp82 (P4 e P8) relacionados na adesão de células HeLa (Pereira et al., 1999) localizados aproximadamente no meio da molécula. O peptídeo P4 aparentou ser mais eficiente na adesão e também exibiu uma maior inibição da invasão a célula pelo *T. cruzi*. Apenas o peptídeo P4 foi encontrado na grande maioria dos membros do grupo I e II, mas o peptídeo P8 não foi encontrado em nenhum dos clones (tabela 7). Como P4 tem importância relevante na adesão e inibição da invasão a célula hospedeira, sendo encontrado na maioria dos membros, possivelmente dados futuros comprovarão o papel deste epítipo.

A posição das seqüências relevantes para a Tc85-45, um dos clones selecionados para estudos mais detalhados, é mostrada nas figuras 13 e 14.

Com a utilização do programa ClustalW (Thompson et al., 1994) para o multialinhamento das seqüências de nucleotídeos, sem as regiões correspondentes aos primers, foram observados dois grupos principais com identidades acima e abaixo de 50%. Os dois grupos estão representados na árvore da análise de Neighbor Joining (figura 7). A partir de uma árvore derivada do alinhamento de seqüências deduzidas de aminoácidos realizou-se a análise de “bootstrap” (Felsenstein, 1985), com busca heurística completa, com 100 e 500 repetições representadas, respectivamente, com valores em caracteres normal e itálico/sublinhado (figura 8).

Durante a comparação das seqüências foram verificadas deleções entre seqüências muito semelhantes. Algumas destas deleções estão mostradas nas figuras 9, 10, 11 e 12. A tabela 8 sumariza as deleções encontradas.

A figura 14 mostra o alinhamento de aminoácidos entre Tc85-45, SA85-1.1, Tc85-11 e gp82, com destaque para as regiões conservadas dos peptídeos P4 e P8 da gp82, envolvidos na invasão a célula hospedeira (Pereira et al., 1999). Como os membros das gp85/trans-sialidases, incluindo as Tc-85, são

glicoproteínas, os sítios putativos de N-glicosilação para a Tc85-45 e Tc85-34 estão indicados na figura 15.

Apesar de ter sido obtida uma banda única de cDNA na construção da BEM/Tc-85, foi possível a caracterização de seqüências gênicas diversas, o que pode indicar que a metodologia de clonagem, aplicando-se primers consensos longos de “média especificidade” e não primers curtos degenerados, pode ser útil no estudo de famílias multigênicas. Com este estudo, demonstrou-se a existência de no mínimo dois grupos de genes. Nos dois grupos, há grandes ou pequenas similaridades, blocos de deleção, sítios de possíveis mutações pontuais, blocos conservados dentre outras características. Uma hipótese para os dados obtidos é a de que as centenas de genes da superfamília das gp85/trans-sialidases resultam em muitas proteínas expostas na superfície celular que, com motivos combinados de várias maneiras, ligam-se especificamente às proteínas do arcabouço extracelular e desligam-se com a eliminação de proteases, dando prosseguimento à infecção.

Tabela 3. Identities entre seqüências de nucleotídeos obtida no programa BioEdit. Os valores estão divididos por 100. Os valores sublinhados indicam os alinhamento entre as seqüências aparentemente idênticas (o alinhamento mostrou diferenças mínimas).

22	43	57	50	13	41	48	10	23	44	45	36	20	19	16	27	35	54	52	32	40	2	29	12	25	14	18	33	26	34	
22	1	0.958	0.976	0.974	0.967	0.938	0.951	0.94	0.62	0.629	0.628	0.619	0.595	0.633	0.633	0.621	0.56	0.621	0.626	0.617	0.5	0.509	0.585	0.5	0.882	0.627	0.651	0.65	0.633	
43	---	1	0.947	0.944	0.942	0.913	0.928	0.914	0.603	0.612	0.61	0.601	0.577	0.617	0.617	0.606	0.576	0.605	0.611	0.601	0.487	0.496	0.571	0.488	0.856	0.609	0.631	0.635	0.618	
57	---	---	1	0.969	0.963	0.962	0.942	0.966	0.937	0.619	0.627	0.626	0.617	0.594	0.632	0.631	0.622	0.559	0.621	0.627	0.615	0.498	0.51	0.587	0.505	0.882	0.628	0.654	0.655	0.635
50	---	---	---	1	0.958	0.958	0.927	0.944	0.947	0.622	0.631	0.629	0.62	0.596	0.633	0.633	0.62	0.559	0.621	0.625	0.62	0.501	0.511	0.582	0.503	0.88	0.625	0.65	0.656	0.634
13	---	---	---	---	1	0.955	0.924	0.942	0.931	0.617	0.622	0.62	0.612	0.588	0.626	0.626	0.615	0.554	0.614	0.62	0.611	0.491	0.503	0.58	0.499	0.877	0.616	0.642	0.647	0.628
41	---	---	---	---	---	1	0.926	0.937	0.932	0.619	0.628	0.628	0.618	0.597	0.633	0.633	0.621	0.561	0.62	0.626	0.618	0.502	0.511	0.584	0.498	0.873	0.62	0.644	0.644	0.628
48	---	---	---	---	---	---	1	0.919	0.904	0.608	0.617	0.616	0.609	0.583	0.622	0.622	0.61	0.546	0.609	0.614	0.604	0.487	0.498	0.586	0.489	0.852	0.621	0.65	0.641	0.62
10	---	---	---	---	---	---	---	1	0.917	0.617	0.627	0.625	0.617	0.593	0.632	0.631	0.621	0.561	0.619	0.625	0.616	0.498	0.508	0.585	0.5	0.87	0.624	0.646	0.648	0.63
23	---	---	---	---	---	---	---	---	1	0.613	0.621	0.621	0.613	0.588	0.624	0.626	0.611	0.55	0.608	0.615	0.61	0.498	0.503	0.581	0.492	0.856	0.614	0.645	0.643	0.629
44	---	---	---	---	---	---	---	---	---	1	0.985	0.979	0.96	0.927	0.926	0.954	0.896	0.82	0.899	0.895	0.881	0.72	0.708	0.7	0.609	0.637	0.464	0.472	0.474	0.464
45	---	---	---	---	---	---	---	---	---	---	1	0.989	0.971	0.936	0.938	0.966	0.907	0.831	0.904	0.905	0.891	0.728	0.713	0.71	0.616	0.645	0.469	0.479	0.481	0.469
36	---	---	---	---	---	---	---	---	---	---	---	1	0.97	0.932	0.934	0.962	0.903	0.828	0.9	0.902	0.889	0.727	0.712	0.708	0.614	0.643	0.469	0.475	0.48	0.47
20	---	---	---	---	---	---	---	---	---	---	---	---	1	0.919	0.917	0.943	0.886	0.813	0.884	0.886	0.873	0.718	0.699	0.701	0.609	0.63	0.464	0.472	0.474	0.469
19	---	---	---	---	---	---	---	---	---	---	---	---	---	1	0.882	0.907	0.858	0.788	0.856	0.854	0.847	0.7	0.686	0.668	0.59	0.612	0.451	0.459	0.454	0.449
16	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1	0.965	0.913	0.834	0.908	0.909	0.897	0.698	0.681	0.716	0.614	0.644	0.47	0.478	0.49	0.474
27	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1	0.917	0.836	0.913	0.913	0.902	0.713	0.695	0.712	0.62	0.645	0.47	0.48	0.485	0.473
35	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1	0.892	0.976	0.971	0.954	0.679	0.663	0.694	0.61	0.625	0.465	0.473	0.47	0.464
54	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1	0.884	0.892	0.873	0.626	0.611	0.628	0.551	0.561	0.421	0.425	0.426	0.421
52	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1	0.973	0.961	0.882	0.663	0.69	0.613	0.624	0.463	0.474	0.472	0.464
32	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1	0.956	0.68	0.66	0.7	0.61	0.628	0.468	0.476	0.476	0.472
40	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1	0.67	0.655	0.688	0.606	0.621	0.463	0.468	0.468	0.464
2	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1	0.682	0.563	0.583	0.522	0.441	0.428	0.439	0.432
29	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1	0.543	0.571	0.525	0.45	0.437	0.438	0.444
12	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1	0.644	0.586	0.455	0.459	0.45	0.439
25	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1	0.515	0.448	0.447	0.429	0.424
14	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1	0.579	0.61	0.596	0.593
18	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1	0.637	0.57	0.586
33	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1	0.562	0.583
26	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1	0.641
34	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1

Tabela 4. pI e massa molecular (Da) teóricos dos membros clonados da BEM/Tc-85.

CLONES	PI	MASSA MOLECULAR
12	5,31	61.250
14	5,47	65.646
16	5,53	66.820
35	5,54	64.247
19	5,56	65.664
25	5,76	65.328
26	5,86	64.372
48	5,86	63.293
44	5,90	64.700
36	5,90	64.716
27	5,90	67.331
13	5,97	65.265
20	6,03	67.370
45	6,17	67.357
32	6,27	64.108
50	6,32	62.633
29	6,32	67.394
22	6,36	65.461
41	6,47	64.943
52	6,76	66.277
54	7,19	60.584
43	7,75	63.743
57	7,76	65.419
33	8,16	63.707
23	8,32	64.908
34	8,39	64.425
02	8,48	66.290
10	8,49	65.879
40	8,76	65.661
18	9,01	65.758

Tabela 5. Análise comparativa das seqüências do grupo I. As seqüências foram submetidas ao GenBank para procura de homologia em bancos de dados não-redundantes, utilizando-se BLASTX (Basic Local Alignment Search Tool). A tabela exhibe os “hits” relevantes para as seqüências sem a região dos primers utilizados na obtenção da biblioteca. Os números sobrescritos indicam o primeiro “hit”, o segundo “hit” e assim por diante. Os valores percentuais mostram a identidade do clone em relação à proteína indicada e está especificada a região da seqüência das proteínas submetidas à análise.

Números de acesso ao GenBank	AAM47176.1 GP-90 (glicoproteína de tripomastigota metacíclico, cepa G)	AAD13347.1 TC85-11 (glicoproteína de tripomastigota, cepa Peru)	AAA30245.1 SA85 1.1 (glicoproteína de tripomastigota e amastigota, cepa CL)	AAA30246.1 Sialidase (glicoproteína de tripomastigota e amastigota, cepa CL)	A45622 GP85	AAA30150.1 GP85 (glicoproteína de tripomastigota, cepa Y)
2	³ 60% 122-696	¹ 59% 75-668	⁴ 58% 72-650	⁶ 55% 34-608	⁵ 58% 92-674	² 58% 55-643
12	¹ 62% 122-706	⁴ 60% 77-644	² 60% 76-645	⁵ 60% 38-605	⁶ 58% 92-666	³ 58% 55-635
16	¹ 69% 122-696	⁴ 64% 75-668	² 64% 72-667	⁶ 64% 34-608	⁵ 64% 92-674	³ 65% 55-643
19	¹ 63% 122-696	³ 58% 75-668	⁴ 59% 72-650	⁶ 59% 92-674	⁵ 59% 92-674	² 59% 55-643
20	¹ 67% 122-696	³ 63% 75-668	⁴ 64% 72-650	⁶ 63% 34-608	⁵ 64% 92-674	² 64% 55-643
25	¹ 63% 122-706	⁵ 62% 77-644	⁴ 62% 76-645	⁶ 62% 38-605	⁵ 60% 92-674	² 61% 55-635
27	¹ 67% 122-696	⁴ 61% 75-668	² 67% 66-667	⁶ 61% 34-608	⁵ 61% 92-674	³ 62% 55-643
29	¹ 60% 122-696	² 57% 75-668	⁴ 57% 72-650	⁶ 56% 34-608	⁵ 56% 92-674	³ 56% 55-643
32	² 62% 122-706	⁴ 64% 75-668	³ 65% 72-667	⁶ 59% 34-608	⁵ 65% 92-674	¹ 66% 55-643
35	² 65% 122-706	⁴ 63% 75-668	¹ 65% 72-667	⁶ 61% 34-608	⁵ 63% 92-674	³ 64% 55-643
36	¹ 67% 122-696	² 63% 75-668	⁴ 63% 72-650	⁶ 62% 34-608	⁵ 62% 92-674	³ 62% 55-643
40	² 65% 122-706	⁴ 63% 75-668	³ 64% 72-667	⁶ 61% 34-608	⁵ 64% 92-674	¹ 65% 55-643
44	¹ 65% 122-696	³ 60% 75-668	⁴ 61% 72-650	⁶ 60% 34-608	⁵ 60% 92-674	² 60% 55-643
45	¹ 66% 122-696	³ 62% 75-668	⁴ 63% 72-650	⁶ 61% 34-608	⁵ 61% 92-674	² 61% 55-643
52	¹ 60% 122-695	⁴ 56% 75-666	³ 58% 72-665	⁶ 55% 34-608	⁵ 58% 92-672	² 58% 55-641
54	² 56% 122-706	⁵ 53% 72-668	³ 56% 72-667	⁶ 51% 34-608	⁴ 56% 92-674	¹ 56% 55-643

Tabela 6. Análise comparativa das seqüências do grupo II. As seqüências foram submetidas ao GenBank para procura de homologia em bancos de dados não-redundantes, utilizando-se BLASTX (Basic Local Alignment Search Tool). A tabela exibe os “hits” relevantes para as seqüências sem a região dos primers utilizados na obtenção da biblioteca. Os números sobrescritos indicam o primeiro “hit”, o segundo “hit” e assim por diante. Os valores percentuais mostram a identidade do clone em relação à proteína indicada e está especificada a região da seqüência das proteínas submetidas à análise.

Números de acesso ao GenBank	AAO84046 ASP-2 (glicoproteína de amastigota, cepa Y, resíduos de 1-692)	AAO84045 ASP-2 (glicoproteína de amastigota, cepa Y, resíduos de 1-742)	AAC47720 ASP-2 (glicoproteína de amastigota, cepa Brazil, resíduos de 1-706)	AAO84044 ASP-2 (glicoproteína de amastigota, cepa Y, resíduos de 1-694)	AAA18827 TSA (trans-sialidase de tripomastigota, cepa Silvio)	AAD10618 TSA-E2 (glicoproteína de tripomastigota, cepa Esmeraldo)
10	¹ 52% 91-676	² 52% 91-676	³ 52% 93-676	⁴ 53% 94-678	⁵ 46% 52-635	⁶ 47% 51-637
13	¹ 59% 91-676	² 59% 91-676	³ 59% 93-676	⁴ 59% 94-678	⁵ 61% 52-635	⁶ 51% 51-637
14	² 58% 91-676	¹ 58% 91-676	³ 56% 93-676	⁴ 57% 94-678	⁵ 50% 52-635	⁶ 48% 51-637
18	¹ 51% 91-676	² 51% 91-676	³ 50% 93-676	⁴ 50% 94-678	⁸ 44% 52-635	⁶ 44% 51-637
22	¹ 61% 91-676	² 61% 91-676	³ 60% 93-676	⁴ 60% 94-678	⁵ 52% 52-635	⁶ 52% 51-637
23	¹ 59% 91-676	² 59% 91-676	⁴ 58% 93-676	³ 59% 94-678	⁵ 51% 52-635	⁶ 51% 51-637
26	¹ 42% 91-676	² 42% 91-676	⁵ 40% 93-676	³ 42% 94-678	⁶ 35% 52-632	⁹ 34% 51-637
33	² 52% 91-676	¹ 52% 91-676	³ 52% 93-676	⁴ 52% 94-678	⁵ 45% 52-635	⁶ 44% 51-637
34	¹ 49% 91-676	² 49% 91-676	³ 47% 93-676	⁴ 47% 94-678	⁵ 42% 52-635	⁷ 41% 51-637
41	¹ 61% 91-676	² 61% 91-676	⁴ 60% 93-676	³ 60% 94-678	⁵ 52% 51-635	⁶ 52% 51-637
43	² 60% 91-676	¹ 60% 91-676	³ 59% 93-676	⁴ 59% 94-678	⁵ 51% 52-635	⁶ 50% 51-637
48	² 59% 91-676	¹ 59% 91-676	³ 58% 93-676	⁴ 59% 94-678	⁵ 52% 52-635	⁶ 51% 51-637
50	¹ 61% 91-676	² 61% 91-676	³ 61% 93-676	⁴ 61% 94-678	⁵ 53% 52-635	⁶ 52% 51-637
57	² 60% 91-676	¹ 60% 91-676	³ 60% 93-676	⁴ 60% 94-678	⁵ 52% 52-635	⁶ 52% 51-637

Tabela 7. Análise da distribuição de seqüências características nos grupos I e II. A tabela mostra a presença (+) e/ou ausência (-) dos motivos ASP 1, ASP 2, ASP “like”, a seqüência RGD, o epítipo do Mab H1A10, variantes de VTV, os peptídeos ligantes de lamina (N-17, N-20 e N-21) e os peptídeos P4 e P8 de adesão à célula. A posição da arginina do motivo RGD e o início dos motivos ASP e VTV estão indicados entre parênteses. Os resíduos indicados por letras maiúsculas são conservados e os resíduos em letra minúscula são degenerados. Os clones não incluídos em nenhum dos grupos têm identidade <90% entre eles. Grupo I e Grupo II possuem identidades ≥ 90% entre seus membros.

Clones	Grupo I	Grupo II	Motivos											
			ASP 1 (209) (SxDxGxTW)	ASP 2 (254) (SxDxGxTW)	ASP “like” (430) (STDxSxxDTW)	(RGD305)	Epítipo do Mab H1A10 (Peptídeo G)	Variantes de VTV (576)	Peptídeos ligantes de lamina N-17	Peptídeos N-20	Peptídeos N-21	Peptídeos de adesão a célula P4	Peptídeos de adesão a célula P8	
16	X		+ (StDsGanW)	+	+	+	+	+ (kGEaPLEIvrfCF)	+ (Vkv)	+	+	+	+	-
27	X		+	+	-	+	+	+ (kGEaPLEIvrfCF)	+ (Vkv)	+	+	+	+	-
36	X		+	+	-	+	+	+ (kGEaPLEIlyFCF)	+ (Vkv)	+	+	+	+	-
44	X		+	+	-	+	+	+ (kGEaPLEIlyFCF)	+ (Vkv)	+	+	+	+	-
45	X		+	+	-	+	+	+ (kGEaPLEIlyFCF)	+ (Vkv)	+	+	+	+	-
19	X		+	+	+	+	+	+ (kGEaPLEIlyFCF)	+ (Vkv)	+	+	+	+	-
20	X		-	+	+	+	+	+ (kGEaPLEIlyFCF)	+ (Vkv)	+	+	+	+	-
32	X		+	+	+	+	+	+ (kGEaPLEIvrfCF)	+	-	+	+	+	-
35	X		+	+	+	+	+	+ (kGEaPLEIvrfCF)	+	-	+	+	+	-
52	X		+	+	+	+	+	+ (kGEaPLEIvrfCF)	+	-	+	+	+	-
54	X		+	-	+	+	+	+ (kGEaPLEIvrfCF)	+	-	+	+	+	-
40	X		+	+	+	+	+	+ (kGEaPLEIvrfCF)	+	-	+	+	+	-
10		X	+ (StDnGknW)	+ (SsDmGtrW)	+	+	-		+ (Vtm)	-	-	-	+	-
14		X	+ (StDnGknW)	+ (SsDmGtrW)	+	+	-		+ (Vtm)	-	-	-	+	-
13		X	+ (StDnGknW)	+ (SsDmGtrW)	+	+	-		+ (Vtm)	-	-	-	+	-
43		X	+ (StDnGknW)	+ (SsDmGtrW)	+	+	-		+ (Vtm)	-	-	-	+	-
22		X	+ (StDnGknW)	+ (SsDmGtrW)	+	+	-		+ (Vtm)	-	-	-	+	-
57		X	+ (StDnGknW)	+ (SsDmGtrW)	+	+	-		+ (Vtm)	-	-	-	+	-
48		X	+ (StDnGknW)	+ (SsDmGtrW)	+	+	-		+ (Vtm)	-	-	-	+	-
41		X	+ (StDdGknW)	+ (SsDmGtrW)	+	+	-		+ (Vtm)	-	-	-	+	-
23		X	+ (StDdGknW)	+ (SsDmGtrW)	-	-	-		+ (Vtm)	-	-	-	+	-
50		X	+ (StDdGknW)	+ (SsDmGtrW)	-	-	-		+ (Vtm)	-	-	-	+	-
29			+ (StatGsTW)	+	-	-	+	+ (kGEaPLEIlyFCF)	+ (Vkv)	-	+	+	+	-
2			+	+	+	+	+	+ (kGEaPLEIlyFCF)	+ (Vkv)	-	-	-	+	-
12			+ (StDnGekW)	+	+	+	+	+ (TGErPLEIvrfCF)	+	-	-	-	+	-
25			+	+	+	+	+	+ (TGErPLEIvrfCF)	+	-	-	-	+	-
18			+ (StDnGknW)	+ (SsDmGtrW)	+	+	-		+ (Vtm)	-	-	-	+	-
33			+ (StDnGknW)	-	+	+	-		+ (Vtm)	-	-	-	+	-
34			-	+ (SsDmGtrW)	-	-	-		+ (Vtm)	-	-	-	-	-
26			-	+ (SsDmGtrW)	-	-	-		+ (Vpm)	-	-	-	-	-

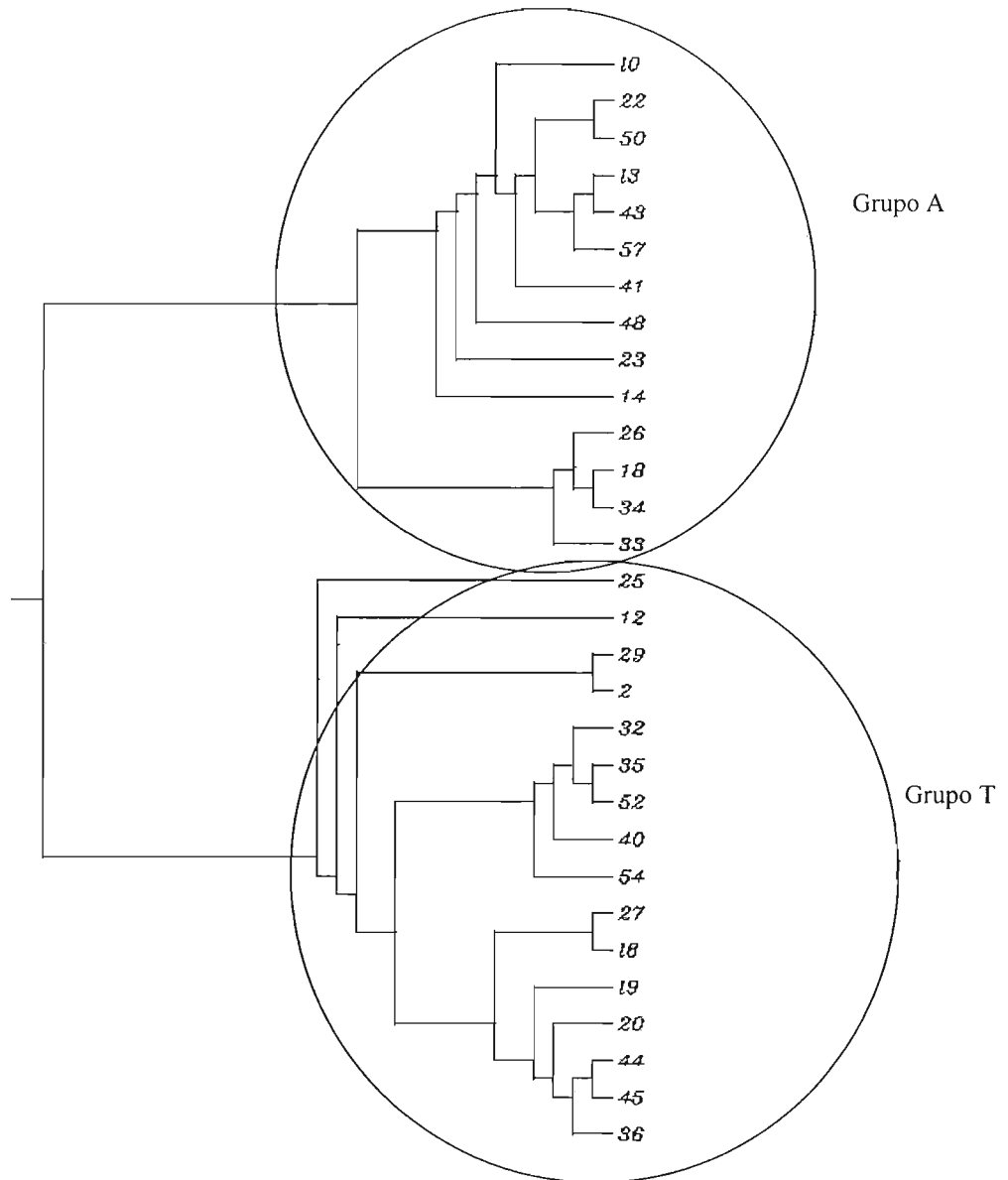


Figura 7. Cladograma obtido com o algoritmo Neighbor joining. O alinhamento utilizado foi somente entre os membros sequenciados. Os grupos T e A estão indicados.

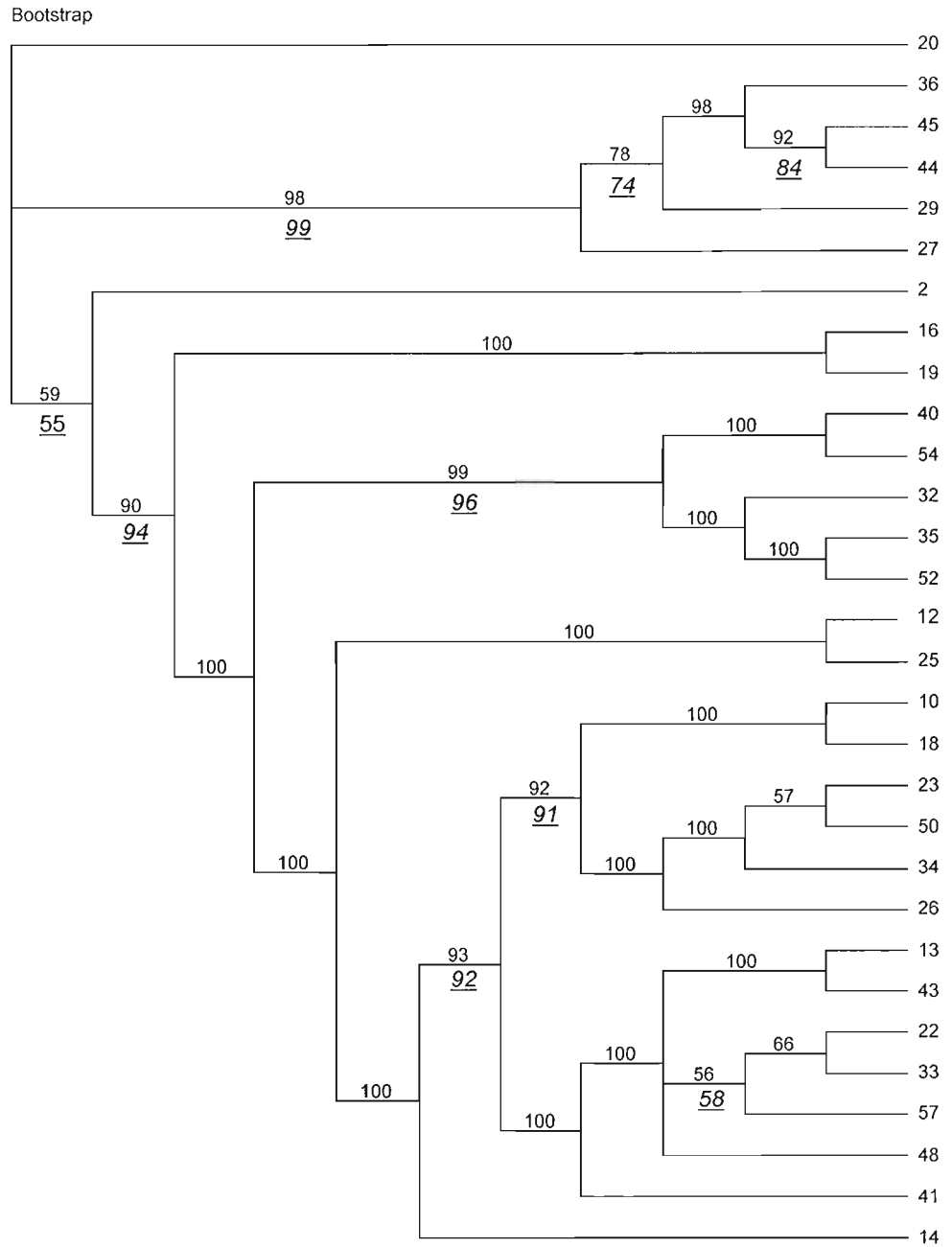


Figura 8. Cladograma obtido no programa Paup 4.0 indicando os valores Bootstrap com busca heurística completa. Os valores mostrados em caracter normal referem-se à análise repetida 100 vezes e os valores em itálico/sublinhado, a 500 repetições da análise.

```

1   GTACCGAAAGGTGGAAACGTCTAACCAAGACTAAGAGGGATTCAATTTGTTTCACCCTCTCTCGTCAGTGCCTGGTGG
1   GTACCGAAAGGTGGAAACGTCTAACCAAGACTAAGAGGGATTCAATTTGTTTCACCCTCTCTCGTCAGTGCCTGGTGG

101  AAGTGTATACCGTTAATGCTGATAATCAACGTAAGAAAACCTCTTCTGATGTTGTTGCGGAGTACATCGACACC
101  AAGTGTATACCGTTAATGCTGATAATCAACGTAAGAAAACCTCTTCTGATGTTGTTGCGGAGTACATCGACACC

201  CAAGGTGAATGAAAGTACATGGAAGGCAAAACACCGTGCCTTGGTACAACGGATGGAG. CCGATAATGGTG. GGGTT
201  CAAGGTGAATGAAAGTACATGGAAGGCAAAACACCGTGCCTTGGTACAACGGATGGAGG CCGATAATGGTG TGGTT

299  AAGGGCAATAAGGTGTTTTCTTTCCGGGAAGC. TTTATGAGTACCAACAA. - C GATGGAAGTCGGAAATATAG
301  AAGGGCAATAAGGTGTTTTCTTTCCGGGAAGCCTTTATGAGTACCAACGA. ACT GATGGAAGTCGGAAATATAG

395  TTGTGGGTGATGTGAGGGAGCCTAC. CAGCAGCGAGCCGACTGAACGGATCAAATGGGGCCAAATCAGGTCTCTG
401  TTGTGGGTGATGTGAGGGAGCCTAC. CAGCAGCGAGCCGACTGAACGGATCAAATGGGGCCAAATCAGGTCTCTG

495  CGAAGGTAATGGACGGGGTTTTCTTGGCCGCTGGTGGCTCAGGTGTTCTGATGGAGGATGGCAGCCTTGTGTTTC
501  CGAAGGTAATGGACGGGGTTTTCTTGGCCGCTGGTGGCTCAGGTGTTCTGATGGAGGATGGCAGCCTTGTGTTTC

595  GACGTTTACTCCATGATCATTACTCGAAGGACGACGGCAGTACCTGGGCGCTCTCAACTGGCGTTTTCCCTGC
801  GACGTTTACTCCATGATCATTACTCGAAGGACGACGGCAGTACCTGGGCGCTCTCAACTGGCGTTTTCCCTGC

895  AATGGGATGGATCACTTCTCATGATTGTTAATTGCAAGTATAGTCAGAG. GGTGTACGAGTCGCGTGACATGGG
701  AATGGGATGGATCACTTCTCATGATTGTTAATTGCAAGTATAGTCAGAG. ....

795  GCTTCCAGGCGTGTGGGTCAACTCANCNATTGTACTTTTGGGNACCTGAGCTTGGCGTGGAAGCCCTTATCA
799  .....

895  ATGCTGTACACACA. GAGAGGGGA. CTTCTTGGGGGAGAAAAGTGAAGAGCGCTCTATCTTTGGGTACGGACAA
753  ..... GA. .... CTTCTTGGGGGAGAAAAGTGAAGAGCGCTCTATCTTTGGGTACGGACAA

995  TTGGCATGGAAAATGCTGTGAAGGAGGAGCTTGGCAGCGCCCTGCTGTACTCGGATGGCAATTTGCATCTGTTA
833  TTGGCATGGAAAATGCTGTGAAGGAGGAGCTTGGCAGCGCCCTGCTGTACTCGGATGGCAATTTGCATCTGTTA

1095  TGACTTGTGCTTTCT. NCCTGG. CCGAGGAAAGT. G. GTACGATCAATTTTCTT. CTGG. AGACCTGGGCGAAAAA. AG
933  TGACTTGTGCTTTCT. CGCCTGA. CCGAGGAAAGT. GA. GTACGATCAATTTTCTT. CTGA. AGACCTGGGCGAAAAA. AGG

1195  CCCAC. GCGGGTCTGGTGGCAGTTTTGTCCGATACGGC. ANGCAATG. G. CACGTGGTACGACGAGTAT. TTTGCT. NT
1033  CCCAC. GCGGGTCTGGTGGCAGTTTTGTCCGATACGGC. CAGCAATG. GA. CACGTGGTACGACGAGTAT. CTTTGGCT. CT

1295  AGGCA. AG. GATGGATTTCAATTGACGGAGCCTAGCTCCAGGGTATCATGGTCCGTGAACACTCGGGTTAATAAT
1133  AGGCA. AG. GATGGATTTCAATTGACGGAGCCTAGCTCCAGGGTATCATGGTCCGTGAACACTCGGGTTAATAAT

1395  CTTCACACTTGTGGCATCGG. T. GATCATCGAA. GA. G. ACTCCGATTGCGGACGCTCCTCT. C. CTGACTGCGACGTTGG
1233  CTTCACACTTGTGGCATCGG. C. GATCATCGAAA. AA. A. ACTCCGATTGCGGACGCTCCTCT. T. CTGACTGCGACGTTGG

1495  GGTCTGTCGTATACCGCTGATAAGAAAGTGGGAGAAAGATGTTGAGGAAAACAAAAAACACGGAGCAGCACTTG
1333  GGTCTGTCGTATACCGCTGATAAGAAAGTGGGAGAAAGATGTTGAGGAAAACAAAAAACACGGAGCAGCACTTG

1595  CACTCATGCTGCAAGGCAAAAAGGCCCTCTGTGCAC. A. TTGATGGTAAAGTCGCTGGGGGAAGAAAGTGGCGTTA
1433  CACTCATGCTGCAAGGCAAAAAGGCCCTCTGTGCAC. G. TTGATGGTAAAGTCGCTGGGGGAAGAAAGTGGCGTTA

1895  ACGCTTTTGTCTTTGGCGCGTGCAGTGAGGATGCCGGCCAAAAAACTAAGGTGACGGTGAAGA. G. C. 1759
1533  ACGCTTTTGTCTTTGGCGCGTGCAGTGAGGATGCCGGCCAAAAAACTAAGGTGACGGTGAAGA. A. C. 1500

```

Figura 9. Comparação entre nucleotídeos dos clones 52 e 54. São membros com 88% de identidade, existindo uma deleção no clone 54. A região sombreada corresponde às bases idênticas.

Tc85-52	1	VPKGGTSN	KTKRDSFVSPSLVSAAGGVIAAF AEGQVYTVNADNQRKKTSSDVVAEYIDTTW	80
Tc85-54	1	-----	KTKRDSFVSPSLVSAAGGVIAAF AEGQVYTVNADNQRKKTSSDVVAEYIDTTW	52
Tc85-52	81	DWS TLVGKVNES	TWKANTVLGTTDGR - IMVGGFLLPTTTTKGNKVFLLAGSFMSTN - NDG	118
Tc85-54	53	DWS TLVGKVNES	TWKANTVLGTTDGDADNGVGGFYYP TTTTKGNKVFLLAGSLYEYORTDG	112
Tc85-52	119	SRKYSGFKLRLELVVGDVREPTSSEPTERIKWGQIRSLLESTIAAHÉGKWTGFLAAGGS		178
Tc85-54	113	SRKYSGFKLRLELVVGDVREPTSSEPTERIKWGQIRSLLESTIAAHÉGKWTGFLAAGGS		172
Tc85-52	179	GVLME DGTLVFP	LMARNAAKDVYSMI IYSKDDGSTWALSTGVSPANCTDPRIT EW DGSLL	238
Tc85-54	173	GVLME DGTLVFP	LMARNAAKDVYSMI IYSKDDGSTWALSTGVSPANCTDPRIT EW DGSLL	232
Tc85-52	239	MIVNCKYSQ	RVYESRDMGTTWTGAAGTLPGVWVNSXXCTFGXLSLRVEALITATIEGRKV	298
Tc85-54	233	MIVNCKYSQ	-----	241
Tc85-52	299	MLYTD	RGDFLGEKSERALYLWVTDNNRSFYFGPVG MENAVKEELASALLYSDGNLHLLQ	358
Tc85-54	241	-----	RGDFLGEKSERALYLWVTDNNRSFYFGPVG MENAVKEELASALLYSDGNLHLLQ	290
Tc85-52	359	RNGEGSDLSLS	XLAEEVGTINFLLETWAKKDAFFFKLSIPTAGLVAVLSDTAXNGTWYD	418
Tc85-54	297	RNGEGSDLSLS	RLT EEVS TINFLLETWAKKDAFFSRLS IPTAGLVAVLSDTASNDTWYD	350
Tc85-52	419	EYFCXNATVRNAP	KARDGFQLTEPSSRVSWSVNTRVNNVRHVSLSHNFTLVASMIIEETP	478
Tc85-54	357	EYLCNATVRNAT	KVKDGFQLTEPSSRVSWSVNTRVNNVRHVSLSHNFTLVASAIIEKETP	410
Tc85-52	479	IADAPLLTATLANNNS	NHTMGLSYTADKKWEKMF EENKTRSS TWVPKKEHQVALMLQGX	538
Tc85-54	417	IADAPLLTATLANNNS	NHTMGLSYTADKKWEKMF EENKTRSS TWVPKKEHQVALMLQGX	470
Tc85-52	539	KASVHI	DGKSLGEEEVPLKGEAPLELVRF CF GACSE DAGQKT KVTVK S	588
Tc85-54	477	KASVHV	DGKSLGEEEVPLKGEAPLELVRF CF GACSE DAGQKT KVTVK	523

Figura 10. Comparação entre as seqüências deduzidas de aminoácidos dos clones 52 e 54. As caixas sombreadas em cinza correspondem aos aminoácidos idênticos e as caixas simples aos aminoácidos similares.

```

43 1  GTGACAAAAGAAGAGTCTGAAATGGAGTGAAATTTGCGTTTCTCACACCCTCTCTCCTCAGTGCTGGTGGGGT
48 1  GTGACAAAAGAAGAGTCTGAAATGGAGTGAAATTTGCGTTTCTCACACCCTCTCTCCTCAGTGCTGGTGGGGT

43 101  TGGCAATCGACTCTCCTCACCGGNGGNGAAA AAAAGGTTCAATTGAAGTGGTTGCGGGGTACGTCAACGCTGCG
48 101  TGGCAATCGACTCTCCTCACCGGGGAAA AAAAGGTTTANCAATTTGAAGTGGTTGCGGGGTACGTCAACGCTGCG

43 201  GGTCAATGATGGTACATGGAGGGGCACACACCCGCGCTTAGTACANBTGAAGGAAGAGATCGTGTGGGTGTTGCGA
48 201  GGTCAATGATGGTACATGGAGGGGCACACACCCGCGCTTAGTACANBTGAAGGAAGAGATCGTGTGGGTGTTGCGA

43 301  AATAATGTGTTTTTTATTGTAGGAAGCCATGAAGAGAAGTATGACAGTGTNACGAAAAGCTGGGAGGCCAGGTGA
48 301  AATAATGTGTTTTTTATTGTAGGAAGCCATGAAGAGAAGTATGACAGTGTNACGAAAAGCTGGGAGGCCAGGTGA

43 401  GGGCCACGCGGTCCACGGACGCCGCAAAGAGTAAAACGATCAACTGGGGTGGACCCACACCGCTTTTAAGACA
48 401  GGGCCACGCGAGTCCACGGACNGGCGTGCAGAGAACNCANTCAACTGGGGTGGACCCNC.....

43 500  TGAGGGAACTTTTGGCCACTGGTGGCGCCGGTCTCTGACGGAAATGGCACGCTCCTGTTTCCCTGGAGGGG
48 468  .....C.....ACTGGTGGCGCCGGTCTCTGACGGAGAAATGGCACGCTCCTGTTTCCCTGGAGGGG

43 600  CGTGATCACTTATTCGACGGACGATGGGAAAAACTGGTATTTCTTGAAGGGAATACCTCCCBCGAAATGCTTTA
48 543  CGTGATCACTTATTCGACGGACGATGGGAAAAACTGGTATTTCTTGAAGGGAATACCTCCCBCGAAATGCTTTA

43 700  GGGCGACGAATTTCTGGTTAGTGATTGTTTTGGTGGTCAGAAATTTGTACGAGTGGAGTGACATGGGGACAAGGTG
48 643  GGGCGACGAATTTCTGGTTAGTGATTGTTTTGGTGGTCAGAAATTTGTACGAGTGGAGTGACATGGGGACAAGGTG

43 800  GCGTGTGGGTCAAGTCAAAATCGAAAGCNAATTTAG.....
48 743  GCGTGTGGGTCAAGTCAAAATCGAAAGCTATTTGGATAACAGCTTCCATTTGGATGCCCTCATCACCGCGACC

43 858  TCAGAGAGGGTACACCTCGGGAGAGAAAGTGGACAAGGCGCTCTGCCTTTGGGTCACGGACGAAAACCGCACGG
48 843  TCAGAGAGGGTACACCTCGGGAGAGAAAGTGGACAAGGCGCTCTGCCTTTGGGTCACGGACGAAAACCGCACGG

43 958  CATGACGTGGAGCATTCCAACGCCCTGCTGTGCTCGGATGGTGGTGGACCTTTTACAGGAGAGGGCCAATGA
48 943  CATGACGTGGAGCATTCCAACGCCCTGCTGTGCTCGGATGGTGGTGGACCTTTTACAGGAGAGGGCCAATGA

43 1058  GCCTGACGGAGGAGCTGAAACAGATCAAGTCCACCCTCAGTACTTGGGGCAGACTGGACGCCCTTCTTCTTCAAG
48 1043  GCCTGACGGAGGAGCTGAAACAGATCAAGTCCACCCTCAGTACTTGGGGCAGACTGGACGCCCTTCTTCTTCAAG

43 1158  TTGGATCCTGTCCGATACGTGTCGTCGGTGGCGACACGTGGATCGACGAGTACCGTTGCGTGAATGCAA|GGTGA
48 1143  TTGGATCCTGTCCGATACGTGTCGTCGGTGGCGACACGTGGATCGACGAGTACCGTTGCGTGAATGCAA|GGTGA

43 1258  TTTAAGTTCACGGCGCCTGGGTCCAGGGCAATGTGGTCTATGAGCAGGTGGGACCTAATAAACAGTACGGCTT
48 1243  TTTAAGTTCACGGCGCCTGGGTCCAGGGCAATGTGGTCTATGAGCAGGTGGGACNCTAATAAACAGTACGGCTT

43 1358  CGACGGTGA|CCATCCACCAGGTTCCGAAGGGGAGCACTCCTCTGCTGGGTGCGGGTCTCGGGGACGGTGC CGG
48 1343  CGACGGTGG|CCATCCACCAGGTTCCGAAGGGGAGCACTCCTCTGCTGGGTGCGGGTCTCGGGGACGGTGC CGG

43 1458  CATGAATAAAAAGTGGGAGAAGGTGTTCAACGGGCACAAAACAGCACCCGGGGAGCACGTGGGAGCCGGGGAGAG
48 1443  CATGAATAAAAAGTGGGAGAAGGTGTTCAACGGGCACAAAACAGCACCCGGGGAGCACGTGGGAGCCGGGGAGAG

43 1558  GACGGCAAAAAGGGCTCCGTGTACGTGGATGGTGTGATTGTGGGGAGCCCGGAGATGATACCAACACTTGAGAC
48 1543  GACGGCAAAAAGGGCTCCGTGTACGTGGATGGTGTGATTGTGGGGAGCCCGGAGATGATACCAACACTTGAGAC

43 1658  ACGTTGGGGGCGACGAGGGAGACATCAACAGCAGCGTGACGATGACGAAC 1707
48 1643  ACGTTGGGGGCGACGAGGGAGACATCAACAGCAGCGTGACGATGACGAAC 1692

```

Figura 11. Comparação entre nucleotídeos dos clones 43 e 48. São membros com 91% de identidade, existindo uma deleção no clone 48 a 5' e no clone 43 a 3'. A região sombreada corresponde às bases idênticas.

Tc85-43 1 MTKEESEIGVKFAFLTPSLLSAGGVMAFAESLVAIDSPHRXKXKGSIEVVAGYVNAAE 80
Tc85-48 1 MTKEESEIGVKFAFLTPSLLSAGGVMAFAESLVAIDSPHRGKKVXQFIEVVAGYVNAAE 80

Tc85-43 81 WSSLVAAVNDGTWRAHTALSTXEGDRVGVAKLPPTTLAKGNVFFIVGSHHEEKYDSVXES 120
Tc85-48 81 WSSLVAAVNDGTWRAHTALSTXEGDRVGVAKLPPTTLAKGNVFFIVGSHHEEKYDSVXES 120

Tc85-43 121 WEARQXXVQLLVGGATRSTDAAKSKTIINWGGPTPLLRQIAPQTQGELELLPTGGAGALT 180
Tc85-48 121 WEARQXXVQLLVGGATQSTDXRAENXXINWGGPX-----TGGAGALT 181

Tc85-43 181 ENGTLLFPLEGRNNGDIVSVITYSTDDGKNWYFLKGIPPAKCFNPRITIEWEXGRRILVS 240
Tc85-48 182 ENGTLLFPLEGRNNGDIVSVITYSTDDGKNWYFLKGIPPAKCFNPRITIEWEXGRRILVS 221

Tc85-43 241 DCFGQQNLYESSDMGTRWTEAVGTLPGVWVKSKSKAXLIG-----KGVMLYI 288
Tc85-48 222 DCFGQQNLYESSDMGTRWTEAVGTLPGVWVKSKSKAII LONSFHLDALITATIEREFMLYI 281

Tc85-43 287 QRGYTSGEKVDKALCLWVTDENRTLHVGFVAMEHDOVEHSNALLCSOGALHLLQERANEKG 348
Tc85-48 282 QRGYTSGEKVDKALCLWVTDENRTLHVGFVAMEHDOVEHSNALLCSOGALHLLQERANEKG 341

Tc85-43 347 EATSLARLTELKTIKSTLSTWARLDAFFFKSSIPPTGLVWILSBTSSGGDTWIDEYRCV 408
Tc85-48 342 GATSLARLTELKTIKSTLSTWARLDAFFSKSSIPPTGLVWILSBTSSGGDTWIDEYRCV 401

Tc85-43 407 NAMVTKAAKVKNKGFKFTAPGSRAMWMSRWDGNKQYGFVNHRRFTLVATVTHQVPKGSTP 488
Tc85-48 402 NATVTKAAKVKNKGFKFTAPGSRAMWMSRWDGNKQYGFVDHRRFTLVATVAHQVPKGSTP 481

Tc85-43 487 LPGAGLGDDGAGKKIIGLSYSMNKKWEKVFNGTKTAPGSTWEPGREHQVALMLQDGGKKS 528
Tc85-48 482 LLGAGLGDDGAGKKIIGLSYSMNKKWEKVFNGTKTAPGSTWEPGREHQVALMLQDGGKKS 521

Tc85-43 527 YVDGVIVGSPEMIPTLETQGFPIQFYVGGDEGDINSSVTMTN 589
Tc85-48 522 YVDGVIVGSPEMIPTLETQGFPIQFYVGGDEGDINSSVTMTN 584

Figura 12. Comparação entre as seqüências deduzidas de aminoácidos dos clones 43 e 48, indicando a permanência das deleções nos dois membros. As caixas sombreadas em cinza correspondem aos aminoácidos idênticos e as caixas simples aos aminoácidos similares.

Tabela 8. Alinhamento de nucleotídeos indicando deleções (Δ) e inserções (I) encontradas. Os números dos clones estão indicados lado a lado. As deleções encontradas eram de cerca de 50 nucleotídeos.

I 19-54 Δ	I 36-20 Δ	I 10-48 Δ
I 20-54 Δ	I 40-20 Δ	I 13-48 Δ
I 27-54 Δ	I 44-20 Δ	I 14-48 Δ
I 32-54 Δ	I 10-43 Δ	I 22-48 Δ
I 35-54 Δ	I 13-43 Δ	I 23-48 Δ
I 36-54 Δ	I 14-43 Δ	I 41-48 Δ
I 40-54 Δ	I 41-43 Δ	I 50-48 Δ
I 45-54 Δ	I 48-43 Δ	
I 27-20 Δ	I 50-43 Δ	
I 32-20 Δ	I 57-43 Δ	
I 35-20 Δ	I 23-43 Δ	

```

Tc85-45 1  GGTSQETKRDSFISPSLVSAAGVIAAFAEGQVYTVNAGRQREKTS SDVVAEYIDATWDWS 80
Asp 1 1  ..... 1
Asp 2 1  ..... 1
RGD 1  ..... 1
Ppideo G 1  ..... 1

Tc85-45 81  TLVGKVNES TWKAHTVLGTTDGTOKRVDF FYPTTTTKGNKVFLLAGSLYKHRQTVGNRT 120
Asp 1 1  ..... 1
Asp 2 1  ..... 1
RGD 1  ..... 1
Ppideo G 1  ..... 1

Tc85-45 121 YSGFNLRLELVVGDVREPTSSEPTERIKWGQIRSL LNESTIAAHEGKWTGFLAAGGSGVL 180
Asp 1 1  ..... 1
Asp 2 1  ..... 1
RGD 1  ..... 1
Ppideo G 1  ..... 1

Tc85-45 181 MEDGTLVFP LMATNAAKDDYSMI IY ASP1
Asp 1 1  ..... 1
Asp 2 1  ..... 1
RGD 1  ..... 1
Ppideo G 1  ..... 1

Tc85-45 241 NCKYSQRVY ASP2
Asp 1 1  ..... 1
Asp 2 1  ..... 1
RGD 1  ..... 1
Ppideo G 1  ..... 1

Tc85-45 301 RGD " RGD "
Asp 1 1  ..... 3
Asp 2 1  ..... 1
RGD 1  ..... 3
Ppideo G 1  ..... 1

Tc85-45 381 EGS DLSLSRLT GELSTIKSVLSTWSQKDAFIF SFSIPPRVWWQYCPMPASNGTWIDEYLC 420
Asp 1 1  ..... 1
Asp 2 1  ..... 1
RGD 1  ..... 1
Ppideo G 1  ..... 1

Tc85-45 421 LNAAVTNATKVKDGFQLTESNSGVLWVFV NTRDDNVRHVSLSHNFTLVASVTIEEVPSNST 480
Asp 1 1  ..... 1
Asp 2 1  ..... 1
RGD 1  ..... 1
Ppideo G 1  ..... 1

Tc85-45 481 LLLTATSANNNSNHTMGLSYTADKKWETMFEENKTRRS TWVPKKEHQVALMLQVKKASV 540
Asp 1 1  ..... 1
Asp 2 1  ..... 1
RGD 1  ..... 1
Ppideo G 1  ..... 1

Tc85-45 541 YIDGNSLGEELPL Pep. G
Asp 1 1  ..... 13
Asp 2 1  ..... 13
RGD 1  ..... 13
Ppideo G 1  ..... 13

```

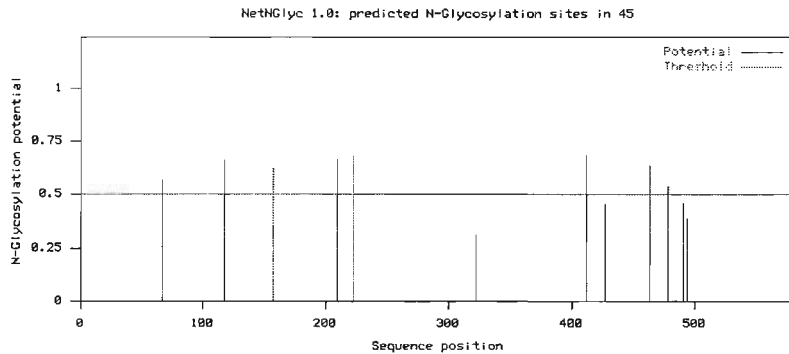
Figura 13. Domínios relevantes na Tc85-45. Os principais domínios da Tc85-45 estão indicados pelas caixas.

SA85-1.1-AAA30245	1	-----MSRRVFAASVLLLVVIMCCGGAAATAQVGSNADASTP-GSA	40
Tc85-11-AF085686	1	-----MSRRVFAASVLLLVVIMCCGGAAATAQVEKATDASTPSGSA	41
Tc85-45	1	-----	1
gp82-AAF02477	1	-----	1
SA85-1.1-AAA30245	41	LTGAIAGEGSSSSGGVEGLQRVDLFVFPQKTQVLPKKGPDSSRRDSFFSPSLVSAGGVIAAF	100
Tc85-11-AF085686	42	LTGAIATAAGSASGSVELPQESILFVFPQTTQVLPKKGPDSSRRDSFFSPSLVSAGGVIAAF	101
Tc85-45	1	-----GGTS-----QETKRDSFISPSLVSAGGVIAAF	27
gp82-AAF02477	1	-----	1
SA85-1.1-AAA30245	101	AEGHINTKN-PHNESAKPFSQVVAAGYIDSAEMWPTLVKVKVSESTWQAHTVVLGKAEQKKS	159
Tc85-11-AF085686	102	AEGRIINAKNTPSTESTKPSQDVVAEYIDSAEMWPTLVKVKVKKKEMWRARTVLGKAEQNESF	161
Tc85-45	28	AEGQVYTVN-AGROREKTSQDVVAEYIDATWDMWPTLVKVKVNESWTKAHTVLTGTTDGTDKR	88
gp82-AAF02477	1	----QYAVD--GKLIKPTSSAVVAEYIDSSMDWPTLVKVKVSESTWMAHTVVLKAEQKGNL	54
SA85-1.1-AAA30245	180	DVVLRLPTTTTKGNKVFLLAG--STDLSYVNWWSMREGSLELKLKLVVGDVTKPPTSSEPT	214
Tc85-11-AF085686	182	DVVVRHPTTITMKGKVFLLVGG--STALSVDNWSMKEGSELEIKLVVGDVTKPPTDSEPSK	217
Tc85-45	87	VDFYYPTTTTKGNKVFLLAGSLYKHRQTVGNRTYSGFNLRLELVVGDVTKPPTSSEPT	146
gp82-AAF02477	55	DVVLRLPTTTMKGKVFLLVGG--SYDMLNESGIMKRDSPDLKLVVGDVTKPSAGGERVD	110
SA85-1.1-AAA30245	215	RIKVGGEIKSLLNESTTIAAQKGLTEFLASGGSGVVMEDGTVFVSLMANNKKGQVFSLI	274
Tc85-11-AF085686	218	RIEIVGEINSLNNGSTLAAHKGLTECLASGGSGVLMEDGALVFSLMANNKKGQVFSMI	277
Tc85-45	140	RIKVGQIRSLNNESTIAAHEGKWTGFLAAGSGVLMEDGTVFVSLMATNAAKD-DYSMI	204
gp82-AAF02477	111	GSTVGTPTS-LNQTTLKIPKAGLKDFYSGGSGVVMEDGTVFVPIAFNAGNA-GFSTT	168
SA85-1.1-AAA30245	275	YSKDNNGSTWLSLEGISPAKGGAPRITWEGSLLMIVDCENDQRVYVSRDMGTTWTEAIGT	334
Tc85-11-AF085686	278	YSKDNNGSTWALSIEDMSPANCTDPRITWEGSLLMIVDCENEQRVYESCDMGTTWTEAIGT	337
Tc85-45	205	YSTDNNGSTWALSSTGVSANCTDPRITWEGSLLMIVNCKYSQRVYESDMGTTWTEAIGT	284
gp82-AAF02477	189	YSTDGGANWMLSNGTPTAECLEPRITWEGSLLMIVDCVVGQRVYESRDMGTTWTEAVGT	228
SA85-1.1-AAA30245	335	LSGVVGS THNWETIGR-RLAMEALITVTFEGRKVMLYTQRGYALGETETTSPLYLWVTDNN	303
Tc85-11-AF085686	338	LPGVVWVNSQSEDPYEGVLRVDALITASIEDRKVMLYTQRGYASGG-EAERALLYLWVTDNN	308
Tc85-45	265	LPGVWTKSRLLSWDL-SLRVEALITATIEGRKVMLYTQRGDFSGE-KSERALLYLWVTDNN	322
gp82-AAF02477	229	LSGVWAKSQSFFRDL-NLRVDAALIAATI EGRKVMLYTQRGYASGE-KRVNPLYLWVTDNN	286
SA85-1.1-AAA30245	394	RSFFVFGPVGMDNAVKGELAGALLYSQGGHLHLQRRDSDGEDSVMSLSRLTEELKTKSVLS	463
Tc85-11-AF085686	397	RSFFVFGPVGMDNAVSGDLTSSLLYSQGGHLHLQRRRGNSESRMSLSRLTEELSTINSVLK	466
Tc85-45	323	RSFFYFGPVGMDNAVRWEFSSLLYSQGGHLHLQRRDNGEGSDLSRLTGEELSTIKSVLS	382
gp82-AAF02477	287	RSFFYFGPIAMGNAAANSMFVSSLLYSQGGHLHLQRRANDKGSMSLARLTELKTKSVLS	346
SA85-1.1-AAA30245	464	TWSQKQVFFSSLSIPTVGLVAVLSDAAGDG-RWYDEYLCNATMTNATKVKDGFQLTEPD	512
Tc85-11-AF085686	467	TWQAQNDIAFFSNLSIPTAGLVAVLSNASASGDTMNDYELCLNATVKNATKVKDGFQLQEPD	516
Tc85-45	383	TWSQKQDAFIFSSSIPRWWQYCPMPASNG-TWYDEYLCNAAATNATKVKDGFQLTESN	441
gp82-AAF02477	347	TWSKLEASFSASSTPTAGLVGLLSNASASGD-AMIDDYRSVNAKVMNAVKWHQDFKFTGFG	405
SA85-1.1-AAA30245	513	SRAVWVNVNIPDGNVRRHISLSHNFTLVASMTIEEAPSGNTPLLTAVLVDAGPEYFMRLSYT	572
Tc85-11-AF085686	517	SRAIIMPVNTQGNVRRHISLSHNFTLVASMTIEEAPSEKT-LLTAVLGNTEPPYIMRLSYT	575
Tc85-45	442	SGVLMVFNTRDDNVRHVSLSHNFTLVASMTIEEVPNSNTLLLTATSANNNSNHTMGLSYT	501
gp82-AAF02477	408	SGAIMPVNNRESNGPHTFVNYNFTLVATMIMHKVPKNSTLLGAVLAEPISLTLFVIGLSYG	485
SA85-1.1-AAA30245	573	ADNKWMTMLKDEKPTTESRPMWAGKEHQVALMLQ-GNKASVYVDDGELLG-EEEVLPTGE	630
Tc85-11-AF085686	578	ADNKWMTMLKDEK--TTRRSTWELKKEYQVALMLQ-GNKRSVYVDDGELLG-EEEVLPTGE	631
Tc85-45	502	ADKKWMTMFEENK-KTRRSTWMPKKEHQVALMLQ-VKKASVYIDGNSLGG-EEELPLTGE	557
gp82-AAF02477	488	TDGTWETVFNGET-TTSGSTWMPGKEYQVALMLQDGNKGSVYVDDGMSVGSLATLPTPEV	523
SA85-1.1-AAA30245	631	KPLEIFAFCFGACKIDGDEEESPEKEIGKKPRVMTVNTVFLYNRP LNSTEMRAIKDRIPVP	690
Tc85-11-AF085686	632	TPLEPFPGFCFGACGEDDDEEESP EEEIGKKPRVMTVNTVFLYNRP LNSTEMTAIKDRKVP	691
Tc85-45	568	APLELLYFCFGACG-----EDAGQKTNVKKV-----	583
gp82-AAF02477	524	RGAEIADFYFVGGE-D-----EEDKKSSTVTKNVFLYNRPLGADLRMVK-----	588
SA85-1.1-AAA30245	681	TRAPEPQVKIAPKPAAP-----AAP-----AGNEETA	717
Tc85-11-AF085686	682	KRAPEPQVKIVPNPVPAPVAVPVPRELPAPGRTTVGRANTQHAPAGRLTSAGNEGTA	751
Tc85-45	583	-----	583
gp82-AAF02477	588	-----K-----ID-----	571
SA85-1.1-AAA30245	718	RETGGGGANGDAVSAYGRVLLPLLLGLWLALATA- 752	
Tc85-11-AF085686	752	REKGGGGANGDAGSAYGRELLPMLLLLGLWLALATA- 788	
Tc85-45	583	-----	583
gp82-AAF02477	571	-----G-SMHGGVSRALLLLLGLCGFAALY 595	

Figura 14. Comparação entre as seqüências da superfamília das gp85/trans-sialidasas (os números de acesso ao Genbank estão ao lado do nome) e a Tc85-45, indicando blocos conservados. As caixas sombreadas em cinza correspondem aos aminoácidos idênticos e as caixas simples aos aminoácidos similares. A linha descontinua refere-se ao peptídeo P4 e a linha sólida ao peptídeo P8 da gp82, ambos envolvidos na invasão a célula hospedeira (Pereira et al., 1999).

(A)

GGTSQETKRDSFISPSLVSAGGVIAAFAEGQVYTVNAGRQREKTSSDVAEYIDATWDWSTLVGKVNNE
STWKAHTVLGTTDGTDKRVDFYYPTTTTKGNKVFLLAGSLYKHRQTVGNRTYSGFNLRLLELVVGD
EPTSSEPTERIKWGQIRSLNESTIAAHEGKWTGFLAAGGSGVLMEDGTLVFPLMATNAAKDDYSMIY
STDNGSTWALSTGVSPANCTDPRITEWDGSLLMIVNCKYSQRVYESRDMGTTWTEAIGTLPGVWTKS
RLLSWDLRLRVEALITATIEGRKVMPLYTQRGDFSGEKSERALYLWVTDNNRSFYFGPVGMDNAVRWE
FPSNLLYSDGKLHLLQQRDNEGESDLSLSRLTGELSTIKSVLSTWSQKDAFIFSIPPRVWWQYCPMPA
SNGTWIDEYLCLNAAVTNATKVKDGFQLTESNSGVLFVNTRDDNVRHVSLSHNFTLVASVTIEEVPS
NSTLLLATSANNNSNHTMGLSYTADKKWETMFEENKKTRRSTWVPKKEHQVALMLQVKKASVYID
GNSLGEEELPLKGEAPLELLYFCFGACGEDAGQKTNVKVK



(B)

VTKEESEIGVKFAFLTPSLLSAGGVMVAFAESLVAIDFPPREKRKSSXEVVAGYVNAEAETWSSLVAAV
NDGTWRAHTALSTVKEEDRVGVAKLPPTLAKGNNVFFIVGSHEEKYDSVTKSWEAGDSVRPDVAGD
TPCTEDEPRMINWGGPIPHSKQTGRQIRGGXXXAPSLTGVGAVTVSGMLPSRPGVRNSGTVSETIFLMV
GGRNXVFLEGIHRGRCXXSXIIAWEKARILVSDCFGGQDLYESSDMGTRWTEAVGTLPGVVKSLSKA
ILDNSFHLDALITATIEREFMLYIQRGYTSGEKVDKALCLWVTDENRTLHVRDAMARGVVLFSDPQFW
VALPHSRGRDRSVEDTLRALQMGRKTRFRTHNTRGDRQASLAGSSXHMIGXFEFLSDTSSGGDTWID
EYRCVNATVTRAQQSKNAFKFGLGTXS VWIGW WRLNKQYCXXTHRFTLVATVTIHQVPGSTPLL
AGLDGAGKKIIGLSYSMNKKWEKVFNGTKTAPGSTWEPGREHQVALMLQDGKKGSVYVDGVIVGS
PEMIPTLETQGFEPQYVGG DEGDMSSVTMTN

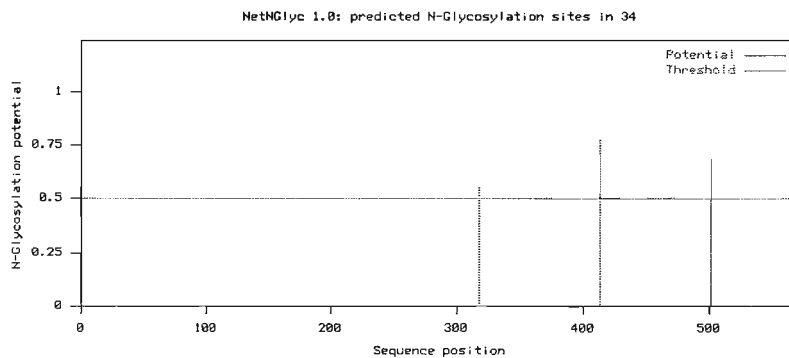


Figura 15. Potenciais sítios de N-glicosilação potenciais da Tc85-45 (A) e Tc85-34 (B). Os sítios de N-glicosilação estão sublinhados na seqüência acima. Os gráficos ilustram esses sítios putativos ao longo da cadeia de proteína (o eixo x representa o comprimento da proteína do N ao C-terminal). As posições potenciais (linhas verticais) que cruzam o limite (“threshold”) – linha horizontal a 0,5- são tidas como glicosiladas.

4.3. Expressão de proteínas recombinantes.

Foram utilizadas bactérias da cepa BL21(DE3)plyS transformadas com os clones 45 e 34, pertencentes aos grupos I e II, respectivamente. Após a indução das bactérias com 1 mM de IPTG, as amostras foram analisadas por SDS-PAGE. Como é observado na figura 16 as proteínas correspondentes aos clones 45 e 34 foram expressos nas frações indicadas. Nas fileiras (A) 2, 4, 5 e (B) 2, 3, 5 são observadas bandas proteicas de aproximadamente 67 kDa, como estimado. As proteínas de fusão foram isoladas por cromatografia de afinidade a Ni²⁺ como descrito em “Material e Métodos”. As fileiras 5 das figuras 16 A e B mostram as proteínas após a purificação. As bandas da fileira 5 são reconhecidas pelo anticorpo monoclonal anti-6His tag, mostrando a degradação parcial da proteína (dados não mostrados).

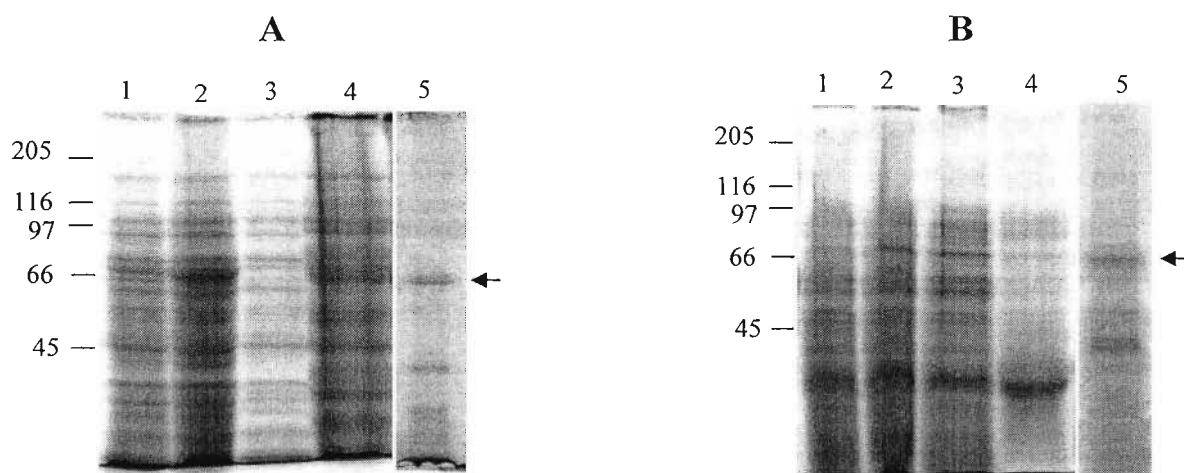


Figura 16. Análise da expressão das recombinantes (A) Tc85-45 e (B) Tc85-34. Extrato de bactérias transformadas, induzidas ou não por 1mM de IPTG: (1) não-induzido, (2) induzido por 4 horas. Após a lise das bactérias, a presença de proteína foi analisada na fração solúvel (3), insolúvel (4) e a proteína purificada por coluna de afinidade a metal (5), como descrito em “Material e Métodos”. A análise foi feita em gel 9% de poliácridamida contendo SDS e corado com Coomassie Blue.

4.4. Análise de ligação de proteínas de fusão a elementos da matriz extracelular por dot blot.

Durante o processo de adesão e invasão do parasita às células é possível que seja relevante a presença de vários epítomos que se ligam a diferentes moléculas da matriz extracelular e lâmina basal. A fim de verificar esta possibilidade, analisamos a interação de novos membros clonados (Tc-85) a componentes de matriz extracelular.

A técnica de Dot Blot nos pareceu mais apropriada para uma varredura inicial de padrões de ligação entre Tc-85 e elementos de matriz extracelular. Um primeiro ensaio de Dot Blot foi realizado com concentrações crescentes de laminina, fibronectina e gelatina e incubadas com 20 µg de Tc85-45. Houve reação visível a partir de 1 µg de laminina e fibronectina. A seguir, foi realizado um Dot Blot variando-se a quantidade de proteína de fusão, observando-se ligação de Tc85-45 a laminina, de maneira dose-dependente (figura 17). Como o fragmento N-terminal da proteína Tc85-11 liga-se a laminina (tese de doutorado, Quelopana, 2003), analisamos por Dot Blot a ligação da N-Tc85-11 e a Tc85-45 a laminina, aparentemente com resultados semelhantes (figura 18). A mesma avaliação foi feita com fibronectina (figura 19) e, da mesma forma, foi observada a ligação de Tc85-45 a fibronectina de maneira dose-dependente.

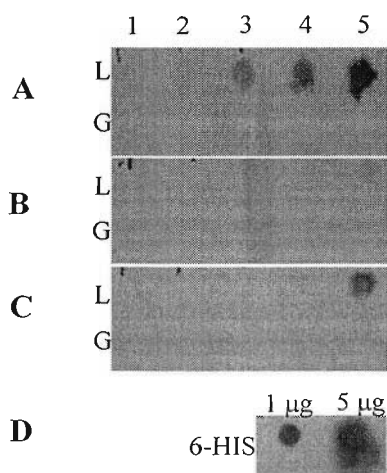


Figura 17. Reatividade da proteína de fusão Tc85-45 com laminina. Foram aplicadas quantidades crescentes de gelatina (G) e laminina (L) em membrana de nitrocelulose na seguinte ordem (1) 0,5 µg, (2) 1 µg, (3) 2 µg, (4) 3 µg e (5) 5 µg. Após o bloqueio com leite desnatado a 3% em TCMM, a membrana foi incubada com 30 µg de Tc85-45 (A), 10 µg de Tc85-45 (B) e sem proteína de fusão (C), seguido de anticorpo anti-6His tag, anti-IgG de camundongo conjugado a peroxidase e revelação com ECL, como descrito em “Material e Métodos”. Uma proteína de fusão não relacionada contendo cauda de 6-histidinas foi utilizada como controle do anticorpo anti-6His tag (D).

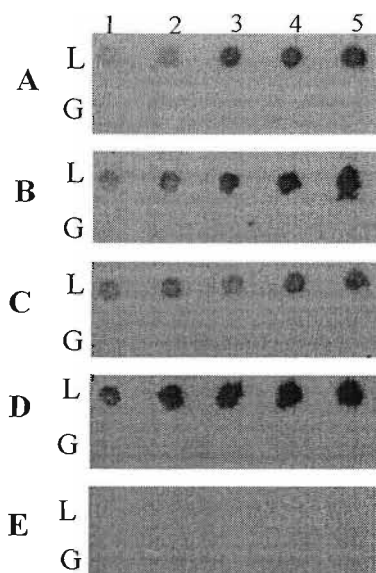


Figura 18. Reatividade da proteína de fusão Tc85-45 com laminina. Foram aplicadas quantidades crescentes de gelatina, e laminina em membrana de nitrocelulose na seguinte ordem (1) 0,5 µg, (2) 1 µg, (3) 2 µg, (4) 3 µg, (5) 5 µg. Após o bloqueio com leite desnatado a 3% em tampão TCMM, a membrana foi incubada com porção N-terminal da Tc85-11, respectivamente: (A) 10 µg, (B) 30 µg, e com Tc85-45, respectivamente: (C) 10 µg, (D) 30 µg e (E) sem proteína de fusão, seguido de anticorpo anti-6His tag, anti-IgG de camundongo conjugado a peroxidase e revelação com ECL, como descrito em “Material e Métodos”.

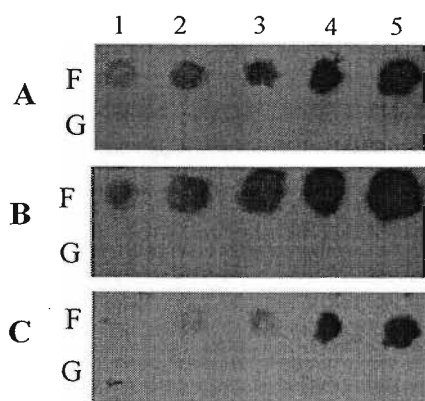


Figura 19. Reatividade da proteína de fusão Tc85-45 com fibronectina. Foram aplicadas quantidades crescentes de gelatina (G) e fibronectina (F) em membrana de nitrocelulose na seguinte ordem (1) 0,5 µg, (2) 1 µg, (3) 2 µg, (4) 3 µg, (5) 5 µg. Após o bloqueio com leite desnatado a 3% em tampão TBS-T, a membrana foi incubada com Tc85-45, respectivamente: (A) 10 µg, (B) 30 µg, e (C) sem proteína seguido de anticorpo anti-6His tag, anti-IgG de camundongo conjugado a peroxidase e revelação com ECL, como descrito em “Material e Métodos”.

Na figura 20 estão mostrados os Dot Blots feitos para avaliar a ligação da laminina e fibronectina com a proteína recombinante Tc85-34. Como pode ser observado, a reação da lamina e fibronectina apenas na incubação com anticorpo monoclonal anti-IgG de camundongo conjugado à peroxidase (20 B e F) é praticamente de mesma intensidade quando se incuba previamente com a proteína recombinante (20 C, D, G e H). Desta maneira, pode-se afirmar que não ocorre ligação entre a Tc85-34 e laminina ou fibronectina. Estes resultados são concordantes com experimentos de ligação feitos com células de mamífero e matriz extracelular (dados do trabalho de Mestrado de Paula Signorini).

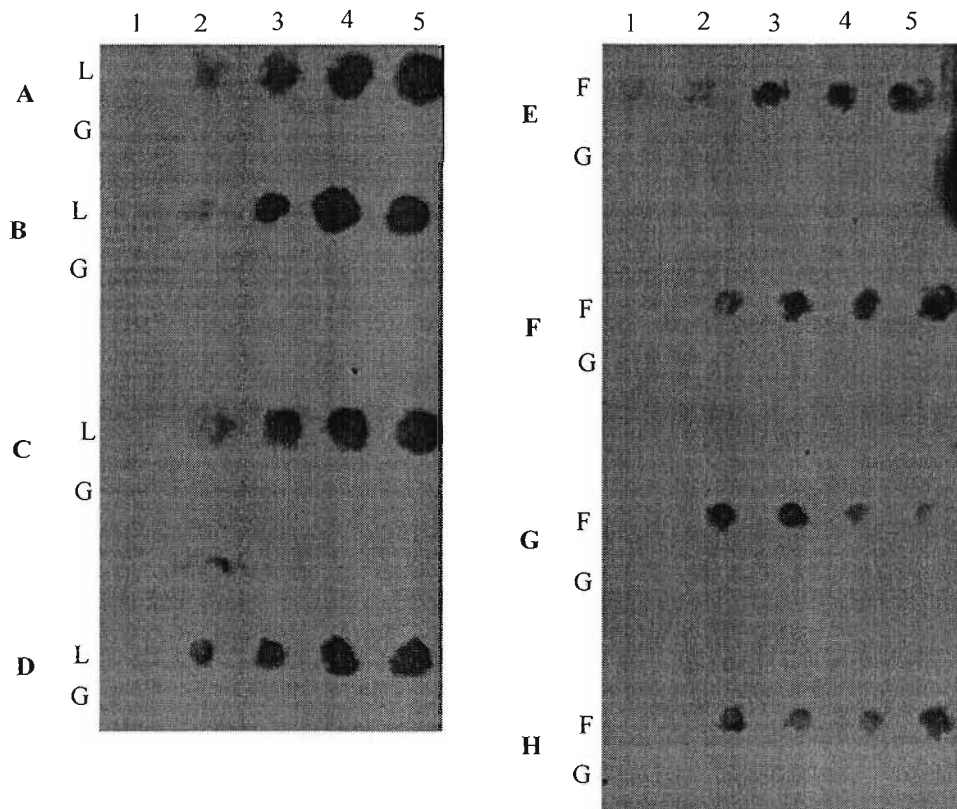


Figura 20. Reatividade da proteína de fusão Tc85-34 com laminina e fibronectina. Foram aplicadas quantidades crescentes de gelatina (G), laminina (L) fibronectina (F) em membrana de nitrocelulose na seguinte ordem (1) 0,5 μg , (2) 1 μg , (3) 2 μg , (4) 3 μg e (5) 5 μg . Após o bloqueio com leite desnatado a 3% em TCMM, a membrana foi incubada com 30 μg de Tc85-34 (D e H), 10 μg de Tc85-34 (C e G), sem proteína de fusão (A e E) e sem anticorpo anti-6His tag (B e F), seguido de anticorpo anti-6His tag, anti-IgG de camundongo conjugado a peroxidase e revelação com ECL, como descrito em “Material e Métodos”.

5. CONSIDERAÇÕES FINAIS

Os resultados obtidos pelos diferentes grupos da área levaram nosso laboratório a trabalhar com a hipótese de que a família da Tc-85 é constituída por diferentes membros que se ligam a diferentes componentes da matriz extracelular, da membrana basal e da superfície celular do hospedeiro. Para ampliar os dados sobre a família e verificar a validade da hipótese acima citada, houve a necessidade de obtenção de diferentes proteínas da família, sendo que o passo inicial foi a construção de uma biblioteca de cDNA enriquecida em genes da família da Tc-85 (BEM/Tc-85). Os objetivos deste projeto visaram aumentar nosso conhecimento sobre a interação do parasita com o meio externo, não restrito à invasão propriamente dita da célula, mas também como forma de locomoção dentro do hospedeiro. O *T. cruzi* necessita interagir com estruturas macromoleculares do hospedeiro para ultrapassar as barreiras impostas pelo organismo e, assim, alcançar a corrente sanguínea e órgãos-alvo. A laminina, ligante de Tc85-11 (Giordano et al., 1999; Quelopana, 2003), é um dos principais componentes da lâmina basal, que envolve vasos e células musculares e nervosas (Bosman & Stamenkovic, 2003), células alvo do *T. cruzi*. A citoqueratina 18 (CK18) foi caracterizada em nosso laboratório como um outro ligante da Tc85-11 nas células epiteliais (Magdesian et al., 2001). Portanto, o laboratório pretende mapear a Tc-85 (extremidades C e N-terminal) buscando domínios em sua estrutura que sejam relevantes no processo infeccioso do parasita. Simultaneamente, tem-se a intenção de caracterizar os receptores de Tc-85 em células de mamíferos, pois sabe-se que a interação parasita-célula é do tipo receptor-ligante.

Em um contexto de várias evidências da performance efetiva de elementos da matriz extracelular no mecanismo de infecção do parasita e o pequeno conhecimento da variabilidade gênica das proteínas de superfície, este trabalho pretendeu demonstrar a importância de alguns sítios de adesão de proteínas da família da Tc-85 de *T. cruzi*. A existência de tal variabilidade genética pode ser relacionada com diferentes padrões de expressão e/ou parte das proteína codificadas são trocadas como forma de adaptação ao meio ambiente.

Além da Tc85-11, ligante de laminina, glicoproteínas de aproximadamente 80.000 a 90.000 Da foram descritas como ligantes de fibronectina (Ouaisi, et al.,

1992) e como moléculas de adesão celular (Lima & Villalta, 1989, Ramirez et al., 1993). Recentemente, foi demonstrado que a ligação do *T. cruzi* a laminina é aumentada por galectina-3 em cerca de 20 vezes (Moody et al., 2000), reforçando o papel de laminina e sua possível modulação no processo de integração com o *T. cruzi*.

Kahn et al., 1990, relataram para glicoproteínas de 85 kDa a existência de no mínimo 100 genes de 2-2,5 kb representando pelo menos 1% do genoma completo do *T. cruzi*. Não está elucidado como é a expressão de diferentes proteínas desta família em diferentes parasitas de uma população e se isso afetaria a infecção em determinado tecido. Em trabalho recente, a expressão concomitante em um parasita de dois membros da família da SA85-1.1, também pertencente a superfamília das gp85/trans-sialidases, já foi detectada (Kahn et al., 1999). Recentemente, foi descrito que as regiões subteloméricas do parasita são enriquecidas em pseudogenes da família das trans-sialidases e da família VIPER de retroelementos. Foi especulado que a localização telomérica dos genes da família das trans-sialidases possa estar relacionada com variantes geradas por recombinação não-homóloga (Chiurillo et al., 1999 e 2002).

A exemplo de outras células, após a ligação, o mecanismo do parasita desconectar-se dos elementos da matriz extracelular poderia ser devido à liberação da glicoproteína de 85 kDa na forma de vesículas membranosas (Abuin et al., 1996; Gonçalves et al., 1991) ou por ação de proteases. Duas proteases foram bem caracterizadas, a cruzipaina (Cazzulo et al., 1997) e a colagenase (Santana et al., 1997), como elementos importantes na infecção de *T. cruzi*. Desta forma, tanto a liberação da glicoproteína de 85 kDa quanto a digestão por proteases poderiam auxiliar a migração do parasita através das estruturas da matriz extracelular e lâmina basal. A cruzipaina foi detectada associada com fibras de fibronectina no miocárdio de camundongos cronicamente infectados, apesar do papel da fibronectina e seus receptores na fisiopatologia da miocardite crônica causada pelo *T. cruzi* não estar esclarecido (Marino et al., 2003).

Nós demonstramos por experimentos *in vitro*, nos quais elementos de matriz extracelular são testados, como ligantes de proteínas recombinantes purificadas, que a proteína codificada pelo inserto de cDNA Tc85-45 é capaz de se ligar de maneira saturável a laminina e a fibronectina, mas não a albumina e gelatina. Isto nos permite

inferir que a seqüência de peptídeos específica para a ligação de diferentes moléculas de matriz extracelular e receptores celulares varia entre os membros da família, constituindo uma família de glicoproteínas multiadesivas, pelo menos ao que se refere ao grupo I.

Várias linhas de pesquisa demonstraram que a fibronectina promove adesão e captura de tripomastigotas por macrófagos e fibroblastos (Wirth & Kierszenbaum, 1984; Ouaiissi et al., 1985) e amastigotas por macrófagos murinos ou humanos (Noisin & Villalta, 1989). As formas tripomastigotas de *T. cruzi* reconhecem especificamente e se ligam à fibronectina com o envolvimento da seqüência RGD e inibem a infecção da célula hospedeira. (Ouaiissi et al., 1986).

A expressão aumentada de elementos da matriz extracelular, incluindo fibronectina, laminina e colágenos tipos III e IV foi descrita no tecido cardíaco de pacientes chagásicos crônicos (Andrade et al., 1989; Calvet et al., 2004). O aumento de laminina também foi descrito *in vitro* em culturas infectadas pelo *T. cruzi* (Marino et al., 2003). Além disso, células de mamíferos infectadas por *T. cruzi* produzem citocinas e quimiocinas que não só participam do controle do parasitismo como também contribuem para o estabelecimento de lesões inflamatórias crônicas em vários tecidos-alvo e frequentemente instalam uma miocardite severa (Gilat et al., 1996).

A grande variação das glicoproteínas de superfície Tc-85 visualizadas em produtos maduros (cDNA) com a manutenção de blocos conservados- ainda que não signifique necessariamente presença de proteína- e sua correlação com a adesão a componentes da matriz extracelular, pode estar conectada à imunoevasão. A manutenção de grande informação genômica desta família durante o processo evolutivo pode estar relacionada com o mecanismo de sobrevivência da espécie, com os parasitas selecionados por sua capacidade de sobreviver e invadir células mais eficientemente, escapando do sistema imune. As VSG (Variant Surface Proteins) são expressas e estão presentes na superfície do *Trypanosoma brucei* como um mecanismo de escape do sistema imune (Cross, 1996). Um possível mecanismo semelhante com a expressão concomitante de vários membros da família na população do *T. cruzi*, responsáveis pela adesão do parasita a elementos do hospedeiro, propiciariam a fixação e entrada do parasita nas células alvo, com o consequente estabelecimento da infecção.

6. CONCLUSÕES

1. A análise da biblioteca de cDNA enriquecida em membros da família da Tc-85 mostrou identidade de 40-90% entre seus membros, sendo que nenhum era idêntico entre si (família multigênica complexa)
2. Os 30 membros obtidos quando comparados por BLASTX dividem-se em dois subgrupos majoritários: (I) maior identidade com proteínas de superfície de tripomastigotas; (II) maior identidade com proteínas de superfície de amastigotas.
3. O estudo comparativo mostra a presença de variantes na região C-terminal conhecida como FLY, a presença de dois motivos Asp box, sendo que um terceiro motivo Asp degenerado foi observado em algumas seqüências (STDxSxxDTW-Asp “like”).
4. Todos os membros do grupo I possuem a seqüência xGExPLExxxFCF, o epítipo do monoclonal H1A10, e a seqüência RGD.
5. Parte dos membros do grupo I contém os peptídeos caracterizados como ligantes de laminina.
6. Ensaios *in vitro* de adesão da proteína recombinante Tc85-45 a proteínas de matriz extracelular sugerem a ligação à laminina e à fibronectina, confirmando a hipótese de que esse grupo nos tripomastigotas está relacionado com adesão a elementos de matriz.
7. Ensaios *in vitro* de adesão da proteína recombinante Tc85-34 a proteínas de matriz extracelular, mostram que esta proteína não se liga à laminina e à fibronectina, sugerindo que este grupo não está relacionado com adesão a elementos de matriz.

7. BIBLIOGRAFIA

- Abuin, G., Couto, A. S., Lederkremer, R. M., Casal, O. L., Galli, C., Colli, W. and Alves, M. J. M. (1996) Trypanosoma cruzi: the Tc-85 surface glycoprotein shed by trypomastigotes bears a modified glycosylphosphatidylinositol anchor. *Exp Parasitol* **82(3)**: 290-7.
- Abuin, G., Colli, W., Alves, M. J. (1989) Turnover and shedding of the Tc-85 surface glycoprotein of *Trypanosoma cruzi* trypomastigotes. *Braz J Med Biol Res*, **29(3)**: 335-341.
- Adams, J. C. (1997) Thrombospondin-1. *Int J Biochem Cell Biol*, **29**: 861-865.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2002) *Molecular Biology of The Cell*. Fourth ed. Garland Publishing. New York and London.
- Almeida-de-Faria M., Freymuller E., Colli W., Alves M. J. M. (1999) *Trypanosoma cruzi*: characterization of an intracellular epimastigote-like form. *Exp Parasitol*. **92(4)**: 263-274.
- Alves, M. J. M., Albuin, G., Kuwajima, V. Y., Colli, W. (1986) Partial inhibition of trypomastigote entry into cultured mammalian cells by monoclonal antibodies against a surface glycoprotein of *Trypanosoma cruzi*. *Mol Biochem Parasitol*, **21**: 75-82.
- Andrade S. G., Grimaud J. A., Stocker-Guerret S. (1989) Sequential changes of the connective matrix components of the myocardium (fibronectin and laminin) and evolution of the cardiac fibrosis in mice infected with *Trypanosoma cruzi*. *Amer J Trop Med Hyg*, **40**: 252-260.
- Andrade S. G. (1974) Caracterização de cepas do *Trypanosoma cruzi* isoladas no Recôncavo Baiano. *Rev Pat Trop*, **1**: 65-121.
- Andrews N. W., Colli W. (1982) Adhesion and interiorization of *Trypanosoma cruzi* in mammalian cells. *J Protozool*, **29(2)**: 264-269.
- Barnes, M. J., Knight, C. G., Farndale, R. W. (1998) The collagen-platelet interaction. *Curr Opin Hematol*, **5(5)**: 314-320.
- Belkin, A. M., Stepp, M. A. (2000) Integrins as receptors for laminins. *Micro Res Tech*, **51**: 280-301.
- Bingle L.E., Eastlake J.L., Bailey M., Gibson W.C. (2001) A novel GFP approach for the analysis of genetic exchange in trypanosomes allowing the *in situ* detection of mating events. *Microbiol*, **147(Pt 12)**: 3231-40.
- Bosman F.T., Stamenkovic I. (2003) Functional structure and composition of the extracellular matrix. *J Pathol*, **200(4)**: 423-8

- Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, **72**: 248-254.
- Brenner, W., Gross, S., Steinbach, F., Horn, S., Hohenfellner R., Thuroff, J. W. (2000) Differential inhibition of renal cancer cell invasion mediated by fibronectin, collagen and laminin. *Cancer Lett*, **155(2)**: 199-205.
- Briones M. R., Souto R. P., Stolf B. S., Zingales B. (1999) The evolution of two *Trypanosoma cruzi* subgroups inferred from rRNA genes can be correlated with the interchange of American mammalian faunas in the Cenozoic and has implications to pathogenicity and host specificity. *Mol Biochem Parasitol*, **104(2)**: 219-232.
- Brisse S., Barnabe C., Tibayrenc M. (2000) Identification of six *Trypanosoma cruzi* phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis. *Int J Parasitol*, **30(1)**: 35-44.
- Burleigh, B. A., Woolsey, A.M. (2002) Cell signaling and *Trypanosoma cruzi* invasion. *Cell microbiol*, **4(11)**: 701-711.
- Burleigh, B. A., Andrews, N. W. (1998) Signaling and host cell invasion by *Trypanosoma cruzi*. *Curr Opin Microbiol*, **1**: 461-465.
- Burleigh, B. A., Andrews, N. W. (1995) The mechanisms of *Trypanosoma cruzi* invasion of mammalian cells. *Annu Rev Microbiol*, **49**: 175-200
- Buscaglia C.A., Di Noia J. M. (2003) *Trypanosoma cruzi* clonal diversity and the epidemiology of Chagas' disease. *Microbes Infect*, **5(5)**:419-27.
- Buscaglia, C. A., Campetella O., Leguizamon M. S., Frasch A. C. (1998) The repetitive domain of *Trypanosoma cruzi* trans-sialidase enhances the immune response against the catalytic domain. *J Infect Dis*, **177**: 431- 436.
- Buschiazzo A., Campetella, O., Frasch, A. C. (1997) *Trypanosoma rangeli* sialidase: cloning, expression and similarity to *T. cruzi* trans-sialidase. *Glycobiol*, **7(8)**: 1167-1173.
- Calvet C. M., Meuser M., Almeida D., Meirelles M. N., Pereira M. C. (2004) *Trypanosoma cruzi*-cardiomyocyte interaction: role of fibronectin in the recognition process and extracellular matrix expression in vitro and in vivo. *Exp Parasitol*, **107(1-2)**: 20-30.
- Cazzulo J. J., Stoka V., Turk V. (1997) Cruzipain, the major cysteine proteinase from the protozoan parasite *Trypanosoma cruzi*. *Biol Chem*, **378(1)**:1-10.
- Chagas, C. (1909) Nova tripanozomiase humana: Estudos sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi* n. gen., n. sp., agente etiológico de nova entidade morbida do homem. *Mem Inst Oswaldo Cruz*, **1(2)**: 159-218.

- Chiquet, M. (1999) Regulation of extracellular matrix gene expression by mechanical stress. *Matrix Biol*, **18(5)**: 417-426.
- Chiurillo, M.A, Cano, I., Franco da Silveira, J., Ramirez, J. L. (1999) Organization of telomeric and sub-telomeric regions of chromosomes from the protozoan parasite *Trypanosoma cruzi*. *Mol Biochem Parasitol*, **100**: 173-183.
- Chiurillo, M.A, Santos, M. R. M., Franco da Silveira, J., Ramirez, J. L. (2002) A general improved approach for cloning and characterization of telomeres: the protozoan parasite *Trypanosoma cruzi* as model organism. *Gene*, **294(1-2)**: 197-204.
- Cheng, C., Shuman, S. (2000) DNA strand transfer catalyzed by vaccinia topoisomerase: ligation of DNAs containing a 3' mononucleotide overhang. *Nucl Acids Res*, **28(9)**: 1893-1898.
- Colli, W. (1993) Trans-sialidase: a unique enzyme activity discovered in the protozoan *Trypanosoma cruzi*. *FASEB J*, **7(13)**: 1257-1264.
- Cooper G. M. (1996) *The Cell: A Molecular Approach*. Washington, D.C: ASM Press/Sinauer Associates.
- Coura J. R., Junqueira A. C., Fernandes O., Valente S. A., Miles M.A. (2002) Emerging Chagas disease in Amazonian Brazil. *Trends Parasitol.*, **18(4)**: 171-176.
- Couto, A. S., Lederkremer, R. M., Colli, W., Alves, M. J. M. (1993) The glycoposphatidylinositol anchor of the trypanomastigote specific Tc-85 glycoprotein from *Trypanosoma cruzi*. *Eur J Biochem*, **217**: 597-602.
- Couto, A. S., Gonçalves, M. F., Colli, W., Lederkremer, R. M. (1990) The N-linked carbohydrate chain of the Tc-85 kilodalton glycoprotein from *Trypanosoma cruzi* trypanomastigotes contains sialyl, fucosyl and galactosyl (α1-3) galactose units. *Mol Biochem Parasitol*, **39**: 101-108.
- Cross, G. A. M., Takle, G. B. (1993) The surface trans-sialidase family of *Trypanosoma cruzi*. *Annu Rev Microbiol*, **47**: 385-411.
- Cross, G.A. (1996) Antigenic variation in trypanosomes: secrets surface slowly. *Bioessays*, **18(4)**: 283-91.
- Da Silveira, J. F. Biologia Molecular do *Trypanosoma cruzi*. (2000) In: *Trypanosoma cruzi e a Doença de Chagas*, segunda ed., Guanabara Koogan, Rio de Janeiro, Brasil.
- De Souza, W. (1984) Cell biology of *Trypanosoma cruzi*. *Int Rev Cytol*, **86**: 197-283..
- Dias, E. (1934) Estudos sobre o *Schizotrypanum cruzi*. *Mem Inst Oswaldo Cruz*, **28**: 1-110.
- Dias, J. C. P. (1992). Epidemiology of Chagas' disease. In: *Chagas' Disease (American Trypanosomiasis): Its Impact on Transfusion and Clinical Medicine* (S. Wendel, Z.

- Brener, M. E. Camargo, A. Rassi, eds.), pp.49-80, São Paulo: International Society of Blood Transfusions.
- Dvorak, J. A., Howe, C. L. (1976) The attraction of *Trypanosoma cruzi* to vertebrate cells *in vitro*. *J Protozool*, **23**: 534-537.
- Ensoli B., Barillari G., Salahuddin S. Z., Gallo R. C., Wong-Staal F. (1990) Tat protein of HIV-1 stimulates growth of cells derived from Kaposi's sarcoma lesions of AIDS patients *Nature*, **345(6270)**: 84-86.
- Erickson, A. C., Couchman, J. R. (2000) Still more complexity in mammalian basement membranes. *J Histochem Cytochem*, **48**: 1291-1306.
- Felsenstein, J. (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, **39**: 783-791.
- Fire A., Xu S., Montgomery M. K., Kostas S. A., Driver S. E., Mello C. C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, **391(6669)**: 806-11.
- Franchin, G., Pereira-Chioccola, V. L., Schenkman, S., Rodrigues, M. M. (1997) Passive transfer of a monoclonal antibody specific for a sialic acid-dependent epitope on the surface of *Trypanosoma cruzi* trypomastigotes reduces infection in mice. *Infect Immun*, **65(7)**: 2548-2554.
- Frasch, A. C. (2000) Functional diversity in the trans-sialidase and mucin families in *Trypanosoma cruzi*. *Parasitol Today*, **16(7)**: 282-286.
- Gaunt M. W., Yeo M., Frame I. A., Stothard J. R., Carrasco H. J., Taylor M. C., Mena S. S., Veazey P., Miles G. A., Acosta N., de Arias A. R., Miles M. A. (2003) Mechanism of genetic exchange in American trypanosomes. *Nature*, **421(6926)**: 936-9.
- Gilat D., Cahalon L., Hershkoviz R. and Lider O. (1996). Interplay of T cells and cytokines in the context of enzymatically modified extracellular matrix. *Immun Today*, **17**: 16-20.
- Giordano, R., Fouts, D. L., Tewari, D., Colli, W., Manning, J. E., Alves, M. J. M. (1999) Cloning of a surface membrane glycoprotein specific for the infective form of *Trypanosoma cruzi* having adhesive properties to laminin. *J Biol Chem*, **274(6)**: 3461-3468.
- Giordano, R., Chammas, R., Veiga, S. S., Colli, W., Alves, M. J. (1994) An acidic component of the heterogenous Tc-85 protein family from the surface of *Trypanosoma cruzi* is a laminin binding glycoprotein. *Mol Biochem Parasitol*, **65(1)**: 85-94.
- Goldfinger, L.E., Hopkinson, S. B., deHart, G. W., Collawn, S., Couchman, J.R., Jones, J.C. (1999) The $\alpha 3$ laminin subunit, $\alpha 6\beta 4$ and $\alpha 3\beta 1$ integrin coordinately regulate wound healing in cultured epithelial cells and in the skin. *J Cell Sci*, **112**: 2615–2629.

- Gonçalves, M. F., Umezawa, E. S., Katzin, A. M., de Souza, W., Alves, M. J. M., Zingales, B., Colli W. (1991) *Trypanosoma cruzi*: shedding of surface antigens as membrane vesicles. *Exp Parasitol*, **72(1)**:43-53.
- Hanahan, D. (1983) Transformation of *Escherichia coli* with plasmids. *J Mol Biol*, **166**: 557-580.
- Hodivala-Dilke, K. M., DiPersio, C. M., Kreidberg, J. A., Hynes RO. (1998) Novel roles for alpha3beta1 integrin as a regulator of cytoskeletal assembly and as a trans-dominant inhibitor of integrin receptor function in mouse keratinocytes. *J Cell Biol*, **142**: 1357–1369.
- Huang X, Madan A. (1999) CAP3: A DNA sequence assembly program. *Genome Res*, **9(9)**: 868-877.
- Huang S., Endo R. I., Nemerow G. R. (1995) Upregulation of integrins alpha v beta 3 and alpha v beta 5 on human monocytes and T lymphocytes facilitates adenovirus-mediated gene delivery. *J Virol*, **69(4)**: 2257-2263.
- Kahn, S. J., Nguyen, D., Norsen, J., Wleklinski, M. Granston, T., Kahn, M. (1999) *Trypanosoma cruzi*: monoclonal antibodies to the surface glycoprotein superfamily differentiate subsets of the 85-kDa surface glycoproteins and confirm simultaneous expression of variant 85-kDa surface glycoproteins. *Exper Parasitol*, **92**: 48-56
- Kahn S., Van Voorhis W. C., Eisen H. (1990) The major 85-kD surface antigen of the mammalian form of *Trypanosoma cruzi* is encoded by a large heterogeneous family of simultaneously expressed genes. *J Exp Med*, **172(2)**: 589-97.
- Katzin, A. M., Colli, W. (1983) Lectin receptors in *Trypanosoma cruzi*, an N-acetyl-D-glucosamine containing surface glycoprotein specific for the trypomastigote stage. *Biochim Biophys Acta*, **727**: 403-411.
- Koch, M., Olson, P. F., Albus, A., Jin, W., Hunter, D. D., Brunken, W. J., Burgeson, R. E., Champlaud, M.F. (1999) Characterization and expression of the laminin γ 3 chain: a novel, non-basement membrane-associated, laminin chain. *J Cell Biol*, **145(3)**: 605-617.
- Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**: 680-685.
- Lea S., Abu-Ghazaleh R., Blakemore W., Curry S., Fry E., Jackson T., King A., Logan D., Newman J., Stuart D. (1995) Structural comparison of two strains of foot-and-mouth disease virus subtype O1 and a laboratory antigenic variant, G67. *Structure*, **3(6)**: 571-580.
- Leininger E., Roberts M., Kenimer J. G., Charles I. G., Fairweather N., Novotny P., Brennan M. J. (1991) Pertactin, an Arg-Gly-Asp-containing *Bordetella pertussis* surface protein that promotes adherence of mammalian cells. *Proc Natl Acad Sci USA*, **88(2)**: 345-349.

- Ley V., Andrews N. W., Robbins E. S., Nussenzweig V. (1988) Amastigotes of *Trypanosoma cruzi* sustain an infective cycle in mammalian cells. *J Exp Med.* **168(2)**: 649-59.
- Lima M. F., Villalta F. (1989) *Trypanosoma cruzi* trypomastigote clones differentially express a parasite cell adhesion molecule. *Mol Biochem Parasitol.* **33(2)**: 159-70.
- Machado C. A., Ayala F. J. (2001) Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. *Proc Natl Acad Sci USA*, **98(13)**: 7396-7401.
- Magdesian M. H., Giordano R., Ulrich H., Juliano M. A., Juliano L., Schumacher R. I., Colli W., Alves M. J. M. (2001) Infection by *Trypanosoma cruzi*- Identification of a parasite ligand and its host cell receptor. *J Biol Chem*, **276(22)**: 19382-19389.
- Mair G., Shi H., Li H., Djikeng A., Aviles H. O., Bishop J. R., Falcone F. H., Gavrilescu C., Montgomery J. L., Santori M. I., Stern L. S., Wang Z., Ullu E., Tschudi C. (2000) A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. *RNA*, **6(2)**: 163-169.
- Marino A. P., Silva A. A., Pinho R. T., Lannes-Vieira J. (2003) *Trypanosoma cruzi* infection: a continuous invader-host cell cross talk with participation of extracellular matrix and adhesion and chemoattractant molecules. *Braz J Med Biol Res*, **36(8)**: 1121-1133
- Melo, R. C., Brenez, Z. (1978) Tissue tropism of different *Trypanosoma cruzi* strains. *J Parasitol*, **64 (3)**: 475-482.
- Ming M., Ewen, M. E., Pereira, M. E. A. (1995) Trypanosome invasion of mammalian cells requires activation of the TGF β signaling pathway. *Cell*, **82**: 287-296.
- Moody, T. N., Ochieng, J., Villalta, F. (2000) Novel mechanism that *Trypanosoma cruzi* uses to adhere to the extracellular matrix mediated by human galectin-3. *FEBS Letters*, **470**: 305-308.
- Mortara R. A. (1991) *Trypanosoma cruzi*: amastigotes and trypomastigotes interact with different structures on the surface of HeLa cells. *Exp Parasitol*, **73(1)**: 1-14.
- Miura, R., Aspberg, A., Ethell, I. M., Hagihara, K., Schnaar, R. L., Rouslahti, E., Yamaguchi, Y. (1999) The proteoglycan lectin domain binds sulfated cell surface glycolipids and promotes cell adhesion. *J Biol Chem*, **274(16)**: 11431-11438.
- Noisin E. L., Villalta F. (1989) Fibronectin increases *Trypanosoma cruzi* amastigote binding to and uptake by murine macrophages and human monocytes. *Infect Immun*, **57**: 1030-1034.
- Ortega-Barria, E., Pereira, M. E. (1991) A novel *T. cruzi* heparin-binding protein promotes fibroblast adhesion and penetration of engineered bacteria and trypanosomes into mammalian cells. *Cell*, **67(2)**: 411-421.

- Ouaissi, A., Cornette J., Schoneck R., Plumas-Marty, B., Taibi A., Loyens M., Capron A. (1992) Fibronectin cleavage fragments provide a growth factor-like activity for the differentiation of *Trypanosoma cruzi* trypomastigotes to amastigotes. *Eur J Cell Biol*, **59(1)**: 68-79.
- Ouaissi M.A., Cornette J., Afchain D., Capron A., Gras-Masse H., Tartar A. (1986). *Trypanosoma cruzi* infection inhibited by peptides modeled from a fibronectin cell attachment domain. *Science*, **234**: 603-607.
- Ouaissi M.A., Cornette J., Capron A. (1985). *Trypanosoma cruzi*: modulation of parasite-cell interaction by plasma fibronectin. *Eur J Immun*, **15**: 1096-1101.
- Ouaissi M. A., Afchain D., Capron A., Grimaud J. A. (1984) Fibronectin receptors on *Trypanosoma cruzi* trypomastigotes and their biological function. *Nature*, **308(5957)**: 380-382
- Pereira, C. M., Favoreto Jr, S., Da Silveira, J. F., Yoshida, N., Castilho, B. A. (1999) Adhesion of *Escherichia coli* to HeLa cells mediated by *Trypanosoma cruzi* surface glycoprotein-derived peptides inserted in the outer membrane protein LamB. *Infect Immun*, **67(9)**: 4908-4911.
- Pereira, M. E., Mejia, J. S., Ortega, B. E., Matzilevich, D., Prioli, R. P. (1991) The *Trypanosoma cruzi* neuraminidase contains sequences similar to bacterial neuraminidases, YWTD repeats of the low-density lipoprotein receptor, and type III modules of fibronectin. *J Exp Med*, **174**: 179-191.
- Quelopana, M. M. *Trypanosoma cruzi* e a interação com a matriz extracelular: Modelagem da proteína Tc85-11 e determinação do sítio de ligação a laminina. *Tese de Doutorado*, Instituto de Química da Universidade de São Paulo, 2003.
- Ramirez, M. I., Boscardin, S. B., Sang, W. H., Paranhos-Baccala, G., Yoshida, N., Kelly, J. M., Mortara, R. A., Da Silveira, J. F. (1999) Heterologous expression of a *Trypanosoma cruzi* surface glycoprotein (gp82) in mammalian cells indicates the existence of different signal sequence requirements and processing. *J Eukaryot Microbiol*, **46(6)**: 557-565.
- Ramirez, M. I., Ruiz, R. C., Araya, J. E., Da Silveira, J. F., Yoshida, N. (1993) Involvement of the stage-specific 82-kilodalton adhesion molecule of *Trypanosoma cruzi* metacyclic trypomastigotes in host cell invasion. *Infect Immun*, **6**: 3636-3641.
- Ribeirão, M., Pereira-Chioccola, V. L., Eichinger, D., Rodrigues, M.M., Schenkman, S. (1997) Temperature differences for trans-glycosylation and hydrolysis reaction reveal an acceptor binding site in the catalytic mechanism of *Trypanosoma cruzi* sialidase. *Glycobiol*, **7(8)**: 1237-1246.
- Rodriguez A., Webster P., Ortego J., Andrews N. W. (1997) Lysosomes behave as Ca^{2+} -regulated exocytic vesicles in fibroblasts and epithelial cells. *J Cell Biol*, **137(1)**: 93-104.

- Roivainen M., Piirainen L., Hovi T., Virtanen I., Riikonen T., Heino J., Hyypia T. (1994) Entry of coxsackievirus A9 into host cells: specific interactions with alpha v beta 3 integrin, the vitronectin receptor. *Virology*, **203(2)**: 357-365.
- Rostand, K. S., Esko, J. D. (1997) Microbial adherence to and invasion through proteoglycans. *Infect Immun*, **65(1)**: 1-8.
- Ruoslahti E. (1996) RGD and other recognition sequences for integrins. *Annu Rev Cell Dev Biol*, **12**: 697-715.
- Ryan, M. C., Christiano, A. M. (1996) The functions of laminins: lessons from *in vivo* studies. *Matrix Biology*, **15**: 369-381.
- Sage, A.H., Bornstein, P. (1991) Extracellular proteins that modulate cell-matrix interactions SPARC, tenascin and thrombospondin. *J Biol Chem*, **266(23)**: 14831-14834.
- Sambrook E. U., Fritsch E. F., Maniatis, Molecular Cloning A Laboratory Manual. Second ed. (1989) Cold Spring Harbour Laboratory Press. New York.
- Santana J. M., Grellier P., Schrevel J., Teixeira A. R. (1997) A *Trypanosoma cruzi*-secreted 80 kDa proteinase with specificity for human collagen types I and IV. *Biochem J*, **325 (Pt 1)**: 129-37.
- Sanger, F., Nicklen, S., Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitor. *Proc Natl Acad Sci USA*, **74**: 5463-5467
- Scharfstein J., Schmitz V., Morandi V., Capella M. M., Lima A. P., Morrot A., Juliano L., Muller-Esterl W. (2000) Host cell invasion by *Trypanosoma cruzi* is potentiated by activation of bradykinin B(2) receptors. *J Exp Med*, **192(9)**: 1289-1300.
- Schenkman, S., Eichinger, D., Pereira, M. E. A., Nussenzweig, V. (1994) Structural and functional properties of *Trypanosoma* trans-sialidase. *Annu Rev Microbiol*, **48**: 499-523.
- ^aSchenkman, S., Jiang, M. S., Hart, G. W., Nussenzweig, V. (1991) A novel cell surface trans-sialidase of *Trypanosoma cruzi* generates a stage-specific epitope required for invasion of mammalian cells. *Cell*, **65**: 1117-1125.
- ^bSchenkman, S., Robbins, E. S., Nussenzweig, V. (1991) Attachment of *Trypanosoma cruzi* to mammalian cells requires parasite energy, and invasion can be independent of target cell cytoskeleton. *Infect Immun*, **59**: 645-654.
- Silva M. B., Schattner M., Ramos C. R., Junqueira-de-Azevedo I. L., Guarnieri M. C., Lazzari M. A., Sampaio C. A., Pozner R. G., Ventura J. S., Ho P. L., Chudzinski-Tavassi A. M. (2003) A prothrombin activator from *Bothrops erythromelas* (jararaca-da-seca) snake venom: characterization and molecular cloning. *Biochem J*, **369(Pt 1)**: 129-139.

- Silveira, A. C. (2000) Current situation with Chagas disease vector control in the Americas. *Cad Saúde Pública*, **16 (sup. 2)**: 35-42
- Smith, L. E., Eichinger, D. (1997) Directed mutagenesis of the *Trypanosoma cruzi* trans-sialidase enzyme identifies two domains involved in its sialyltransferase activity. *Glycobiol*, **7(3)**: 445-451.
- Smith, T.F., Waterman, M.S. (1981) Identification of common molecular subsequences. *J Mol Biol*, **147**: 195-197.
- Tardieux, I., Nathanson, M. H., Andrews, N. A. (1994) Role in host cell invasion of *Trypanosoma cruzi*-induced cytosolic free Ca^{2+} transients. *J Exp Med*, **179**: 1017-1022.
- Tardieux, I., Webster, P., Ravesloot, J., Boron, W., Lunn, J. A., Heuser, J. E., Andrews, N. A. (1992) Lysosome recruitment and fusion are early events required for Trypanosome invasion of mammalian cells. *Cell*, **71**: 1117-1130.
- Tashiro, K., Monji, A., Yoshida, I., Hayashi, Y., Matsuda, K., Tashiro, N., Mitsuyama, Y. (1999) An IKLLI-containing peptide derived from the laminin $\alpha 1$ chain mediating heparan-binding, cell adhesion, neurite outgrowth and proliferation, represents a binding site for integrin $\alpha 3\beta 1$ and heparan sulphate proteoglycan. *Biochem J*, **340**: 119-126.
- Teixeira S. M., daRocha W. D. (2003) Control of gene expression and genetic manipulation in the Trypanosomatidae. *Genet Mol Res*, **2(1)**: 148-58.
- Timpl R., Rohde H., Robey P. G., Rennard S. I., Foidart J. M., Martin G. R. (1979) Laminin- a glycoprotein from basement membranes. *J Biol Chem*, **254(19)**: 9933-9937.
- Tomlinson, S., Raper, J. (1998) Natural human immunity to trypanosomes. *Parasitol Today*, **14**: 354-359.
- Tonelli R. R., Silber A. M., Almeida-de-Faria M., Hirata I. Y., Colli W., Alves M. J. M. (2004) L-Proline is essential for the intracellular differentiation of *Trypanosoma cruzi*. *Cel Microbiol*, **6(8)**: 733-741.
- Thompson J. D., Higgins D. G., Gibson T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, **22(22)**: 4673-4680.
- Towbin, H., Staehelin, T., Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrilamide gels to nitrocellulose sheets: Procedures and some applications. *Proc Natl Acad Sci USA*, **76**: 4350-4354.
- Truyens C., Rivera M. T., Ouassiss A., Carlier Y. (1995) High circulating levels of fibronectin and antibodies against its RGD adhesion site during mouse *Trypanosoma cruzi* infection: relation to survival. *Exp Parasitol*, **80(3)**: 499-506.

- Ullu E., Djikeng A., Shi H., Tschudi C. (2002) RNA interference: advances and questions. *Philos Trans R Soc Lond B Biol Sci*, **357(1417)**: 65-70.
- Van Nhieu G. T., Isberg R. R. (1991) The *Yersinia pseudotuberculosis* invasin protein and human fibronectin bind to mutually exclusive sites on the alpha 5 beta 1 integrin receptor. *J Biol Chem*, **266(36)**: 24367-24375.
- Wilkowsky S. E., Barbieri M. A., Stahl P. D., Isola E. L. (2002) Regulation of *Trypanosoma cruzi* invasion of nonphagocytic cells by the endocytically active GTPases dynamin, Rab5, and Rab7. *Biochem Biophys Res Commun*, **291(3)**: 516-521.
- Wirth J. J., Kierszenbaum F. (1984). Fibronectin enhances macrophage association with invasive forms of *Trypanosoma cruzi*. *J Immunol*, **133**: 460-464.
- WHO Expert Committee. (2002) Control of Chagas disease. *World Health Organ Tech Rep Ser*, **905, i-vi**: 1-109.
- Woolsey A. M., Sunwoo L., Petersen C. A., Brachmann S. M., Cantley L. C., Burleigh B. A. (2003) Novel PI 3-kinase-dependent mechanisms of trypanosome invasion and vacuole maturation. *J Cell Sci*, **116(Pt 17)**: 3611-3622.
- Zingales B., Souto R. P., Mangia R. H., Lisboa C. V., Campbell D. A., Coura J. R., Jansen A., Fernandes O. (1988) Molecular epidemiology of American trypanosomiasis in Brazil based on dimorphisms of rRNA and mini-exon gene sequences. *Int J Parasitol*, **28(1)**: 105-112.

ANEXO 1. Alinhamento global das seqüências nucleotídicas.

10	1	GTGACAAAAGAAAGAGTGTGAAATTTGAGTGA...GTTTCTCAACCCCTCTCTCTCAGTGTCTG	87
12	1	CTGCCGGAAGGTGAGGTACTCCAAGGACGAAAGCGGATGCA...TTTGTTCACCCTCTCTCTCAGTGTCTG	70
13	1	GTGACAAAAGAAAGAGTGTGAAATTTGAGTGA...GTTTCTCAACCCCTCTCTCTCAGTGTCTG	87
14	1	GTGACAAAAGAAAGAGTGTGAAATTTGAGTGA...GTTTCTCAACCCCTCTCTCTCAGTGTCTG	87
16	1	GTACCGAAAAGGTGAAAGTCTAAGAGACTAAGAGGGATTCA...TTTATTTACCCTCTCTCTCAGTGTCTG	70
18	1	GTACCGAAAAGGTGAAAGTCTAAGAGACTAAGAGGGATTCA...TTTATTTACCCTCTCTCTCAGTGTCTG	87
19	1	GTACCGAAAAGGTGAAAGTCTAAGAGACTAAGAGGGATTCA...TTTATTTACCCTCTCTCTCAGTGTCTG	70
2	1	GTGCCATAAGGGGGTACCAGTCAAGAGACGAAACGTTACT...CTTCCTCTCTCTCAGTGTCTG	70
20	1	GTACCGAAAAGGTGAAAGTCTAAGAGACTAAGAGGGATTCA...TTTATTTACCCTCTCTCTCAGTGTCTG	70
22	1	GTGACAAAAGAAAGAGTGTGAAATTTGAGTGA...GTTTCTCAACCCCTCTCTCTCAGTGTCTG	87
23	1	GTGACAAAAGAAAGAGTGTGAAATTTGAGTGA...GTTTCTCAACCCCTCTCTCTCAGTGTCTG	87
25	1	CTGCCGGAAGGTGAGGGACTCCAAGGTACGAAAGAGAGAC...GCTTCTGATCCCAAGCCTTGTCTCGGGG	70
26	1	GTGACAAAAGGTGAGAGTGAATAGGGGTGAATTCG...CATTCCTTACTCCGTCATTTGCTATCGGG	87
27	1	GTACCGAAAAGGTGAAAGTCTAAGAGACTAAGAGGGATTCA...TTTATTTACCCTCTCTCTCAGTGTCTG	70
29	1	GTGCCATAAGGGGGTACCAGTCAAGAGACGAAACGTTACT...CTTCCTCTCTCTCAGTGTCTG	70
32	1	GTACCGAAAAGGTGAAAGTCTAAGAGACTAAGAGGGATTCA...TTTATTTACCCTCTCTCTCAGTGTCTG	70
33	1	GTGACCAAGGAAAGAGAGCGAGATTGGGTCAATTTG...GTTTCTCAACCCCTAGCCTACTGTCTGGCG	87
34	1	GTAAACAAAAGAGAAATCTGAAATAGGGTCAAGTTTG...GTTTCTGACTCCCTCTACTGTCTGGCG	87
35	1	GTACCGAAAAGGTGAAAGTCTAAGAGACTAAGAGGGATTCA...TTTATTTACCCTCTCTCTCAGTGTCTG	70
36	1	GTACCGAAAAGGTGAAAGTCTAAGAGACTAAGAGGGATTCA...TTTATTTACCCTCTCTCTCAGTGTCTG	70
40	1	GTGACAAAAGAAAGAGTGTGAAATTTGAGTGA...GTTTCTCAACCCCTCTCTCTCAGTGTCTG	87
41	1	GTGACAAAAGAAAGAGTGTGAAATTTGAGTGA...GTTTCTCAACCCCTCTCTCTCAGTGTCTG	87
43	1	GTGACAAAAGAAAGAGTGTGAAATTTGAGTGA...GTTTCTCAACCCCTCTCTCTCAGTGTCTG	87
44	1	GTACCGAAAAGGTGAAAGTCTAAGAGACTAAGAGGGATTCA...TTTATTTACCCTCTCTCTCAGTGTCTG	70
45	1	GTACCGAAAAGGTGAAAGTCTAAGAGACTAAGAGGGATTCA...TTTATTTACCCTCTCTCTCAGTGTCTG	70
48	1	GTGACAAAAGAAAGAGTGTGAAATTTGAGTGA...GTTTCTCAACCCCTCTCTCTCAGTGTCTG	87
50	1	GTGACAAAAGAAAGAGTGTGAAATTTGAGTGA...GTTTCTCAACCCCTCTCTCTCAGTGTCTG	87
52	1	GTACCGAAAAGGTGAAAGTCTAAGAGACTAAGAGGGATTCA...TTTATTTACCCTCTCTCTCAGTGTCTG	70
54	1	GTACCGAAAAGGTGAAAGTCTAAGAGACTAAGAGGGATTCA...TTTATTTACCCTCTCTCTCAGTGTCTG	70
57	1	GTGACAAAAGAAAGAGTGTGAAATTTGAGTGA...GTTTCTCAACCCCTCTCTCTCAGTGTCTG	87
10	88	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	135
12	71	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	140
13	88	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	135
14	88	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	134
16	71	GTGGAGTAAATGCTGCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	138
18	88	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	135
19	71	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	138
2	71	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	137
20	71	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	138
22	88	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	135
23	88	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	135
25	71	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	140
26	88	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	134
27	71	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	138
29	71	GTGGAGTAAATGCTGCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	138
32	71	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	138
33	88	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	134
34	88	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	134
35	71	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	138
36	71	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	138
40	71	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	138
41	88	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	135
43	88	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	135
44	71	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	138
45	71	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	138
48	88	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	135
50	88	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	135
52	71	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	138
54	71	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	138
57	88	GTGGGTGATGGTTCCTTTTGGTGAAGAGTCTTGTGBCAAT...GATTTCTCTC-ACCGGGGGA-AAAAG	135
10	130	G...TTCAATTGAAAGTGTGTTGCGGGGTACAGTCAACGCT...GCGGAGACGTGGTGTCTCTGTCTGCG	201
12	141	GCCAAATTTCTTCTGATGTTGTTGCGGGGTACATCTGACTCT...TTGTTGGATGGTCCACTCTGTGTTGCCAAG	210
13	130	G...TTCAATTGAAAGTGTGTTGCGGGGTACAGTCAACGCT...GCGGAGACGTGGTGTCTCTGTCTGCG	201
14	135	GTTTGGCAANTTGAAGTGTGTTGCGGGGTACAGTCAACGCT...GCGGAGACGTGGTGTCTCTGTCTGCG	204
16	139	A...CTTCTTCGATGTTGTTGCGGGGTACATCTGACTCT...GCGGAGACGTGGTGTCTCTGTCTGCG	204
18	138	GT...TTCAATTGAAAGTGTGTTGCGGGGTACAGTCAACGCT...GCGGAGACGTGGTGTCTCTGTCTGCG	201
19	139	A...CTTCTTCGATGTTGTTGCGGGGTACATCTGACTCT...TTGTTGGATGGTCCACTCTGTGTTGCCAAG	204
2	138	AA...CATCGTCTGATGTTGTTGCGGGGTACATCTGACTCT...TTGTTGGATGGTCCACTCTGTGTTGCCAAG	204
20	139	A...CTTCTTCGATGTTGTTGCGGGGTACATCTGACTCT...TTGTTGGATGGTCCACTCTGTGTTGCCAAG	204
22	138	G...TTCAATTGAAAGTGTGTTGCGGGGTACAGTCAACGCT...GCGGAGACGTGGTGTCTCTGTCTGCG	201
23	130	G...TTCAATTGAAAGTGTGTTGCGGGGTACAGTCAACGCT...GCGGAGACGTGGTGTCTCTGTCTGCG	201
25	141	GCCAAATTTCTTCTGATGTTGTTGCGGGGTACATCTGACTCT...TTGTTGGATGGTCCACTCTGTGTTGCCAAG	210
26	135	GT...TTAGCAATTGTTGCGGGGTACAGTCAACGCT...TTGGACTGTCTGTAGTTGGTGT	198
27	139	A...CTTCTTCGATGTTGTTGCGGGGTACATCTGACTCT...TTGTTGGATGGTCCACTCTGTGTTGCCAAG	204
29	139	A...CTTCTTCGATGTTGTTGCGGGGTACATCTGACTCT...TTGTTGGATGGTCCACTCTGTGTTGCCAAG	204
32	139	A...CTTCTTCGATGTTGTTGCGGGGTACATCTGACTCT...TTGTTGGATGGTCCACTCTGTGTTGCCAAG	204
33	135	GT...ACAATTCTGAAAGTGTGTTGCGGGGTACATCTGACTCT...TTGTTGGATGGTCCACTCTGTGTTGCCAAG	201
34	135	GT...CCAATTTGAAAGTGTGTTGCGGGGTACATCTGACTCT...TTGTTGGATGGTCCACTCTGTGTTGCCAAG	201
35	139	A...CTTCTTCGATGTTGTTGCGGGGTACATCTGACTCT...TTGTTGGATGGTCCACTCTGTGTTGCCAAG	204
36	139	A...CTTCTTCGATGTTGTTGCGGGGTACATCTGACTCT...TTGTTGGATGGTCCACTCTGTGTTGCCAAG	204
40	139	A...CTTCTTCGATGTTGTTGCGGGGTACATCTGACTCT...TTGTTGGATGGTCCACTCTGTGTTGCCAAG	204
41	136	G...TTCAATTGAAAGTGTGTTGCGGGGTACATCTGACTCT...TTGTTGGATGGTCCACTCTGTGTTGCCAAG	201
43	130	G...TTCAATTGAAAGTGTGTTGCGGGGTACATCTGACTCT...TTGTTGGATGGTCCACTCTGTGTTGCCAAG	201
44	139	A...CTTCTTCGATGTTGTTGCGGGGTACATCTGACTCT...TTGTTGGATGGTCCACTCTGTGTTGCCAAG	204
45	139	A...CTTCTTCGATGTTGTTGCGGGGTACATCTGACTCT...TTGTTGGATGGTCCACTCTGTGTTGCCAAG	204
48	138	T...ANCAATTGAAAGTGTGTTGCGGGGTACATCTGACTCT...TTGTTGGATGGTCCACTCTGTGTTGCCAAG	201
50	136	G...TTCAATTGAAAGTGTGTTGCGGGGTACATCTGACTCT...TTGTTGGATGGTCCACTCTGTGTTGCCAAG	201
52	139	A...CTTCTTCGATGTTGTTGCGGGGTACATCTGACTCT...TTGTTGGATGGTCCACTCTGTGTTGCCAAG	204

54 130 A . . . CTTCTTCTGATGTTGTTGCGGAGTACATCGACACCGTGGGATGGTCCACTCTTGTGCGCAAG 204
57 130 G . . . TTCAATTGAAAGTGGTTGCGGAGTACGTCACCCCTGGGAGACGTGGTCTCTTTGCTCTGCG 201

10 202 GTC AATGATGGTACATGGAGGGCCACACACCGCGCTTAGTACAG-TGAAGGAAGAGG--ATCGTGTGGGTG 208
12 211 GTCAATGATGGTACATGGAGGGCCACACACCGCGCTTAGTACACCGATGGAAACGGATAATCGTGTGGGTG 280
13 202 GTC AATGATGGTACATGGAGGGCCACACACCGCGCTTAGTACANGTGAAGGAAGAG--ATCGTGTGGGTG 208
14 205 GTC AATGATGGTACATGGAGGGCCACACACCGCGCTTAGTACANGTGAAGGAAGAGAT--CGTGTGGGTG 271
16 205 GTC AATGAAAGTACATGGAGGGCCAAACACCGTCTTGGTACACCGATGGAGCGGATAATCGTGTGGAT 274
18 202 GTGAATGATGGCAGCTGGCGAGCCATTACTGTCATGTCGACTGTTAGGGGAAG--ACCCTGTAGGAG 208
19 205 GTGAATGAAAGTACATGGAGGGCCACACACTGTGCTTGGTACACCGATGGAAACGGATAAACGTGTGGAT 274
2 205 GTC AATGCAAGTACATGGAGGGCCACACCGCGCTTAGTACACCGATGGAAACGGATAAACGTGTGGAT 271
20 205 GTGAATGAAAGTACATGGAGGGCCACACACTGTGCTTGGTACACCGATGGAAACGGATAAACGTGTGGAT 274
22 202 GTC AATGATGGTACATGGAGGGCCACACACCGCGCTTAGTACANGTGAAGGAAGAG--ATCGTGTGGGTG 208
23 202 GTCAATGATGGTACATGGAGGGCCACACACCGCGCTTAGTACANGTGAAGGAAGAG--ATCGTGTGGGTG 208
25 211 GTGTCAGATCTACCTGGAAAGGCTAATACAGTCTCTGGACCAACTGATGGTACTGATAATCGTGTGGAA 280
26 199 GTCAATGCTGGTATATGGAGGCTACCGCACAGAGCTACTCCG--AAGGCGTGG--TTCTGTGGGTG 202
27 205 GTGAATGAAAGTACATGGAGGGCCACACACTGTGCTTGGTACACCGATGGAAACGGATAAACGTGTGGAT 274
29 205 G TAAATGAGTCACTTGGAAAGCCATACCGTATTAGGTACACCGATGGACCACAGAGGTGTGAT 274
32 205 GTGAATGAAAGTACATGGAGGGCCAAACACCGTCTTGGTACACCGATGGAAACGGATAAACGTGTGGAT 272
33 202 GTGAATGATGGGACCTTGGCGAGCCATTACGGCGCTTCCACTGTTAAAGGAGAG--ATCGCTTGGCG 208
34 202 GTCAATGCTGGTACTTGGAGGCTACACACCGCGCTTAGTACANGTGAAGGAAGAG--ACAGTGTGGGTG 208
35 205 GTGAATGAAAGTACATGGAGGGCCAAACACCGTCTTGGTACACCGATGGAAACGGATAAACGTGTGGAT 272
36 205 GTGAATGAAAGTACATGGAGGGCCACACACTGTGCTTGGTACACCGATGGAAACGGATAAACGTGTGGAT 274
40 205 GTCAATGAAAGTACATGGAGGGCCAAACACCGTCTTGGTACACCGATGGAAACGGATAAACGTGTGGAT 272
41 202 GTCAATGATGGTACATGGAGGGCCACACACCGCGCTTAGTACANGTGAAGGAAGAG--ATCGTGTGGGTG 208
43 202 GTCAATGATGGTACATGGAGGGCCACACACCGCGCTTAGTACANGTGAAGGAAGAG--ATCGTGTGGGTG 208
44 205 GTGAATGAAAGTACATGGAGGGCCACACACTGTGCTTGGTACACCGATGGAAACGGATAAACGTGTGGAT 274
45 205 GTGAATGAAAGTACATGGAGGGCCACACACTGTGCTTGGTACACCGATGGAAACGGATAAACGTGTGGAT 274
48 202 GTCAATGATGGTACATGGAGGGCCACACACCGCGCTTAGTACANGTGAAGGAAGAG--ATCGTGTGGGTG 208
50 202 GTCAATGATGGTACATGGAGGGCCACACACCGCGCTTAGTACANGTGAAGGAAGAG--ATCGTGTGGGTG 208
52 205 GTGAATGAAAGTACATGGAGGGCCAAACACCGTCTTGGTACACCGATGGAAACGGATAAACGTGTGGAT 272
54 205 GTGAATGAAAGTACATGGAGGGCCAAACACCGTCTTGGTACACCGATGGAAACGGATAAACGTGTGGAT 274
57 202 GTC AATGATGGTACATGGAGGGCCACACACCGCGCTTAGTACANGTGAAGGAAGAG--ATCGTGTGGGTG 288

10 200 TTTGGCAA-CTACCACGACATTGGCGAAGGGCAATAATGTGTTTTTTATTTG--GAA- 321
12 281 TTTTACCGCGCTACGGACAACCAAAGTAAATGANNTT-CACATGGNGGC- 331
13 200 TTTGCGAAGCTACCCACGGACATTGGCGAAGGGCAATAATGTGTTTTTTATTTGTAGGAA- 325
14 272 TTTGCGAAGCTACCCACGGACATTGGCGAAGGGCAATAATGTGTTTTTTATTTGTAGGAA- 328
16 275 TTTTACTACCCCAACCACCAAGGGCAATAAGGTGTTCTTTCTGCGGAA- 331
18 200 TTTGCAAATCCGCCACGGACCTGGCGAAGGGCAACAAATGTGTTCTTTATTTGTGGAAATAAACTCGGT 338
19 275 TTTTACTACCCCAACCACCAAGGGCAATAAGGTGTTTTCTTTCTTGGCGGAA- 331
2 272 TTTTTACTTCCCAATACCCATGAGGGCCGGGAAAGTTCNT- 325
20 275 TTTTTACTACCCCAACCACCAAGGGCAATAAGGTGTTTTCTTTCTGCGGAA- 331
22 200 TTTGCGAAGCTACCCACGGACATTGGCGAAGGGCAATAATGTGTTTTTTATTTGTAGGAA- 325
23 200 TTTGCGAAGCTACCCACGGACATTGGCGAAGGGCAATAATGTGTTTTTTATTTGTAGGAA- 325
25 281 TCTTTTATAGACTTACGACGACTTAAAGGGCAATAAGGTGTTCTTTGTTGCGGTTJA- 337
26 283 CTGTGGAGCCCACTGATANTAGCGAAGGGCAAGTGGTTCNTACTGNTAGGAA- 319
27 275 TTTTTACTACCCCAACCACCAAGGGCAATAAGGTGTTTTCTTTCTGCGGAA- 331
29 275 TTTTTACTACCCCAACCACCAAGGGCAATAAGGTGTTTTTTACTTGCAGGTT- 331
32 273 TTTTTACTACCC-ACAACCACCAAGGGCAATAAGGTGTTTTCTTTCTGCGGAA- 328
33 200 TCTGTAATAATGGCAACAACCTTGGCTAAAGGGTAACAAGCTATTTTCAATAGTGGAT- 325
34 200 TTTGCAAATCCCAACAACCTTGGCTAAAGGGCAATAAGGTGTTTTTTATTTGTAGGAA- 325
35 273 TTTTTACTACCC-ACAACCACCAAGGGCAATAAGGTGTTTTCTTTCTGCGGAA- 328
36 273 TTTTTACTACCCCAACCACCAAGGGCAATAAGGTGTTTTCTTTCTGCGGAA- 331
40 273 TTTTTACTACCC-ACAACCACCAAGGGCAATAAGGTGTTTTCTTTCTGCGGAA- 328
41 200 TTTGCGAAGCTACCCACGGACATTGGCGAAGGGCAATAAGGTGTTTTTTATTTGTAGGAA- 325
43 200 TTTGCGAAGCTACCCACGGACATTGGCGAAGGGCAATAAGGTGTTTTTTATTTGTAGGAA- 325
44 275 TTTTTACTACCCCAACCACCAAGGGCAATAAGGTGTTTTCTTTCTGCGGAA- 331
45 275 TTTTTACTACCCCAACCACCAAGGGCAATAAGGTGTTTTCTTTCTGCGGAA- 331
48 200 TTTGCGAAGCTACCCACGGACATTGGCGAAGGGCAATAATGTGTTTTTTATTTGTAGGAA- 325
50 200 TTTGCGAAGCTACCCACGGACATTGGCGAAGGGCAATAATGTGTTTTTTATTTGTAGGAA- 325
52 273 TTTTTACTACCC-ACAACCACCAAGGGCAATAAGGTGTTTTCTTTCTGCGGAA- 328
54 275 TTTTTACTACCCCAACCACCAAGGGCAATAAGGTGTTTTCTTTCTGCGGAA- 331
57 200 TTTGCGAAGCTACCCACGGACATTGGCGAAGGGCAATAATGTGTTTTTTATTTGTAGGAA- 325

10 321 -GCCATGAGAGAGATATGA-CAG- TGTACCACAA-AAGCTGGGAG-GCAGGTA- CT- . . 372
12 331 -ACACCCACATTTTAAA- -CCA- ATA- 353
13 325 -GCCATGAGAGAGATATGA-CAG- TGTNACGA-AAGCTGGGAG-GCAGGTA- CNTC- 378
14 328 -GCCATGAGAGAGATATGA-CAG- TGTGTTA-CGAAAGCTGGGAG-GCAGGTA- CNTC- 381
16 331 -GCTTTATGAGTACAAACA-AAC- TGTGGAAATGGACATATAGCGGATTCAA- ACTA- 387
18 330 ATTTCAAAAAGCTGGAAGGAA- GGCTGATTAGCAAGTTCGTGGGGTTCAGTATGGACCGA 397
19 331 -GCTTTATGAGTACAGACA-AAC- TGTGGAAATGGACATATAGCGGATTCAA- CCTA- 387
2 325 -ACCGTTACAGACA- AAC- TGTGGAAAGTCTG--CATAA-GAATCAA- CCAA- 371
20 331 -GCTTTATGAGTACAGACA-AAC- TGTGGAAATGGACATATAGCGGATTCAA- CCTA- 387
22 325 -GCCATGAGAGAGATATGA-CAG- TGTNACGA-AAGCTGGGAG-GCAGGTA- CNTC- 378
23 325 -GCCATGAGAGAGATATGA-CAG- TGTNACCA-AAGTTGGGAG-GCAGGTA- CNAG- 378
25 337 -GTTTCAAGATGCTGAAAGA-GTC- GGGTGGCAACTGGA--AAAG-GGACTCA- TTG- 387
26 319 -ACCGTTGAGAGAGGATATGGCAGANATTACTCCGTACTAA-TTCCTGGGAGGCTGGAA-GATTG- 383
27 331 -GCTTTATGAGTACAGACA-AAC- TGTGGAAATGGACATATAGCGGATTCAA- CCTA- 387
29 331 -CCTTTATGAGTACAGACA-AAC- GGTAGGAAATGGACATATAGCGGATTCAA- CTG- 387
32 328 -GC-TTTATGAGTACCAACA-A- CGATGGAAAGTGGAAATATAGCGGATTCAA- ACTA- 381
33 325 -CTCATGAGAGAAATATGA- TNCCTGTACCAA-TCATGGGAGGCGGGT--GACT- 376
34 325 -CGCACGAAAGAAAGTATG- ACTCGGTACCAA-ATCATGGGAGGAGGG--GATAG- 377
35 328 -GCTTTATGAGTACCAACG-AAC- TGTNAGGAGTGGAAATATAGCGGATTCAA- ACTA- 384
36 331 -GCTTTATGAGTACCAACA-AAC- TGTGGAAATGGACATATAGCGGATTCAA- CCTA- 387
40 328 -GC-TTTATGAGTACCAACA-A- CGATGGAAAGTGGAAATATAGCGGATTCAA- ACTA- 381
41 325 -GCCATGAGAGAGATATGA-CAG- TGTNACCA-AAGCTGGGAG-GCAGGTA- CNTC- 378
43 325 -GCCATGAGAGAGATATGA-CAG- TGTNACGA-AAGCTGGGAG-GCAGGTA- CNTC- 378
44 331 -GCTTTATGAGTACAGACA-AAC- TGTGGAAATGGACATATAGCGGATTCAA- CCTA- 387
45 331 -GCTTTATGAGTACAGACA-AAC- TGTGGAAATGGACATATAGCGGATTCAA- CCTA- 387

48 325 -GCCATGAGAGAGATATGA - CAG - - - - - TGGTTNACGA - AAGCTGGGAG - GCGAGGTGA - - - - - CNTC - 378
50 325 -GCCATGAGAGAGATATGA - CAG - - - - - TGGTTNACGA - AAGCTGGGAG - GCGAGGTGA - - - - - CNTC - 378
52 328 -GCCATTATGAGTACCAACA - A - - - - - CGATTGGAAGTCGAAATATAGCGGATTCAA - - - - - ACTA - 381
54 331 -GCCATTATGAGTACCAACG - AAC - - - - - TGAATGGAAGTCGAAATATAGCGGATTCAA - - - - - ACTA - 387
57 325 -GCCATGAGAGAGATATGA - CAG - - - - - TGGTTCCAAA - AAGCTGGGAG - GCGAGGTGA - - - - - CNTC - 378

10 372 - - - - - CATGTCCAACTGGTTGT - GGGTGAGGCCACCCCGCCACCGGACCG - - - - - CGTGCAAAAATA 427
12 353 - - - - - TACTTTAAA - - - - - GACACCAA - - - - - ACTGAA - - - - - TNTGAAGG - - - - - ATT 388
13 378 - - - - - NATGTCCAACTGGTTGT - GGG - - - - - GGGCCACCGGTTCCACGGACGG - - - - - CGTGCAAGAN 430
14 381 - - - - - NATGTCCAACTGGTTGT - GGGTGAGGCCACCCCGTCCACGGACGG - - - - - CGTGCA - - - - - GAG 434
16 387 - - - - - CGAATGCGCTGGTTGT - GGGTGATGTGAGGGAGCCTACAGAGCAG - - - - - CGAGCC - - - - - GAC 440
18 398 GAAACGGTGACAAGTCGAAAAGATAAA - GGGTGAGGGAGCCCTCATCGGTT - - - - - AAAAAG - - - - - ATC 461
19 387 - - - - - CGACTTGAACCTGGTTGT - GGGTGATGTGAGGGAGCCTACAGAGCAG - - - - - CGAGCC - - - - - GAC 440
2 371 - - - - - CGACGTGGACCGGGTGC - GGGTGATGTGAGGGAGCCTACAGAGCAG - - - - - CGAGCC - - - - - TCA 428
20 387 - - - - - CGACTTGAACCTGGTTGT - GGGTGATGTGAGGGAGCCTACAGAGCAG - - - - - CGAGCC - - - - - GAC 440
22 378 - - - - - NATGTCCAACTGGTTGT - GGGTGAGGCCACCCCGTCCACGGACGG - - - - - CGTGCAAG - - - - - 431
23 378 - - - - - NCTGCCCAACAGTTGCT - GGGTGATGTGAGGGAGCCTACAGAGCAG - - - - - CGAGCC - - - - - 431
25 387 - - - - - GACCTCAAGTTAGTCGCT - GGGTGATGTGAGGGAGCCTACAGAGCAG - - - - - TCTGAGCA - - - - - ATC 440
26 383 - - - - - TGCCCAACAGGGTGC - GGGTGATGTGAGGGAGCCTACAGAGCAG - - - - - GAGAGC - - - - - AGC 431
27 387 - - - - - CGACTTGAACCTGGTTGT - GGGTGATGTGAGGGAGCCTACAGAGCAG - - - - - CGAGCC - - - - - GAC 440
29 387 - - - - - AGGCTAGAACCTAGTGT - GGGTGATGTGAGGGAGCCTACAGAGCAG - - - - - TCGAGCC - - - - - CCAC 440
32 381 - - - - - CGACTTGAACCTGGTTGT - GGGTGATGTGAGGGAGCCTACAGAGCAG - - - - - CGAGCC - - - - - GAC 434
33 376 - - - - - CGCTACAGGATGT - GGGTGAGGCCACCCCGTCCACGGACGG - - - - - GAGAGC - - - - - ANA 428
34 377 - - - - - CGTCCGACCGGATGT - GGGTGAGGCCACCCCGTCCACGGACGG - - - - - GAGAGC - - - - - AGC 424
35 384 - - - - - CGACTTGAACCTGGTTGT - GGGTGATGTGAGGGAGCCTACAGAGCAG - - - - - CGAGCC - - - - - GAC 437
36 387 - - - - - CGACTTGAACCTGGTTGT - GGGTGATGTGAGGGAGCCTACAGAGCAG - - - - - CGAGCC - - - - - GAC 440
40 381 - - - - - CGACTTGAACCTGGTTGT - GGGTGATGTGAGGGAGCCTACAGAGCAG - - - - - CGAGCC - - - - - GAC 434
41 378 - - - - - NATGTCCAACTGGTTGT - GGGTGAGGCCACCCCGTCCACGGACGG - - - - - CGTGCAAG - - - - - 431
43 378 - - - - - NATGTCCAACTGGTTGT - GGGTGAGGCCACCCCGTCCACGGACGG - - - - - GCAAGAGTA 433
44 387 - - - - - CGACTTGAACCTGGTTGT - GGGTGATGTGAGGGAGCCTACAGAGCAG - - - - - CGAGCC - - - - - GAC 440
45 387 - - - - - CGACTTGAACCTGGTTGT - GGGTGATGTGAGGGAGCCTACAGAGCAG - - - - - CGAGCC - - - - - GAC 440
48 378 - - - - - NATGTCCAACTGGTTGT - GGGTGAGGCCACCCCGTCCACGGACGG - - - - - CGTGCAAG - - - - - 431
50 378 - - - - - NATGTCCAACTGGTTGT - GGGTGAGGCCACCCCGTCCACGGACGG - - - - - CGTGCAAG - - - - - 431
52 381 - - - - - CGACTTGAACCTGGTTGT - GGGTGATGTGAGGGAGCCTACAGAGCAG - - - - - CGAGCC - - - - - GAC 434
54 387 - - - - - CGACTTGAACCTGGTTGT - GGGTGATGTGAGGGAGCCTACAGAGCAG - - - - - CGAGCC - - - - - GAC 440
57 378 - - - - - NATGTCCAACTGGTTGT - GGGTGAGGCCACCCCGTCCACGGACGG - - - - - CGTGCAAG - - - - - 431

10 427 - - - - - AAACAATCAACTTGGGGTGGACCCACACCCGTTTAAG - - - - - ACAGATC - - - - - CCCCC 478
12 389 T - - - - - GCTTATTTTGG - - - - - TGGCTCAG - - - - - GTG - - - - - TTATGATGG 421
13 431 GTAAAAAGATCAACT - - - - - GGGGTGGACCCACACCCGTTTAAG - - - - - ACAGATC - - - - - CCCCAC 484
14 435 TA - - - - - AAAGATCAACTGGGGTGGACCCAC - - - - - ACCGCTTTTAAG - - - - - ACAGATC - - - - - CCCCAC 487
16 441 TG - - - - - AACGATCAAAATGGGCCAAATCAG - - - - - GTCACTGTAAA - - - - - CGAGAGTACTATAG 403
18 462 CCCCCATAAGCCACA - - - - - ABGGTGGATTGAGGGCAAT - - - - - T - - - - - TTGTCCATGG 508
19 441 TG - - - - - AGCGATCAAAATGGGGCCAAATCAG - - - - - GTCACTGTAAA - - - - - CGAGAGTACTATAG 403
2 429 CA - - - - - GA - - - - - ATCAAAATGGGGTCAAGTACG - - - - - GGCTTGGTGAA - - - - - CGAGAGAGCATTT 478
20 441 TG - - - - - GBCGATCAAAATGGGGCCAAATCAG - - - - - GTCACTGGTANAACGAGAGTACTATAACGAGAGTACTATAG 508
22 431 - - - - - AACGATCAACT - - - - - GGGTGACCCACCCTGTAAA - - - - - ACAGATC - - - - - CCCCAC 480
23 431 - - - - - AATGATCACT - - - - - GGGTGACCCACACCCGTTTAAG - - - - - ACAGATC - - - - - CCCCAC 481
25 441 TA - - - - - AATGGATTAACCTAGGGCATAACA - - - - - TTTACTCAG - - - - - CCAAAATACCTAA 400
26 432 AGAGAACGATCAACT - - - - - GGGTGACCCACACCCGTTTAAG - - - - - C - - - - - GATAGACCTAA 483
27 441 TG - - - - - AGCGATCAAAATGGGGCCAAATCAG - - - - - GTCACTGTAAA - - - - - CGAGAGTACTATAG 403
29 441 CG - - - - - AACGATCAAAATGGGGCAGATTCCG - - - - - CABCTGTAAA - - - - - TGAAGTACTATAG 403
32 435 TG - - - - - AACGATCAAAATGGGGCCAAATCAG - - - - - GTCTGTAAA - - - - - CGAGAGTACTATAG 487
33 420 AA - - - - - ATGATCACT - - - - - AGGGTGGACTCA - - - - - - - - - - CGCTACCA 485
34 425 CGAGAAGATCAACT - - - - - GGGTGGACCCACACCCGTTTAAG - - - - - C - - - - - AGACCAGGA 477
35 438 TG - - - - - AACGATCAAAATGGGGCCAAATCAG - - - - - GTCTGTAAA - - - - - CGAGAGTACTATAG 490
36 444 TG - - - - - AGCGATCAAAATGGGGCCAAATCAG - - - - - GTCACTGTAAA - - - - - CGAGAGTACTATAG 408
40 435 TG - - - - - AACGATCAAAATGGGGCCAAATCAG - - - - - GTCTGTAAA - - - - - CGAGAGTACTATAG 487
41 431 - - - - - AACGATCAACT - - - - - GGGTGGACCCACACCCGTTTAAG - - - - - ACAGATC - - - - - CCCCAC 481
43 433 - - - - - AAAGATCAACT - - - - - GGGTGGACCCACACCCGTTTAAG - - - - - ACAGATC - - - - - CCCCAC 484
44 441 TG - - - - - AGCGATCAAAATGGGGCCAAATCAG - - - - - GTCACTGTAAA - - - - - CGAGAGTACTATAG 403
45 441 TG - - - - - AGCGATCAAAATGGGGCCAAATCAG - - - - - GTCACTGTAAA - - - - - CGAGAGTACTATAG 403
48 434 - - - - - CNCANTCAACT - - - - - GGGTGGA - - - - - ACC - - - - - NC - - - - - - - - - - 468
50 431 - - - - - AACGATCAACT - - - - - GGGTGGACCCACACCCGTTTAAG - - - - - ACAGATC - - - - - CCCCAC 481
52 435 TG - - - - - AACGATCAAAATGGGGCCAAATCAG - - - - - GTCTGTAAA - - - - - CGAGAGTACTATAG 487
54 441 TG - - - - - AACGATCAAAATGGGGCCAAATCAG - - - - - GTCTGTAAA - - - - - CGAGAGTACTATAG 403
57 431 - - - - - AACGATCAACT - - - - - GGGTGGACCCACACCCGTTTAAG - - - - - ACAGATC - - - - - CCCCAC 481

10 479 AAAACC - - - - - AGGGTGAATTGAGGAACTTT - - - - - GCCAETGGGCGCC - - - - - CGGTGT - - - - - CTGACGAAAGATGG 543
12 422 AGA - - - - - - - - - - ATGGCACA - - - - - TTGTT - - - - - TT - - - - - - - - - - 440
13 485 AGACCC - - - - - AGGGTGAATTGAGGAACTTT - - - - - TCCACTGGTGGCGCC - - - - - GGTGCT - - - - - CTGACGGAGAA - - - - - GG 548
14 488 AGACCC - - - - - AGGGTGAATTGAGGAACTTT - - - - - TCCACTGGTGGCGCC - - - - - GGTGCT - - - - - CTGACGGAGAA - - - - - GG 551
16 494 CTGCTACGGAAGGTAAATGACGGGGTTCCTTGCC - - - - - CGCTGGTGGCTCA - - - - - GGTGTT - - - - - CTGATGGAGATGG 580
18 506 TGGCTCC - - - - - AGGGTGGCTAACTGAG - - - - - - - - - - AAATGCA - - - - - ACCTGC - - - - - CCGTTTCCATGG 558
19 494 CTGCTACGGAAGGTAAATGACGGGGTTCCTTGCC - - - - - CGCTGGTGGCTCA - - - - - GGTGTT - - - - - CTGATGGAGATGG 560
2 479 CAGCAACGGAAGGCAAGTGGACCGGGTTCCTTGCC - - - - - TGCTGGGCTTCA - - - - - GGTGTT - - - - - CTGATGGAGATGG 545
20 509 CTGCTACGGAAGGTAAATGACGGGGTTCCTTGCC - - - - - CGCTGGTGGCTCA - - - - - GGTGTT - - - - - CTGATGGAGATGG 575
22 481 AGACCC - - - - - AGGGTGAATTGAGGAACTTT - - - - - TCCACTGGTGGCGCC - - - - - GGTGCT - - - - - CTGACGGAAA - - - - - GG 545
23 482 TGACCC - - - - - CTATGGATGAAAGGACATTT - - - - - GCGCATGGTGGAGGA - - - - - GATGCT - - - - - CCGACGGGGATGG 546
25 491 AGACCCCTTAAAGGCGGCTTAAAGGACTTCGCCCT - - - - - ATCCGGGCTTCA - - - - - TCAGTT - - - - - NTGGTGGGCTGG 557
26 494 CAGACC - - - - - TGGATGGATGAAAGACTTCGCCCT - - - - - CTAGCTCCGCGCC - - - - - GCGGCT - - - - - TTAATGAGAACTGG 548
27 494 CTGCTACGGAAGGTAAATGACGGGGTTCCTTGCC - - - - - CGCTGGTGGCTCA - - - - - GGTGTT - - - - - CTGATGGAGATGG 580
29 494 CTGCACTAGGAGGTAAATGACAGGATTTCTTGC - - - - - GGCAGGGGCTCA - - - - - GGAGTA - - - - - TTATGAGAGAGG 580
32 488 CTGCTACGGAAGGTAAATGACGGGGTTCCTTGCC - - - - - CGCTGGTGGCTCA - - - - - GGTGTT - - - - - CTGATGGAGATGG 554
33 486 CAGATCG - - - - - CCAATACCNANAGAA - - - - - - - - - - TGGCATG - - - - - NCTCTT - - - - - GTGNTCCACCTGG 515
34 478 CAGATCC - - - - - GGGTGGANTANGAGGGCACTT - - - - - GCTCACGGGCTA - - - - - GGTGTA - - - - - GTGACGTAAGTGG 542
35 491 CTGCTACGGAAGGTAAATGACGGGGTTCCTTGCC - - - - - CGCTGGTGGCTCA - - - - - GGTGTT - - - - - CTGATGGAGATGG 557
36 497 CTGCTACGGAAGGTAAATGACGGGGTTCCTTGCC - - - - - CGCTGGTGGCTCA - - - - - GGTGTT - - - - - CTGATGGAGATGG 583
40 488 CTGCTACGGAAGGTAAATGACGGGGTTCCTTGCC - - - - - CGCTGGTGGCTCA - - - - - GGTGTT - - - - - CTGATGGAGATGG 554
41 482 ACACCA - - - - - ANGTGAAGGAGGACTTT - - - - - TCCACTGGGCGCC - - - - - GGTGCT - - - - - CTGACGGAAA - - - - - GG 545

43 486 AGACCC...AGGGTGAATTGAGGGAACTTTT-GCCCACTGGTGGGGCC-GGTGCCTCTGACGGAAATGG 548
44 494 CTGCTCACGAAAGGTAAATGGACGGGGTTTCTTGC-CGCTGGTGGCTCA-GGTGTT-CTBATGGAGGATGG 560
45 494 CTGCTCACGAAAGGTAAATGGACGGGGTTTCTTGC-CGCTGGTGGCTCA-GGTGTT-CTBATGGAGGATGG 560
48 468-CACTGGTGGGGCC-GGTGCCTCTGACGGAGAAATGG 491
50 482 AGACCC...AGGGTGAATTGAGGGAACTTTT-GCCCACTGGTGGGGCC-GGTGCCTCTGACGGAGAAATGG 546
52 488 CTGCTCACGAAAGGTAAATGGACGGGGTTTCTTGC-CGCTGGTGGCTCA-GGTGTT-CTBATGGAGGATGG 554
54 494 CTGCTCACGAAAGGTAAATGGACGGGGTTTCTTGC-CGCTGGTGGCTCA-GGTGTT-CTBATGGAGGATGG 560
57 482 AGACCC...AGGGTGAATTGAGGGAACTTNT-GCCCACTGGTGGGGCC-GGTGCCTCTGACGGAGAAATGG 546

10 544 CACGCTCCTGTTTCCCTGGAGGGGAGGAAATACCAATGGTGAATTTGCTCCGTG-ATCATTATTTCG 012
12 440-T-CTCTGACGGCGAAGABTGAA-GCAGAAAGCGTTTGCTCCATG-ATCATTTACTCG 485
13 549 CACGCTCCTGTTT-TCCTGGAGGGGAGGAAATAC-AATGGTGAATTTGCTCCGTG-ATCATTATTTCG 015
14 552 CACGCTCCTGTTT-CCCTGGAGGGGAGGAAATAC-AATGGTGAATTTGCTCCGTG-ATCATTATTTCG 018
16 561 CACGCTCCTGTTT-CCACTGATGGCGACAAATGCA-GCAAAAGACGATTACTCCATG-ATCATTTATTTCG 027
18 559 CCGCG-.....-AGGAATCCCAATGGT-..A-...-TATTATCTCCGGGAGGACCTTACTCTC 009
19 581 CACGCTGTTGTTT-CCACTGATGGCGACAAATGCA-GCAAAAGACGATTACTCCATG-ATCATTTATTTCG 027
2 546 AACACTGGTATTC-CAATGATGGCACTAAAGCC-GCTAAAAGACTACTCCATG-ATCATTTATTTCG 012
20 576 CACGCTGTTGTTT-NCACGATGGCGACAAATGCA-GCAAAAGACGATTACTCCATG-ATCATTTATTTCG 042
22 548 CACGCTCCTGTTT-CCCTGGAGGGGAGGAAATAC-AATGGTGAATTTGCTCCGTG-ATCATTATTTCG 012
23 548 CACGCTCCTGTTT-CCCTGGAGGGGAGGAAATAC-AATGGTGAATTTGCTCCGTG-ATCATTATTTCG 012
25 547-CTCACTGATGGCGACAAATGCA-GCAAAAGACGATTACTCCATG-ATCATTTATTTCG 012
26 549 CACACTACTTTTTCCTGGGGTGGAGAACAGT-..GG-..GFACTGCTCCAGG-ACCATTTGTTTG 012
27 581 CACGCTGTTGTTT-CCACTGATGGCGACAAATGCA-GCAAAAGACGATTACTCCATG-ATCATTTATTTCG 027
29 581 GACGTTCTGTTT-CCACTGATGGAACTAAATGCA-GCAAAAGGCTGACTATTTCTATG-ATCATTTATTTCG 027
32 555 CACGCTCCTGTTT-CCACTGATGGCGACAAATGCA-GCAAAAGACGATTACTCCATG-ATCATTTATTTCG 021
33 518 CCGGNG-.....-AGGATACAAATGNGT-..AA-...-TATTAGCCCGTTTACTCCATG-ATCATTTATTTCG 584
34 543 CATGCT-CCCGTCTCCCGGGGGTGGAGAACAGT-..GG-...-TACTGTCTCAGAG-ACCATTTGTTTG 003
35 558 CACGCTGTTGTTT-CCACTGATGGCGAGGAAATGCA-GCAAAAGACGATTACTCCATG-ATCATTTATTTCG 024
36 584 CACGCTGTTGTTT-CCACTGATGGCGAGGAAATGCA-GCAAAAGACGATTACTCCATG-ATCATTTATTTCG 030
40 555 CACGCTGTTGTTT-CCACTGATGGCGAGGAAATGCA-GCAAAAGACGATTACTCCATG-ATCATTTATTTCG 021
41 548 CACGCTGTTGTTT-CCCTGGAGGGGAGGAAATAC-AATGGTGAATTTGCTCCGTG-ATCATTATTTCG 015
43 549 CACGCTCCTGTTT-CCCTGGAGGGGAGGAAATAC-AATGGTGAATTTGCTCCGTG-ATCATTATTTCG 015
44 581 CACGCTGTTGTTT-CCACTGATGGCGAGGAAATGCA-GCAAAAGACGATTACTCCATG-ATCATTTATTTCG 027
45 581 CACGCTGTTGTTT-CCACTGATGGCGAGGAAATGCA-GCAAAAGACGATTACTCCATG-ATCATTTATTTCG 027
48 482 CACGCTCCTGTTT-CCCTGGAGGGGAGGAAATAC-AATGGTGAATTTGCTCCGTG-ATCATTATTTCG 012
50 548 CACGCTCCTGTTT-CCCTGGAGGGGAGGAAATAC-...-TGGTATTGCTCCGTG-ATCATTATTTCG 008
52 555 CACGCTGTTGTTT-CCACTGATGGCGAGGAAATGCA-GCAAAAGACGATTACTCCATG-ATCATTTATTTCG 021
54 581 CACGCTGTTGTTT-CCACTGATGGCGAGGAAATGCA-GCAAAAGACGATTACTCCATG-ATCATTTATTTCG 027
57 548 CACGCTCCTGTTT-CCCTGGAGGGGAGGAAATAC-AATGGTGAATTTGCTCCGTG-ATCATTATTTCG 012

10 013 ACCGGACGATGGGAAAACTGGTATTTCTTGAAGGAAATACCTCCCGGAAATGGCTT-.....CATCA 074
12 490 ACCGGACACGGTGA AAAATGGATGCTTTGCAATGGAACTGCACCTGAGAGATGGGGTGACCCCCCATCA 585
13 016 ACCGGACGATGGGAAAACTGGTATTTCTTGAAGGAAATACCTCCCGGAAATGGCTTAAACCCCCCATCA 085
14 019 ACCGGACGATGGGAAAACTGGTATTTCTTGAAGGAAATACCTCCCGGAAATGGCTTAAACCCCCCATCA 088
16 028 ACCGGACGCGGGCAAACTGGTATTTCTTGAAGGAACTGCTCCCTGCAAAATGGACTGACCCCCCATCA 097
18 007 ACCGATGACGGGAAAACTGGTATTTCTTGAAGGAAATACCTCCCGGAAATGGCTTAAACCTTCGATTA 070
19 028 ACCGGACACGGCAGTACTGGGGCTTCACTGGTGTGTTTCCCTGCGAATGGACTGGCCCCGCATCA 097
2 013 ACTGATACCGTAGTACTGGGGCTTACGACGGGGTATCCGAGCGAATGGACTGATCCGAGATCA 082
20 043 ACCGGACACGGCAGTACTGGGGCTTCAACTGGTGTGTTCCCTGCGAATGGACTGACCCCCCATCA 712
22 013 ACCGGACGATGGGAAAACTGGTATTTCTTGAAGGAAATACCTCCCGGAAATGGCTTAAACCCCCCATCA 082
23 013 ACCGGACGATGGGAAAACTGGTATTTCTTGAAGGAAATACCTCCCGGAAATGGCTTAAACCCCCCATCA 079
25 013 AAAGACGATGGGAAATCTTGGGCATTTAGTCCGGAGTATGCTGCGAATGGACTGACCCCCGCATTA 082
26 013 ATGGCCGGTGGGAAAACTGGNTATCTTGAAGGAACTCTAGAGAGATGTACCACCCGACCATAG 082
27 028 ACCGGACACGGCAGTACTGGGGCTTCACTGGTGTGTTTCCCTGCGAATGGACTGACCCCCCATCA 097
29 028 ACAGCGGACAGGGTCAACTGGGGCTTCACTGGGCGTATCAGCGGCAATGGACTGACCCCCGCATTA 097
32 022 AAAGACGATGGGAAAACTGGGCGCTTCACTGGGCGTTCCTTGGGAAATGGACTGACCCCCGCATTA 091
33 565 ACCGGACGCGGGTAAAGAACTGGTATTTCTTGAAGGAAATACCTCCCGGAAATGGCTTAAACCGGCAATTA 034
34 004 ATGGTCTGGTGGGAAAACTGGTATTTCTTGAAGGAAATACCTAGAGGAAATACCTAGGGGAGAGGNTCNCCT 073
35 025 AAAGACGATGGGAAATCTGGGGCTTCAACTGGGCGTTCCTTGGGAAATGGACTGACCCCCCATCA 094
36 031 ACCGGACACGGCAGTACTGGGGCTTCAACTGGTGTGTTCCCTGCGAATGGACTGACCCCCCATCA 700
40 022 AAAGACGATGGGAAATCTGGGGCTTCAACTGGGCGTTCCTTGGGAAATGGACTGACCCCCGCATTA 088
41 013 ACCGGACGATGGGAAAACTGGTATTTCTTGAAGGAAATACCTCCCGGAAATGGCTTAAACCCCCCATCA 082
43 016 ACCGGACGATGGGAAAACTGGTATTTCTTGAAGGAAATACCTCCCGGAAATGGCTTAAACCCCCCATCA 085
44 028 ACCGGACACGGCAGTACTGGGGCTTCAACTGGTGTGTTCCCTGCGAATGGACTGACCCCCCATCA 097
45 028 ACCGGACACGGCAGTACTGGGGCTTCAACTGGTGTGTTCCCTGCGAATGGACTGACCCCCCATCA 097
48 559 ACCGGACGATGGGAAAACTGGTATTTCTTGAAGGAAATACCTCCCGGAAATGGCTTAAACCCCCCATCA 028
50 007 ACCGGACGATGGGAAAACTGGTATTTCTTGAAGGAAATACCTCCCGGAAATGGCTTAAACCCCCCATCA 073
52 022 AAAGACGACGGCAGTACTGGGGCTTCAACTGGGCGTTCCTTGGGAAATGGACTGACCCCCGCATTA 091
54 028 AAAGACGACGGCAGTACTGGGGCTTCAACTGGGCGTTCCTTGGGAAATGGACTGACCCCCGCATTA 097
57 013 ACCGGACGATGGGAAAACTGGTATTTCTTGAAGGAAATACCTCCCGGAAATGGCTTAAACCCCCCATCA 082

10 075 CCGAATGGGAGG-..GGGGCGGAGAAATTTAGTGGTTAGTGAATGTTTGGTGGTTCAGAAATTTGTACGAGT 743
12 588 CCGAATGGGAGG-..GATCAC-..TTCTCATGATGTTGACTGCGAGAAATGGCCAGAGGTTGTACGAGT 829
13 080 CCGAATGGGAGNAGGGCGGACGAAATTT-..GGTTAGTGAATGTTTGGTGGTTCAGAAATTTGTACGAGT 752
14 080 CCGAATGGGAGG-..GGGGCGGAAATTTAGTGGTTAGTGAATGTTTGGTGGTTCAGAAATTTGTACGAGT 755
16 098 CCGAATGGGATG-..GATCAC-..TTCTCATGATGTTAAATGCAAGTATAGTTCAGAGGTTGTACGAGT 781
18 077 CCGAATGGGAG-..GGTCGGGAAATTTAGTGGTTAGTGAATGTTTGGTGGTTCAGAAATTTGTACGAGT 743
19 098 CCGAATGGGATG-..GATCAC-..TTCTCATGATGTTAAATGCAAAATG-.....-GGAGT 748
2 083 CCGAATGGGATG-..GATCT-..TACTGATGATGTTAACTGCAAGTATTCGACGCGGTTTACGAGT 748
20 713 CCGAATGGGATG-..GATCAC-..TTCTCATGATGTTAAATGCAAGTATAGTTCAGAGGTTGTACGAGT 778
22 083 CCGAATGGGAGNAGGGCGGACGAAATTT-..GGTTAGTGAATGTTTGGTGGTTCAGAAATTTGTACGAGT 748
23 080 CCGAATGGGAGG-..GGGGCGGAAATTTAGTGGTTAGTGAATGTTTGGTGGTTCAGAAATTTGTACGAGT 748
25 083 CCGAATGGGATG-..GATCAC-..TTCTCATGATGTTAAATGCAAGTATTCGACGCGGTTTACGAGT 748
26 083 GANTGGGGAA-..GGAAAGGCACTCTCGTGGTTCTGATGCTTCGGGGGCGAGAAATTTATGAGAG 749
27 098 CCGAATGGGATG-..GATCAC-..TTCTCATGATGTTAAATGCAAGTATAGTTCAGAGGTTGTACGAGT 781
29 098 CCGAATGGGATG-..GAAGT-..TTCTCATGATGTTAAATGCAAAATCTCCAGAGGTTTATGAGAG 781
32 092 CCGAATGGGATG-..GATCAC-..TTCTCATGATGTTAAATGCAAGTATAGTTCAGAGGTTGTACGAGT 755
33 035 CCGAATGGGAGG-..GGGTGACNGATTC-..TGGTCTGTAATAGTCCGTTGTTGATGATGATGATGAG 701
34 074 TCGCATGGGAG-..AAGGCGGAAATTTGGTGGTTCGCACTGTTTGGGGGCGACGACCTGATGAGAG 740
35 095 CCGAATGGGATG-..GATCAC-..TTCTCATGATGTTAAATGCAAGTATAGTTCAGAGGTTGTACGAGT 758

35 701 CCGAATGGGATG...GATCAC...TTCTCATGATTTGTAATGCAAGTAATGTCAGAGGGTACCGAGTC 754
40 099 CCGAATGGGATG...GATCAC...TTCTCATGATTTGTT...TGNAAAGTATGAGAGGGGTACCGAGTC 740
41 082 CCGAATGGGAGNAGG...CCGACGAAATTCT...GGTTAGTGATTTGTTTGGTGGTCAGAAATTGACCGAGTC 740
43 095 CCGAATGGGAGNAGG...CCGACGAAATTCT...GGTTAGTGATTTGTTTGGTGGTCAGAAATTGACCGAGTC 752
44 088 CCGAATGGGATG...GATCAC...TTCTCATGATTTGTAATGCAAGTAATGTCAGAGGGTACCGAGTC 751
45 098 CCGAATGGGATG...GATCAC...TTCTCATGATTTGTAATGCAAGTAATGTCAGAGGGTACCGAGTC 701
46 029 CCGAATGGGAGNAGG...CCGACGAAATTCT...GGTTAGTGATTTGTTTGGTGGTCAGAAATTGACCGAGTC 695
48 074 CCGAATGGGAG...GGGGCCGAAATTC...GGTTAGTGATTTGTTTGGTGGTCAGAAATTGACCGAGTC 740
52 092 CCGAATGGGATG...GATCAC...TTCTCATGATTTGTAATGCAAGTAATGTCAGAGGGTACCGAGTC 753
54 098 CCGAATGGGATG...GATCAC...TTCTCATGATTTGTAATGCAAGTAATGTCAGAGGGTACCGAGTC 740
57 092 CCGAATGGGAGNAGG...CCGACGAAATTCT...GGTTAGTGATTTGTTTGGTGGTCAGAAATTGACCGAGTC 740

10 744 GAGTGCATGCGGACAAAGCTGGACGGAGGGCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 813
12 020 GCGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 808
13 753 GAGTGCATGCGGACAAAGCTGGACGGAGGGCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 822
14 750 GCGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 825
16 762 GCGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 829
18 744 GTCTGATATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 813
19 747 GCGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 814
2 747 TACTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 814
27 777 GCGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 844
22 750 GAGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 810
23 747 GAGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 810
25 747 GCGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 810
26 753 TTTGATGATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 810
27 762 GCGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 820
28 763 GAGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 820
32 750 GCGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 825
33 702 GCGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 771
34 741 CTCGATGATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 807
35 759 GCGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 827
36 765 GCGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 832
41 750 GCGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 810
41 750 GAGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 810
43 753 GAGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 822
44 762 GCGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 825
45 762 GCGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 825
45 095 GAGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 765
50 741 GAGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 810
52 755 GCGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 825
54 740 --- 740
57 750 GAGTGCATGCGGACAAAGCTGGACGGAGGGCCTGTGGGACACTCCGAGGCGTGTGGTCAAATCAAAATCG 810

10 814 AAAGCCATTGTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 880
12 099 AGAAGTCCCCCAATGAAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 788
13 823 AAAGCCATTGTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 880
14 828 AAAGCCATTGTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 885
16 933 TATACTTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 888
18 814 AAAGCCATTGTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 880
19 815 TATTACTTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 883
2 815 TGCTTACTTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 883
20 940 TATTACTTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 883
22 826 AAAGCCATTGTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 888
23 817 AAAGCCATTGTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 883
25 817 TTAAGCTTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 888
26 828 AAAGCCATTGTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 888
27 830 TATACTTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 888
29 830 TACTACTTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 894
32 828 ACTTTGGGACCTTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 895
33 772 AAAGCCATTGTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 874
34 808 AAAGCCATTGTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 878
35 928 GNTACTTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 921
36 833 TATTACTTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 885
40 828 TGTACTTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 888
41 828 AAAGCCATTGTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 888
43 823 AAAGCCATTGTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 847
44 830 TATTACTTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 888
45 930 TATTACTTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 928
48 788 AAAGCCATTGTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 832
50 811 AAAGCCATTGTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 877
52 826 TGTACTTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 895
54 740 --- 753
57 821 AAAGCCATTGTTGGTAACAGCTTCCATTGGATGCCCTCATCCCGGACGCTT...GAAAGGGGTGTC 898

10 881 T-GC-TGT-ACATTCAGAGAGGGTACAA...CCTCGGAGAGAAATGGACAGGCGCTCTGCTTTGGGT 944
12 787 T-GC-TGT-ACATTCAGAGAGGGTACAA...CCTCGGAGAGAAATGGACAGGCGCTCTGCTTTGGGT 930
13 890 T-GC-TGT-ACATTCAGAGAGGGTACAA...CCTCGGAGAGAAATGGACAGGCGCTCTGCTTTGGGT 953
14 898 T-GC-TGT-ACATTCAGAGAGGGTACAA...CCTCGGAGAGAAATGGACAGGCGCTCTGCTTTGGGT 958
16 899 T-GC-TGT-ACATTCAGAGAGGGTACAA...CCTCGGAGAGAAATGGACAGGCGCTCTGCTTTGGGT 952
18 891 T-GC-TGT-ACATTCAGAGAGGGTACAA...CCTCGGAGAGAAATGGACAGGCGCTCTGCTTTGGGT 944
19 894 T-GC-TGT-ACATTCAGAGAGGGTACAA...CCTCGGAGAGAAATGGACAGGCGCTCTGCTTTGGGT 947
2 894 T-GC-TGT-ACATTCAGAGAGGGTACAA...CCTCGGAGAGAAATGGACAGGCGCTCTGCTTTGGGT 947
20 914 T-GC-TGT-ACATTCAGAGAGGGTACAA...CCTCGGAGAGAAATGGACAGGCGCTCTGCTTTGGGT 977
22 887 T-GC-TGT-ACATTCAGAGAGGGTACAA...CCTCGGAGAGAAATGGACAGGCGCTCTGCTTTGGGT 950
23 894 T-GC-TGT-ACATTCAGAGAGGGTACAA...CCTCGGAGAGAAATGGACAGGCGCTCTGCTTTGGGT 947
25 887 T-GC-TGT-ACATTCAGAGAGGGTACAA...CCTCGGAGAGAAATGGACAGGCGCTCTGCTTTGGGT 950
26 889 T-GC-TGT-ACATTCAGAGAGGGTACAA...CCTCGGAGAGAAATGGACAGGCGCTCTGCTTTGGGT 950
27 899 T-GC-TGT-ACATTCAGAGAGGGTACAA...CCTCGGAGAGAAATGGACAGGCGCTCTGCTTTGGGT 982
29 805 T-GC-TGT-ACATTCAGAGAGGGTACAA...CCTCGGAGAGAAATGGACAGGCGCTCTGCTTTGGGT 950
32 898 T-GC-TGT-ACATTCAGAGAGGGTACAA...CCTCGGAGAGAAATGGACAGGCGCTCTGCTTTGGGT 950

33 839 A - GC - TGT TGCACATTAGAGGGCACNC - - C T A C G G G A C G T A A G T G G A C T A A - C G C T C A G C T T T T G G G T 902
34 875 T - GT - TAT - ACATTCAACGAGGATATA - - C T A G T G G C G A G A A G G T G G A T A A G G C A T T A T G C C T A T G G G T 938
35 880 T - GC - TGT - ACACACAGAGAGGGGACT - - T C T T G G G G A G A A A A G T G A A A G A G C G C T C T A T C T T T G G G T 959
36 902 T - GC - TGT - ACACCTCAGCGAGGGGACT - - T C T C G G G G A G A A A A G T G A A A G A G C G C T C T A C C T T T G G G T 985
40 887 T - GC - TGT - ACACACAGAGAGGGGACT - - T C T T G G G G G A G A A A A G T G A A A G A G C G C T C T A T C T T T G G G T 950
41 887 T - GC - TGT - ACATTTCAGAGAGGGGTACA - - C C T C G G G A G A G A A A G T G A C A A G G C G C T C T G C C T T T G G G T 950
43 848 T - GC - TGT - ACATTTCAGAGAGGGGTACA - - C C T C G G G A G A G A A A G T G A C A A G G C G C T C T G C C T T T G G G T 911
44 880 T - GC - TGT - ACACCTCAGCGAGGGGACT - - T C T C G G G G A G A A A A G T G A A A G A G C G C T C T A C C T T T G G G T 982
45 889 T - GC - TGT - ACACCTCAGCGAGGGGACT - - T C T C G G G G A G A A A A G T G A A A G A G C G C T C T A C C T T T G G G T 982
48 833 T - GC - TGT - ACATTTCAGAGAGGGGTACA - - C C T C G G G A G A G A A A G T G A C A A G G C G C T C T G C C T T T G G G T 898
50 878 T - GC - TGT - ACATTTCAGAGAGGGGTACA - - C C T C G G G A G A G A A A G T G A C A A G G C G C T C T G C C T T T G G G T 941
52 880 T - GC - TGT - ACACACAGAGAGGGGACT - - T C T T G G G G A G A A A A G T G A A A G A G C G C T C T A T C T T T G G G T 959
54 753 T - GC - TGT - A C A T T C A G A G A G G G T A C A - - C C T C G G G G A G A A A A G T G A A A G A G C G C T C T A T C T T T G G G T 797
57 887 T - GC - TGT - A C A T T C A G A G A G G G T A C A - - C C T C G G G G A G A A A A G T G A C A A G G C G C T C T G C C T T T G G G T 950

10 945 CACGGACGAA AACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGAGCATGCA - TA - - ACCTGGAGCAT 1010
12 831 CACGGACAA CAACCCTCGCTTTCTCTTGGACAGGGTGGCATGGACAATGCTGGCA - AT TGGGAGTTC 897
13 954 CACGGACGAA AACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGAGCATG - - - - - ACCTGGAGCAT 1014
14 880 CACGGACGAA AACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGAGCATG - - - - - ACCTGGAGCAT 1020
16 983 CACGGACAA CAACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGAAATGCTGTGA - - AGTGGAGCTT 1029
18 945 AACAGACGAA AACCCTCGCTTTAAGTTGTTGGACCGGTTGCTATGGAGCATG - - - - - GCGTGGTGCCTTTGAGCAC 1014
19 948 CACGGACAA CAACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGACAATGCTGTGA - - GGTGGAGTTT 1014
2 948 CACGGATAA AACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGACAATGCTGTGA - - GGTGGAGTTC 1014
20 978 CACGGACAA CAACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGACAATGCTGTGA - - GGTGGAGTT 1044
22 951 CACGGACGAA AACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGAGCATG - - - - - ACCTGGAGCAT 1011
23 948 CACGGACGAA AACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGAGCATG - - - - - GCGTGGAGCAT 1008
25 951 TACGGATAA AACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGACAATGCTGTGA - - GGTGGAGTTT 1017
26 957 CACGGACGAA AACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGAGCATG - - - - - GCGTGGAGCAT 1018
27 983 CACGGACAA CAACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGAAATGCTGTGA - - AGTGGAGCTT 1020
29 980 GACTGACAA CAACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGACAATGCTGTGA - - GGTGGAGTTC 1026
32 980 CACGGACAA CAACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGAAATGCTGTGA - - AGGAGGCTT 1026
33 903 CTTGACAA AACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGACAATGCTGTGA - - GGTGGAGTT 983
34 938 AACTGATGAA AACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGACAATGCTGTGA - - GGTGGAGTT 988
35 980 CACGGACAA CAACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGAAATGCTGTGA - - AGGAGGCTT 1026
36 988 CACGGACAA CAACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGACAATGCTGTGA - - GGTGGAGTT 1032
40 951 CACGGACAA CAACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGAAATGCTGTGA - - AGGAGGCTT 1017
41 951 CACGGACGAA AACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGAGCATG - - - - - ACCTGGAGCAT 1011
43 912 CACGGACGAA AACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGAGCATG - - - - - ACCTGGAGCAT 972
44 983 CACGGACAA CAACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGACAATGCTGTGA - - AGTGGAGTT 1020
45 983 CACGGACAA CAACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGACAATGCTGTGA - - GGTGGAGTT 1029
48 887 CACGGACGAA AACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGAGCATG - - - - - ACCTGGAGCAT 957
50 942 CACGGACGAA AACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGAGCATG - - - - - GCGTGGAGCAT 1002
52 980 CACGGACAA CAACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGAAATGCTGTGA - - AGGAGGCTT 1020
54 788 CACGGACAA CAACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGAAATGCTGTGA - - AGGAGGCTT 884
57 951 CACGGACGAA AACCCTCGCTTTCTATGTTGGACCGGTTGCTATGGAGCATG - - - - - ACCTGGAGCAT 1011

10 1011 TCCAAACGC CCGTGTGCTGCTGCTGGATGGTGGCTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1080
12 898 ACCAACGCAG CCGTGTGCTG - TACTCGGATGGTGGCAATTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 988
13 1015 TCCAAACGC CCGTGTGCTG - TGCTCGGATGGTGGCTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1083
14 1021 ACCAAACCC CCGTGTGCTG - TGCTCGGATGGTGGCTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1089
16 1030 GCGAGCAGC CCGTGTGCTG - TACTCGGATGGTGGCAATTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1098
18 1015 AGTAAACGC CCGTGTGCTG - TGCTCGGATGGTGGCAATTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1083
19 1015 CCGTAAACCC CCGTGTGCTG - TACTCGGATGGTGGCAATTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1079
2 1015 GATCAAAAC CCGTGTGCTG - TACTCGGATGGTGGCAATTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1083
20 1045 CCGAGCAA CCGTGTGCTG - TACTCGGATGGTGGCAATTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1113
22 1012 TCCAAACGC CCGTGTGCTG - TGCTCGGATGGTGGCTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1080
23 1009 TCCAAACGC CCGTGTGCTG - TGCTCGGATGGTGGCTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1077
25 1018 ACAAACGC CCGTGTGCTG - TATAGTGGTGGTAACTTACACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1088
26 1017 TCCAGAGAC CCGTGTGCTG - TACTCGGATGGTGGCAATTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1083
27 1030 GCGAGCAA CCGTGTGCTG - TACTCGGATGGTGGCAATTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1098
29 1027 CCAATCGAA CCGTGTGCTG - TATTCAGATGGTGGCAATTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1095
32 1027 GCGAGCAGC CCGTGTGCTG - TACTCGGATGGTGGCAATTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1085
33 984 AGTAAACGC CCGTGTGCTG - TGCTCGGATGGTGGCTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1032
34 987 TCCAGCAGC CCGTGTGCTG - TT - TTGG - TTGGCTTGGCCCAATTTACAGGAGAGGGGCAATGAAAAGGCGAG 1059
35 1027 GCGAGCAGC CCGTGTGCTG - TACTCGGATGGTGGCAATTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1095
36 1033 CCGAGCAA CCGTGTGCTG - TACTCGGATGGTGGCAATTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1101
40 1018 GCGAGCAGC CCGTGTGCTG - TACTCGGATGGTGGCAATTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1083
41 1012 TCCAAACGC CCGTGTGCTG - TGCTCGGATGGTGGCTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1080
43 973 TCCAAACGC CCGTGTGCTG - TGCTCGGATGGTGGCTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1041
44 1030 CCGAGCAA CCGTGTGCTG - TACTCGGATGGTGGCAATTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1098
45 1030 CCGAGCAA CCGTGTGCTG - TACTCGGATGGTGGCAATTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1098
48 958 TCCAAACGC CCGTGTGCTG - TGCTCGGATGGTGGCTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1026
50 1003 TCCAAACGC CCGTGTGCTG - TGCTCGGATGGTGGCTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1071
52 1027 GCGAGCAGC CCGTGTGCTG - TACTCGGATGGTGGCAATTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1095
54 885 GCGAGCAGC CCGTGTGCTG - TACTCGGATGGTGGCAATTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 933
57 1012 TCCAAACGC CCGTGTGCTG - TGCTCGGATGGTGGCTTGGACCTTTTACAGGAGAGGGGCAATGAAAAGGCGAG 1080

10 1081 GGCATTTT - CACTTGGCCGCGC - TGACGAGAGGAGCTGAGAGGATCAAGTCCACCTT - - - - - CAGTA - - - - 1138
12 987 GGCATTTT - CACTTGGCCGCA - TGACGAGAGGAGCTGAGAGGATCAAGTCCACCTT - - - - - CAGTA - - - - 1024
13 1084 GGCATTTT - CACTTGGCCGCGC - TGACGAGAGGAGCTGAGAGGATCAAGTCCACCTT - - - - - CAGTA - - - - 1141
14 1080 GGCATTTT - CACTTGGCCGCGC - TGACGAGAGGAGCTGAGAGGATCAAGTCCACCTT - - - - - TAGTA - - - - 1147
16 1099 GGCATTTT - CACTTGGCCGCGC - TGACGAGAGGAGCTGAGAGGATCAAGTCCACCTT - - - - - CAGTA - - - - 1158
18 1084 GGCATTTT - CACTTGGCCGCGC - TGACGAGAGGAGCTGAGAGGATCAAGTCCACCTT - - - - - GCGA - - - - 1141
19 1080 GGCATTTT - CACTTGGCCGCGC - TGACGAGAGGAGCTGAGAGGATCAAGTCCACCTT - - - - - GAGG - - - - 1135
2 1084 GGCATTTT - CACTTGGCCGCGC - TGACGAGAGGAGCTGAGAGGATCAAGTCCACCTT - - - - - CAGTA - - - - 1143
20 1114 GGCATTTT - CACTTGGCCGCGC - TGACGAGAGGAGCTGAGAGGATCAAGTCCACCTT - - - - - CAGTA - - - - 1171
22 1081 GGCATTTT - CACTTGGCCGCGC - TGACGAGAGGAGCTGAGAGGATCAAGTCCACCTT - - - - - CAGTA - - - - 1138
23 1078 GGCATTTT - CACTTGGCCGCGC - TGACGAGAGGAGCTGAGAGGATCAAGTCCACCTT - - - - - CAGTA - - - - 1135
25 1087 GGCATTTT - CACTTGGCCGCGC - TGACGAGAGGAGCTGAGAGGATCAAGTCCACCTT - - - - - CAGTA - - - - 1154
26 1084 GGCATTTT - CACTTGGCCGCGC - TGACGAGAGGAGCTGAGAGGATCAAGTCCACCTT - - - - - CAGTA - - - - 1141

27 1099 GACTTGT. CGCTTCTCGCC. TGACGGAGGACTGAGTACAATCAAGTCCGTCCT. CAGTA. ... 1150
29 1098 GATCTGAG. CCTTTCTGTAG. TGACGGAGGGACAGAGTACAATCAAGTCCAGCC. CAATA. ... 1154
32 1096 GACTTGT. CGCTTCTCGCC. TGACGGAGGAGTGAAGTACAATCAATTTCTCT. GAATA. ... 1153
33 1033 GCTATCT. CCTTGGCCAGGC. TGACGGAGGAGCTAAGAGATAAAATCAACATT. GTCGA. ... 1090
34 1080 GACTTGT. TACGTTGCCCTCC. AGATGGGGGGCCGGAAGACGACCAAGTTCACCCA. CAGTA. ... 1117
35 1096 GACTTGT. CGCTTCTNCCC. TGACGGAGGAGTGGGTACGATCAATTTCTCTCT. GGAAA. ... 1153
36 1102 GACTTGT. CGCTTCTGCTC. TGACGGGGGAACTGAGTACAATCAAGTCCGTCCT. CAGTA. ... 1150
40 1084 GCGTTGT. CGCTTCTCGCC. TGACGGAGGAGTGAAGTACGATCAATTTCTCTCT. GAAGA. ... 1141
41 1081 GCAATTT. CACTTGGCCGCC. TGACGGAGGAGCTGAAGAGGATCAAGTCCACCCT. CAGTA. ... 1138
43 1042 GCAATTT. CACTTGGCCGCC. TGACGGAGGAGCTGAAGAGGATCAAGTCCACCCT. CAGTA. ... 1099
44 1099 GACTTGT. CGCTTCTGCTC. TGACGGGGGAACTGAGTACAATCAAGTCCGTCCT. CAGTA. ... 1158
45 1099 GACTTGT. CGCTTCTGCTC. TGACGGGGGAACTGAGTACAATCAAGTCCGTCCT. CAGTA. ... 1158
48 1027 GCAATTT. CACTTGGCCGCC. TGACGGAGGAGCTGAAGAGGATCAAGTCCACCCT. CAGTA. ... 1084
50 1072 GCAATTT. CACTTGGCCGCC. TGACGGAGGAGCTGAAGAGGATCAAGTCCACCCT. CAGTA. ... 1129
52 1096 GACTTGT. CGCTTCTNCCC. TGACGGAGGAGTGGGTACGATCAATTTCTCTCT. GGAGA. ... 1153
54 934 GACTTGT. CGCTTCTGCTC. TGACGGAGGAGTGAAGTACGATCAATTTCTCTCT. CAGTA. ... 091
57 1081 GCAATTT. CACTTGGCCGCC. TGACGGAGGAGCTGAAGAGGATCAAGTCCACCCT. CAGTA. ... 1138

10 1138 - CTGGGCGAGACTGGACGCCCTTCTTCTCAAGTCG. TCCATAACCCACGACTGGTCTGGT. ... ATTC. CT 1202
12 1024 - CTGGGCGGAAAGGACATTTCTTCTCCAGCTTT. TCTATACCCACGGCGGGTCTGGTGGCAGT. TT. TT 1091
13 1141 - CTGGGCGGAGACTGGACGCCCTTCTTCTCCAAGTCG. TCCATAACCCACGACTGGTCTGGT. ... ATC. CT 1205
14 1147 - CGTGGTCTCAGAAAGGATGCTTTTATATTCAGT. TCCATAACCCACGGCGGGTCTGGTGGCAGT. TT. TT 1214
16 1156 - CTGGGCGAAAGGACGCCCTTCTTCTCCAGCTTT. TCTATACCCACGGCGGGTCTGGTGGCAGT. TT. TT 1223
18 1141 - CATGGGCTAGGTTGGACGCCCTTCTTCAAGTCA. AGTATACCCACGACTGGTCTGGTATG. GATTTT 1208
19 1135 - AGCAGTCCGAAAGGATGCTTTTATCTTCAAT. T. AGTATACCCACGGCG. ... CTAGTGGCTTC. TC 1190
2 1143 - CTAGGAAA. GAAAGGCGGCTCTGTATTTGCAACT. A. TCTATACCCACGGAGGTTGGTGGT. GCGCTCTTA 1209
20 1171 - CTGGTCCGAAAGGACGCCCTTCTTCTCCAGCTTT. TCTATACCCACGGCGGGTCTGGTGGCAGT. TT. TT 1238
22 1138 - CTGGGCGGAGACTGGACGCCCTTCTTCTCAAGTCG. TCCATAACCCACGACTGGTCTGGT. TTGGATC. CT 1205
23 1135 - CTGGGCGGAGACTGGACGCCCTTCTTCCAAGTCG. TCCATAACCCACGACTGGTCTGGT. TTGGATC. CT 1202
25 1155 - CTGGGCGGAAAGGACCTTTTCTTCCGCTGCTTCTCTATACCCACGGCGGGTCTGGTGGCAGT. TT. TT 1223
26 1141 - CTAGGCGGAAAGGCGGACTTCTGCTGCAAGTAGT. TCCATAACCCACGGAGTGGTCTGGT. TGA. ... TCCA 1205
27 1150 - CTGGGCGAAAGGACGCCCTTCTTCTCAGCTTT. TCTATACCCACGGCGGGTCTGGTGGCAGT. TT. TT 1222
29 1154 - TCTAGT. ... GCTGGAAAGTACACTTCAAGTCA. TCTATACCCACGGCGGGTCTGGTGGCAGT. TT. TT 1219
32 1153 - CTGGGCGAAAGGACGCCCTTCTTCTCAGGTTG. TCTATACCCACGGCGGGTCTGGTGGCAGT. TT. TT 1220
33 1080 - CTGGGCGGCTTAGATGCTTTCTTCTCCAAGTCG. TCCATAACCCACGACTGGTCTGGTGG. ATACNT 1157
34 1117 - CTAGGCGGAGACTGGACGCCCTTCTTCTGCAAGT. TCCATAACCCACGACTGGTCTGGTGG. ATACNT 1184
35 1153 - CTGGGCGGAAAGGACGCCCTTCTTCTCAGGTTG. TCTATACCCACGGCGGGTCTGGTGGCAGT. TT. TT 1220
36 1159 - CTGGTCCGAAAGGACGCCCTTCTTCTCAGCTTT. TCTATACCCACGGCGGGTCTGGTGGCAGT. TT. TT 1227
40 1141 - CTGGGCGGAAAGGACGCCCTTCTTCTCAGGTTG. TCTATACCCACGGCGGGTCTGGTGGCAGT. TT. TT 1208
41 1138 - CTGGGCGGAGACTGGACGCCCTTCTTCTCAAACTG. TCCATAACCCACGACTGGTCTGGT. TTGGATC. CT 1205
43 1099 - CTGGGCGGAGACTGGACGCCCTTCTTCTCAAGTCG. TCCATAACCCACGACTGGTCTGGT. TTGGATC. CT 1108
44 1156 - CTGGTCCGAAAGGACGCCCTTCTTCTCAGCTTT. TCTAT. CCCACGGCGGGTCTGGTGGCAGT. TT. TT 1223
45 1158 - CTGGTCCGAAAGGACGCCCTTCTTCTCAGCTTT. TCTAT. CCCACGGCGGGTCTGGTGGCAGT. TT. TT 1222
48 1084 - CTGGGCGGAGACTGGACGCCCTTCTTCTCAAGTCG. TCCATAACCCACGACTGGTCTGGT. TTGGATC. CT 1151
50 1120 - CTGGGCGGAGACTGGACGCCCTTCTTCTCAAGTCG. TCCATAACCCACGACTGGTCTGGT. TTGGATC. CT 1196
52 1153 - CTGGGCGGAAAGGACGCCCTTCTTCTCAAGTCG. TCTATACCCACGGCGGGTCTGGTGGCAGT. TT. TT 1220
54 991 - CTGGGCGGAAAGGACGCCCTTCTTCTCAGGTTG. TCTATACCCACGGCGGGTCTGGTGGCAGT. TT. TT 1058
57 1138 - TTTGGGCGGAGACTGGACGCCCTTCTTCTCAAGTCG. TCCATAACCCACGACTGGTCTGGT. TTGGATC. CT 1205

10 1202 GTCCGATACGTCGTCCGGTGGCGACACGTGGATCGACGAGTACCGCTGCGTGAATGCAACGGTCAACGAA 1272
12 1092 GTCCG. ATGC. GGCCAGTGTGACACGTGGATCGACGAGTACCTTTTCCCTGAATGCAACGGTCAACGAA 1158
13 1206 GTCCGATACGTCGTCCGGTGGCGACACGTGGATCGACGAGTACCGCTGCGTGAATGCAACGGTCAACGAA 1275
14 1215 GTCCGATACGTCGTCCGGTGGCGACACGTGGATCGACGAGTACCGCTGCGTGAATGCAACGGTCAACGAA 1284
16 1224 GTCCG. ATGC. GGCCAGTGTGACACGTGGATCGACGAGTACCTTTTCCCTGAATGCAACGGTCAACGAA 1290
18 1200 ATCAGACACGAGCAGTGGCGGGATACCTGGATGACGAGTATCGTTGCTGAATGCAACGGTCAACGAA 1278
19 1200 ACCGG. ACCGGTGC. AGTAAACGAGACGTGGATCGACGAGTACCTTTTCCCTGAATGCAACGGTCAACGAA 1286
2 1210 AGCC. ACCG. AGCATCTAATGGACATGGATGAGTGAATGCTTTGCTTAAACGGCGGGTCAACGAA 1275
20 1230 GTCCG. ATGCCGCT. CGAATGGACATGGATGAGTGAATTTGCTTAAATGCAACGGTCAACGAA 1305
22 1208 GTCCGATACGTCGTCCGGTGGCGACACGTGGATCGACGAGTACCGCTGCGTGAATGCAACGGTCAACGAA 1275
23 1203 GTCCGATACGTCGTCCGGTGGCGACACGTGGATCGACGAGTACCGCTGCGTGAATGCAACGGTCAACGAA 1272
25 1224 GTCCG. ACCG. AGCATCAGACGATACGTGGATCGACGAGTAC. ... TGC. ... AATGCGCCGTTACGAA 1284
26 1206 GTACGATATGCTGTACGATGGCGCCACGTGGATCGACGAGTACCGCTGCGTGAATGCAACGGTCAACGAA 1273
27 1223 GTCCG. ATGCCGCT. CGAATGGACATGGATGAGTGAATTTGCTTAAATGCAACGGTCAACGAA 1290
29 1220 GTCCG. ACCG. TGGCTCTAATGGACATGGATGAGTGAATTTGCTTAAATGCAACGGTCAACGAA 1287
32 1158 GTCCG. ATAC. GGCCAGTGTGACACGTGGATCGACGAGTACCTTTTCCCTGAATGCAACGGTCAACGAA 1287
33 1221 CAGTGACACTTCTAGTGGTGGAGACACGTGGATCGACGAGTACAGATGCTTAAATGCAACGGTCAACGAA 1227
34 1185 GAGTGACACGAGTGTGGCGGGATACCTGGATGAGTGAATGCTTTGCTTAAATGCAACGGTCAACGAA 1251
35 1221 GTCCG. ATAC. GGCCAGTGTGACACGTGGATCGACGAGTACCTTTTCCCTGAATGCAACGGTCAACGAA 1287
36 1228 GTCCG. ATGCCGCT. ... ACAATGGACACGTGGATCGACGAGTACCTTTTCCCTGAATGCAACGGTCAACGAA 1293
40 1209 GTCCG. ATCC. GGCCAGTGTGACACGTGGATCGACGAGTACCTTTTCCCTGAATGCAACGGTCAACGAA 1275
41 1208 GTCCGATACGTCGTCCGGTGGCGACACGTGGATCGACGAGTACCGCTGCGTGAATGCAACGGTCAACGAA 1275
43 1167 GTCCGATACGTCGTCCGGTGGCGACACGTGGATCGACGAGTACCGCTGCGTGAATGCAACGGTCAACGAA 1236
44 1224 GTCCG. TGGCGGCT. AGCAATGGACATGGATGAGTGAATTTGCTTAAATGCAACGGTCAACGAA 1290
45 1223 GTCCG. ATGCCGCT. CGAATGGACATGGATGAGTGAATTTGCTTAAATGCAACGGTCAACGAA 1290
48 1152 GTCCGATACGTCGTCCGGTGGCGACACGTGGATCGACGAGTACCGCTGCGTGAATGCAACGGTCAACGAA 1221
50 1107 GTCCGATACGTCGTCCGGTGGCGACACGTGGATCGACGAGTACCGCTGCGTGAATGCAACGGTCAACGAA 1286
52 1221 GTCCG. ATAC. GGCCAGTGTGACACGTGGATCGACGAGTACCTTTTCCCTGAATGCAACGGTCAACGAA 1287
54 1050 GTCCG. ATAC. GGCCAGTGTGACACGTGGATCGACGAGTACCTTTTCCCTGAATGCAACGGTCAACGAA 1125
57 1206 GTCCGATACGTCGTCCGGTGGCGACACGTGGATCGACGAGTACCGCTGCGTGAATGCAACGGTCAACGAA 1275

10 1273 GCA. GCGAAGGTCAAATAATGGCTTTAAGTTACAGGCGGCTGGTCCAGGGCAATGTTGGTCTATGAGC. ... 1338
12 1159 GCA. AAGAAGGTCAAAGATGGCTTTCAATGACAGGAGGCTCAACTCCGGGGTAAATGGTCTGTGAA. ... 1224
13 1276 GCA. GCGAAGTCAAATAATGGCTTTAAGTTACAGGCGGCTGGTCCNAGGCAATGTTGGTCTATGAGC. ... 1341
14 1285 GCT. ACCCAAAGTTAAAGACGGATTTCAAGTCACTGAAAGTAACTCTGGGGTACTCTGGTTGTCAAC. ... 1350
16 1291 GCG. ACGAAGGTCAAAGATGGATTTCAATGACAGGAGCTTAACTCCGGGGTACTATGGTTGTGAAT. ... 1356
18 1279 GCGAGCGAAGGACAAAATGGCTTT. ... GTCACGGTGCATGGGTC. CGCGCAATGTTGGTCTGTGAGC. ... 1342
19 1287 GCG. ACGAAGGTCAAAGATGGATTTCAATGACAGGAGCTTAACTCCGGGGTACTGTGGTTGTGAA. ... 1332
2 1276 GCG. ACTAAGGTCAAAGATGGATTTCAAGTCACTGAAAGTAACTCTGGGGTGTCTTGGTTGTGAAT. ... 1341
20 1306 GCG. ACGAAGGTCAAAGATGGATTTCAATGACAGGAGCTTAACTCCGGGGTACTGTGGTTGTGAA. ... 1371
22 1276 GCA. GCGAAGGTCAAATAATGGCTTTAAGTTACAGGCGGCTGGTCCAGGGCAATGTTGGTCTATGAGC. ... 1341

23 1273 NCA-ACCAAGGTCAAAAATGGCTTTAAATTCACCGGCGCCGGGTCCAGGGCAATGTGGTCTATGAGC... 1338
25 1285 GCA-AAAAAGTCAAGATGGGTTTCAAGCTAACGGAACTAACCTCCGGGTAAATGTGGCCGGTTTAA... 1350
26 1274 GAA-GTAAAGGTCAAGAACTAGCGTGTGTCATCCCGCTCCCGGG-CAGGCAATGTGGTCCGATGTG... 1339
27 1291 GTC-ACGAAGGTCAAGGATGGATTTCAATTCACGGAGTCCAACCTCCGGGTACTGTGGTGTGTAAC... 1356
29 1288 GCT-ACAAAGGTAAAGATGGTITTTCAACTAACTGAGTCCAACCTCCGGGTGTCTGTGGTTCGTGAAC... 1353
32 1288 GCC-ACGAAGGTCAAGGATGGATTTCAATTCACGGAGTCCAACCTCCGGGTACTGTGGTTCGTGAAC... 1353
33 1228 GCT-GCAAAAGTAAAAAACGGGTTTAAAGTTCACAGCACCAGGATCCAGAGGATGTGGTCCGATGAGC... 1294
34 1252 AGA-GCACAA-CAATCAAAAACGGTTCAGGTTCCGGT--GGA-TTAGGCACGNGGTCAAGTGGATT... 1314
35 1288 GCC-ACGAAGGTCAAGGATGGATTTCAATTCACGGAGCTTAGCTCCAGGGTATCATGGTCCGTGAAC... 1353
36 1294 GCC-ACGAAGGTCAAGGATGGATTTCAATTCACGGAGTCCAACCTCCGGGTACTGTGGTGTGTAAC... 1359
40 1276 GCC-ACGAAGGTCAAGGATGGATTTCAATTCACGGAGCTAACCTCCAGGGTATGGGCCCTGAAC... 1341
41 1276 GCA-GGAAGGTCAAAAATGGCTTTAAGTTCACGGCCCTGGGTCCAGGGCAATGTGGTCTATGAAC... 1342
43 1237 GCA-GGAAGGTCAAAAATGGCTTTAAGTTCACGGCCCTGGGTCCAGGGCAATGTGGTCTATGAAC... 1302
44 1291 CCC-CCAAAGTCAAGGATGGATTTAAATTCACGGAGTCCAACCTCCGGGTACTGTGGTGTGTAAC... 1358
45 1291 GCC-ACGAAGGTCAAGGATGGATTTCAATTCACGGAGTCCAACCTCCGGGTACTGTGGTGTGTAAC... 1358
48 1222 GCA-GGAAGGTCAAAAATGGCTTTAAGTTCACGGCCCTGGGTCCAGGGCAATGTGGTCTATGAAC... 1287
50 1287 GCA-GCAAGGTCAAAAATGGCTTTAAGTTCACGGCCCTGGGTCCAGGGCAATGTGGTCTATGAAC... 1332
52 1288 GCC-ACGAAGGTCAAGGATGGATTTCAATTCACGGAGCTTAGCTCCAGGGTATCATGGTCCGTGAAC... 1353
54 1126 GCC-ACGAAGGTCAAGGATGGATTTCAATTCACGGAGCTTAGCTCCAGGGTATCATGGTCCGTGAAC... 1191
57 1276 GCA-GGAAGGTCAAAAATGGCTTTAAGTTCACGGCCCTGGGTCCAGGGCAATGTGGTCTATGAAC... 1341

10 1338 ---AGGTGGGACCFAATAAAGAGTACGGCTTTGTGAACACAGATTCACCTTTGTGGGGAAGGTGACCA 1405
12 1224 ---ATTCGGGATGATAATGTGCGTCAATATCCTTGAGCCACAACTTCACACTTTTGGCCGTGGTGACCA 1291
13 1341 ---AGGTGGGACTFAATAAAGAGTACGGCTTTGTGAACACAGATTCACCTTTGTGGGGAAGGTGACCA 1408
14 1350 ---ACACGGGACGATAAAGTGTAGCCAGTATCATTGAGCCACAACTTCACCTTTGTGGGGAAGGTGACCA 1417
16 1356 ---ACTCGGGATGATAAAGTGTGCGCAATGATCATTGAGCCACAACTTCACACTTTGTGGGGAAGGTGACCA 1423
18 1342 ---GGTGGGTCCTAATAAACAATAGGGTTCGTGAGTCCACAGATTCACCTTTGTGGGGAAGGTGACCA 1408
19 1332 ---ACTCGGGATGATAAAGTGTGCGTCAATGATCATTGAGCCACAACTTCACACTTTGTGGGGAAGGTGACCA 1399
2 1341 ---ACACGGGACGATAAAGTGTGCGTCAATGATCATTGAGCCACAACTTCACACTTTGTGGGGAAGGTGACCA 1408
20 1371 ---ACTCGGGATGATAAAGTGTGCGTCAATGATCATTGAGCCACAACTTCACACTTTGTGGGGAAGGTGACCA 1438
22 1341 ---AGGTGGGACCTAATAAAGAGTACGGCTTTGTGAACACAGATTCACCTTTGTGGGGAAGGTGACCA 1408
23 1338 ---AGGTGGGACCTAATAAAGAGTACGGCTTTGTGAACACAGATTCACCTTTGTGGGGAAGGTGACCA 1405
25 1350 ---ATTCGGGACGATAAAGTGTAGGCAATGATCATTGAGCCACAACTTCACACTTTGTGGGGAAGGTGACCA 1417
26 1339 ---GATGGGCCCTAACAACAATAGGGTTCGTGAGTCCACAGGTTAGACTCGTGGGCAAGGTGACCA 1405
27 1356 ---ACTCGGGATGATAAAGTGTGCGTCAATGATCATTGAGCCACAACTTCACACTTTGTGGGGAAGGTGACCA 1423
29 1353 ---ACTCGGGACGATAAAGTGTGCGTCAATGATCATTGAGCCACAACTTCACACTTTGTGGGGAAGGTGACCA 1420
32 1353 ---ACTCGGGTAAATAAAGTGTGCGTCAATGATCATTGAGCCACAACTTCACACTTTGTGGGGAAGGTGACCA 1420
33 1294 ---GGTGGGACCTAATAAACAATAGGGTTCGTGAGTCCACAGGTTAGTTCGTGGGCAAGGTGACCA 1380
34 1315 GGCCTGGTGGCGACTAATAAAGAGTACTGCGNCTTACGCATCGCTTCACACTTTGTGGGGAAGGTGACCA 1384
35 1353 ---ACTCGGGTAAATAAAGTGTGCGTCAATGATCATTGAGCCACAACTTCACACTTTGTGGGGAAGGTGACCA 1420
36 1359 ---ACTCGGGATGATAAAGTGTGCGTCAATGATCATTGAGCCACAACTTCACACTTTGTGGGGAAGGTGACCA 1426
40 1341 ---ACTCGGGTAAATAAAGTGTGCGTCAATGATCATTGAGCCACAACTTCACACTTTGTGGGGAAGGTGACCA 1408
41 1342 ---AGGTGGGACCTAATAAACAATAGGGTTCGTGAGTCCACAGGTTAGTTCGTGGGCAAGGTGACCA 1408
43 1302 ---AGGTGGGACCTAATAAACAATAGGGTTCGTGAGTCCACAGGTTAGTTCGTGGGCAAGGTGACCA 1369
44 1356 ---ACTCGGGATGATAAAGTGTGCGTCAATGATCATTGAGCCACAACTTCACACTTTGTGGGGAAGGTGACCA 1423
45 1356 ---ACTCGGGATGATAAAGTGTGCGTCAATGATCATTGAGCCACAACTTCACACTTTGTGGGGAAGGTGACCA 1423
48 1287 ---AGGTGGGACNCTAATAAACAATAGGGTTCGTGAGTCCACAGGTTAGTTCGTGGGCAAGGTGACCA 1354
50 1332 ---AGGTGGGACCTAATAAACAATAGGGTTCGTGAGTCCACAGGTTAGTTCGTGGGCAAGGTGACCA 1399
52 1353 ---ACTCGGGTAAATAAAGTGTGCGTCAATGATCATTGAGCCACAACTTCACACTTTGTGGGGAAGGTGACCA 1420
54 1191 ---ACTCGGGTAAATAAAGTGTGCGTCAATGATCATTGAGCCACAACTTCACACTTTGTGGGGAAGGTGACCA 1258
57 1341 ---AGGTGGGACCTAATAAACAATAGGGTTCGTGAGTCCACAGGTTAGTTCGTGGGCAAGGTGACCA 1408

10 1406 TCCACCAGGTTCGGAAAGGGGAGCCTCTCTGCTGGGTGGGGTCTCCGGGACGGTCCGGGCAAGGAGAT 1475
12 1292 TCGAAGTGGTTCGGAGTGAGAAAACCTCTCTACTGGGTGGGACGCTTGGGGACACTAATCCCGGCATAC 1381
13 1409 TCCACCAGGTTCGGAAAGGGGAGCCTCTCTGCTGGGTGGGGTCTCCGGGACGGTCCGGGCAAGGAGAT 1478
14 1418 TCCACCAGGTTCGGAAAGGGGAGCCTCTCTGCTGGGTGGGGTCTCCGGGACGGTCCGGGCAAGGAGAT 1487
16 1424 TCGAAGATGTTCGGAGTAATAGCACCTCTACTGGGTGGGATGTGGGAAACAATAATCCAATCATAC 1493
18 1409 TCCACCAGGTTCGGAAAGGGTTCGACCCCTACTTGGAGGGGCTTGGGGACACTAATTCGAATCATAC 1478
19 1400 TCGAAGAGGTTCGGAGTAATAGCACCTCTACTACTGACGGGACGTTGGGAAACAATAATTCGAATCATAC 1489
2 1409 TCGAAGAGGTTCGGTTCGAACAGTACGGTACTCTGACAGGGCACGTTGGGCAACAACAATTCGAATCATAC 1478
20 1439 TCGAAGAGGTTCGGAGTAATAGCACCTCTACTGACGGGACGTTGGGCAACAATAATTCGAATCATAC 1508
22 1409 TCCACCAGGTTCGGAAAGGGGAGCCTCTCTGCTGGGTGGGGTCTCCGGGACGGTCCGGGCAAGGAGAT 1478
23 1406 TCCACCAGGTTCGGAAAGGGGAGCCTCTCTGCTGGGTGGGGTCTCCGGGACGGTCCGGGCAAGGAGAT 1475
25 1418 TCGAAGTACCGCGCTGTAAGAAAGACCCCTTGTAGGGGTACACTAGGTGATACCAATAGTCCACACAC 1487
26 1406 TCCACCAGGTTCGGAAAGGGTAGTACCCCTACTTGGTGGTGGGCTCCGGGACGGGAGGGGAAAGAGAT 1475
27 1424 TCGAAGAGGTTCGGAGTAATAGCACCTCTACTGGGTGGGATGTGGGAAACAATAATTCGAATCATAC 1493
29 1421 TCGAAGAGGTTCGGAGTAATAGCACCTCTCTTGGAGGGGACGTTAGCCAAACAACAATAATTCGAATCATAC 1490
32 1421 TCGAAGAGGACTCCGATTTGGGACCTCTCTCTGACTGGGACATGGGCAACAATAATTCGAATCATAC 1490
33 1361 TCCACCAGGTTCGGAAAGGGTAGTACCCCTGTTAGGGGACGGTCTGGGGATGGTCCAGGGCAAGGAAAT 1430
34 1385 TCCACCAGGTTCGGAAAGGGTCTACTCTCTATAGGAGCCGGTTCGGGGATGGTCCGGGCAAGGAAAT 1454
35 1421 TCGAAGAGGACTCCGATTTGGGACCTCTCTCTACTGACTGGGACGTTGGGCAACAATAATTCGAATCATAC 1490
36 1427 TCGAAGAGGACTTCGGAGTAATAGCACCTCTACTGACTGGGACGTTGGGCAACAATAATTCGAATCATAC 1490
40 1409 TCGAAGAGGACTCCGATTTGGGACCTCTCTCTGACTGGGACGTTGGGCAACAATAATTCGAATCATAC 1478
41 1409 TCCACCAGGTTCGGAAAGGGGAGCCTCTCTGCTGGGTGGGGTCTCCGGGACGGTCCGGGCAAGGAGAT 1478
43 1370 TCCACCAGGTTCGGAAAGGGGAGCCTCTCTGCGGGTGGGGTCTCCGGGACGGTCCGGGCAAGGAGAT 1439
44 1424 TCGAAGAGGTTCGGAGTAATAGCACCTCTACTGACTGGGACGTTGGGCAACAATAATTCGAATCATAC 1493
45 1424 TCGAAGAGGTTCGGAGTAATAGCACCTCTACTGACTGGGACGTTGGGCAACAATAATTCGAATCATAC 1493
48 1355 TCCACCAGGTTCGGAAAGGGGAGCCTCTCTGCTGGGTGGGGTCTCCGGGACGGTCCGGGCAAGGAGAT 1424
50 1400 TCCACCAGGTTCGGAAAGGGGAGCCTCTCTGCTGGGTGGGGTCTCCGGGACGGTCCGGGCAAGGAGAT 1489
52 1421 TCGAAGAGGACTCCGATTTGGGACCTCTCTCTGACTGGGACGTTGGGCAACAATAATTCGAATCATAC 1490
54 1259 TCGAAGAAAACCTCCGATTTGGGACCTCTCTTGGAGTGGGACGTTGGGCAACAATAATTCGAATCATAC 1328
57 1409 TCCACCAGGTTCGGAAAGGGGAGCCTCTCTGCTGGGTGGGGTCTCCGGGACGGTCCGGGCAAGGAGAT 1478

10 1476 CATTTGGGCTGTGCTACAGCATGATAAAAAGTGGGAGAGGTTGTTCAACGGCAACAAAACAGGACGGGG 1546
12 1382 CATTTGGGCTGTGCTACAGCATGATAAAAAGTGGGAGAGGTTGTTCAACGGCAACAAAACAGGACGGGG 1428
13 1479 CATTTGGGCTGTGCTACAGCATGATAAAAAGTGGGAGAGGTTGTTCAACGGCAACAAAACAGGACGGGG 1548
14 1488 CATTTGGGCTGTGCTACAGCATGATAAAAAGTGGGAGAGGTTGTTCAACGGCAACAAAACAGGACGGGG 1557
16 1404 CATTTGGGCTGTGCTACAGCATGATAAAAAGTGGGAGAGGTTGTTGAGGAAAACAAAACAGGACGGGG 1583
18 1479 CATTTGGGCTGTGCTACAGCATGATAAAAAGTGGGAGAGGTTGTTAATGGCAACAAAACAGGACGGGG 1548
19 1470 CATTTGGGCTGTGCTACAGCATGATAAAAAGTGGGAGAGGTTGTTGAGGAAAACAAAACAGGACGGGG 1539

2 1479 AATGGGACTCTCTCTATAACAGCGGAA[CA]AAAGTGGGAGACGATGTTTGAAGAGAA[AA]GAA[GA]GACAGACGA 1548
20 1509 CATGGGTCTGTCGTATACCGCTGATAAAGTGGGAGACGATGTTTCAAGGAAACAAAAAACAGCGCGGAGG 1578
22 1479 CATGGGCTGTCGTACAGCATGAATAAAAGTGGGAGAAAGTGTTCACCGCA[CA]AAAAACAGCACCGGGG 1548
23 1476 CATGGGCTGTCGTACAGCATGAATAAAAGTGGGAGAAAGTGTTCACCGCA[CA]AAAAACAGCACCGGGG 1545
25 1498 AATGGGTATTAATACTATACACAGCAGAC[CA]AAAGTGGGTTACBACCTTCAACBA - - TAA[GA]ACTCCGAGTCC 1554
26 1476 AATCGGCTGTCATACAGCATGAATAAAAGTGGGAGAAAGTGTTCACCGCA[CA]AAAAACAGCACCGGGG 1545
27 1484 CATGGGACTGTCGTATACCGCTGATAAAGTGGGAGACGATGTTTCAAGGAAACAAAAAACAGCACCGGAGC 1503
29 1401 TATGGGCTGTCGTACAGCATGAATAAAAGTGGGAGAAAGTGTTCACCGCA[CA]AAAAACAGCACCGGAGC 1500
32 1401 CATGGGTCTGTCGTATACCGCTGATAAAGTGGGAGAAAGTGTTCACCGCA[CA]AAAAACAGCACCGGAGC 1500
33 1431 AATGGGTAAAGTACTCGATGAATAAAAGTGGGAGAAAGTGTTCACCGCA[CA]AAAAACAGCACCGGAGC 1500
34 1465 TATCGGCTTTCCTACTCAATGAATAAAAGTGGGAGAAAGTGTTCACCGCA[CA]AAAAACAGCACCGGAGC 1524
35 1491 CATGGGTCTGTCGTATACCGCTGATAAAGTGGGAGAAAGTGTTCACCGCA[CA]AAAAACAGCACCGGAGC 1500
36 1497 CATGGGTCTGTCGTATACCGCTGATAAAGTGGGAGAAAGTGTTCACCGCA[CA]AAAAACAGCACCGGAGC 1500
40 1479 CATGGGTCTGTCGTATACCGCTGATAAAGTGGGAGAAAGTGTTCACCGCA[CA]AAAAACAGCACCGGAGC 1549
41 1479 CATGGGTCTGTCGTATACCGCTGATAAAGTGGGAGAAAGTGTTCACCGCA[CA]AAAAACAGCACCGGAGC 1548
43 1440 CATGGGCTGTCGTACAGCATGAATAAAAGTGGGAGAAAGTGTTCACCGCA[CA]AAAAACAGCACCGGGG 1509
44 1404 CATGGGTCTGTCGTATACCGCTGATAAAGTGGGAGACGATGTTTCAAGGAAACAAAAAACAGCACCGGAGC 1503
45 1404 CATGGGTCTGTCGTATACCGCTGATAAAGTGGGAGACGATGTTTCAAGGAAACAAAAAACAGCACCGGAGC 1503
48 1425 CATGGCTTGTCTACAGCATGAATAAAAGTGGGAGAAAGTGTTCACCGCA[CA]AAAAACAGCACCGGGG 1484
50 1470 CATGGGCTGTCGTACAGCATGAATAAAAGTGGGAGAAAGTGTTCACCGCA[CA]AAAAACAGCACCGGGG 1539
52 1401 CATGGGTCTGTCGTATACCGCTGATAAAGTGGGAGAAAGTGTTCACCGCA[CA]AAAAACAGCACCGGAGC 1500
54 1329 CATGGGTCTGTCGTATACCGCTGATAAAGTGGGAGAAAGTGTTCACCGCA[CA]AAAAACAGCACCGGAGC 1388
57 1479 CATGGGCTGTCGTACAGCATGAATAAAAGTGGGAGAAAGTGTTCACCGCA[CA]AAAAACAGCACCGGGG 1548

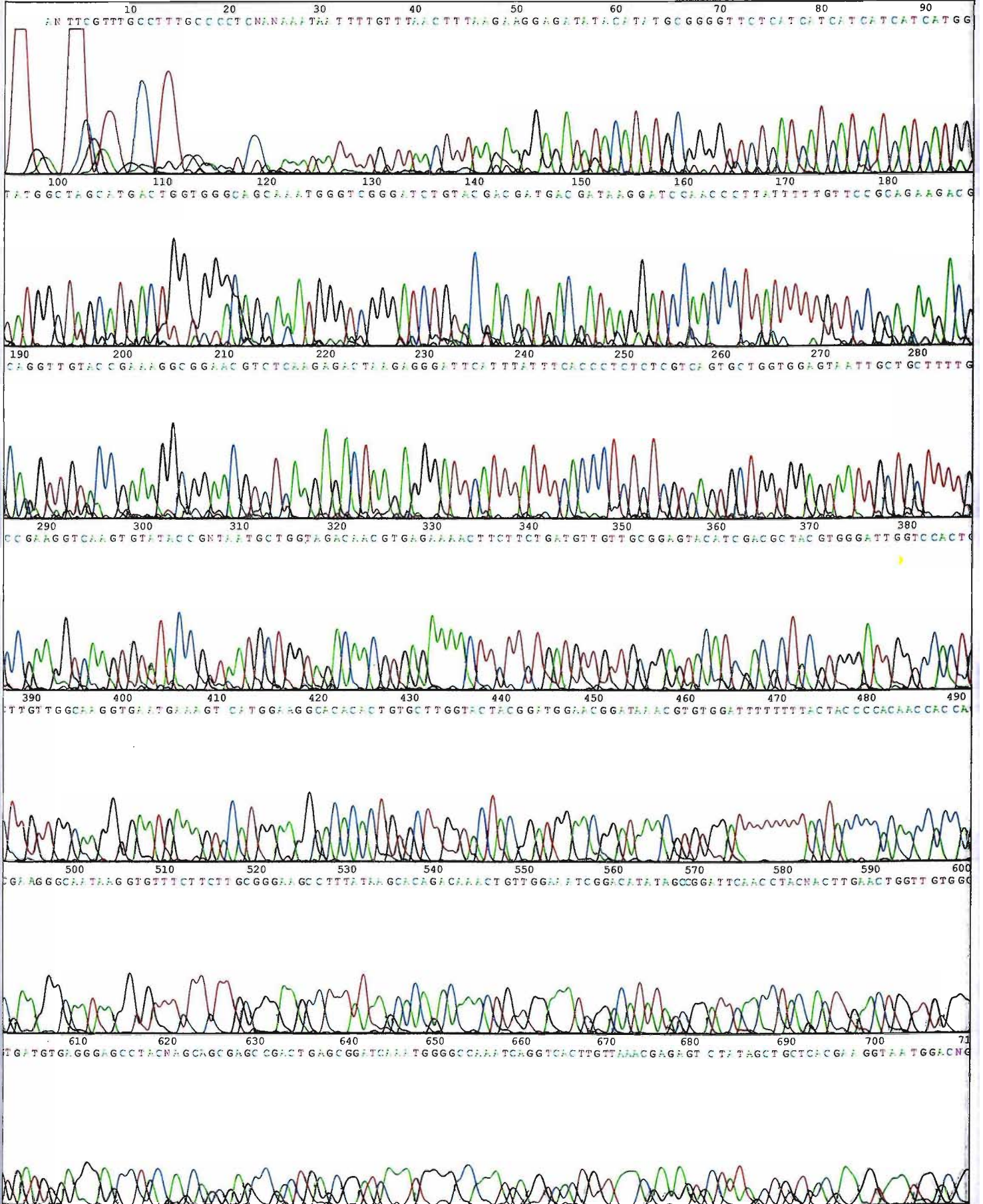
10 1540 AGCACGTTGGGAGCCGGGAGAGAACACCAAGTGGCGCTCATGCTGCAGGACGGCAAAAAGGGCTCCGTTGT 1615
12 1429 GGCACCTTGGGAGTTCGGGAAAGAGTACCAAGTGGCACTCATGCTGCAGG - - - GCAAAAAGGGCTCTGTGT 1495
13 1540 AGCACGTTGGGAGCCGGGAGAGAACACCAAGTGGCGCTCATGCTGCAGGACGGCAAAAAGGGCTCCGTTGT 1618
14 1558 AGCACGTTGGGAGCCGGGAGAGAACATAGGTGGCGCTCATGCTGCAGGACGGCAAAAAGGGCTCCGTTGT 1627
16 1584 AGCACCTTGGGAGCCGGAAGAAAGAACACCAAGTGGCGCTCATGCTGCAGG - - - GCAAC[AG]AGGCTCTGTGT 1830
18 1549 TCAACCTTGGGAAACCGGAAGAGGAGCATTCAAGTGGCGCTTATGCTTCAAGATGGCAAAAAGGGATCCGTTGT 1618
19 1540 AGCACCTTGGGAGTCCGGAAGAAAGAACACCAAGTGGCGCTCATGCTGCAGG - - - TCAAAAAGGGCTCTGTGT 1608
2 1540 AGTACTTGGGTCCTCAAAAGAAAGAACATCAAGTGGCACTCATGCTGCAGG - - - TCAAAAAGGGCTAGCGTTGT 1615
20 1579 AGCACCTTGGGAGCCGGAAGAAAGAACACCAAGTGGCGCTCATGCTGCAGG - - - TCAAAAAGGGCTCTGTGT 1645
22 1549 AGCACGTTGGGAGCCGGGAGAGAACACCAAGTGGCGCTCATGCTGCAGGACGGCAAAAAGGGCTCCGTTGT 1618
23 1548 AGCACGTTGGGAGCCGGGAGAGAACACCAAGTGGCGCTCATGCTGCAGGACGGCAAAAAGGGCTCCGTTGT 1615
25 1555 GGCACATGGGAAATCTGGAAGAGAGTACCAAGTGGCACTCATGCTGCAGG - - - GCAAAAAGGGCTCCGTTGT 1621
26 1548 AGCACCTTGGGAGCCGGAAGAAAGAACACCAAGTGGCGCTTATGCTTCAAGATGGTAAAAAGGGATCACTGTGT 1615
27 1584 AGTACTTGGGAGCCGGAAGAAAGAACACCAAGTGGCGCTCATGCTGCAGG - - - GCAAC[AG]AGGCTCTGTGT 1630
29 1581 TCCACGTTGGGTGCCGAAGAAAGAGCATTCAAGTGGCGCTTATGCTTCAAGG - - - TCAAAAAGGGCAAGTGTAT 1627
32 1581 AGCACCTTGGGTACC[GA]AGAAAGAACACCAAGTGGCACTCATGCTGCAGG - - - GCAAAAAGGGCTCTGTGT 1627
33 1501 TCCACCTTGGGAAACCCGGAGGAGAACACCAAGTGGCGCTTATGCTTCAAGGATGGAAAGAAAGGATCTGTGT 1570
34 1525 AGCACCTTGGGAAACCCGGTCTGTAACACCAAGTGGCGCTCATGCTGCAGGACGGCAAAAAGGGCTCTGTGT 1594
35 1581 AGCACCTTGGGTACC[GA]AGAAAGAACACCAAGTGGCACTCATGCTGCAGG - - - GCAAAAAGGGCTCTGTGT 1627
36 1587 AGCACCTTGGGTGCCGAAGAAAGAACACCAAGTGGCGCTCATGCTGCAGG - - - TCAAAAAGGGCTCTGTGT 1633
40 1549 AGCACCTTGGGTACC[GA]AGAAAGAACACCAAGTGGCACTCATGCTGCAGG - - - GCAAAAAGGGCTCTGTGT 1615
41 1548 AGCACGTTGGGAGCCGGGAGAGAACACCAAGTGGCGCTCATGCTGCAGGACGGCAAAAAGGGCTCCGTTGT 1618
43 1510 AGCACGTTGGGAGCCGGGAGAGAACACCAAGTGGCGCTCATGCTGCAGGACGGCAAAAAGGGCTCCGTTGT 1570
44 1584 AGCACCTTGGGTGCCGAAGAAAGAACACCAAGTGGCGCTCATGCTGCAGG - - - TCAAAAAGGGCTCTGTGT 1630
45 1584 AGCACCTTGGGTGCCGAAGAAAGAACACCAAGTGGCGCTCATGCTGCAGG - - - TCAAAAAGGGCTCTGTGT 1630
48 1495 AGCACGTTGGGAGCCGGGAGAGAACACCAAGTGGCGCTCATGCTGCAGGACGGCAAAAAGGGCTCCGTTGT 1584
50 1540 AGCACGTTGGGAGCCGGGAGAGAACACCAAGTGGCGCTCATGCTGCAGGACGGCAAAAAGGGCTCCGTTGT 1609
52 1581 AGCACCTTGGGTACC[GA]AGAAAGAACACCAAGTGGCACTCATGCTGCAGG - - - GCAAAAAGGGCTCTGTGT 1627
54 1399 AGCACCTTGGGTACC[GA]AGAAAGAACACCAAGTGGCACTCATGCTGCAGG - - - GCAAAAAGGGCTCTGTGT 1485
57 1540 AGCACGTTGGGAGCCGGGAGAGAACACCAAGTGGCGCTCATGCTGCAGGACGGCAAAAAGGGCTCCGTTGT 1618

10 1616 ACGTGGATGGTGTGATTGTGGGAGCCCGGAGATGATACCAACTTGAGAGGAGGGGTTTGAATCC 1685
12 1408 ACATTTGATGGTCACTGATGGGGGAGAGGAGAGATGATACCAAGGTTGAGAGGAGCA - - - CTTGAGTCTGT 1582
13 1619 ACGTGGATGGTGTGATTGTGGGAGCCCGGAGATGATACCAACTTGAGAGGAGGGGTTTGAATCCC 1688
14 1628 ACGTGGATGGTGTGATTGTGGGAGCCCGGAGATGATACCAACTTGAGAGGAGGGGTTTGAATCCC 1697
16 1631 ACATTTGATGGTAATTCGAGGGGAGGAAAGTGGCGTAAAGGGTGAAGGCCCA - - - CTTGAGTCTGT 1697
18 1619 ACGTGGATGGTGTGATTGTGGGATCTCCGGAAATGATTCATATTAGAGACTCAGGGATCGAAATCCC 1688
19 1607 ACATTTGATGGTAATTCGAGGGGAGGAAAGTGGCGTAAAGGGTGAAGGCCCA - - - CTTGAGTCTGT 1673
2 1616 ACATTTGATGGTAACCTGTTAGGGGAGAGAGGAGTTCGCGTAAAGGGGAGGGGCC - - - CTCGAGCTTT 1682
20 1640 ACATTTGATGGTAATTCGAGGGGAGGAAAGTGGCGTAAAGGGTGAAGGCCCA - - - CTTGAGCTTT 1712
22 1619 ACGTGGATGGTGTGATTGTGGGAGCCCGGAGATGATACCAACTTGAGAGGAGGGGTTTGAATCCC 1688
23 1616 ACGTGGATGGTGTGATTGTGGGAGCCCGGAGATGATACCAACTTGAGAGGAGGGGTTTGAATCCC 1685
25 1622 ACATTTGATGGTCAAGAGCTTGGGAGGAGAAACCAATGCTGACTGGTGAAGGTTCA - - - CTGAGATCTGT 1688
26 1616 ACGTGGATGGTAACTGTTAGGGGAGAGAGGAGTTCGCGTAAAGGGGAGGGGCC - - - CTCGAGCTTT 1685
27 1631 ACATTTGATGGTAATTCGAGGGGAGGAAAGTGGCGTAAAGGGTGAAGGCCCA - - - CTTGAGCTTT 1697
29 1628 ATATCGATGGCAACTCACTGGGGAAGAGGAGTGGCGTAAAGGGGAGAGGCCCA - - - CTCGAGCTTT 1694
32 1628 ACATTTGATGGTAAAGTCGTTGGGGAAGAGAGAGTGGCGTAAAGGGTGAAGGCCCA - - - CTTGAGCTTT 1694
33 1571 ATGTTGAGAGGGGGTGTGTTGCGGAGGCTTGAAGTATCCCACTTGAAGACAGAGGGTTTGAATCCC 1640
34 1595 ACGTGGAGCGGCGTTATCGTGGTTTCACTGAGATGATCCACTGAGAGAGGAGGTTTGAATCCC 1684
35 1628 ACATTTGATGGTAAGTCGTTGGGGAAGAGAGTGGCGTAAAGGGTGAAGGCCCA - - - CTGAG - TGT 1693
36 1634 ACATTTGATGGTAATTCGAGGGGAGGAAAGTGGCGTAAAGGGTGAAGGCCCA - - - CTTGAGCTTT 1700
40 1616 ACATTTGATGGTAAGTCGTTGGGGAAGAGAGAGTGGCGTAAAGGGTGAAGGCCCA - - - CTTGAGCTTT 1682
41 1619 ACGTGGATGGTGTGATTGTGGGAGCCCGGAGATGATACCAACTTGAGAGGAGGGGTTTGAATCCC 1688
43 1580 ACGTGGATGGTGTGATTGTGGGAGCCCGGAGATGATACCAACTTGAGAGGAGGGGTTTGAATCCC 1649
44 1631 ACATTTGATGGTAATTCGAGGGGAGGAAAGTGGCGTAAAGGGTGAAGGCCCA - - - CTTGAGCTTT 1697
45 1631 ACATTTGATGGTAATTCGAGGGGAGGAAAGTGGCGTAAAGGGTGAAGGCCCA - - - CTTGAGCTTT 1697
48 1585 ACGTGGATGGTGTGATTGTGGGAGCCCGGAGATGATACCAACTTGAGAGGAGGGGTTTGAATCCC 1634
50 1610 ACGTGGATGGTGTGATTGTGGGAGCCCGGAGATGATACCAACTTGAGAGGAGGGGTTTGAATCCC 1679
52 1628 ACATTTGATGGTAAGTCGTTGGGGAAGAGAGAGTGGCGTAAAGGGTGAAGGCCCA - - - CTTGAGCTTT 1694
54 1488 ACGTGGATGGTAAGTCGTTGGGGAAGAGAGAGTGGCGTAAAGGGTGAAGGCCCA - - - CTTGAGCTTT 1532
57 1619 ACGTGGATGGTGTGATTGTGGGAGCCCGGAGATGATACCAACTTGAGAGGAGGGGTTTGAATCCC 1688

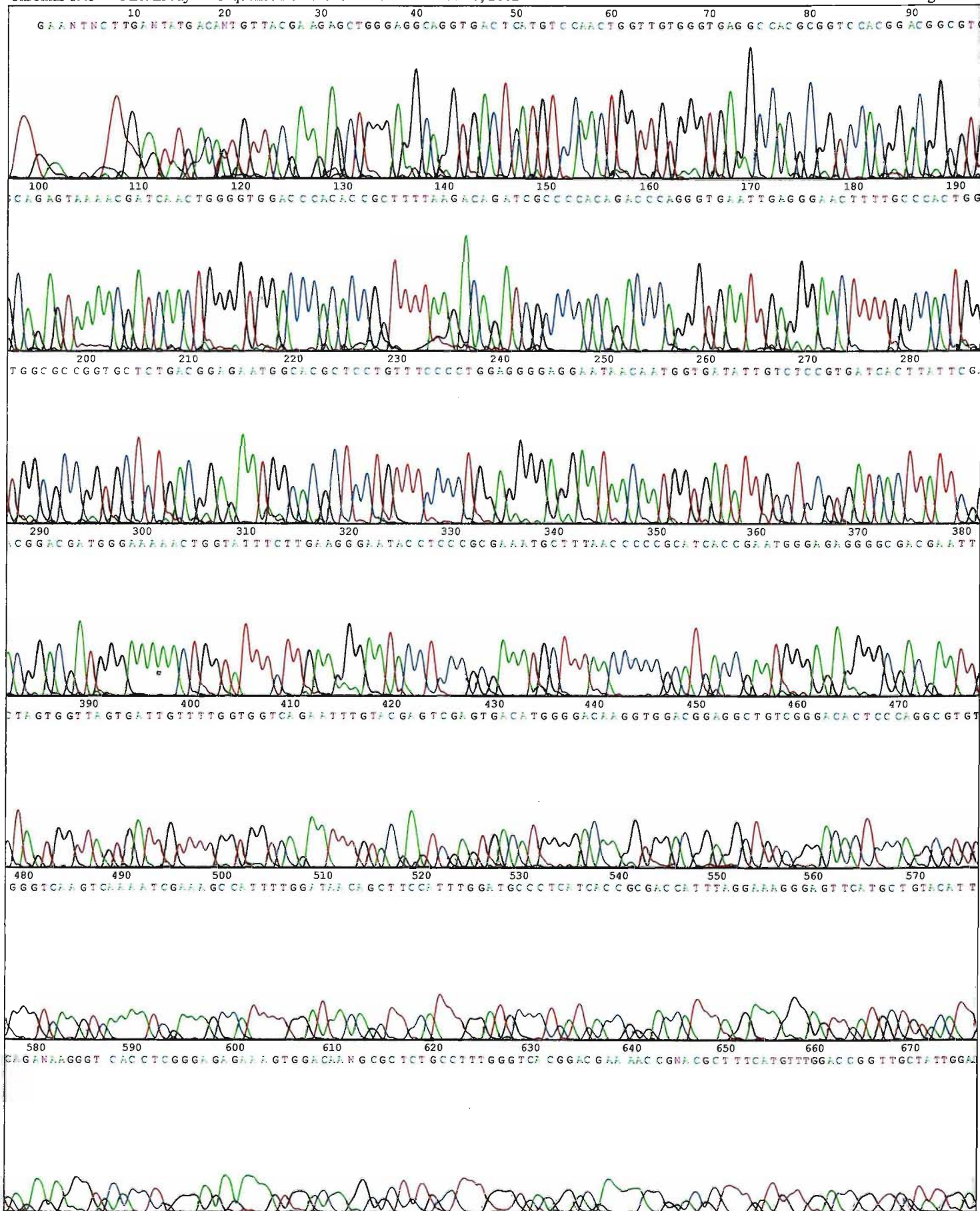
10 1688 ACAATTTT - ACGTTGGGGGGGAGCGAGGAGACAT - - - - CAACAGCAGCGGTGACGGTGAAGAAC 1743
12 1583 ACACCTTT - GCTTTGGCGCGTGC - - - AAAAATGC - - - - CAACCTTCTGTGACGGTGAAGAAC 1617
13 1688 ACAATTTT - ACGTTGGGGGGGAGCGAGGAGACAT - - - - CAACAGCAGCGGTGACGGTGAAGAAC 1740
14 1688 ACAATTTT - ACGTTGGGGGGGAGCGAGGAGACAT - - - - CAACAGCAGCGGTGACGGTGAAGAAC 1755

16 1698 ACGCTTTT-GCTTTGGCCGCTGCGGTAGGATGCCGGCCAGAAAACTAATGTGAAGGTGAAAAAC 1761
18 1699 GCAATCTT-ACGTAGGAGGGGATGAGGDTGACAT-----AAATTCCTCCGTAACTATGACCAAC 1746
19 1674 ATACTTTT-GCTTTGGTGCCTGCGGTGAGGATGCCGGCCAGAAAACTAATGTGAAGGTGAAAAAC 1737
2 1683 ATACTTTT-GTTTCGGTGCCTGCGGCGAAGATGCTGGCCAGAAACAAACGTAAAGTAAAAAAT 1746
20 1713 ATACTTTT-GCTTTGGTGCCTGCGGTGAGGATGCCGGCCAGAAAACTAATGTGAAGGTGAAAAAC 1776
22 1699 ACAATTCT-ACGTTGGGGGCGACGAGGGAGACAT-----CAACAGCAGCGGTGACGATGACGAAC 1746
23 1680 ACAATTCT-ACGTTGGGGGCGACGAGGGAGACAT-----CAACAGCAGCGGTGACGATGACGAAC 1743
25 1690 ACACCTCT-GCTTCGGAGCATGT---AAAATGC-----ATAATTCCTCCGTACAGTAAAAAC 1743
26 1698 ACAATTTT-ATGTTGGCCGAGATGAGGGCAGCAT-----AAATTCCTCTGTACTATGACAAAT 1743
27 1698 ACGCTTTT-GCTTTGGCCGCTGCGGTAGGATGCCGGCCAGAAAACTAATGTGAAGGTGAAAAAC 1761
29 1695 ATACTTCT-GCTTTGGTGCCTGCGGAGGAGGACGCCGGCCAGAAAGACAAACGTAAAGGTAAAAAT 1758
32 1695 ACGCTTTT-GCTTTGGCCGCTGCGAGTGAAGATGCCGGCCAGAAAACTAAGGTGACGGTGAAGAAC 1758
33 1641 GCAATCTT-ATGTAGGCGGTGACGAAGGCATAT-----CAACTCTAGTGTAACTATGACTAAT 1698
34 1665 RCAGT---ACGTCGGAGGAGACGAGGGGATAT-----TAACTCTTCTGTAACTATGACCAAT 1719
35 1694 ACGCTNTT-GCTTTGGCCGCTGCGAGTGAAGATGCCGGCCAGAAAACTAAGGTGACGGTGAAGAAC 1758
36 1701 ATACTTTT-GCTTTGGTGCCTGCGGTGAGGATGCCGGCCAGAAAACTAATGTGAAGGTGAAAAAC 1764
40 1683 ACGCTTTT-GCTTTGGCCGCTGCGAGTGAAGATGCCGGCCAGAAAACTAAGGTGACGGTGAAGAAC 1746
41 1680 ACAATTCT-ACGTTGGGGGCGACGAGGGAGACAT-----CAACAGCAGCGGTGACGATGACGAAC 1746
43 1650 ACAATTCT-ACGTTGGGGGCGACGAGGGAGACAT-----CAACAGCAGCGGTGACGATGACGAAC 1707
44 1698 ATACTTTT-GCTTTGGTGCCTGCGGTGAGGATGCCGGCCAGAAAACTAATGTGAAGGTGAAAAAC 1761
45 1698 ATACTTCT-GCTTTGGTGCCTGCGGAGGAGGACGCCGGCCAGAAAACTAATGTGAAGGTGAAAAAC 1761
48 1635 ACAATTCT-ACGTTGGGGGCGACGAGGGAGACAT-----CAACAGCAGCGGTGACGATGACGAAC 1692
50 1680 ACAATTCT-ACGTTGGGGGCGACGAGGGAGACAT-----CAACAGCAGCGGTGACGATGACGAAC 1737
52 1695 ACGCTTTT-GCTTTGGCCGCTGCGAGTGAAGATGCCGGCCAGAAAACTAAGGTGACGGTGAAGAAC 1758
54 1533 ACGCTTTT-GCTTTGGCCGCTGCGAGTGAAGATGCCGGCCAGAAAACTAAGGTGACGGTGAAGAAC 1596
57 1690 ACAATCTT-ACGTTGGGGGCGACGAGGGAGACAT-----CAACAGCAGCGGTGACGATGACGAAC 1746

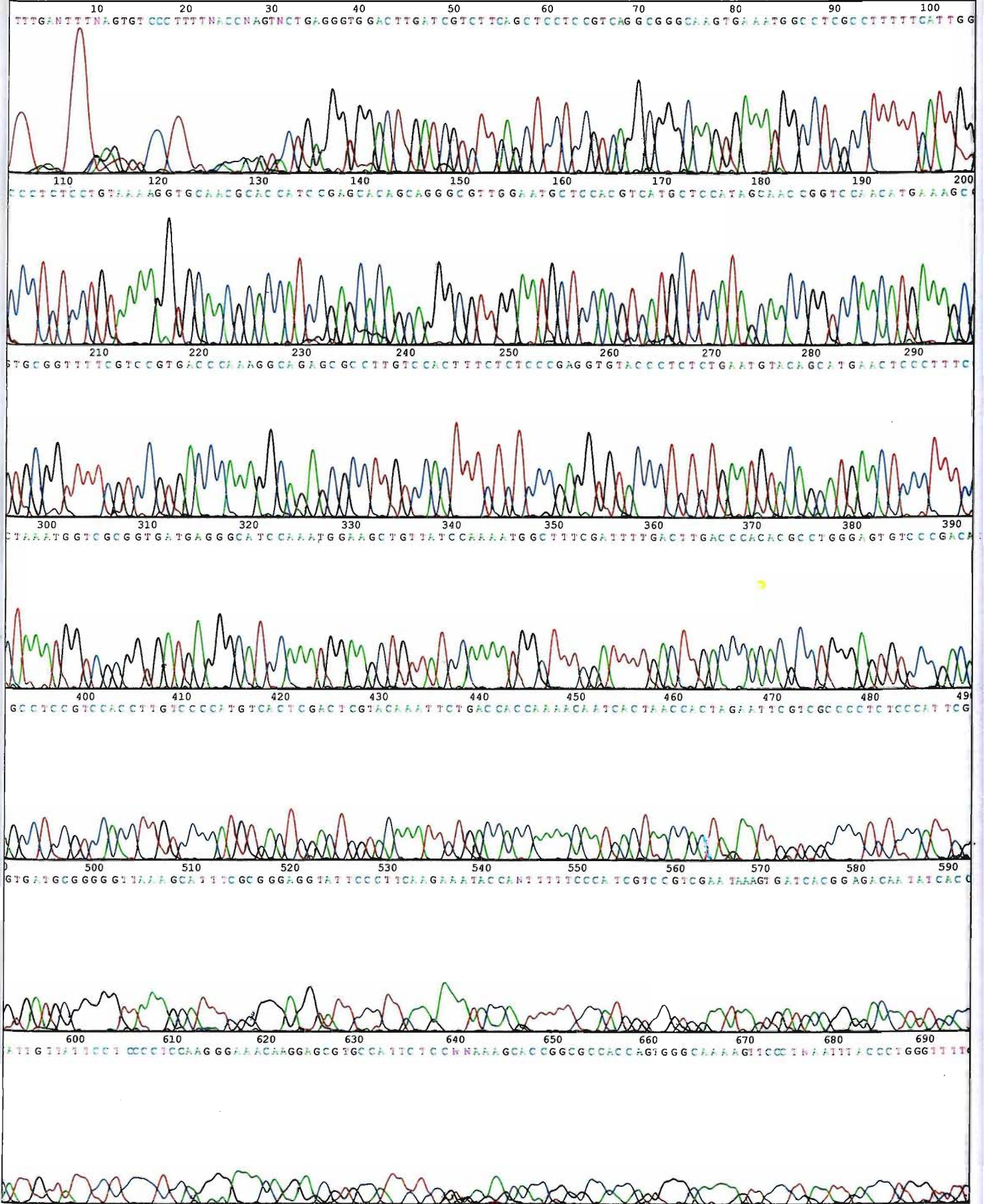
Anexo 2. Alguns cromatogramas obtidos. Legenda dos primers: (A) T7 forward; (B) Tc300-5B; (C) cDNA-3A.



(A)



(B)



(C)

54 125 LVVGVVREP TTS SEP - TE - - R I K W G Q I R S L L - - - - N E S T I A A H E G K W T - G F L A A G G S G V L M E - D G T L M F P 184
57 130 LVVGVGATRS TDR - V - QR - - T L N W G G P P L L - - - - R Q I A P Q - T G G L R - E L X P T G G A G A L T E - N G T L L F P 187

10 187 WRGGITNGD IVSVITYS TDDGKNMYFLKGI PPAKCFHHRMGE - GRRILVMSDFGGGNLYESSDMGTRW 254
12 149 LTAKEAEEDVCSMI IYSTONGEKMLSNGTSPERC GAPRITIEWE GSLLMIND - GENGORVYESSROMGTTW 217
13 189 LEGRNNNGDIVSVITYS TDDGKNMYFLKGI PPAKCFNPRITIEWEXGRRILVMSDFGGGNLYESSDMGTRW 258
14 190 LEGRNNNGDIVSVITYS TDDGKNMYFLKGI PPAKCFNPRITIEWERGRILVMSDFGGGNLYESSDMGTRW 259
16 192 LMATNAAKDDYSMI IYSTONGANWMLSNGTSPANC TDPRI TEWOGSLLMIND - CKYSQRYESSROMGTTW 280
18 185 WRGGTPMVLSPWSTYS TDDGKNMYFLKGI PPAKCFNPRITIEWEGRRILVMSDFGGGNLYESSDMGTRW 254
19 192 LMATNAAKDDYSMI IYSTONGSTWALSTGVS PANC TDPRI TEWOGSLLMIND - CKY - - - - ESROMGTTW 255
2 188 LMATNAAKDDYSMI IYSTONGSTWALSTGVS PANC TDPRI TEWOGSLLMIND - CKYSQRYESSROMGTTW 258
20 198 LMATNAAKDDYSMI IYSTONGSTWALSTGVS PANC TRPRI TEWOGSLLMIND - CKYSQRYESSROMGTTW 286
22 188 LEGRNNNGDIVSVITYS TDDGKNMYFLKGI PPAKCFNPRITIEWEXGRRILVMSDFGGGNLYESSDMGTRW 257
23 189 LEGRNNNGGXCLXXITYS TDDGKNMYFLKGI PPAKCF - PRITIEWEGARILVMSDFGGGNLYESSDMGTRW 250
25 188 LMARNAAKDDYSMI IYSKDDGSTWALSTGVS PANC TDPRI TEWOGSLLMIND - CKYSQRYESSROMGTTW 258
26 189 FGG - EEQWGTASGTICLMAGGKNMFXFLGPTLERCYHRTIGXGEGRRILVMSDFGGGNLYESSDMGTRW 257
27 192 LMATNAAKDDYSMI IYSTONGSTWALSTGVS PANC TDPRI TEWOGSLLMIND - CKYSQRYESSROMGTTW 280
29 192 LMETNAAKDDYSMI IYSTATGSTWALSTGVM SHANC TDPRI TEWOGSLLMIND - CKYSQRYESSROMGTTW 280
32 191 LMARNAAKDDYSMI IYSKDDGSTWALSTGVS PANC TDPRI TEWOGSLLMIND - CKYSQRYESSROMGTTW 259
33 178 MX - - - - NISPVITYS TDDGKNMYFLKGI PPAKCFNPRITIEWERGDXVLARNSPGCDLHKDQMGPF 241
34 187 PG - VRNSGTVSETIFLMVGGRNXFLEGIHRGRXXSXILAWEKARILVMSDFGGGNLYESSDMGTRW 254
35 192 LMARNAAKDDYSMI IYSKDDGSTWALSTGVS PANC TDPRI TEWOGSLLMIND - CKYSQRYESSROMGTTW 280
36 192 LMATNAAKDDYSMI IYSTONGSTWALSTGVS PANC TDPRI TEWOGSLLMIND - CKYSQRYESSROMGTTW 280
40 191 LMARNAAKDDYSMI IYSKDDGSTWALSTGVS PANC T - PRITIEWOGSLLMIND - CXXIXRMYESSROMGTTW 257
41 188 LEGRNNNGDIVSVITYS TDDGKNMYFLKGI PPAKCFNPRITIEWEXGRRILVMSDFGGGNLYESSDMGTRW 257
43 189 LEGRNNNGDIVSVITYS TDDGKNMYFLKGI PPAKCFNPRITIEWEXGRRILVMSDFGGGNLYESSDMGTRW 258
44 190 LMATNAAKDDYSMI IYSTONGSTWALSTGVS PANC TDPRI TEWOGSLLMIND - CKYSQRYESSROMGTTW 258
45 190 LMATNAAKDDYSMI IYSTONGSTWALSTGVS PANC TDPRI TEWOGSLLMIND - CKYSQRYESSROMGTTW 258
48 170 LEGRNNNGDIVSVITYS TDDGKNMYFLKGI PPAKCFNPRITIEWEXGRRILVMSDFGGGNLYESSDMGTRW 239
50 188 LEGRNNNG - IVSVITYS TDDGKNMYFLKGI PPAKCF - PRITIEWEGARILVMSDFGGGNLYESSDMGTRW 254
52 191 LMARNAAKDDYSMI IYSKDDGSTWALSTGVS PANC TDPRI TEWOGSLLMIND - CKYSQRYESSROMGTTW 259
54 185 LMARNAAKDDYSMI IYSKDDGSTWALSTGVS PANC TDPRI TEWOGSLLMIND - CKYSQRYESSROMGTTW 241
57 188 LEGRNNNGDIVSVITYS TDDGKNMYFLKGI PPAKCFNPRITIEWEXGRRILVMSDFGGGNLYESSDMGTRW 257

10 255 TEAMGTLPGVWVKSKSKAILDN - SFHLDALITATI - EREFMLYI QRGYTS - EKVDKALCLWVTDENRRTL 321
12 218 TEAIGTLPGVWAK - LQPEVPO - MKLRVEALITATI EGKVMFLYI QRGYASR - EKRRATALLY LWVTDENRNSF 284
13 259 TEAMGTLPGVWVKSKSKAILDN - SFHLDALITATI - EREFMLYI QREYTS - EKVDKALCLWVTDENRRTL 325
14 280 TETVGTLPVWVKSKSKAILDN - SFHLDALITATI - WEREFMLYI QRGYTS - EKVDKALCLWVTDENRRTL 327
16 281 TEAIGTLPGVWVNSR - LYFWDL - LRVEALITATI EGKVMFLYI QRGDFLG - EKSERALYLWVTDENRNSF 327
18 255 TEAMGTLPGVWVKSKSKAILDN - SFHLDALITATI - EREFMLYI QRGYTS - EKVDKALCLWVTDENRRTL 321
19 258 TEAIGTLPGVWTKSR - LLSWDL - LRVEALITATI EGKVMFLYI QRGDFS - EKSERALYLWVTDENRNSF 322
2 257 TEAIGTLPGVWTKSR - LLSWDL - LRVEALITATI EGKVMFLYI QRGDFS - EKSERALYLWVTDENRNSF 323
20 207 TEAIGTLPGVWTKSR - LLSWDL - LRVEALITATI EGKVMFLYI QRGDFS - EKSERALYLWVTDENRNSF 333
22 258 TEAMGTLPGVWVKSKSKAILDN - SFHLDALITATI - EREFMLYI QRGYTS - EKVDKALCLWVTDENRRTL 324
23 257 TEAMGTLPGVWVKSKSKAILDN - SFHLDALITATI - EREFMLYI QRGYTS - EKVDKALCLWVTDENRRTL 323
25 257 TGAAGTLPGVWVNSXALXFGX - L - LRVEALITATI EGKVMFLYI QRGDFLG - EKSERALYLWVTDENRNSF 324
26 258 TEAMGTLPGVWVKSKSKAILDN - SFHLDALITATI - EXXNSXLYI QRGYVX - EKVDKALCLWVTDENRRTL 326
27 281 TEAIGTLPGVWVNSR - LYFWDL - LRVEALITATI EGKVMFLYI QRGDFLG - EKSERALYLWVTDENRNSF 327
29 281 TEAIGTLPGVWTKSR - LLSWDL - LRVEALITATI EGKVMFLYI QRGDFS - EKSERALYLWVTDENRNSF 326
32 280 TGAAGTLPGVWVNSDG - TLGPEXG - LRVEALITATI EGKVMFLYI QRGDFLG - EKSERALYLWVTDENRNSF 320
33 242 TPKRSKRSQSSRLRVPESQTILOY - SLHLDALMSATFQXRVELLHRI - GHXHG - NVSGLTLLSFWVSKDKDRAF 308
34 255 TEAMGTLPGV - VKSKSKAILDN - SFHLDALITATI - EREFMLYI QRGYTS - EKVDKALCLWVTDENRRTL 320
35 201 TGAAGTLPGVSKTHDXLLGPE - - LAVEALITATI EGKVMFLYI QRGDFLG - EKSERALYLWVTDENRNSF 327
36 281 TEAIGTLPGVWTKSR - LLSWDL - LRVEALITATI EGKVMFLYI QRGDFS - EKSERALYLWVTDENRNSF 327
40 258 TGAAGTLPGVWVNSXCTFGX - L - LRVEALITATI - EREFMLYI QRGDFLG - EKSERALYLWVTDENRNSF 324
41 258 TEAMGTLPGVWVKSKSKAILDN - SFHLDALITATI - EREFMLYI QRGYTS - EKVDKALCLWVTDENRRTL 324
43 259 TEAMGTLPGVWVKSKSKAXLG - - - - - - - - - - K - GVMFLYI QRGYTS - EKVDKALCLWVTDENRRTL 311
44 259 TEAIGTLPGVWTKSR - LLSWDL - LRVEALITATI EGKVMFLYI QRGDFS - EKSERALYLWVTDENRNSF 325
45 259 TEAIGTLPGVWTKSR - LLSWDL - LRVEALITATI EGKVMFLYI QRGDFS - EKSERALYLWVTDENRNSF 325
48 240 TEAMGTLPGVWVKSKSKAILDN - SFHLDALITATI - EREFMLYI QRGYTS - EKVDKALCLWVTDENRRTL 305
50 255 TEAMGTLPGVWVKSKSKAILDN - SFHLDALITATI - EREFMLYI QRGYTS - EKVDKALCLWVTDENRRTL 321
52 280 TGAAGTLPGVWVNSXCTFGX - L - LRVEALITATI EGKVMFLYI QRGDFLG - EKSERALYLWVTDENRNSF 327
54 241 TGAAGTLPGVWVNSXCTFGX - L - LRVEALITATI EGKVMFLYI QRGDFLG - EKSERALYLWVTDENRNSF 285
57 258 TEAMGTLPGVWVKSKSKAILDN - SFHLDALITATI - EREFMLYI QRGYTS - EKVDKALCLWVTDENRRTL 324

10 322 SCWTCGYGAWI - TW - - S IPTPCCC GAGALHLLQERANEKGEAISLARL - TEELKTIKSTLST - - - - WAKL 383
12 285 SLGQ - - - - VGMDNANWVFTSSLLY - DGNLHLLQRRDQKGEAISLSRM - TEELSJKSVLST - - - - WAKK 348
13 320 HVGP - - - - VAMEHDV - - EHSNALLC - DGALHLLQERANEKGEAISLARL - TEELKTIKSTLST - - - - WAKL 385
14 328 HVGP - - - - DSVDYD - - - - EHTNPILL - DGALHLLQERANEKGEAISLARL - TEELKTIKSTLST - - - - WSK 387
16 328 YFGP - - - - VGMDNANWVFTSSLLY - DGNLHLLQRRDQKGEAISLSRL - TEELSJKSVLST - - - - WAKK 389
18 322 SCWTCGYGAWRGA - - - - EHSNALLC - DGALHLLQERANEKGEAISLARL - TEELKTIKSTLST - - - - WAKL 384
19 323 YFGP - - - - VGMDNANWVFTSSLLY - DGNLHLLQRRDQKGEAISLSRL - TEELSJKSVLST - - - - WAKK 382
2 324 YFGP - - - - VGMDNANWVFTSSLLY - DGNLHLLQRRDQKGEAISLSRL - TEELSJKSVLST - - - - CRKK 385
20 334 YFGP - - - - VGMDNANWVFTSSLLY - DGNLHLLQRRDQKGEAISLSRL - TEELSJKSVLST - - - - WSK 395
22 325 HVGP - - - - VAMEHDV - - EHSNALLC - DGALHLLQERANEKGEAISLARL - TEELKTIKSTLST - - - - WAKL 384
23 324 HVGP - - - - VAMEHDV - - EHSNALLC - DGALHLLQERANEKGEAISLARL - TEELKTIKSTLST - - - - WAKL 383
25 325 YFGP - - - - VGMDNANWVFTSSLLY - DGNLHLLQRRDQKGEAISLSRM - TEELSJKSVLST - - - - WSK 390
26 327 ALDR - - - - MLWRLE - - - - GVS - ETSX - LCGVPLSLERANGKQVVI - LVRL - MAAPKTRYI - LNT - - - - RRNG 385
27 328 YFGP - - - - VGMDNANWVFTSSLLY - DGNLHLLQRRDQKGEAISLSRL - TEELSJKSVLST - - - - WAKK 389
29 327 YFGP - - - - VGMDNANWVFTSSLLY - DGNLHLLQRRDQKGEAISLSRL - TEELSJKSVLST - - - - LVAG 388
32 328 YFGP - - - - VGMDNANWVFTSSLLY - DGNLHLLQRRDQKGEAISLSRL - TEELSJKSVLST - - - - WAKK 389
33 309 HVGP - - - - VAMEHDV - - EHSNALLC - DGALHLLQERANEKGEAISLARL - TEELKTIKSTLST - - - - WAKL 388
34 321 HVDR - - - - AMAR - - - - G - VVLFSD - PQFVWALPHSRGRORSVEOTLRAL - DMGRKTRFRTHNT - - - - RGDR 377
35 328 YFGP - - - - VGMDNANWVFTSSLLY - DGNLHLLQRRDQKGEAISLSRL - TEELSJKSVLST - - - - WSK 389
36 328 YFGP - - - - VGMDNANWVFTSSLLY - DGNLHLLQRRDQKGEAISLSRL - TEELSJKSVLST - - - - WSK 389
40 325 YFGP - - - - VGMDNANWVFTSSLLY - DGNLHLLQRRDQKGEAISLSRL - TEELSJKSVLST - - - - WAKK 385
41 325 HVGP - - - - VAMEHDV - - EHSNALLC - DGALHLLQERANEKGEAISLARL - TEELKTIKSTLST - - - - WAKL 384
43 312 HVGP - - - - VAMEHDV - - EHSNALLC - DGALHLLQERANEKGEAISLARL - TEELKTIKSTLST - - - - WAKL 371
44 328 YFGP - - - - VGMDNANWVFTSSLLY - DGNLHLLQRRDQKGEAISLSRL - TEELSJKSVLST - - - - WSK 387
45 326 YFGP - - - - VGMDNANWVFTSSLLY - DGNLHLLQRRDQKGEAISLSRL - TEELSJKSVLST - - - - WSK 387

48 307 HVGP. --MAMEHDV. -EHSNALLCSOGALHLLQDERANEKGGAI S LARL TEELKTI KSTLST. --MARL 386
50 322 HVGP. --MAMEHDV. -EHSNALLCSOGALHLLQDERANEKGGAI S LARL TEELKTI KSTLST. --MARL 381
52 328 YFGP. --VGMENAVKEELASALLYSDGNLHLLQQRDNGEGSOLSLXL AEEMGTINFLLET. --MAKK 389
54 280 YFGP. --VGMENAVKEELASALLYSDGNLHLLQQRDNGEGSOLSLXL TEEMSTINFLKTI. --MAKK 327
57 325 HVGP. --MAMEHDV. -EHSNALLCSOGALHLLQDERANEKGGAI S LARL TEELKTI KSTLST. --MARL 384

10 384 DAF FSKS IPTTGLV. ILS DTS SGGD TWIDEYRCV NATVTKAAKVKNG. --FKFTAPGSRAMMSMSRWGP 440
12 347 DAF FSKS IPTTAGLVAVLS DAASDD. TWIDDYLC LNATV TNAKVKKDG. --FOLTEPNSGVLMFVNIPDD 412
13 385 DAF FSKS IPTTGLV. ILS DTS SGGD TWIDKSRFVTQTGPKAAEVKNG. --FKFTXAGSXAMXMSRWGL 461
14 388 DAF FSKS IPACLVAVLSNAXXSNGTWIDEYLC LNAAVTNATKVKDG. --FOLTESNSGVLMFVNTRDD 464
16 390 DAF FSKS IPTTAGLVAVLS DAASNG. TWIDEYLC LNAAVTNATKVKDG. --FHLTELNSGVLMFVNTRDD 465
18 385 DAF FSKS IPTTGLVAVLS DTS SGGD TWIDEYRCV NATVTKGSEGGK. --MLVHGAWVRAMMSMSRWGP 460
19 383 DAF FSKS IPTTAGLVAVLS PDGASNG. TWIDEYLC LNAAVTNATKVKDG. --FOLTESNSGVLMFVNTRDD 447
2 380 GRLYLQLS IPTTEVAVLSHAASNG. TWIDEYLC LNAAVTNATKVKDG. --FOLTESNSGVLMFVNTRDD 461
20 386 DAF FSKS IPTTAGLVAVLS DAASNG. TWIDEYLC LNAAVTNATKVKDG. --FOLTESNSGVLMFVNTRDD 401
22 385 DAF FSKS IPTTGLVAVLS DTS SGGD TWIDEYRCV NATVTKAAKVKNG. --FKFTAPGSRAMMSMSRWGP 451
23 384 DAF FSKS IPTTGLVAVLS DTS SGGD TWIDEYRCV NATVTKXTKVKNG. --FKFTGAGSRAMMSMSRWGP 460
25 301 TFFPFAFS IPGEXLVPVLS DAASDD. TWIDDY. -CNAPVTNAKVKKDG. --FOLTEPNSGVLMFVNIPDD 464
26 385 ATTAASXTPMIGLVIDYDMSYDGAIT. -ACAAEECNDGGGSKGXASVVIAAPG. DAMMSMSRWGP 460
27 300 DAF FSKS IPFRVWVWQYCPMPASNG. TWIDEYLC LNAAVTNATKVKDG. --FOLTESNSGVLMFVNTRDD 465
29 389 RLHLQFIYIYPRRVWVWQYCPMPASNG. TWIDEYLC LNAAVTNATKVKDG. --FOLTESNSGVLMFVNTRDD 464
32 390 DAF FSKS IPTTAGLVAVLS DTS SGGD TWIDEYRCV NATVTKAAKVKNG. --FOLTEPNSRVSMMSNRVN 465
33 389 DAF FSKS IPTTAGLVAVLS DTS SGGD TWIDEYRCV NATVTKAAKVKNG. --FKFTAPGSRAMMSMSRWGP 435
34 378 QASLAGS XHMI GXFELSDTS SGGD TWIDEYRCV NATVTRAQDQSK. -AFKFGGLG-TXSVMIGVWVRL 443
35 390 DAF FSKS IPTTAGLVAVLS DTS SGGD TWIDEYRCV NATVTKAAKVKNG. --FOLTESNSRVSMMSNRVN 465
36 390 DAF FSKS IPTTAGLVAVLS DTS SGGD TWIDEYRCV NATVTKAAKVKNG. --FOLTESNSGVLMFVNTRDD 465
40 385 DAF FSKS IPTTAGLVAVLS DTS SGGD TWIDEYRCV NATVTKAAKVKNG. --FOLTEPNSRVSMMSNRVN 461
41 385 DAF FSKS IPTTAGLVAVLS DTS SGGD TWIDEYRCV NATVTKAAKVKNG. --FKFTAPGSRAMMSMNQVGT 461
43 372 DAF FSKS IPTTAGLVAVLS DTS SGGD TWIDEYRCV NATVTKAAKVKNG. --FKFTAPGSRAMMSMSRWGP 438
44 388 DAF FSKS IPFRVWVWQYCPMPASNG. TWIDDY. -NAPVTNAKVKKDG. --FKLTESNSRVSMMSNRVN 463
45 388 DAF FSKS IPFRVWVWQYCPMPASNG. TWIDEYLC LNAAVTNATKVKDG. --FOLTESNSGVLMFVNTRDD 463
48 387 DAF FSKS IPTTAGLVAVLS DTS SGGD TWIDEYRCV NATVTKAAKVKNG. --FKFTAPGSRAMMSMSRWGP 433
50 382 DAF FSKS IPTTAGLVAVLS DTS SGGD TWIDEYRCV NATVTKAAKVKNG. --FKFTAPGSRAMMSINRWGP 448
52 390 DAF FSKS IPTTAGLVAVLS DTS SGGD TWIDEYRCV NATVTKAAKVKNG. --FOLTEPNSRVSMMSNRVN 465
54 328 DAF FSKS IPTTAGLVAVLS DTS SGGD TWIDEYRCV NATVTKAAKVKNG. --FOLTEPNSRVSMMSNRVN 393
57 385 DAF FSKS IPTTAGLVAVLS DTS SGGD TWIDEYRCV NATVTKAAKVKNG. --FKFTGPGSRAMMSISRWGP 461

10 460 NKQYGFVNHRRFTLVATVTHQVFKGSTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 519
12 413 NVRHISLHNFLLASVTEVAPSEKTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 481
13 462 NKQYGFVNHRRFTLVATVTHQVFKGSTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 521
14 465 NVRHISLHNFLLASVTEVAPSEKTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 524
16 460 NVRHISLHNFLLASVTEVAPSEKTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 525
18 461 NKQYGFVNHRRFTLVATVTHQVFKGSTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 520
19 448 NVRHISLHNFLLASVTEVAPSEKTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 517
2 462 NVRHISLHNFLLASVTEVAPSEKTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 521
20 462 NVRHISLHNFLLASVTEVAPSEKTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 531
22 462 NKQYGFVNHRRFTLVATVTHQVFKGSTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 521
23 461 NKQYGFVNHRRFTLVATVTHQVFKGSTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 520
25 465 NVRHISLHNFLLASVTEVAPSEKTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 523
26 461 NKQYGFVNHRRFTLVATVTHQVFKGSTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 520
27 466 NVRHISLHNFLLASVTEVAPSEKTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 525
29 465 NVRHISLHNFLLASVTEVAPSEKTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 524
32 466 NVRHISLHNFLLASVTEVAPSEKTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 525
33 430 NKQYGFVNHRRFTLVATVTHQVFKGSTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 505
34 444 NKQYGFVNHRRFTLVATVTHQVFKGSTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 513
35 460 NVRHISLHNFLLASVTEVAPSEKTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 525
36 466 NVRHISLHNFLLASVTEVAPSEKTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 525
40 462 NVRHISLHNFLLASVTEVAPSEKTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 521
41 462 TVRGLGFVNHRRFTLVATVTHQVFKGSTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 521
43 439 NKQYGFVNHRRFTLVATVTHQVFKGSTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 508
44 464 NVRHISLHNFLLASVTEVAPSEKTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 523
45 464 NVRHISLHNFLLASVTEVAPSEKTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 523
48 434 NKQYGFVNHRRFTLVATVTHQVFKGSTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 503
50 449 NKQYGFVNHRRFTLVATVTHQVFKGSTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 518
52 466 NVRHISLHNFLLASVTEVAPSEKTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 525
54 394 NVRHISLHNFLLASVTEVAPSEKTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 483
57 462 NKQYGFVNHRRFTLVATVTHQVFKGSTPLLGAGLGDGAGKIIIGLSYS MNKKWEKVFNGTKTAPGSTWEP 521

10 520 GREHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 580
12 482 GKEHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 539
13 522 GREHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 582
14 525 GREHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 585
16 526 KKEHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 585
18 521 GREHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 580
19 518 KKEHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 577
2 522 KKEHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 582
20 532 KKEHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 592
22 522 GREHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 582
23 521 GREHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 581
25 524 GKEHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 581
26 521 GREHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 581
27 526 KKEHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 585
29 525 KKEHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 584
32 526 KKEHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 588
33 506 GREHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 566
34 514 GREHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 573
35 526 KKEHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 588
36 526 KKEHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 585
40 522 KKEHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 582
41 522 GREHQVALMLQDGGKGSVYVGGVIVGSP EMIPTLETQGF EIPQFYVGGDEGDIIN. -SVMTMN 582

```

43 509 GREHQAVALMLQDGKKGSVYMDGVI VGSPEMIP TLE TQGF EIPQFYVGGDEGDIN - SSVTMTN 509
44 524 KKEHQVALMLQ-VKKASVYIDGNSLGEEELP LKGEAP - LELLYFCFGACGEDAGQKTNNKVK - 583
45 524 KKEHQVALMLQ-VKKASVYIDGNSLGEEELP LKGEAP - LELLYFCFGACGEDAGQKTNNKVK - 583
4d 504 GREHQAVALMLQDGKKGSVYMDGVI VGSPEMIP TLE TQGF EIPQFYVGGDEGDIN - SSVTMTN 504
50 519 GREHQAVALMLQDGKKGSVYMDGVI VGSPEMIP TLE TQGF EIPQFYVGGDEGDIN - SSVTMTN 519
52 520 KKEHQVALMLQ-GKKASVHIDGKSLGEEEMP LKGEAP - LELLVRF CFGACSE DAGQKT KMTVK S 580
5d 484 KKEHQVALMLQ-GKKASVHV D GKSLGEEEMP LKGEAP - LELLVRF CFGACSE DAGQKT KMTVK - 523
57 522 GREHQAVALMLQDGKKGSVYMDGVI VGSPEMIP TLE TQGF EIPQFYVGGDEGDIN - SSVTMTN 582

```