The complete set of *Toxoplasma gondii* ribosomal protein genes contains two conserved promoter elements

N. F. J. VAN POPPEL*, J. WELAGEN, A. N. VERMEULEN *and* D. SCHAAP

Department of Parasitology R&D, Intervet International BV, P.O. Box 31, 5830 AA Boxmeer, The Netherlands

(*Received 10 October 2005; revised 14 December 2005; accepted 28 December 2005; first published online 4 May 2006*)

SUMMARY

Recently we showed that *de novo* ribosome biosynthesis is transcriptionally regulated in Coccidia, depending on their life-cycle stage. Since the expression of ribosomal protein genes is likely coordinated, the transcriptional control of all *Toxoplasma gondii* ribosomal protein (RP) genes was analysed. Therefore, the complete set of all cytoplasmic RPs was defined, containing 79 different RPs in *T. gondii*. RP genes were randomly distributed over the genome, each with a unique upstream region with the exception of 8 RP genes which were paired in a head-to-head orientation. To study if the RP genes share conserved promoter elements, a database was made containing upstream sequences of all *T. gondii* RP genes. Promoter activity was confirmed for the upstream sequences of 8 RP genes, some of which are comparable in strength to the alpha-tubulin promoter. In the complete set of RP upstream sequences 2 novel and highly conserved elements were identified, named *Toxoplasma* Ribosomal Protein (TRP)-1 (consensus: TCGGCTTATATTCGG) and TRP-2 ([T/C] GCATGC[G/A]). TRP-1 and/or TRP-2 were present in 95% of all RP upstream sequences and moreover, were specifically localized in a small region near the presumptive transcriptional start site (10–330 bp upstream). Although TRP elements were mostly absent in known *T. gondii* promoters, they are present elsewhere in the *T. gondii* genome suggesting that they operate not only in RP genes but in a larger set of genes. The identification of TRP elements creates a basis to further study the underlying mechanism by which RP transcription is controlled in *T. gondii*.†

Key words: ribosome biosynthesis, transcriptional regulation, promoter, TRP element.

INTRODUCTION

The eukaryotic ribosome is a complex structure composed of about 80 different ribosomal proteins (RP) and 4 structural rRNAs, which are assembled into a small and large subunit. Together these subunits form the translational machinery of a cell. Ribosomes are essential for a cell and its constituents are highly expressed in rapidly dividing cells. Expression of RP and rRNA genes in eukaryotes is coordinately regulated in response to stress and growth stimuli, which permits the cell to adjust the number of ribosomes and overall protein synthetic capacity to environmental conditions (Pearson and Haber, 1980; Ju and Warner, 1994; Meyuhas, 2000). For higher eukaryotes, ribosome biosynthesis is regulated at the level of translation (Meyuhas, 2000). In contrast, it is primarily transcriptionally regulated in *Saccharomyces cerevisiae*, where transcription is switched off during stress (e.g. nutrient deficiency) and spore formation (Warner, 1999). Transcription of RP genes is then coordinately regulated by a combination of transcription factors and specific DNA elements present in their promoters. At the moment, ribosome biosynthesis has scarcely been studied in parasites. Recently, we demonstrated that coccidian parasites also transcriptionally regulate *de novo* ribosome biosynthesis (Schaap *et al*. 2005). For example, in *Eimeria tenella* rRNA and RP genes are abundantly transcribed in rapidly growing merozoite stages and almost not transcribed in dormant oocyst stages. Similarly, *Toxoplasma gondii* showed a 100-fold difference in RP transcription, when rapidly growing tachyzoites were compared with oocyst stages (Schaap *et al*. 2005). Thus, Coccidia can transcriptionally regulate ribosome biosynthesis, dependent on their life-cycle stage.

Since transcription of RP genes was simultaneously regulated in Coccidia, we anticipate that this is coordinated for the complete set of RP genes by an underlying control mechanism. Therefore, we made a start to analyse the transcriptional regulation of the complete set of RP genes as a whole in *T. gondii*. By clustering of publicly available expressed sequence tags (ESTs), all cytoplasmic *T. gondii* RPs were deduced. Although a large number of ESTs is available for this parasite and EST clusters were published (Ajioka *et al*. 1998; Li *et al*. 2003), these proteins were not identified previously.

* Corresponding author: Tel: +31 485 587967. Fax: +31 485 585313. E-mail: nicole.vanpoppel@intervet.com

† The nucleotide sequence data reported in this paper are available in the Third Party Annotation Section of the DDBJ/EMBL/GenBank databases under the Accession numbers TPA: BK004896-BK004974.

doi:10.1017/S0031182006009954 Printed in the United Kingdom

To determine if and how transcription of all *T. gondii* RP genes may be coordinately regulated, these genes were studied for their genomic organization. Moreover, their upstream sequences were analysed for the presence of conserved DNA elements as well as promoter strength. The identification of 2 novel conserved elements in upstream sequences of the complete set of *T. gondii* RP genes allows us to study transcriptional control of RP genes.

MATERIALS AND METHODS

Identification of the complete set of T. gondii *cytoplasmic RPs*

For the identification of the coding sequences of *T. gondii* cytoplasmic RPs publicly accessible ESTs annotated for RPs (2192 ESTs, dd. 24 April 2003) were collected from NCBI/Nucleotide (http://www.ncbi.nlm.nih.gov). Clustering of ESTs into contigs was performed with Sequencher 4.1.4® software, after which the consensus of each contig was exported to CloneManager 6.0 to determine open reading frames (ORFs). Protein predictions were verified for the consensus of each contig by BLAST search against the database of *T. gondii* Twinscan2 predicted proteins on www.toxodb.org. BLAST searches for similarity to RPs in other organisms were performed at NCBI/BLAST (swiss-prot db). Pfam domain searches were performed at http://www.sanger.ac.uk/Software/Pfam and were used for the annotation of each *T. gondii* RP-specific ORF. The coding sequences of the complete set of *T. gondii* RPs reported in this paper are available in the Third Party Annotation Section of the DDBJ/EMBL/GenBank databases under the accession numbers TPA: BK004896-BK004974 (protein sequences at DAA04986-DAA05064).

Comparative analysis of RP upstream sequences

A database was made consisting of the upstream sequences of 79 *T. gondii* RP genes. About 1000 bp of genomic sequence upstream of the start of each RP EST contig were collected, assuming that the start of each RP EST contig roughly corresponds to the presumptive transcriptional start site of the RP gene. These preliminary genomic sequence data were obtained from The Institute for Genomic Research (TIGR) website at http://www.tigr.org by blastn using each cDNA consensus from the final EST database as query. Searches for conserved motifs were carried out by comparison of these RP upstream sequences using the program Multiple Em for Motif Elicitation (MEME) (http://meme.sdsc.edu). MEME was run under the following settings: a motif width ranging from 6 to 50 bases, a discovery limit of 10 motifs and the total number found per site ranging from 2 to 150.

Parasite strain and culturing

T. gondii RHΔHXGPRT (Donald *et al.* 1996) tachyzoites were maintained in culture at 5% CO_2 and 37 °C by serial passage in Vero cells or human foreskin fibroblasts (HFF), grown in Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% heat-inactivated fetal calf serum and 2 mM L-glutamine.

Molecular techniques

Ten different LacZ expression vectors were made by placing a *LacZ* reporter gene under the control of either the *T. gondii* alpha-tubulin promoter (TUB) or upstream sequences of a *T. gondii* RP gene (*p*RPS3, *p*RPS10, *p*RPS13, *p*RPS25, *p*RPS29, *p*RPL9, *p*RPL13 and *p*RPL38). The different LacZ constructs were based on the p*CAT-GFP* plasmid (Striepen *et al.* 1998), which contains a fusion of chloramphenicol acetyltransferase gene (*CAT*) to green fluorescent protein (*GFP*) gene driven by the dihydrofolate reductase (DHFR) promoter and flanked by the DHFR 3′ untranslated region (UTR). First, *GFP* was replaced with the *LacZ* gene, which was derived from genomic DNA of *Escherichia coli* BL21 and amplified by polymerase chain reaction (PCR) using the primers (restriction sites are underlined, start- and stopcodons are in bold); LacZ-AvrII (fw): 5′-CGATCCTAGG**ATG**ACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTG-3′ and LacZ-PstI (rv): 5′-CGATCTGCAG**TTA**TTTTTGACACCAGACCAACTGG-3′. The PCR product was digested with AvrII/PstI and inserted in p*CAT-GFP* (AvrII/PstI digested) resulting in p*CAT-LacZ*. The TUB and RP upstream sequences (except *p*RPS13) and their complete 5′ UTR were PCR amplified from *T. gondii* RHΔHXGPRT genomic DNA using the primers depicted in Table 1. The PCR products were digested with HindIII/AvrII or KpnI/AvrII and inserted in p*CAT-LacZ* by replacing the p*CAT* sequence. The resulting constructs were named *p*TUB[AvrII]*LacZ* or *p*RP(xx)*LacZ* (xx refers to the name of RP used).

Since RPS13 upstream sequences contain an internal AvrII restriction site, the construct *p*RPS13*LacZ* was made in a different way together with a *p*TUB*LacZ* construct. These constructs were also based on p*CAT-GFP* replacing *CAT-GFP* with the *LacZ* gene. *LacZ* was PCR amplified using LacZ-BglII (fw): 5′-CGATAGATCT**ATG**ACCATGATTACGGATTCACTG-3′ and LacZ-PstI (rv), digested with BglII/PstI and inserted in p*CAT-GFP* (BglII/PstI digested). The resulting construct was named *p*DHFR*LacZ*. The *p*RPS13 and *p*TUB were PCR amplified as described above for the other promoters. The PCR products were digested with HindIII/BglII and inserted in *p*DHFR*LacZ* by

https://doi.org/10.1017/S0031182006009954 Published online by Cambridge University Press

Table 1. List of primers used to generate *p*RP*LacZ* reporter constructs

(The restriction sites HindIII, KpnI, AvrII and BglII are underlined.)

Construct	Forward primer	Reverse primer
*p*RPS3 *LacZ*	5'-CGATGGTACCAAGGTGAACATGCCTTCCCCAATTTTGCC-3'	5'-CGATCCTAGGCTTGGACAAAACAGGGATGAGAAC-3'
*p*RPS10 *LacZ*	5'-CGATAAGCTTTTCCCTGTCGGATTTCTACTTTTG-3'	5'-CGATCCTAGGTCCAGGGACCGTCATTTTTTGG-3'
*p*RPS13 *LacZ*	5'-CGATAAGCTTGTCAGTGCATGACACAACCGCG-3'	5'-CGATAGATCTGGTGACGGAGAAAAATCGAAATTGG-3'
*p*RPS25 *LacZ*	5'-CGATGGTACCGAAACGGGAAGGTCTCATGGCGCGG-3'	5'-CGATCCTAGGGGTGAATCTGCAAACAGGGACG-3'
*p*RPS29 *LacZ*	5'-CGATAAGCTTCCAGCCCAGGTCGGCGAGTGC-3'	5'-CGATCCTAGGTTTGCGAGGCGCCTCAGAAAAACA-3'
*p*RPL9 *LacZ*	5'-CGATAAGCTTACAGAGAAAGAGGTCTGCATCC-3'	5'-CGATCCTAGGGGTGGCGACGGAGCGGGAGAGGG-3'
*p*RPL13 *LacZ*	5'-CGATGGTACCGAACCGACGGGCAGTTTCACG-3'	5'-CGATCCTAGGTTTGACAGAAGAGAAGCAGACG-3'
*p*RPL38 *LacZ*	5'-CGATAAGCTTTTTGTCGTCCAGTGGGGTTCAGTG-3'	5'-CGATCCTAGGGTTGTCTACTTAGAATATTCCTGCG-3'
*p*TUB*LacZ* [AvrII]	5'-CGAAGCTTGGGGGATCCACTAGTTCTAG-3'	5'-CGCCTAGGGGGAATTCAAGAAAAAATGCCAACG-3'
*p*TUB*LacZ* [BglII]	5'-CGAAGCTTGGGGGATCCACTAGTTCTAG-3'	5'-CGAGATCTGGGAATTCAAGAAAAAATGCCAACG-3'

replacing the DHFR promoter (HindIII/BglII digested). The constructs were named *p*TUB[BglII] *LacZ* and *p*RPS13*LacZ*.

Electroporation of T. gondii *RHΔHXGPRT tachyzoites and CPRG assay*

Transient transfections were carried out by electroporation as described previously (Soldati and Boothroyd, 1993) using freshly harvested *T. gondii* RHΔHXGPRT tachyzoites (10^7) and 20 μg sterilized circular plasmid DNA (QIAGEN Plasmid maxi kit, Qiagen) in a 2 mm gap cuvette (BTX electroporator; 1·8 kV, 100 Ω, 25 μF) in a total volume of 400 μl electroporation buffer, which was composed of 120 mM KCl, 0·15 mM $CaCl_2$, 10 mM K_2HPO_4/KH_2PO_4 pH 7·6, 25 mM HEPES, 2 mM EDTA, 5 mM $MgCl_2$. Immediately prior to use, fresh 2 mM ATP and 5 mM glutathione (GSH) were supplemented to the buffer and sterilized by filtration through a 0·22 μm filter. After electroporation triplicates (50 μl) of each sample were added to separate wells (24-well plates) containing a confluent monolayer of HFF cells. Parasites were cultured overnight (37 °C, 5% CO_2) and LacZ activity was determined 16–24 h post-infection by a chlorophenol red-β-D-galactopyranoside (CPRG) assay as described previously (Seeber and Boothroyd, 1996). Briefly, infected monolayers were lysed by adding 200 μl of assay buffer (100 mM HEPES pH 8·0, 1 mM $MgSO_4$, 1% Triton X-100, 5 mM dithiothreitol) per well and were then incubated for 1 h at 50 °C. An aliquot (50 μl) of the cleared lysate was diluted into assay buffer (100 μl final volume) and mixed with an equal volume of assay buffer containing 2 mM CPRG (Roche) as substrate. Substrate conversion was carried out at 30 °C (~4 h) and measured at 570 nm using a microplate reader.

Three independent assays were carried out as described above using HFF cells infected with untransfected *T. gondii* RHΔHXGPRT tachyzoites as a negative control. Within an individual assay all data were subtracted by the LacZ level of this control sample, after which the data were converted to percentages. For conversion into percentages, the average of the *p*TUB*LacZ*[BglII] triplicate was set on 100%, after which all samples were related to this sample. All triplicates of the 3 independent assays were used to calculate the average of each tested construct with its standard deviation. Since equal concentrations of construct DNA were transfected, the obtained data were converted to equimolar amounts.

RESULTS

The complete set of T. gondii *cytoplasmic RPs*

To identify the complete set of RP coding sequences, *T. gondii* RH tachyzoite ESTs were initially collected, which had been annotated as RP. This resulted in a dataset containing 2192 ESTs which were assembled into 84 contigs. These contigs were checked for sequence inconsistencies and the consensus of each contig was used to deduce ORFs. These ORFs showed high similarity to 31 eukaryotic small subunit RPs (containing 2 partial contigs similar to RPS23 which could not yet be merged), 43 eukaryotic large subunit RPs and the eukaryotic ribosomal proteins SA, P0, P1 and P2, as determined by standard protein-protein BLAST. In addition, ORFs of 5 contigs showed similarity with prokaryotic RPs, namely RPS17, RPL13, RPL22, RPL24 and RPL28. These proteins are probably components of the plastid or mitochondrion and for that reason they were excluded from the dataset. Subsequently, the consensus of each contig was used in a new search at NCBI (tblastn; est_others db) to identify additional ESTs which had either not been annotated or falsely annotated. A total of 588 additional *T. gondii* RP ESTs were identified which were added to the initial dataset, resulting in a final dataset consisting of 2780 ESTs assembled in 78 contigs. Compared to the complete set of RPs present in human or yeast (Mager *et al.* 1997; Kenmochi *et al.* 1998; Planta and Mager, 1998; Yoshihama *et al.* 2002), only the small RPL41 of *T. gondii* was missing in the dataset. Using human RPL41 (GenBank Accession no. P28751) as query, 18 homologous *T. gondii* ESTs were obtained which clustered into 1 contig. With the RPL41 contig included,

https://doi.org/10.1017/S0031182006009954 Published online by Cambridge University Press

79 different contigs were generated each containing a full-length ORF similar to a eukaryotic RP. Protein predictions were confirmed by BLAST search against the database of *T. gondii* Twinscan2 predicted proteins. In general, coding sequences of these putative *T. gondii* RPs were highly similar to their human and yeast homologue (shown in Table 2), except for the putative *T. gondii* RPL28 (8·00E-03). *T. gondii* RPL28 was 21% identical with human RPL28 and 44% identical with *E. tenella* RPL28 (which was identified from 10 assembled ESTs, results not shown).

Since the nomenclature of RPs is often ambiguous and differing between species, all *T. gondii* RP-specific ORFs were classified by their similarity to human RPs. This classification is in agreement with the annotations for conserved eukaryotic RP domains as defined in the Pfam database (Sanger). As is obvious from Table 2, this annotation sometimes differs from yeast annotated RPs. In summary, the complete set of *T. gondii* cytoplasmic RPs was identified and consists of 31 small subunit RPs, 44 large subunit RPs and the proteins SA, P0, P1 and P2, being highly similar to RPs present in human and yeast.

Clustering of RP genes on the T. gondii *genome*

Previously, we showed that RP transcription is coordinately regulated in *E. tenella* which was also suggested for *T. gondii* (Schaap *et al.* 2005). Furthermore, genes encoding proteins which display similar functions or are required in specific tissues are sometimes clustered on the genome to allow coordinate regulated transcription (van Driel *et al.* 2003). Therefore, it was investigated whether RP genes are clustered on the *T. gondii* genome. Coding sequences of all *T. gondii* RPs were used as a query in a BLAST search against the Twinscan database. All RP coding sequences were detected within the Twinscan database except for RPL41. Analysis of their positions revealed that 10 RP genes were paired on the genome, being (1) *RPS5* with *RPS29*, (2) *RPS16* with *RPL13*, (3) *RPS24* with *RPL10A*, (4) *RPL11* with *P2*, and (5) *SA* with *RPL31* (Fig. 1). In the first 4 pairs, RP genes were arranged in a head-to-head orientation with an intergenic region ranging from 280 bp to 380 bp. If a promoter is limited to the intergenic region, this would indicate that these RP putative promoters are at most 280 to 380 bps long. Moreover, these intergenic regions then contain either 2 small promoters or one bidirectional promoter. In pair 5, the genes were arranged in a tail-to-tail orientation being spaced 3764 bp apart. Apart from the above-described gene clusters, all other *T. gondii* RP genes were spaced more than 10 kb apart. Thus, most RP genes are randomly distributed as individual genes over the genome indicating that their transcription is individually regulated.

Generation of a T. gondii *database containing RP upstream sequences*

Since *T. gondii* RP genes are not clustered on the genome, transcription of these genes must be individually regulated though in a concerted manner. Similarly to *S. cerevisiae*, we investigated if transcription of the complete set of RP genes is coordinately regulated in *T. gondii* and whether conserved promoter elements are present in this set of *T. gondii* genes. Therefore, a RP putative promoter database was made consisting of genomic sequences immediately upstream of the presumptive transcriptional start site of all *T. gondii* RP genes. It should be noted that transcriptional start sites of the 79 RP genes were not experimentally determined, but were based on the start of each RP EST contig. Since these contigs contain many ESTs (35 ESTs on average), the true transcriptional start sites will likely be close to the presumptive transcriptional start sites we used. For 2 RP genes (*RPS13* and *RPL9*) the transcriptional start sites were experimentally determined, which indeed correlated well to the start of the RP EST contigs (van Poppel, manuscript submitted). The *T. gondii* genomic sequences immediately upstream of the presumptive transcriptional start site of all RP genes were obtained by blastn (TIGR) using the consensus of each RP EST contig as query. If available, 1000 bp of genomic upstream sequences were collected for each RP; for 69 RPs 915–1000 bp were collected while for the remaining 10 RPs between 160 and 844 bp of upstream sequences were obtained. Although the lengths of promoter regions for RP genes are not determined, the upstream sequences present in this database were considered as the putative promoters.

T. gondii *RP upstream sequences contain two highly conserved and localized DNA elements*

The complete set of *T. gondii* RP upstream sequences present in the database was used to search for conserved DNA elements. The upstream sequences of all *T. gondii* RP genes were compared with each other in multiple searches using the program MEME. This program indicates which sequence elements are overrepresented in a database. Three specific DNA elements were identified which were highly enriched within the RP putative promoter database, being TCGGCTTATATTCGG, [T/C]GCATGC[G/A] and polypyrimidine tracts.

The first sequence element, TCGGCTTATATTCGG (15 bp), was named *Toxoplasma* Ribosomal Protein-1 element (TRP-1). This conserved and novel element was identified 58 times in the *T. gondii* RP putative promoter database. TRP-1 elements showed limited variation (see Table 3), and the percentages of conservation per nucleotide are as follows $T_{69}\underline{\mathbf{C}}_{97}\underline{\mathbf{G}}_{93}\underline{\mathbf{G}}_{83}\underline{\mathbf{C}}_{95}\underline{\mathbf{T}}_{86}\underline{\mathbf{T}}_{90}\underline{\mathbf{A}}_{86}\underline{\mathbf{T}}_{91}\underline{\mathbf{A}}_{88}T_{60}T_{59}C_{67}\underline{\mathbf{G}}_{79}G_{69}$

https://doi.org/10.1017/S0031182006009954 Published online by Cambridge University Press

Table 2. *Toxoplasma gondii* cytoplasmic ribosomal proteins

(Shown from the left to the right are: identification number (ID), the name for each RP protein in *T. gondii* (TgRP) and features for each *T. gondii* RP such as GenBank Accession number of the coding sequence (acc nr), its calculated molecular weight (MW), total number of amino acids (AA) and the number of ESTs present per RP contig. This is followed by Twinscan protein predictions including *T. gondii* chromosome (Chr.) location, and followed by the Pfam domains, which showed homology to parts of each *T. gondii* RP together with the E-value. Furthermore, the homology of *T. gondii* RPs with the human and yeast RPs are given together with the E-value. (a) Region of the *T. gondii* RP protein in which the Pfam domain is present. (b) The complete protein consists of an N-terminal fusion of ubiquitin to the RP. (c) ESTs showing similarity to ubiquitin were excluded. (d) Designated previously as ribosomal protein S37, now as ribosomal protein S31 (Mager *et al.* 1997). (e) Instead of standard blastp, protein BLAST search for short nearly exact matches was used for RPS30 and RPL41. (f) Incomplete coding sequence; missing part (25 amino acids) at the C-terminus was determined using *T. gondii* genomic DNA sequence data (TIGR) according to similarity with human and yeast RPL3 protein sequences. (g) Ambiguous start of translation; human and yeast RPL4 consists of 140 additional amino acids at the N-terminus. Although their accompanying coding sequences were present on the *T. gondii* genome, no ESTs were detected for that extension. Therefore, the largest ORF as deduced from assembled ESTs is given. (h) These proteins are smaller in size compared to their human homologues, however, similar in size compared to homologues in yeast and *Eimeria tenella*.)

T. gondii ribosomal proteins						TgTwinScan			Pfam homology			Human homology			Yeast homology		
ID	TgRP	acc.nr	MW	AA	EST	Chr.	Number	E-value	Part[a]	Domain	E-value	protein	acc.nr	E-value	protein	acc.nr	E-value
1	S2	BK004974	29335·4	269	45	IX	1960	1·50E−146	83–149	Ribosomal_S5	1·60E−37	S2	P15880	3·00E−93	S2	P25443	6·00E−71
2	S3	BK004896	26076	235	47	VIII	5861	2·10E−122	106–190	Ribosomal_S3_C	1·90E−26	S3	P23396	2·00E−75	S3	P05750	5·00E−64
3	S3A	BK004897	29371	259	40	VIII	5895	2·70E−138	11–222	Ribosomal_S3Ae	3·20E−130	S3A	P49241	4·00E−80	S1-B	P23248	3·00E−64
															S1-A	P33442	8·00E−64
4	S4	BK004898	30143·1	263	26	Ib	0830	4·10E−142	94–170	Ribosomal_S4e	4·10E−34	S4 (X)	P12750	5·00E−94	S4	P05753	8·00E−89
												S4 (Y)	P22090	1·00E−91			
5	S5	BK004899	21612·8	192	42	LG14	6233	1·00E−99	39–192	Ribosomal_S7	3·60E−62	S5	P46782	2·00E−71	S5	P26783	2·00E−59
6	S6	BK004900	28955·8	256	40	IV	1004	6·20E−132	1–127	Ribosomal_S6e	9·40E−66	S6	P10660	4·00E−55	S6	P02365	8·00E−55
7	S7	BK004901	22587·3	196	58	LG14	6043	3·00E−100	4–196	Ribosomal_S7e	3·30E−67	S7	P23821	5·00E−30	S7-B	P48164	1·00E−23
															S7-A	P26786	1·00E−22
8	S8	BK004902	23462	205	123	XII	4023	1·90E−107	1–187	Ribosomal_S8e	2·90E−82	S8	P09058	1·00E−55	S8	P05754	1·00E−55
9	S9	BK004903	21683·9	188	39	XII	4236	9·30E−97	4–106	Ribosomal_S4	1·20E−32	S9	P46781	2·00E−57	S9-B	P05755	4·00E−55
									107–154	S4	1·40E−12				S9-A	O13516	5·00E−55
10	S10	BK004904	17397·8	152	22	III	1710	1·70E−81	11–108	S10_plectin	8·00E−50	S10	P46783	2·00E−24	S10-A	Q08745	1·00E−17
															S10-B	P46784	5·00E−17
11	S11	BK004905	18676·7	161	55	X	5495	5·40E−85				S11	P04643	1·00E−44	S11	P26781	2·00E−41
12	S12	BK004906	15398·4	142	24	VII	3929	1·40E−72	26–121	Ribosomal_L7Ae	3·00E−33	S12	P25398	2·00E−24	S12	P48589	3·00E−16
13	S13	BK004907	17176	151	35	VIII	7307	6·80E−78	61–151	Ribosomal_S15	8·20E−32	S13	Q02546	4·00E−56	S13	P05756	4·00E−53
14	S14	BK004908	16301·2	156	19	VI	7064	1·00E−60				S14	P06366	3·00E−47	S14-B	P39516	8·00E−43
															S14-A	P06367	3·00E−42
15	S15	BK004909	17156·9	150	25	V	1122	1·10E−77	52–133	Ribosomal_S19	2·50E−34	S15	P11174	3·00E−37	S15	Q01855	3·00E−32
16	S15A	BK004910	14947·4	130	32	X	2669	3·40E−67	5–130	Ribosomal_S8	9·50E−50	S15A	P39027	1·00E−51	S22	P04648	4·00E−48
17	S16	BK004911	16555·2	148	25	VI	6994	5·80E−81	16–148	Ribosomal_S9	3·00E−69	S16	P17008	2·00E−49	S16	P40213	1·00E−42
18	S17	BK004912	15228·4	132	33	Ib	0856	2·80E−65	1–122	Ribosomal_S17e	3·70E−58	S17	P08708	3·00E−28	S17-B	P14127	7·00E−27
															S17-A	P02407	3·00E−26
19	S18	BK004913	17722·3	156	40	X	5318	3·90E−82	16–144	Ribosomal_S13	3·80E−67	S18	P25232	1·00E−46	S18	P35271	2·00E−42
20	S19	BK004914	18379·9	160	43	X	2470	1·30E−85	17–157	Ribosomal_S19e	2·20E−66	S19	P39019	1·00E−30	S19-B	P07281	3·00E−27
															S19-A	P07280	4·00E−27
21	S20	BK004915	13367·4	118	42	II	2638	3·00E−59	18–113	Ribosomal_S10	1·20E−52	S20	P17075	4·00E−39	S20	P38701	2·00E−22
22	S21	BK004916	9034·1	82	25	VI	6740	4·00E−41	1–81	Ribosomal_S21e	2·60E−41	S21	P35265	5·00E−18	S21	P05760	2·00E−20
23	S23	BK004917	15801·3	143	26	VIII	5672	7·30E−74	7–142	Ribosomal_S12	8·00E−74	S23	P39028	2·00E−57	S23	P32827	5·00E−52
24	S24	BK004918	15194·6	135	24	X	1289	8·00E−68	25–108	Ribosomal_S24e	5·30E−22	S24	P16632	3·00E−15	S24	P26782	3·00E−20
25	S25	BK004919	12166·9	108	55	VIII	5774	2·90E−54	1–106	Ribosomal_S25	1·90E−50	S25	P25111	1·00E−15	S25	P07282	3·00E−08
26	S26	BK004920	12725·4	112	18	LG14	6325	3·00E−59	1–112	Ribosomal_S26e	3·10E−52	S26	P02383	3·00E−21	S26-A	P39938	1·00E−18
															S26-B	P39939	2·00E−18

Table 2. (*Cont.*)

T. gondii ribosomal proteins						TgTwinScan			Pfam homology			Human homology			Yeast homology		
ID	TgRP	acc.nr	MW	AA	EST	Chr.	Number	E-value	Part[a]	Domain	E-value	protein	acc.nr	E-value	protein	acc.nr	E-value
27	S27	BK004921	9091·7	82	25	XII	1453	2·40E−43	26–80	Ribosomal_S27e	1·70E−31	S27	P42677	7·00E−22	S27-A	P35997	2·00E−23
															S27-B	P38711	2·00E−23
28	S27A[b]	BK004922	17320·8	154	43[c]	XII	4040	1·20E−80	1–74	Ubiquitin	1·10E−19	S27A	P14798	2·00E−12	S37[d]	P05759	8·00E−11
									101–147	Ribosomal_S27	6·30E−31	Ubiquitin	P02248	2·00E−11	Ubiquitin	P04838	2·00E−10
29	S28	BK004923	7587·7	68	40	Ib	0952	1·10E−34	1–68	Ribosomal_S28e	7·00E−33	S28	P25112	5·00E−08	S28	P02380	7·00E−08
30	S29	BK004924	6519·7	54	49	LG14	6234	1·30E−26	1–54	Ribosomal_S14	4·70E−06	S29	P30054	1·00E−10	S29-B	P41058	4·00E−10
															S29-A	P41057	7·00E−10
31	S30	BK004925	6599·7	59	57	IX	0294	4·40E−28	2–58	Ribosomal_S30	2·70E−29	S30[e]	Q05472	3·00E−24	S30[e]	Q12087	1·00E−17
32	L3[f]	BK004926	44176·1	389	30	X	5519	9·10E−201	50–338	Ribosomal_L3	4·40E−160	L3	P39023	e-134	L3	P14126	e-115
												L3-like	Q92901	e-116			
33	L4[g]	BK004927	31097·7	279	21	XI	4515	2·70E−161	2–127	Ribosomal_L4	9·20E−11	L4	P36578	2·00E−42	L4-A	P10664	2·00E−46
															L4-B	P49626	2·00E−46
34	L5	BK004928	35341·1	310	37	IV	2240	3·80E−164	26–172	Ribosomal_L18p	3·90E−57	L5	P46777	1·00E−68	L5	P26321	1·00E−60
35	L6[h]	BK004929	21519·1	193	15	XI	4864	9·30E−97	78–193	Ribosomal_L6e	5·60E−40	L6	Q02878	1·00E−22	L6-A	Q02326	3·00E−22
															L6-B	P05739	1·00E−21
36	L7	BK004930	30121·2	258	35	XI	4985	1·20E−130	92–144	Ribosomal_L30	1·60E−18	L7	P18124	1·00E−45	L7-B	Q12213	3·00E−50
															L7-A	P05737	2·00E−49
37	L7A	BK004931	31031·3	276	45	VI	6882	2·10E−145				L7A	P11518	2·00E−56	L8-B	P29453	1·00E−51
															L8-A	P17076	3·00E−51
38	L8	BK004932	28254·3	260	45	VII	3845	4·30E−140	11–90	Ribosomal_L2	6·60E−23	L8	P25120	2·00E−82	L2	P05736	3·00E−80
									96–231	Ribosomal_L2_C	4·50E−67						
39	L9	BK004933	21423·8	190	24	V	3457	5·60E−99	12–86	Ribosomal_L6	8·30E−17	L9	P32969	5·00E−47	L9-B	P51401	4·00E−34
									98–177	Ribosomal_L6	7·70E−16				L9-A	P05738	4·00E−34
40	L10	BK004934	24800·5	221	22	IX	0098	2·50E−119	1–176	Ribosomal_L10e	1·20E−134	L10	P27635	1·00E−84	L10	P41805	5·00E−74
41	L10A	BK004935	24602·8	217	29	X	1290	3·20E−112	4–212	Ribosomal_L1	9·90E−65	L10A	P53025	7·00E−70	L1	P53030	4·00E−61
42	L11	BK004936	20199	175	78	XI	4558	1·30E−92	8–61	Ribosomal_L5	5·50E−25	L11	P39026	5·00E−68	L11	P06380	9·00E−64
									65–164	Ribosomal_L5_C	1·60E−60						
43	L12	BK004937	17820·4	165	37	III	3017	8·80E−85	12–70	Ribosomal_L11_N	1·50E−19	L12	P30050	8·00E−60	L12	P17079	8·00E−54
									75–143	Ribosomal_L11	7·80E−22						
44	L13	BK004938	23975·8	212	79	VI	6995	5·90E−111	6–189	Ribosomal_L13e	9·20E−72	L13	P26373	5·00E−30	L13-A	Q12690	4·00E−24
															L13-B	P40212	7·00E−24
45	L13A	BK004939	23603·5	203	26	IX	0312	3·50E−106				L13A	P40429	9·00E−52	L16-A	P26784	7·00E−51
															L16-B	P26785	2·00E−48
46	L14	BK004940	15462·1	134	48	IX	3229	1·70E−67	48–124	Ribosomal_L14e	1·40E−26	L14	P50914	3·00E−21	L14-B	P38754	2·00E−16
															L14-A	P36105	2·00E−16
47	L15	BK004941	24325·3	205	24	X	5602	8·40E−112	2–194	Ribosomal_L15e	6·10E−122	L15	P39030	6·00E−57	L15-B	P54780	7·00E−54
															L15-A	P05748	7·00E−54
48	L17	BK004942	22268·7	195	39	III	0662	5·80E−104	17–158	Ribosomal_L22	1·00E−60	L17	P18621	4·00E−52	L17-A	P05740	8·00E−39
															L17-B	P46990	1·00E−38
49	L18	BK004943	21169·6	187	61	XII	0691	7·30E−97	11–187	Ribosomal_L18e	2·90E−101	L18	Q07020	3·00E−40	L18	P07279	1·00E−32

50	L18A	BK004944	21329·8	183	40	VI	6959	1·40E−95	10–182	Ribosomal_L18ae		L18A	Q02543	2·00E−38	L20	P47913	7·00E−39
51	L19	BK004945	22159·9	187	44	IX	0166	6·60E−96	1–148	Ribosomal_L19e	1·20E−85	L19	P14118	3·00E−30	L19	P05735	8·00E−31
52	L21	BK004946	17893·6	157	12	XII	4045	1·30E−83	2–100	Ribosomal_L21e	3·90E−46	L21	P46778	4·00E−29	L21-B	Q12672	3·00E−30
															L21-A	Q02753	4·00E−30
53	L22	BK004947	15073·4	133	36	LG14	6091	6·20E−68	21–131	Ribosomal_L22e	3·20E−47	L22	P35268	7·00E−18	L22-B	P56628	3·00E−08
															L22-A	P05749	6·00E−08
54	L23	BK004948	14818·2	139	16	XI	4741	1·50E−73	18–139	Ribosomal_L14	2·00E−46	L23	P23131	3·00E−49	L23	P04451	1·00E−47
55	L23A	BK004949	18561·4	167	47	LG14	5984	3·80E−84	14–75	Ribosomal_L23eN	5·10E−03	L23A	P29316	6·00E−32	L25	P04456	2·00E−27
									85–164	Ribosomal_L23	5·80E−28						
56	L24	BK004950	17597·5	155	32	LG14	6391	2·80E−79	8–78	Ribosomal_L24e	2·40E−19	L24	P38663	6·00E−14	L24-B	P24000	5·00E−11
															L24-A	P04449	5·00E−11
57	L26	BK004951	16308·8	141	28	XII	4228	1·60E−71				L26-like	Q9UNX3	3·00E−25	L26-A	P05743	2·00E−21
												L26	Q02877	4·00E−25	L26-B	P53221	2·00E−21
58	L27	BK004952	16802	146	51	VI	6961	8·10E−75	51–146	Ribosomal_L27e	1·20E−19	L27	P08526	3·00E−17	L27	P38706	1·00E−20
59	L27A	BK004953	16344	147	32	XI	4621	5·90E−79	109–141	L15	2·70E−14	L27A	P46776	1·00E−30	L28	P02406	2·00E−21
60	L28	BK004954	14752·4	129	28	VIII	5639	6·40E−66	3–123	Ribosomal_L28e	1·70E−07	L28	P46779	8·00E−03	—	—	—
61	L29[h]	BK004955	6366·5	54	29	X	2679	6·40E−27	3–42	Ribosomal_L29e	4·20E−18	L29	P47914	3·00E−05	L29	P05747	0·016
62	L30	BK004956	11661·4	108	32	VIII	5855	4·10E−55	11–105	Ribosomal_L7Ae	1·30E−34	L30	P04645	2·00E−35	L30	P14120	5·00E−27
63	L31	BK004957	14355·4	120	12	IX	3168	4·20E−62	11–105	Ribosomal_L31e	5·70E−51	L31	P12947	1·00E−24	L31	P04649	1·00E−17
64	L32	BK004958	15617	134	29	IX	3250	1·80E−68	15–124	Ribosomal_L32e	1·30E−61	L32	P02433	2·00E−31	L32	P38061	1·00E−24
65	L34	BK004959	15382·9	134	24	X	5532	2·90E−70	5–101	Ribosomal_L34e	6·70E−41	L34	P49207	2·00E−11	L34-B	P40525	1·00E−16
															L34-A	P87262	1·00E−16
66	L35	BK004960	14193·8	123	18	XII	4406	3·50E−58	6–69	Ribosomal_L29	1·20E−16	L35	P42766	5·00E−15	L35	P39741	3·00E−11
67	L35A	BK004961	12846·9	112	22	XII	4298	1·20E−57	12–106	Ribosomal_L35Ae	3·70E−48	L35A	P18077	3·00E−23	L33-A	P05744	2·00E−26
															L33-B	P41056	6·00E−26
68	L36	BK004962	11407·4	101	25	LG14	6006	2·50E−48	2–100	Ribosomal_L36e	1·30E−33	L36	Q9Y3U8	1·00E−10	L36-B	O14455	4·00E−08
															L36-A	P05745	4·00E−08
69	L37	BK004963	11220·9	98	29	LG14	6054	4·70E−54	5–58	Ribosomal_L37e	4·30E−32	L37	P02403	1·00E−10	L37-A	P49166	2·00E−12
															L37-B	P51402	3·00E−12
70	L37A	BK004964	10404·2	96	44	XII	0707	6·20E−45	2–92	Ribosomal_L37ae		L37A	P12751	6·00E−19	L43	P49631	2·00E−16
71	L38	BK004965	9583·2	84	37	VIII	5769	1·10E−40	2–80	Ribosomal_L38e	2·50E−27	L38	P23411	1·00E−08	L38	P49167	8·00E−07
72	L39	BK004966	6153·4	51	40	III	3033	2·30E−15	9–51	Ribosomal_L39	4·70E−22	L39-like	Q96EH5	4·00E−07	L39	P04650	6·00E−03
												L39	P02404	2·00E−06			
73	L40[b]	BK004967	14685	129	23[c]	IX	0182	2·70E−67	1–74	Ubiquitin	7·30E−43	Ubiquitin	P02248	4·00E−35	Ubiquitin	P04838	8·00E−34
									77–128	Ribosomal_L40e	1·30E−32	L40	P14793	1·00E−15	L40	P14796	9·00E−17
74	L41	BK004968	5098·1	39	18				15–39	Ribosomal_L41	5·10E−06	L41[e]	P28751	4·00E−11	L41[e]	P05746	2·00E−09
75	L44	BK004969	12062·4	105	16	VII	3807	1·20E−55	19–95	Ribosomal_L44	7·90E−46	L44	P09896	8·00E−23	L42	P02405	3·00E−23
76	SA	BK004970	31511·4	287	27	IX	3167	2·60E−156				SA	P08865	8·00E−63	S0-A	P32905	6·00E−64
															S0-B	P46654	1·00E−63
77	P0	BK004971	34156·8	314	32	XII	1516	3·00E−162	6–108	Ribosomal_L10	2·90E−34	P0	P05388	2·00E−72	P0	P05317	1·00E−57
									225–313	60s_Ribosomal	3·00E−15						
78	P1	BK004972	12127·3	118	40	VI	6797	1·80E−54	32–117	60s_Ribosomal	1·10E−30	P1	P05386	1·00E−09	P1-alpha	P05318	0·001
															P1-beta	P10622	0·005
79	P2	BK004973	11768·1	113	23	XI	4557	6·70E−55	19–112	60s_Ribosomal	7·10E−30	P2	P05387	6·00E−09	P2	P02400	1·00E−05

Table 3. Conservation of TRP-1 consensus in *Toxoplasma gondii* RP upstream sequences

(Summarized in this Table are the 58 TRP-1 elements which have been identified in *T. gondii* RP upstream sequences. For each position, frequency of each individual nucleotide (T, A, G or C) is given in percentages. The most prevalent nucleotide is indicated in bold. On the bottom the consensus obtained from the 58 TRP-1 elements is shown.)

	Position														
Nucleotide	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
T	**69**	0	0	9	0	**86**	**90**	12	**91**	0	**60**	**59**	10	3	0
A	2	0	0	2	0	5	9	**86**	7	**88**	14	31	17	7	29
G	9	3	**93**	**83**	5	7	0	2	0	2	19	10	5	**79**	**69**
C	21	**97**	7	7	**95**	2	2	0	2	10	7	0	**67**	10	2
Consensus	**T**	**C**	**G**	**G**	**C**	**T**	**T**	**A**	**T**	**A**	**T**	**T**	**C**	**G**	**G**

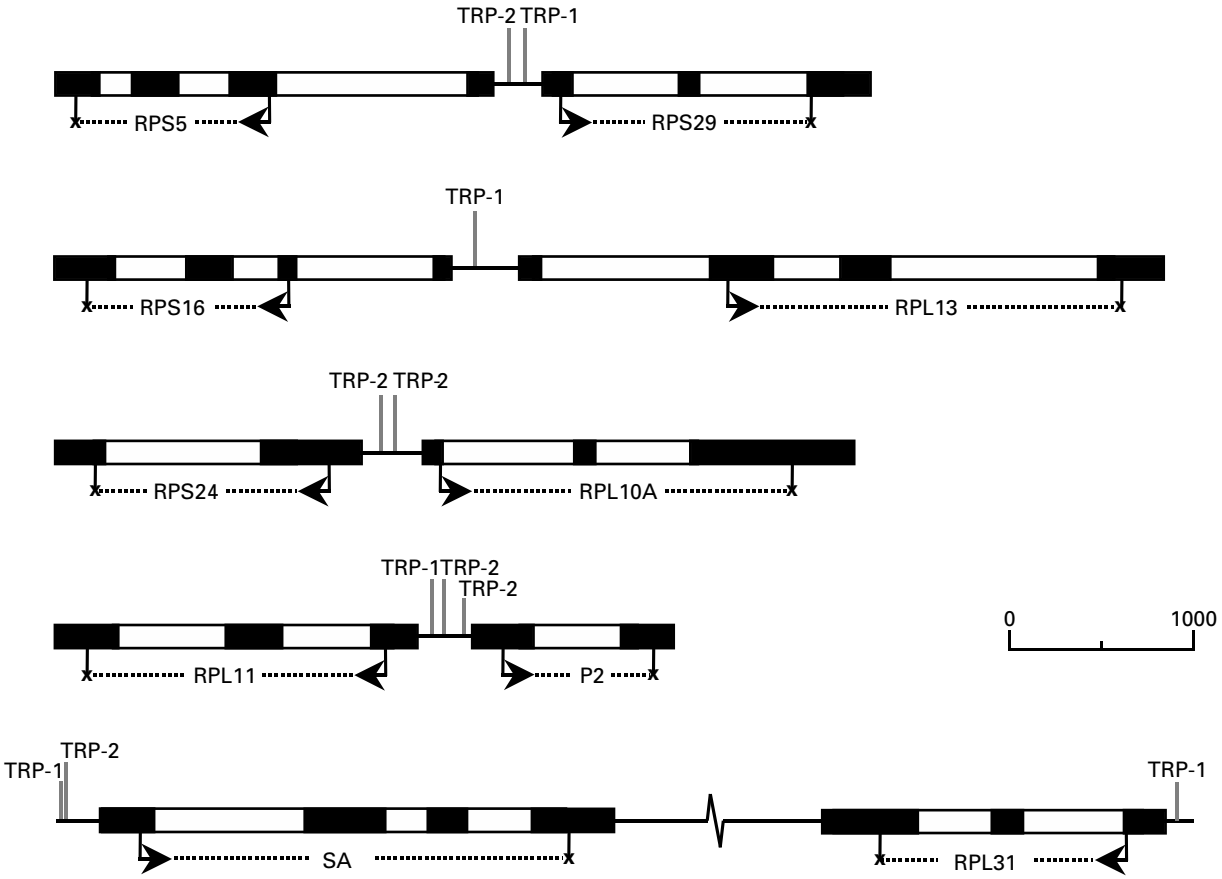

Fig. 1. Pairs of RP genes on the *Toxoplasma gondii* genome. Shown are 5 clusters of paired *T. gondii* RP genes on scale, with the exception of the intergenic region between the genes *SA* and *RPL31* (being 3764 bp). Exons are shown as black boxes and introns as open boxes. Intergenic regions are shown as solid lines and TRP elements are indicated as vertical lines at their position within the sequence. Arrows depict start codons and X depicts stop codons. For *RPS5*, *RPS16* and *RPL13* the start codon is preceded by an intron. For *RPS5* the last 119 bps of the 3′ UTR are not depicted and for ribosomal protein *P2* the last 26 bps of the 3′ UTR are not shown. The RP genes which are arranged in a head to head orientation contain intergenic regions between 280–380 bp. The genomic coordinates of these clusters on the *T. gondii* chromosomes are as follows; RPS5/RPS29: LG14 – 1·95 Mb; RPS16/RPL13: VI – 0·85 Mb; RPS24/RPL10A: X – 6·91 Mb; RPL11/P2: XI – 0·82 Mb; SA/RPL31: IX – 1·31 Mb.

(bold and underlined indicates a conservation $\geqslant$79%). TRP-1 was present in both orientations and mostly once per RP upstream sequence. TRP-1 was observed twice in the upstream sequences of *RPS4*, *RPL6*, *RPL35* and *ubiquitin-RPL40*, and 3 times in the upstream sequence of *RPL30*. The second identified sequence element, [T/C]GCATGC[G/A], was also novel, containing a reversible sequence of 8 nucleotides and was named *Toxoplasma* Ribosomal Protein-2 element (TRP-2). The TRP-2 element was found 73 times in the database, mostly once per RP upstream sequence. Two TRP-2 elements were

https://doi.org/10.1017/S0031182006009954 Published online by Cambridge University Press

detected in the upstream sequences of *RPS11*, *RPS13*, *RPS24*, *RPS28*, *RPL10A*, *RPL11*, *RPL18A*, *RPL32*, *RPL41* and *P2*. Three elements were observed in the upstream sequences of *RPS18*, *RPS27* and *RPL27*. The third identified sequence element was a polypyrimidine tract consisting of T/C stretches of 5–30 nucleotides. Polypyrimidine tracts have previously been shown to be present in eukaryotic RP genes, but in the 5′ UTR instead of the upstream sequences (Meyuhas, 2000). These tracts were observed 73 times in the total database.

To further characterize these elements, their positions in relation to the start of transcription were analysed. TRP-1 and TRP-2 elements were almost all highly localized to a specific region, 10–330 bp upstream of the presumptive transcriptional start site (Fig. 2A,B). In contrast, the polypyrimidine tracts were randomly distributed over the RP upstream sequences and no specific common localization was observed (Fig. 2C). TRP-1 and/or TRP-2 were present in 95% of all RP upstream sequences; 41 contained 1 TRP element, 34 contained both TRP elements and for only 4 no TRP element was identified. No co-localization of both types of TRP elements was observed.

To determine if TRP elements are specifically associated with *T. gondii* RP upstream sequences, the presence of these elements in RP upstream sequences was compared to their occurrences in the *T. gondii* genome. Since TRP elements were mainly present within the 330 bp upstream of the presumptive transcriptional start site, their enrichment was determined for this region. For TRP-1 a more restricted consensus was used, being CGGCTTA-TANNNG, to which 20 TRP-1 elements fully complied. Based on random chance we calculated that a TRP-1 element would be present 0·050 times within a region of 330 bps (2 (sense/antisense) × 1/1048576 (random chance) × 330 (bp) × 79 (RPs) = 0·050). Thus, TRP-1 elements were 400-fold enriched (20/0·050 = 400) in the 330 bp upstream of the presumptive transcriptional start site compared to random chance. Similar calculations were performed for the presence of TRP-2 elements showing a 40-fold enrichment in the 330 bp region of all 79 RP putative promoters. Subsequently, the presence of both TRP elements was determined within the *T. gondii* genome, using publicly accessible *Toxoplasma* genomic sequence data (being $\sim 6{\cdot}4 \times 10^7$ bp). In the genomic sequence data TRP-1 and TRP-2 were respectively present 351 times and 30 856 times, being a 3-fold and 8-fold enrichment compared to random chance. Since 7588 genes are identified in the *T. gondii* genome (Twinscan database) and assuming that each of the identified TRP elements is localized within a single gene promoter, RP putative promoter regions are significantly overrepresented with TRP-1 elements (20/79) compared to the genome (351/7588; χ^2-test = 64·23; $P \ll 0{\cdot}005$). However, this

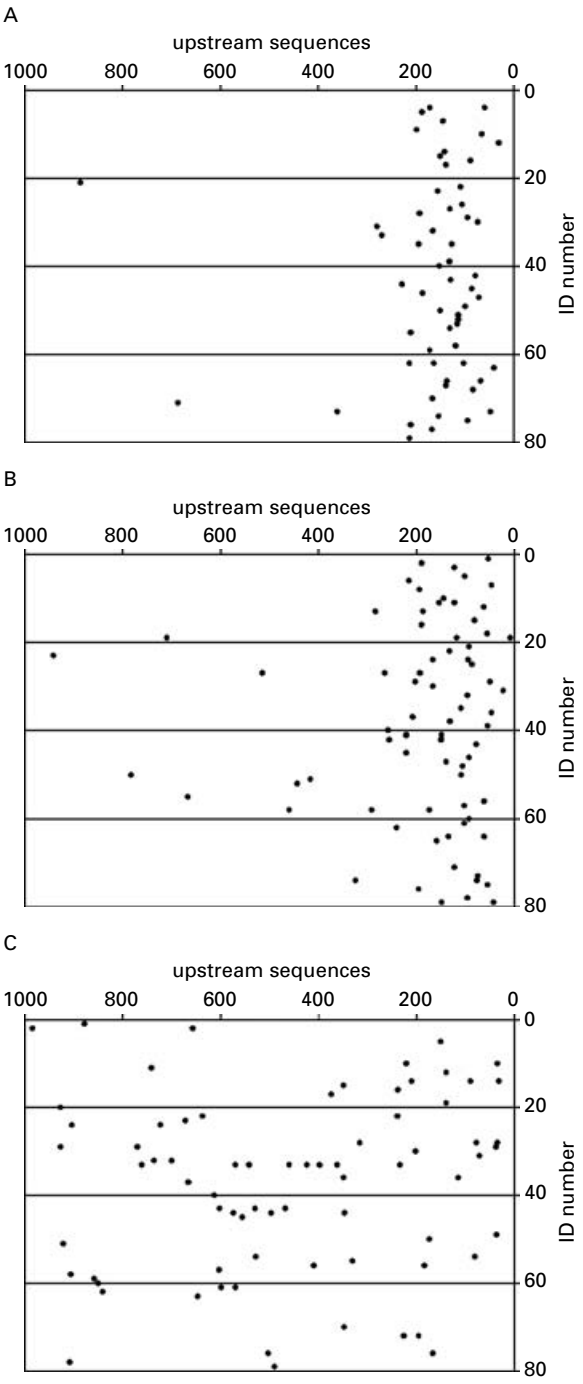

Fig. 2. Localization of sequence elements in RP upstream sequences. Shown are the locations of TRP-1 elements (A), TRP-2 elements (B) and polypyrimidine tracts (C) in *T. gondii* RP upstream sequences. Each dot represents a DNA sequence element. The position of the element within the RP upstream sequence is depicted on the X-axis, where the presumptive transcriptional start site corresponds to position 0 and position 1000 represents 1000 bp upstream from the presumptive transcriptional start site. The ID number of the respective *T. gondii* RP is indicated on the Y-axis. Of each RP gene 1000 bps of upstream sequence were analysed except for: *RPS17* (753 bp), *RPS18* (844 bp), *RPL3* (839 bp), *RPL7* (536 bp), *RPL11* (828 bp), *RPL36* (556 bp), *RPL37* (160 bp), *RPL39* (760 bp), *RPL41* (390 bp) and *P0* (291 bp).

https://doi.org/10.1017/S0031182006009954 Published online by Cambridge University Press

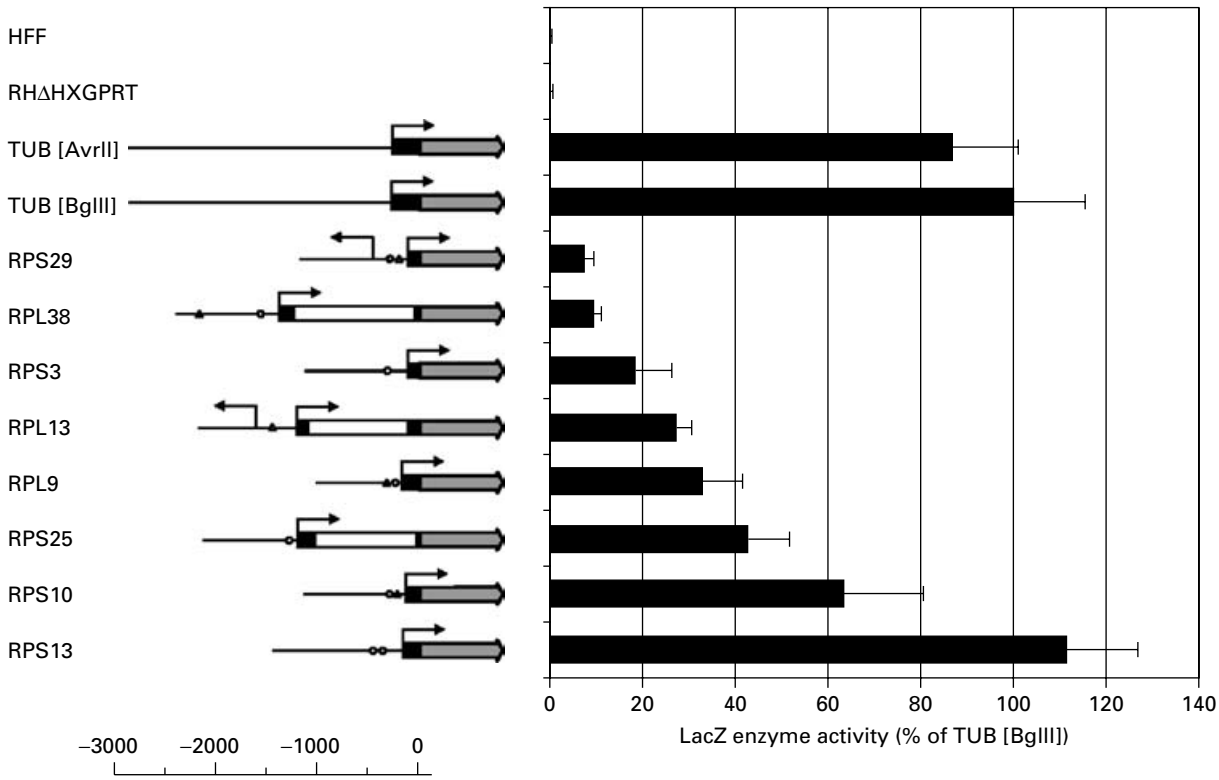

Fig. 3. LacZ expression driven by different RP promoters. Shown on the left are upstream sequences of different RP genes with their 5′ UTR fused to *LacZ*, which were tested for expression. The RP upstream sequences are indicated with a solid line, followed by black boxes for exons and open boxes for introns of the respective RP 5′ UTRs (on scale). Roughly 1 kb of RP upstream sequence was used; *p*RPS29 (1009 bp); *p*RPL38 (953 bp); *p*RPS3 (1000 bp); *p*RPL13 (954 bp); *p*RPL9 (828 bp); *p*RPS25 (941 bp); *p*RPS10 (1000 bp) and *p*RPS13 (1276 bp). The construct *p*RPS10*LacZ* contains 5 codons of *RPS10* prior to *LacZ*. As a control the TUB promoter and its 5′ UTR were used to drive expression. The *LacZ* ORF is indicated as a grey arrow (not on scale; 3·1 kb). The localization of TRP-1 elements is indicated as open triangles and of TRP-2 elements as open circles. Presumptive transcriptional start sites of RP genes and their flanking genes are indicated by a thin arrow. Each construct was analysed with 3 independent CPRG assays. On the right the LacZ enzyme activity as determined for each construct is related to the average LacZ enzyme activity obtained by *p*TUB*LacZ*[BglII] and shown as an average percentage with its standard deviation. Results are shown as equimolar amounts of transfected plasmid DNA.

could not be concluded for occurrence of TRP-2 elements.

T. gondii *RP promoters have variable promoter strength*

To determine whether TRP elements could be correlated to gene expression, a study was performed in which 8 different RP genes were randomly selected which contained either TRP-1, TRP-2 or both TRP elements in their putative promoter regions, being *RPS3*, *RPS10*, *RPS13*, *RPS25*, *RPS29*, *RPL9*, *RPL13* and *RPL38*. Of these RP genes, upstream sequences together with their 5′ UTR were cloned immediately upstream of the *LacZ* reporter gene and compared with the *T. gondii* TUB promoter with 5′ UTR for their strength to drive expression. Using the *LacZ* reporter gene as readout, expression levels were determined with a CPRG assay. Three independent assays were performed in which each construct was transiently transfected in the *T. gondii* strain RHΔHXGPRT. These experiments showed reproducible expression patterns for each individual construct, which are shown in Fig. 3. As a positive control *p*TUB*LacZ* was used. This construct was cloned twice in which the 5′ UTR of TUB was separated from the *LacZ* startcodon by either an AvrII or a BglII restriction site. Although this region is important for translation efficiency (Seeber, 1997), the difference in restriction sites did not affect expression levels (see Fig. 3), allowing us to use both restriction sites in these two constructs for integration of RP sequences. LacZ was expressed at different levels, when driven by different RP promoters with their 5′ UTR. Highest expression was observed for the RPS13 promoter (including 5′ UTR), which was comparable in strength to the strong *T. gondii* TUB promoter, whereas 15-fold lower expression was obtained with the RPS29 promoter (including 5′ UTR) being the weakest tested RP promoter. These data showed that RP promoters and their 5′ UTR differ in their strength to drive

https://doi.org/10.1017/S0031182006009954 Published online by Cambridge University Press

expression of a heterologous gene. No direct correlation could be observed between expression levels and the presence of a TRP element in the upstream sequences for the analysed RPs.

DISCUSSION

Ribosomes are responsible for protein synthesis and as such are essential for growth in all living organisms (Warner, 1999; Meyuhas, 2000; Schaap *et al*. 2005). Synthesis of *de novo* ribosomes requires expression of 4 different rRNAs and a large set of RPs. The 4 different rRNAs were previously described for *T. gondii* (Guay *et al*. 1992; Gagnon *et al*. 1996). Here, the complete set of *T. gondii* cytoplasmic RPs is identified consisting of 79 different proteins. These proteins are highly similar in both numbers and protein sequence to higher eukaryotes like human, suggesting a conserved ribosome complexity.

Since synthesis of these ribosomal components consumes a large proportion of the cell's energy, it is tightly and coordinately regulated in eukaryotes. In most eukaryotes, synthesis of the set of RPs is primarily regulated at the level of translation (Meyuhas, 2000). In contrast, in *S. cerevisiae* and Coccidia like *T. gondii* and *E. tenella*, synthesis of RPs is regulated at the level of transcription and differs dependent on its life-cycle stage (Warner, 1999; Schaap *et al*. 2005). For *T. gondii* we previously showed that the complete set of RP genes is highly transcribed in tachyzoites, whereas transcripts were almost absent in the oocyst stages (Schaap *et al*. 2005). Since these gene products are functionally linked and also simultaneously transcribed, it was suggested that they are coordinately regulated.

Coordinated transcriptional control of a set of genes is usually regulated by one or more transcription factors binding to specific promoter elements. In addition, gene clusters exist in higher eukaryotes for functionally related genes, such as α-globin genes, β-globin genes, histone genes and Hox genes (van Driel *et al*. 2003). RPs are also functionally related proteins and in *T. gondii* 8 RP genes are located paired on the genome. The paired RP genes are arranged in a head to head orientation with small intergenic regions of 280–380 bps. Their small size suggests these regions may operate as bidirectional promoters. However, most RP genes (71 resp.) were randomly distributed in the genome and over the different *T. gondii* chromosomes suggesting that a conserved promoter structure should control the coordinate expression of these genes.

To study if transcription of the set of individual RP genes is coordinated in *T. gondii* by a combination of promoter elements and transcription factors, their upstream sequences were compared for conserved DNA elements. Since *T. gondii* RP promoters have not been defined before, 1000 bps of upstream sequences were selected for all RP genes. By comparative analysis two novel highly conserved DNA elements were identified in the upstream sequences of nearly all *T. gondii* RP genes, named TRP-1 (consensus TCGGCTTATATTCGG) and TRP-2 ([T/C]GCATGC[G/A]). Both elements were specifically localized in a region 10–330 bps upstream of the presumptive transcriptional start site of these RP genes. In addition, TRP-1 and TRP-2 elements were highly over-represented in these RP upstream regions compared to random chance, respectively 400- and 40-fold. Due to their specific localization these TRP elements likely operate as promoter elements in the coordinated transcriptional control of *T. gondii* RP genes.

Comparative analysis of upstream sequences of RP genes from the coccidian parasite *E. tenella* showed no enrichment for these TRP elements. Instead a different element was identified, being GGGCT-G[T/C]GGGG[G/C][G/T]GC (results not shown) which was similarly positioned to the TRP elements in *T. gondii*. These findings suggest that a comparable control mechanism may be present for transcription of RP genes in both parasites, but that genus specific DNA elements have evolved.

To determine if TRP elements are exclusively associated with RP genes, known *T. gondii* promoters were analysed, including GRA1, GRA3, GRA5, GRA6, GRA7, SAG1, TUB1. With one exception, TRP elements were not present in these promoters suggesting that TRP elements are not general promoter elements in *T. gondii*. In addition, a genome-wide analysis for TRP elements was performed, whereby we did not select for gene promoters only. In total 351 TRP-1 elements were detected within the *T. gondii* genome, suggesting that this element may operate in transcriptional control of a larger set of gene products, including RP genes. TRP-2 elements were detected 30856 times within the genome. This number is higher than the estimated 7588 genes present in the *T. gondii* genome (ref. Twinscan) being too high to function as a key element for regulation. Therefore, we expect that, apart from acting in combination with transcription factors, TRP-2 elements may operate in combination with additional gene control such as DNA accessibility. It is well known that chromatin remodelling by histone modifications (acetylation, methylation, phosphorylation) can alter DNA accessibility and thereby influences gene control by the transcription machinery (Horn and Peterson, 2002).

The identification of TRP elements in the set of *T. gondii* RP genes allowed us to further investigate their role. In a first study, the promoter strength of upstream sequences of 8 different RP genes was analysed and correlated with the presence of TRP-1 and/or TRP-2. No direct correlation could be detected between RP promoter strength and the presence of any of the TRP elements. Remarkably, up to a 15-fold difference in strength was observed

https://doi.org/10.1017/S0031182006009954 Published online by Cambridge University Press

between the *T. gondii* RP promoters. Since ribosome assembly is based on equimolar usage of RPs, another level of regulation must be involved to maintain the stoichiometry among the *T. gondii* ribosomal components. Such regulation could operate at different levels including mRNA processing and stability, translational efficiency and protein turnover. In this respect it may be relevant that RP genes in *T. gondii* frequently contain introns in their 5′ UTRs. In *S. cerevisiae* such introns were suggested to be involved in autoregulation (Warner *et al*. 1985).

Since *S. cerevisiae* also transcriptionally regulates RP genes, it is informative to compare *T. gondii* with *S. cerevisiae*. In *S. cerevisiae* RP transcription is dependent on external stimuli such as carbon sources and nutrients, which can trigger signal transduction pathways and thereby activate or inhibit the function of transcription factors. Complicated transcriptional networks are present in *S. cerevisiae* whereby multiple transcription factors (including Fhl1, Rap1, Yap5, Crf1, Ifh1) bind to different promoter elements, that control transcriptional induction or silencing of *S. cerevisiae* RP genes (Lee *et al*. 2002; Martin *et al*. 2004; Wade *et al*. 2004). In *S. cerevisiae* the identified promoter elements are present in many, but not in all RP gene promoters, similar to the presence of TRP elements in *T. gondii* RP genes. Thus, a set of multiple promoter elements and accompanying transcription factors is required for coordinated regulation of RP transcription. Moreover, in *S. cerevisiae* the conserved promoter elements in RP genes are also not limited to the select group of RP genes but appeared to be involved in transcriptional control of several other genes as well. No direct homologues of the above-described yeast transcription factors could be identified in the *T. gondii* genome. However, the identification of TRP elements in *T. gondii* should allow the characterization of their respective transcription factors.

In summary, the complete set of *T. gondii* RPs has been identified. Their upstream sequences are enriched with TRP-1 and TRP-2 elements which are specifically localized in front of the transcriptional start sites and therefore expected to be involved in coordinated transcription of RP genes. The identification of these elements creates a basis to further study the underlying mechanism by which RP transcription is controlled in *T. gondii*. Since ribosome biosynthesis is directly linked to cell growth in all organisms, understanding the transcriptional control of *T. gondii* RP genes will simultaneously explain how growth is regulated at the molecular level in parasites.

We would like to acknowledge Dr Erik de Vries for help with the genome analysis of TRP elements. EST data were obtained from NCBI at http://www.ncbi.nlm.nih.gov. Preliminary genomic and/or cDNA sequence data was accessed via http://ToxoDB.org and/or http://www.tigr.org/tdb/t_gondii/. Genomic data were provided by The Institute for Genomic Research (supported by the NIH grant no. AI05093), and by the Sanger Center (Wellcome Trust). EST sequences were generated by Washington University (NIH grant no. 1R01AI045806-01A1).

REFERENCES

Ajioka, J. W., Boothroyd, J. C., Brunk, B. P., Hehl, A., Hillier, L., Manger, I. D., Marra, M., Overton, G. C., Roos, D. S., Wan, K. L., Waterston, R. and Sibley, L. D. (1998). Gene discovery by EST sequencing in *Toxoplasma gondii* reveals sequences restricted to the Apicomplexa. *Genome Research* **8**, 18–28.

Donald, R. G., Carter, D., Ullman, B. and Roos, D. S. (1996). Insertional tagging, cloning, and expression of the *Toxoplasma gondii* hypoxanthine-xanthine-guanine phosphoribosyltransferase gene. Use as a selectable marker for stable transformation. *Journal of Biological Chemistry* **271**, 14010–14019.

Gagnon, S., Bourbeau, D. and Levesque, R. C. (1996). Secondary structures and features of the 18S, 5.8S and 26S ribosomal RNAs from the Apicomplexan parasite *Toxoplasma gondii*. *Gene* **173**, 129–135.

Guay, J. M., Huot, A., Gagnon, S., Tremblay, A. and Levesque, R. C. (1992). Physical and genetic mapping of cloned ribosomal DNA from *Toxoplasma gondii*: primary and secondary structure of the 5S gene. *Gene* **114**, 165–171.

Horn, P. J. and Peterson, C. L. (2002). Molecular biology. Chromatin higher order folding–wrapping up transcription. *Science* **297**, 1824–1827.

Ju, Q. and Warner, J. R. (1994). Ribosome synthesis during the growth cycle of *Saccharomyces cerevisiae*. *Yeast* **10**, 151–157.

Kenmochi, N., Kawaguchi, T., Rozen, S., Davis, E., Goodman, N., Hudson, T. J., Tanaka, T. and Page, D. C. (1998). A map of 75 human ribosomal protein genes. *Genome Research* **8**, 509–523.

Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., Zeitlinger, J., Jennings, E. G., Murray, H. L., Gordon, D. B., Ren, B., Wyrick, J. J., Tagne, J. B., Volkert, T. L., Fraenkel, E., Gifford, D. K. and Young, R. A. (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799–804.

Li, L., Brunk, B. P., Kissinger, J. C., Pape, D., Tang, K., Cole, R. H., Martin, J., Wylie, T., Dante, M., Fogarty, S. J., Howe, D. K., Liberator, P., Diaz, C., Anderson, J., White, M., Jerome, M. E., Johnson, E. A., Radke, J. A., Stoeckert, C. J. Jr**, Waterston, R. H., Clifton, S. W., Roos, D. S. and Sibley, L. D.** (2003). Gene discovery in the apicomplexa as revealed by EST sequencing and assembly of a comparative gene database. *Genome Research* **13**, 443–454.

Mager, W. H., Planta, R. J., Ballesta, J. G., Lee, J. C., Mizuta, K., Suzuki, K., Warner, J. R. and Woolford, J. (1997). A new nomenclature for the cytoplasmic ribosomal proteins of *Saccharomyces cerevisiae*. *Nucleic Acids Research* **25**, 4872–4875.

Martin, D. E., Soulard, A. and Hall, M. N. (2004). TOR regulates ribosomal protein gene expression via PKA

and the Forkhead transcription factor FHL1. *Cell* **119**, 969–979.

Meyuhas, O. (2000). Synthesis of the translational apparatus is regulated at the translational level. *European Journal of Biochemistry* **267**, 6321–6330.

Pearson, N. J. and Haber, J. E. (1980). Changes in regulation of ribosomal protein synthesis during vegetative growth and sporulation of *Saccharomyces cerevisiae*. *Journal of Bacteriology* **143**, 1411–1419.

Planta, R. J. and Mager, W. H. (1998). The list of cytoplasmic ribosomal proteins of *Saccharomyces cerevisiae*. *Yeast* **14**, 471–477.

Schaap, D., Arts, G., van Poppel, N. F. J. and Vermeulen, A. N. (2005). De novo ribosome biosynthesis is transcriptionally regulated in *Eimeria tenella*, dependent on its life cycle stage. *Molecular and Biochemical Parasitology* **139**, 239–248.

Seeber, F. (1997). Consensus sequence of translational initiation sites from *Toxoplasma gondii* genes. *Parasitology Research* **83**, 309–311.

Seeber, F. and Boothroyd, J. C. (1996). *Escherichia coli* beta-galactosidase as an in vitro and in vivo reporter enzyme and stable transfection marker in the intracellular protozoan parasite *Toxoplasma gondii*. *Gene* **169**, 39–45.

Soldati, D. and Boothroyd, J. C. (1993). Transient transfection and expression in the obligate intracellular parasite *Toxoplasma gondii*. *Science* **260**, 349–352.

Striepen, B., He, C. Y., Matrajt, M., Soldati, D. and Roos, D. S. (1998). Expression, selection, and organellar targeting of the green fluorescent protein in *Toxoplasma gondii*. *Molecular and Biochemical Parasitology* **92**, 325–338.

van Driel, R., Fransz, P. F. and Verschure, P. J. (2003). The eukaryotic genome: a system regulated at different hierarchical levels. *Journal of Cell Science* **116**, 4067–4075.

Wade, J. T., Hall, D. B. and Struhl, K. (2004). The transcription factor Ifh1 is a key regulator of yeast ribosomal protein genes. *Nature, London* **432**, 1054–1058.

Warner, J. R., Mitra, G., Schwindinger, W. F., Studeny, M. and Fried, H. M. (1985). *Saccharomyces cerevisiae* coordinates accumulation of yeast ribosomal proteins by modulating mRNA splicing, translational initiation, and protein turnover. *Molecular and Cellular Biology* **5**, 1512–1521.

Warner, J. R. (1999). The economics of ribosome biosynthesis in yeast. *Trends in Biochemical Sciences* **24**, 437–440.

Yoshihama, M., Uechi, T., Asakawa, S., Kawasaki, K., Kato, S., Higa, S., Maeda, N., Minoshima, S., Tanaka, T., Shimizu, N. and Kenmochi, N. (2002). The human ribosomal protein genes: sequencing and comparative analysis of 73 genes. *Genome Research* **12**, 379–390.