Ribosomal RNA Transcription Machineries in Intestinal Protozoan Parasites: A Bioinformatic Analysis

Francisco Alejandro Lagunas-Rangel

Abstract

Purpose Ribosome biogenesis is a key process in all living organisms, energetically expensive and tightly regulated. Currently, little is known about the components of the ribosomal RNA (rRNA) transcription machinery that are present in intestinal parasites, such as *Giardia duodenalis*, *Cryptosporidium parvum*, and *Entamoeba histolytica*. Thus, in the present work, an analysis was carried out looking for the components of the rRNA transcription machinery that are conserved in intestinal parasites and if these could be used to design new treatment strategies.

Methods The different components of the rRNA transcription machinery were searched in the studied parasites with the NCBI BLAST tool in the EuPathDB Bioinformatics Resource Center database. The sequences of the RRN3 and POLR1F orthologs were aligned and important regions identified. Subsequently, three-dimensional models were built with different bioinformatic tools and a structural analysis was performed.

Results Among the protozoa examined, *C. parvum* is the parasite with the fewest identifiable components of the rRNA transcription machinery. TBP, RRN3, POLR1A, POLR1B, POLR1C, POLR1D, POLR1F, POLR1H, POLR2E, POLR2F and POLR2H subunits were identified in all species studied. Furthermore, the interaction regions between RRN3 and POLR1F were found to be conserved and could be used to design drugs that inhibit rRNA transcription in the parasites studied.

Conclusion The inhibition of the rRNA transcription machinery in parasites might be a new therapeutic strategy against these microorganisms.

Keywords RNA polymerase I · SL1 · UBF · RRN3 · POLR1F

Introduction

Intestinal protozoal infections are a major health problem, especially in developing countries, where poor household hygiene practices, inadequate sanitary facilities, and low socioeconomic conditions favor their spread [1]. In particular, protozoa are responsible for important intestinal diseases in humans, with high morbidity and, in some cases, mortality [2]. *Giardia duodenalis* and *Cryptosporidium parvum* are the most common pathogenic intestinal protozoan parasites with an annual incidence of about 10,000 cases each in the United States and Europe alone, whereas for *Entamoeba histolytica*, a worldwide annual incidence of 100 million cases is estimated [3]. In general, these parasites cause poor digestion, impair absorption and increase nutrient loss, among other things. Indeed, even in asymptomatic infections, subtle damage and disturbances of intestinal function may occur [4]. One aspect to highlight is the increase in treatment failure and the appearance of strains resistant to current drugs due to their massive and inappropriate use, which has led us to the need to devise new treatment strategies [2, 5, 6].

On the other hand, the efficient growth and proliferation of parasites require a balanced production of ribosomes for protein synthesis [7–10]. Notably, the rate-limiting step of ribosome biogenesis is the synthesis of ribosomal RNA (rRNA) by RNA polymerase I (Pol I) [10–12]. The rRNA transcription machinery comprises three main components: the Pol I enzyme, the TBP (TATA-binding protein)-TAF (TBP-associated factor) complex SL1 (selectivity factor 1) and the trans-activator protein UBF (upstream binding factor) [13]. Currently, little is known about the components of the rRNA transcription machinery that are present...
in intestinal parasites, but it is known that if any of these do not function properly, the parasites die due to cell cycle arrest and apoptosis [14–17]. In this sense, the rRNA transcription machinery becomes a feasible target for the design of new anti-parasitic drugs. The proposal of this work was to identify the putative components of the ribosomal RNA transcription machinery in the three most prominent intestinal protozoan pathogens, *G. duodenalis*, *C. parvum*, and *E. histolytica*. Furthermore, special emphasis was placed on the interaction between RRN3 and POLR1F, which is a key step to link Pol I with the rest of the components of the transcriptional machinery and where anti-parasitic drugs might be designed.

**Materials and Methods**

**Database Screening**

The amino acid sequences of the proteins involved in the initiation of rRNA transcription in humans were obtained from the UniProt Knowledgebase (UniProtKB) [18] using the name of each protein. To find orthologs of human proteins, the whole genome sequences of the intestinal parasites *G. duodenalis* (Assemblage A_isolate_WB), *C. parvum* (Iowa II) and *E. histolytica* (HM1-IMSS) were examined in the corresponding databases of the EuPathDB Bioinformatics Resource Center [19] and using the NCBI BLAST tool [20] with default search parameters. The rRNA transcription machinery of *Saccharomyces cerevisiae* and the genome of *Entamoeba dispar* were also analyzed for comparative purposes.

**Sequence Analysis**

The structural domains of the rRNA transcription machinery proteins present in all organisms were predicted and analyzed using InterPro [21]. Multiple sequence alignments of RRN3 orthologs and POLR1F orthologs were performed using Clustal Omega in CLC Genomics Workbench 21 (Qiagen Bioinformatics, Aarhus C, Denmark). Based on these alignments, the identity and similarity percentages between the orthologs were calculated. Three-dimensional structures were predicted using SWISS-MODEL [22], and illustrations were made using UCSF Chimera software [23].

**Results**

**TBP is the Only Subunit of the TBP-TAF Complex SL1 that has Identifiable Orthologs in All Species Analyzed. Pol I-Specific Factor RRN3 is also Conserved**

The only component of the TBP-TAF complex SL1 that was found in all the organisms analyzed was TBP, with one ortholog in *G. duodenalis*, two in *C. parvum* and three in both *E. histolytica* and *E. parvum*. Since TAF12 orthologs were identified only in *E. histolytica* and *E. dispar*, *G. duodenalis* and *C. parvum* were the species with the fewest identifiable subunits of the SL1 transcription factor. RRN3 orthologs were found in the genome of all organisms, but the sequences of *E. histolytica* and *E. dispar* orthologs diverged widely. Regarding the UBF transcription activator, two orthologs were identified in *G. duodenalis* and three in both *E. histolytica* and *E. dispar*. No orthologs for this protein were found in *C. parvum*. The summary of the results is presented in Tables 1 and 2.

**Intestinal Protozoan Parasites Lack Orthologs of the POLR1E, POLR1G, and RPA3 Subunits in RNA Polymerase I Complex**

Ortholog search for the 14 major subunits of RNA polymerase I was also performed. Thus, it was found that for the subunits POLR1A, POLR1B, POLR1C, POLR1D, POLR1F, POLR1H, POLR2E, POLR2F and POLR2H there is at least one ortholog in each species. In contrast, for the POLR1E, POLR1G and RPA3 subunits, no orthologs were found in any of the analyzed parasites. Orthologs of the POLR2K subunit were identified in *G. duodenalis* and *C. parvum*, but not in *E. histolytica* and *E. dispar*. Interestingly, no orthologs were found for the POLR2L subunit in *E. dispar*. In this way, *E. dispar* is the organism with the fewest identifiable components of RNA polymerase I. The summary of the results is presented in Tables 1 and 2.

**The Conserved Subunits of the rRNA Transcription Machinery Show Differences Between Them**

Figure 1 shows the rRNA transcription machinery in each species analyzed according to our bioinformatic analysis. All the orthologs of TBP, POLR1D and POLR1F identified in parasites were proteins smaller than those in humans. The reduction in the number of amino acids was between 26.5% and 42.5% for TBP, between 15.8% and 26.3% for POLR1D, and between 24.3% and 37.3% for POLR1F. Except for *C. parvum*, the RRN3 orthologs had between 14.6% and 28.4% fewer amino acids than the human protein, but RRN3 HEAT repeats are conserved in all species studied. POLR1A orthologs in *G. duodenalis* and *C. parvum* had between 4.9% and 24.5% more amino acids than the human protein, but for the *E. histolytica* and *E. dispar* orthologs the number of amino acids was between 4.9% and 19.8% less than the human counterpart. Meanwhile, POLR1B orthologs from *G. duodenalis* and *C. parvum* were proteins with 6% to 13.4% more amino acids than human protein, but almost...
### Table 1: Prediction of rRNA transcription machineries in *Giardia duodenalis* and *Cryptosporidium parvum*

| Homo sapiens | Saccharomyces cerevisiae | *Giardia duodenalis* (Assemblage A_isolate_WB) | Cryptosporidium parvum (Iowa II) |
|---------------|--------------------------|-----------------------------------------------|----------------------------------|
| **Gene**      | **UniProtKB\(^a\)**    | **Size\(^b\)** | **Gene** | **UniProtKB\(^a\)** | **Size\(^b\)** | **E-value** | **Gene** | **UniProtKB\(^a\)** | **Size\(^b\)** | **E-value** |
| TBP-TAF complex SL1 |                 |               |          |                     |               |           |          |                     |               |           |
| TBP           | P62380       | 339           | TBP      | P13393    | 240           | GL50803_1721 | E2R70    | 200       | 6e−05      | 25          | cgd8_2030   | Q6SEL4   | 249       | 4e−67      | 51          |
| TAF1C         | Q15572       | 869           | Rm6      | P32786    | 894           | –               | –           | –               | –               | –               | cgd8_210   | Q5CPZ4   | 195       | 7e−14      | 32          |
| TAF1B         | Q53T94       | 588           | Rm7      | P40992    | 514           | –               | –           | –               | –               | –               | –               | –               | –               | –               | –               |
| TAF1A         | Q15573       | 450           | Rm11     | Q04712    | 507           | –               | –           | –               | –               | –               | –               | –               | –               | –               | –               |
| TAF1D         | Q9H5J8       | 278           | –        | –         | –               | –               | –           | –               | –               | –               | –               | –               | –               | –               | –               |
| TAF12         | Q16514       | 161           | –        | –         | –               | –               | –           | –               | –               | –               | –               | –               | –               | –               | –               |
| RNR3          | Q0NY6Y       | 651           | Rm3      | P36070    | 627           | GL50803_11742  | A8B8V3   | 556       | 2e−05      | 23          | cgd6_2810   | Q5CX25   | 882       | 1e−04      | 27          |
| **Transcription activator UBF** |             |               |          |           |               |               |           |          |                     |               |           |
| UBF           | P17480       | 764           | UAF3O   | Q08747    | 228           | GL50803_17626  | A8BW36   | 204       | 1e−05      | 30          | –               | –               | –               | –               | –               | –               |
|              |              |               | Rm5     | Q02983    | 363           | GL50803_3349   | A8BDD1   | 183       | 5e−05      | 30          | –               | –               | –               | –               | –               | –               |
|              |              |               | Rm9     | P53437    | 365           | –               | –           | –               | –               | –               | –               | –               | –               | –               | –               |
|              |              |               | Rm10    | P38204    | 145           | –               | –           | –               | –               | –               | –               | –               | –               | –               | –               |
| **RNA polymerase I complex** |             |               |          |           |               |               |           |          |                     |               |           |
| POLR1A        | Q95602       | 1720          | A190    | P10964    | 1664          | GL50803_16223  | A8B4P1   | 1714      | 1e−78      | 28          | cgd3_2620   | Q5CU2    | 1895      | 9e−109     | 35          |
| POLR1B        | Q9F9Y6       | 1135          | A135    | P22138    | 1203          | GL50803_23496  | A8B7R7   | 2142      | 3e−70      | 26          | cgd6_3290   | Q5CWY0   | 1871      | 2e−102     | 35          |
| POLR1C        | O15160       | 346           | AC40    | P07703    | 335           | GL50803_17187  | A8B2N4   | 1288      | 3e−102     | 30          | cgd7_3720   | Q5CY45   | 1177      | 3e−122     | 29          |
| POLR1D        | P0DPB6       | 133           | AC19    | P28000    | 142           | GL50803_10055  | A8B9C8   | 350       | 5e−47      | 33          | cgd8_300    | Q5CPY6   | 344       | 5e−69      | 40          |
| POLR1E        | Q9GZS1       | 414           | A49     | Q01080    | 415           | GL50803_7474   | A2RY7    | 326       | 5e−16      | 28          | cgd1_2710   | Q5CH3    | 355       | 5e−29      | 29          |
| POLR1F        | Q3B726       | 338           | A43     | P46669    | 326           | GL50803_17422  | A8BD91   | 229       | 1e−08      | 24          | cgd1_1620   | Q5CSR5   | 256       | 7e−07      | 27          |
| POLR1G        | Q15446       | 510           | A34.5   | P47006    | 233           | GL50803_8518   | A8BBB2   | 123       | 1e−12      | 41          | cgd7_503    | F0X620   | 106       | 7e−08      | 27          |
| POLR1H        | Q9PIU0       | 126           | A12.2   | P32529    | 125           | GL50803_10840  | A8BTA8   | 101       | 1e−15      | 59          | cgd4_3200   | A3FQL8   | 98        | 4e−15      | 60          |
| POLR2E        | P19388       | 210           | Rbp5    | P20434    | 215           | GL50803_137609 | E2R58    | 229       | 7e−26      | 32          | cgd2_980    | Q5CTZ0   | 205       | 7e−62      | 45          |
| POLR2F        | P61218       | 127           | Rbp6    | P20435    | 155           | GL50803_15955  | E2RT0    | 104       | 4e−26      | 32          | cgd7_4770   | Q5CVX6   | 129       | 4e−36      | 72          |
| POLR2H        | P52434       | 150           | Rbp8    | P20436    | 146           | GL50803_15144  | E2R32    | 150       | 1e−07      | 27          | cgd1_2260   | Q5CSK9   | 144       | 3e−22      | 35          |
| POLR2K        | P53803       | 58            | Rbp7    | P40422    | 70            | GL50803_9509   | E2R91    | 54        | 1e−04      | 33          | cgd7_3240   | A3FP1    | 71        | 9e−12      | 43          |
| POLR2L        | P62875       | 67            | Rbp10   | P22139    | 70            | GL50803_14413  | A8BAL6   | 122       | 1e−22      | 59          | cgd4_3260   | A3FQL9   | 72        | 1e−28      | 66          |
| RPA3          | P35244       | 121           | RPA14   | P50106    | 137           | –               | –           | –               | –               | –               | –               | –               | –               | –               | –               |

\(^a\) Access number in the UniProt Knowledgebase  
\(^b\) Number of amino acids  
\(^{c}\) (Identity) values expressed in percentage (%)
Table 2  Prediction of rRNA transcription machineries in *Entamoeba histolytica* and *Entamoeba dispar*

| Homo sapiens | Saccharomyces cerevisiae | *Entamoeba histolytica* (HM1-IMSS) | E-value | Entamoeba dispar (SAW760) |
|--------------|--------------------------|-----------------------------------|---------|--------------------------|
| **Gene**     | **UniProtKB**a | **Size**b | **Gene**     | **UniProtKB**a | **Size**b | **E-value** | **Gene**     | **UniProtKB**a | **Size**b | **E-value** |
| TBP-TAF complex SL1 | | | | | | | | |
| TBP          | P62380       | 339   | TBP          | P13393       | 240   | EHI_077240  | A7UCF2       | 216   | 2e−66 | 56 | EDI_292240  | B0EHD5       | 216   | 6e−66 | 56 |
|              |             |       |              |              |       | EHI_112050  | C4M7H7       | 212   | 2e−64 | 56 | EDI_260400  | B0EIX2       | 212   | 1e−64 | 56 |
|              |             |       |              |              |       | EHI_020610  | P52653       | 234   | 7e−64 | 56 | EDI_172550  | B0EIP7       | 234   | 4e−64 | 56 |
| TAF1C        | Q15572       | 869   | Rrn6         | P32786       | 894   | –           | –           | –     | –     | –   | –             | –            | –     | –     | –   |
| TAF1B        | Q53T94       | 588   | Rrn7         | P40992       | 514   | –           | –           | –     | –     | –   | –             | –            | –     | –     | –   |
| TAF1A        | Q15573       | 450   | Rrn11        | Q04712       | 507   | –           | –           | –     | –     | –   | –             | –            | –     | –     | –   |
| TAF1D        | Q9H5J8       | 278   | –            | –            | –     | –           | –           | –     | –     | –   | –             | –            | –     | –     | –   |
| TAF12        | Q16514       | 161   | –            | –            | –     | –           | –           | –     | –     | –   | –             | –            | –     | –     | –   |
| RRN3         | Q9NYV6       | 651   | Rrn3         | P36070       | 627   | EHI_035130  | C4M4Z3       | 466   | 3e−25d | 15 | EDI_198050  | B0EPH1       | 466   | 3e−25d | 15 |
| Transcription activator UBF | | | | | | | | | |
| UBF          | P17480       | 764   | UAF30        | Q08747       | 228   | EHI_045480  | C4LT9F       | 111   | 3e−6  | 31 | EDI_340970  | B0EV32       | 111   | 3e−6  | 31 |
|              |             |       | Rrn5         | Q02983       | 363   | EHI_093800  | C4LYH1       | 114   | 5e−6  | 30 | EDI_049480  | B0EFF8       | 112   | 3e−6  | 31 |
|              |             |       | Rrn9         | P53437       | 365   | EHI_179340  | C4MX9X4      | 384   | 4e−4  | 35 | EDI_110640  | B0EK44       | 395   | 3e−4  | 35 |
|              |             |       | Rrn10        | P38204       | 145   | EHI_130080  | C4L089A2     | 470   | 5e−4  | 35 | EDI_085740  | B0EF23       | 106   | 6e−4  | 40 |
| RNA polymerase I complex | | | | | | | | | |
| POLR1A       | Q95602       | 1720  | A190         | P10964       | 1664  | EHI_095890  | C4M3E6       | 1570  | 5e−106| 36 | EDI_337480  | B0ED6        | 1568  | 4e−102| 36 |
|              |             |       |              |              |       | EHI_121760  | Q6IUR3       | 1636  | 5e−106| 30 | EDI_169920  | B0E6R8       | 1587  | 4e−102| 30 |
|              |             |       |              |              |       | EHI_125350  | C4M626       | 1379  | 3e−97 | 35 | EDI_116030  | B0E6G6       | 1379  | 5e−98 | 35 |
| POLR1B       | Q9H9Y6       | 1135  | A135         | P22138       | 1203  | EHI_180620  | C4MAB3       | 1106  | 1e−123| 39 | EDI_303700  | B0ERH0       | 1106  | 2e−124| 40 |
|              |             |       |              |              |       | EHI_095860  | C4M3E3       | 1122  | 1e−123| 31 | EDI_337350  | B0ED3        | 1075  | 2e−124| 32 |
|              |             |       |              |              |       | EHI_022940  | C4LUK7       | 1170  | 1e−119| 28 | EDI_248550  | B0EG45       | 1170  | 1e−119| 28 |
Table 2 (continued)

| Homo sapiens | Saccharomyces cerevisiae | Entamoeba histolytica (HM1-IMSS) | Entamoeba dispar (SAW760) |
|--------------|--------------------------|---------------------------------|--------------------------|
| Gene         | UniProtKB<sup>a</sup>   | Size<sup>b</sup> | Gene | UniProtKB<sup>a</sup> | Size<sup>b</sup> | E-value | f<sup>c</sup> | Gene | UniProtKB<sup>a</sup> | Size<sup>b</sup> | E-value | f<sup>c</sup> |
| POLR1C       | O15160                   | 346               | AC40 | P07703                  | 335               | EHI_178010 | C4M554   | 283  | 5e−24     | 27       | EDI_002920 | B0EJB9  | 291 | 1e−73  | 44     |
| POLR1D       | P0DPB6                   | 133               | AC19 | P28000                  | 142               | EHI_087360 | C4LXS9   | 112  | 9e−12     | 34       | EDI_085890 | B0E866  | 112 | 3e−11  | 33     |
| POLR1E       | Q9GZS1                   | 414               | A49  | Q01080                  | 415               | –            | –        | –    | –         | –        | –            | –      | –    | –      | –      |
| POLR1F       | Q3B726                   | 338               | A43  | P46669                  | 326               | EHI_124360 | B1N352   | 212  | 2e−05     | 29       | EDI_023800 | B0EHA3  | 212 | 2e−05  | 28     |
| POLR1G       | O15446                   | 510               | A34.5| P47006                  | 233               | –            | –        | –    | –         | –        | –            | –      | –    | –      | –      |
| POLR1H       | Q9P1U0                   | 126               | A12.2| P32529                  | 125               | EHI_044620 | C4LT84   | 122  | 1e−16     | 40       | EDI_323220 | B0EEN5  | 122 | 1e−16  | 40     |
| POLR2E       | P19388                   | 210               | Rpb5 | P20434                  | 215               | EHI_142090 | C4LW54   | 204  | 1e−47     | 39       | EDI_338140 | B0EA74  | 204 | 6e−48  | 39     |
| POLR2F       | P61218                   | 127               | Rpb6 | P20435                  | 155               | EHI_088230 | C4M6S1   | 122  | 2e−32     | 73       | EDI_088050 | B0EM77  | 122 | 6e−32  | 72     |
| POLR2H       | P52434                   | 150               | Rpb8 | P20436                  | 146               | EHI_038570 | C4LZP1   | 143  | 6e−15     | 31       | EDI_259550 | B0EB74  | 143 | 3e−15  | 31     |
| POLR2K       | P53803                   | 58                | Rpb12| P40422                  | 70                | –            | –        | –    | –         | –        | –            | –      | –    | –      | –      |
| POLR2L       | P62875                   | 67                | Rpb10| P22139                  | 70                | EHI_122780 | C4M5M4   | 73   | 2e−19     | 56       | –            | –      | –    | –      | –      |
| RPA3         | P35244                   | 121               | RPA14| P50106                  | 137               | –            | –        | –    | –         | –        | –            | –      | –    | –      | –      |

<sup>a</sup>Access number in the UniProt Knowledgebase

<sup>b</sup>Number of amino acids

<sup>c</sup><i>I</i> (Identity) values expressed in percentage (%)

<sup>d</sup>Srivastava et al. [31]
all orthologs from *E. histolytica* and *E. dispar* had similar amounts. In contrast, POLR1C orthologs in *G. duodenalis* and *C. parvum* maintained a similar number of amino acids as human protein, but orthologs in *E. histolytica* and *E. dispar* were proteins with 15.9% to 18% less amino acids. Most of the identified orthologs of POLR1H, POLR2E, POLR2F, and POLR2H in parasites were very similar in size to human and yeast proteins. A schematic representation of these data appears in Supplementary Fig. 1.
The RRN3 and POLR1F Orthologs Maintain Important Residues for Their Interaction in All the Species Studied

Despite poor sequence conservation (Table 3), sequence alignments of POLR1F orthologs showed that the region mediating its interaction with the RRN3 subunit is conserved in intestinal protozoan parasites (Fig. 2A). Furthermore, three-dimensional predictions of these parasitic proteins revealed strong structural similarity to their human and yeast counterparts, where the binding area remains exposed in all cases to facilitate their interaction with RRN3 (Fig. 2B).

On the other hand, sequence alignment analysis of the RRN3 orthologs also revealed low conservation (Table 3), but residues of a serine patch that serve this protein to bind POLR1F had high conservation (Fig. 3A), particularly those corresponding to residues S101, S102 and S185 of yeast RRN3. With the three-dimensional predictions, a high structural similarity of the RRN3 orthologs was observed, where the characteristic HEAT repeat fold (repeats of alpha helices joined by a short loop) is maintained. In addition, the identified serine patch residues are exposed in all cases and would allow their interaction with POLR1F orthologs (Fig. 3B). These findings suggest that the interaction points between RRN3 and POLR1F are conserved in different species, including intestinal protozoan parasites.

Discussion

Transcription of rRNA by Pol I is the key regulatory step in ribosome production and is tightly controlled by an intricate network of signaling pathways and epigenetic mechanisms [24]. The transcription by Pol I requires the formation of a preinitiation complex (PIC) that directs promoter-specific transcription of rDNA and whose components are the Pol I enzyme, the TBP-TAF complex SL1 and UBF [25]. The only subunit of the TBP-TAF complex SL1 that was identified in all the species analyzed was TBP. This responds to the fact that TBP is considered the most conserved initiation factor in archaeo-eukaryotic transcription initiation complexes [26]. No orthologs of the three Pol I-specific TAFs were identified, but all species have other members of the TFIID (transcription factor II D) family protein encoded in their genome that could carry out this function. UBF plays an essential role in maintaining a state of euchromatin on rDNA and enhancing rRNA expression [27], and that is why it is interesting that in C. parvum, a UBF ortholog was not found. UBF is not essential for the initiation of transcription in vitro, but it is essential for the formation of PIC in vivo and functions in the pre- and post-initiation steps [25, 27]. However, the genome of C. parvum exhibits other proteins with HMG (high mobility group) boxes whose specific function has not yet been characterized. Almost all the core

Table 3  Identity and similarity between the orthologs of RRN3 and POLRF1 in the species analyzed

|      | H. sapiens Q9NYV6 | S. cerevisiae P36070 | G. duodenalis A8B8V3 | C. parvum Q5CX25 | E. histolytica C4M4Z3 | E. dispar B0EPH1 |
|------|-------------------|----------------------|----------------------|------------------|---------------------|-----------------|
| RRN3 | H. sapiens Q9NYV6 | 20.06                | 12.82                | 11.37            | 6.09                | 5.72            |
|      | S. cerevisiae P36070 | 35.62               | 13.29                | 10.11            | 4.06                | 4.06            |
|      | G. duodenalis A8B8V3 | 27.56               | 27.89                | 7.89             | 5.22                | 4.81            |
|      | C. parvum Q5CX25 | 23.88                | 23.05                | 18.29            | 8.92                | 8.92            |
|      | E. histolytica C4M4Z3 | 12.94              | 11.53                | 14.97            | 17.94               | 91.20           |
|      | E. dispar B0EPH1 | 12.56                | 11.15                | 14.01            | 18.05               | 96.14           |

|      | H. sapiens Q3B726 | S. cerevisiae P46669 | G. duodenalis A8BD91 | C. parvum Q5CSR5 | E. histolytica B1N352 | E. dispar B0EHA3 |
|------|-------------------|----------------------|----------------------|------------------|---------------------|-----------------|
| POLR1F | H. sapiens Q3B726 | 13.98                | 10.12                | 13.30            | 10.54               | 10.26           |
|      | S. cerevisiae P46669 | 29.29               | 10.51                | 10.40            | 7.76                | 8.05            |
|      | G. duodenalis A8BD91 | 26.01               | 24.32                | 19.18            | 11.15               | 11.15           |
|      | C. parvum Q5CSR5 | 23.14                | 21.07                | 21.25            | 27.72               | 98.58           |
|      | E. histolytica B1N352 | 20.51              | 18.97                | 21.25            | 27.72               | 98.58           |
|      | E. dispar B0EHA3 | 20.51                | 18.97                | 21.25            | 27.72               | 100             |

Cells above and to the right of the central diagonal indicate percent amino acid identity, while cells below and to the left indicate percent similarity.
subunits of Pol I were identified in intestinal protozoan parasites, only in *E. histolytica* no orthologs of POLR2K were found and in *E. dispar* orthologs of POLR2K and POLR2L. These last two mentioned subunits are shared with the other two DNA-directed RNA polymerases (RNA Pol I and RNA Pol III). Notably, the heterodimer POLR1E-POLR1G, which is required for RNA elongation by Pol I [28], was not found in any of the analyzed parasite species. Therefore, the question arises how these organisms carry out rRNA elongation. The heterodimer formed by the POLR1F and RPA3 subunits plays a role in recruiting Pol I to the promoter region [29]. In the analysis, only orthologs were found for the POLR1F subunit, thus all intestinal protozoan parasites lacking identifiable RPA3 orthologs. RPA3 prevents DNA rehybridization during transcription and, in parallel, recruits and activates different proteins and complexes [30]. For proteins in which orthologs were not found, it does not necessarily mean that these are not present in these organisms. It may be that their identity is very low and a special search is required to find them, as was the case with RRN3 of *E. histolytica* [31]. Using the BLAST tool with default search parameters, RRN3 orthologs were found in the genome of

![Fig. 2](image_url)

**Fig. 2** POLR1F orthologs in intestinal parasites preserve the region that interacts with the RRN3 subunit. A Sequence alignments of the POLR1F orthologs identified in the different species studied. The degree of conservation is shown in colors and a red line is placed to indicate the amino acids that are known to mediate the interaction of this subunit with RRN3. Predicted structures of human POLR1F (B) and their orthologs in *Saccharomyces cerevisiae* (C), *Giardia duodenalis* (D), *Cryptosporidium parvum* (E), *Entamoeba histolytica* (F), and *Entamoeba dispar* (G). The region that interacts with RRN3 is marked with red arrows (color figure online)
(a) Conservation

(b) *H. sapiens* Q9NY6

(c) *S. cerevisiae* P36070

(d) *G. duodenalis* A88BV3

(e) *C. parvum* Q5CX25

(f) *E. histolytica* Q5CX25

(g) *E. dispar* B0EPH1
G. duodenalis and C. parvum, but not in the genomes of E. histolytica and E. dispar. However, previously Srivastava et al. [31] reported a putative E. histolytica ortholog of RRN3 and of which there is a homologue in E. dispar. Differences in the components of the rRNA transcription machinery may be due to differences in the promoter, how the rDNA is organized in the genome of the parasites, and the number of copies, among other things. The way in which rDNA is organized in intestinal protozoan parasites varies between species, for G. duodenalis and C. parvum the classical conformation of repeats in tandem is maintained [32, 33], but for E. histolytica, these genes are located on extrachromosomal circular DNA molecules [34]. Regarding the number of rRNA copies, G. duodenalis has approximately 86 copies [32], C. parvum 5 copies [33] and E. histolytica approximately 200 copies [35]. Furthermore, the rDNA promoters of G. duodenalis and C. parvum have not been identified, but that of E. histolytica has [36]. Interestingly, in G. duodenalis, the presence of binding sequences for TBP and TAF in the intergenic region of the rDNA were identified [8].

The interaction between RRN3 and the A43 subunit (POLR1F in humans) is essential for the recruitment of Pol I into the preinitiation complex in the rDNA promoter [37, 38]. Important to this interaction is a conserved region of 22 amino acids in A43 [37] and a conserved serine patch on the surface of RRN3 which is formed by residues S101, S102, S109, S110, S145, S146, S185 and S186 [39]. This interaction is also regulated by phosphorylation of both proteins [39, 40]. In the conducted research, both regions were found to be partially conserved in intestinal protozoan parasites, with 8 residues (out of 22) highly conserved in the A43 counterparts and the residues corresponding to S101, S102 and S185 in the RRN3 counterparts (Figs. 2A, 3A). There are currently two drugs in cancer clinical trials that target the RNA polymerase I transcription (CX-5461 and CX-3543) and, in particular, CX-5461 does this by preventing the interaction between SL1 and Pol I in the rRNA promoter [41]. In this way, based on our analysis, molecules similar to CX-5461 could be designed against intestinal parasite rRNA transcription machineries as a new treatment strategy. Although it should also be considered that the sub-species and variants of the mentioned protozoa may have differences in sequence and structure. Given the differences between human and parasitic proteins, it may be possible to design molecules that specifically inhibit this machinery in parasites (and thus not affect the hosts), where the residues and regions that stand out in this work can be taken as a starting point.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11686-022-00612-7.

Funding Open access funding provided by Uppsala University. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethics statement This material is the original work of the author and has not been previously published elsewhere. Bioinformatic work does not require an ethical permit and the goal is to reduce work with animals.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Theel ES, Pritt BS (2016) Parasites. Microbiol Spectr. https://doi.org/10.1128/microbiolspec.DMIH2-0013-2015
2. El-Taweel HA (2015) Understanding drug resistance in human intestinal protozoa. Parasitol Res 114:1647–1659. https://doi.org/10.1007/s00436-015-4423-1
3. Hemphill A, Müller N, Müller J (2019) Comparative pathobiology of the intestinal protozoan parasites Giardia lamblia, Entamoeba histolytica, and Cryptosporidium parvum. Pathogens 8:116. https://doi.org/10.3390/pathogens8030116
4. Chifunda K, Kelly P (2019) Parasitic infections of the gut in children. Paediatr Int Child Health 39:65–72. https://doi.org/10.1080/20469047.2018.1479055
5. Pramanik PK, Alam MN, Roy Chowdhury D, Chakraborti T (2019) Drug resistance in protozoan parasites: an incessant wrestle for survival. J Glob Antimicrob Resist 18:1–11. https://doi.org/10.1016/j.jgar.2019.01.023
6. Upcroft P, Upcroft JA (2001) Drug targets and mechanisms of resistance in the anaerobic protozoa. Clin Microbiol Rev 14:150–164. https://doi.org/10.1128/CMR.14.1.150-164.2001
7. Grummt I (2013) The nucleolus—guardian of cellular homeostasis and genome integrity. Chromosoma 122:487–497. https://doi.org/10.1007/s00412-013-0430-0
8. Lagunas-Rangel FA, Bazán-Tejeda ML, Bermúdez-Cruz RM (2021) Ribosomal DNA in the protozoan parasite Giardia duodenalis has a differential chromatin distribution and epigenetic markings across the subunits. Acta Trop 217:105872. https://doi.org/10.1016/j.actatropica.2021.105872
9. Lagunas-Rangel FA, Yee J, Bermúdez-Cruz RM (2021) An update on cell division of Giardia duodenalis trophozoites. Microbiol Res 250:126807. https://doi.org/10.1016/j.micres.2021.126807
10. Mancio-Silva L, Lopez-Rubio JJ, Claes A, Scherf A (2013) Sir2a regulates rDNA transcription and multiplication rate in the human malaria parasite Plasmodium falciparum. Nat Commun 4:1530. https://doi.org/10.1038/ncomms2539
11. McStay B, Grummt I (2008) The epigenetics of rRNA genes: from molecular to chromosomal biology. Annu Rev Cell Dev Biol 24:131–157. https://doi.org/10.1146/annurev.cellbio.24.110707.175259
12. Lagunas-Rangel FA, Yee J, Bazán-Tejeda ML et al (2021) Sirtuin DExD/H protein Hel66 is involved in ribosome biogenesis in Trypanosoma brucei. Sci Rep 11:18325. https://doi.org/10.1038/s41598-021-97020-0
13. Jaremko D, Ciganda M, Christen L, Williams N (2019) Trypanosoma brucei L11 is essential to ribosome biogenesis and interacts with the kinetoplastid-specific proteins P34 and P37. mSphere. https://doi.org/10.1128/mSphere.00475-19
14. Nepomuceno-Mejía T, Florencio-Martínez LE, Pineda-García I, Martínez-Calvillo S (2022) Identification of factors involved in ribosome assembly in the protozoan parasite Leishmania major. Acta Trop 228:106315. https://doi.org/10.1016/j.actatropica.2022.106315
15. Lagunas-Rangel FA, Bazán-Tejeda ML, García-Villa E, Bermúdez-Cruz RM (2020) Nicotinamide induces G2 cell cycle arrest in Giardia duodenalis trophozoites and promotes changes in sirtuins transcriptional expression. Exp Parasitol 209:107822. https://doi.org/10.1016/j.exppara.2019.107822
16. Bateman A, Martin M-J, Orchard S et al (2021) UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res 49:D480–D489. https://doi.org/10.1093/nar/gkaa1100
17. Aurrecoechea C, Barreto A, Basenko EY et al (2017) EuPathDB: the eukaryotic pathogen genomics database resource. Nucleic Acids Res 45:D581–D591. https://doi.org/10.1093/nar/gkw1105
18. Camacho C, Coulouris G, Avagyan V et al (2009) BLAST+: architecture and applications. BMC Bioinform 10:421. https://doi.org/10.1186/1471-2105-10-421
19. Blum M, Chang H-Y, Chuguransky S et al (2021) The InterPro architecture and applications. BMC Bioinform 10:421. https://doi.org/10.1186/1471-2105-10-421
20. Waterhouse A, Bertoni M, Bienert S et al (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46:W296–W303. https://doi.org/10.1093/nar/gky427
21. Pettersen EF, Goddard TD, Huang CC et al (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. https://doi.org/10.1002/jcc.20084
22. Srivastava R, Srivastava R, Ahn SH (2016) The epigenetic pathways to ribosomal DNA silencing. Microbiol Mol Biol Rev 80:545–563. https://doi.org/10.1128/MMBR.00005-16