Secondary structure models of the nuclear internal transcribed spacer regions and 5.8S rRNA in Calciodinelloideae (Peridiniaceae) and other dinoflagellates

Marc Gottschling* and Jörg Plötner1

Freie Universität Berlin, Institut für Geologische Wissenschaften–Fachrichtung Paläontologie, Malteserstraße 74–100, D-12249 Berlin, Germany and 1Humboldt Universität Berlin, Museum für Naturkunde, Institut für Systematische Zoologie, Invalidenstraße 43, D-10015 Berlin, Germany

ABSTRACT

Secondary structure models of the 5.8S rRNA and both internal transcribed spacers (ITS1 and ITS2) are proposed for Calciodinelloideae (Peridiniaceae) and are also plausible for other dinoflagellates. The secondary structure of the 5.8S rRNA corresponds to previously developed models, with two internal paired regions and at least one 5.8S rRNA–28S rRNA interaction. A general secondary structure model of ITS1 for Calciodinelloideae (and other dinoflagellates), consisting of an open multibranch loop with three major helices, is proposed. The homology of these paired regions with those found in other taxa, published in previous studies (e.g. yeast, green algae and Platyhelmithes) remains to be determined. Finally, a general secondary structure model of ITS2 for Calciodinelloideae (and other dinoflagellates) is reconstructed. Based on the 5.8S rRNA–28S rRNA interaction, it consists of a closed multibranch loop, with four major helices. At least helix III and IV have homology with paired regions found in other eukaryotic taxa (e.g. yeast, green algae and vertebrates). Since the secondary structures of both ITS regions are more conserved than the nucleotide sequences, their analysis helps in understanding molecular evolution and increases the number of structural characters. Thus, the structure models developed in this study may be generally useful for future phylogenetic analyses.

INTRODUCTION

Studies on the secondary structures of rRNAs and adjacent regions have increased since the beginning of DNA sequencing in the late 1970s. rRNAs occur as multiple tandem repeats in nuclear DNA. Each repeat is transcribed as a single rRNA precursor, which is subsequently cleaved by a series of nucleolar events leading to the mature small subunit rRNA (SSU), the mature 5.8S rRNA and the mature large subunit rRNA (LSU). The SSU is separated from the 5.8S rRNA by the first internal transcribed spacer (ITS1), and the second internal transcribed spacer (ITS2) is located between the 5.8S rRNA and the LSU (1–3).

General models of the secondary structure have been proposed for the different rRNAs including SSU (4,5), 5.8S rRNA (6) and LSU (3,7). They appear to be conserved among all eukaryotes. To identify functional domains in the ribosomal structure, rRNA–protein interactions have recently been investigated (8,9). In the processing events during the maturation of rRNAs, both ITS1 and ITS2 also play important roles (10–12), which evidently require higher order (secondary) structures, in spite of dramatic nucleotide sequence variation.

Several authors have emphasized that the secondary, not the primary structure (i.e. nucleotide sequence itself) of the ITS regions is conserved at higher systematic levels (12–14). Unfortunately, comparative studies of the secondary structure of these evolutionarily highly divergent regions are still rare, although such work could add significantly to the number of structural characters available for phylogenetic analyses (14–16).

In this study, the secondary structures of ITS1 and ITS2 together with the 5.8S rRNA of selected dinoflagellates (as representatives of basal eukaryotes) were comprehensively investigated. Such case studies are relevant to provision of basic data, both for reconstructing molecular evolution in broader phylogenetic contexts and for analyzing function in ribosome biogenesis. We focused on the Calciodinelloideae (Peridiniaceae; main representatives: Scrippsiella, Ensiculifera and Pentapharsodinium) that produce calcareous cysts. Since the production of calcareous structures is unique among alveolates (but also found in the vegetative stages of Thoracosphaera), it is considered as an apomorphy and evidence for the monophyly of the corresponding group. Secondary structure models can be used for improving alignments at higher systematic levels even with strongly divergent regions such as the ITSs (11,17), and the framework

*To whom correspondence should be addressed. Tel: +49 30 838 70 808; Fax: +49 30 838 70 245; Email: caix@zedat.fu-berlin.de

*AY327462, AY427463

Nucleic Acids Research, Vol. 32 No. 1 © Oxford University Press 2004; all rights reserved
dictated by the secondary structure is considered as a tool for expanding the preliminary molecular phylogenies of Calcidinelloideae published by D’Onofrio et al. (18) and Montresor et al. (19).

MATERIALS AND METHODS

Approximately 150 dinoflagellate ITS1 and ITS2 sequences deposited in GenBank were investigated, of which six sequences were studied in detail together with new sequences from Calcidinelloideae operosum and Thoracosphaera heimii. These (and other species mentioned in the text) are listed in Table 1.

Fresh material of C. operosum and T. heimii cultured at the University of Bremen (Germany), Fachbereich Geowissenschaften, was used for sequencing, following standard protocols as described previously (24). After harvesting cells by centrifugation, the genomic DNA was extracted and purified using a DNeasy Plant Mini Kit (Qiagen). The universal primers ITS1 and ITS4 were those used by White et al. (25) designed at the 3’ end of SSU and at the 5’ end of LSU, respectively. The DNA was amplified using a Taq PCR Core Kit (Qiagen), and the PCR products were purified with QiAquick purification columns (Qiagen). A BigDye™ Ready Reaction Sequencing Kit (Applied Biosystems) was used for the sequencing reaction, which was run on a GeneAmp PCR System 2400 (Perkin Elmer). Sequence analysis (electrophoresis and detection of the fluorescent dye-labeled nucleotide fragments) was performed with an automatic DNA sequencer (ABI Prism™ 377, Perkin Elmer).

Common structural elements were initially recognized with the help of mFOLD (26,27) by screening for thermodynamically optimal and suboptimal secondary structures (default settings, with \( T = 25^\circ\text{C} \)). Additionally, paired regions (helices or ‘hairpins’) were identified both by mutual comparison of sequences in a manual alignment [use of the m-fold tool (13)] and by complementary base substitutions (e.g. paired G–C into paired A–U) as proposed in previous studies (6,13). Energy levels of the presumptive secondary structures were then calculated with mFOLD (26,27).

RESULTS

General topology

In the nuclear genome, the SSU is separated from the 5.8S rRNA by ITS1, and ITS2 is located between the 5.8S rRNA and the LSU (Fig. 1). While the rRNAs are conserved at higher systematic levels with respect to both their sequences and their secondary structures, the ITSs are very divergent in their nucleotide sequences.

A more or less extended amount (3–20) of optimal and suboptimal secondary structures is yielded by mFOLD. However, different parameter settings of, for example, temperature (\( T = 10^\circ\text{C}, T = 20^\circ\text{C}, T = 30^\circ\text{C} \)) do not affect the general architecture, but result in different energy levels for the secondary structures (data not shown).

5.8S rRNA structure

The divergence rate in the 5.8S rRNA is low, thus even only distantly related species (e.g. Calcidinelloideae compared with species of Symbiodinium) have similar sequences. The secondary structure of the 5.8S rRNA shows at least three paired regions (Fig. 2): first, a region with a basal pairing (B) and a small loop with two helices (C and D) and, secondly, a single helix (F). Thirdly, the 5’ end of the 5.8S rRNA pairs with the 3’ end of the LSU (A’, 5.8S–LSU interaction).

ITS1 structure

Several secondary structure predictions of the ITS1 transcript found in Calcidinelloideae and other dinoflagellates are proposed in Figure 3A–H. These are described numerically...
and statistically in Table 2, and are the basis for reconstructing a general secondary structure model with a multibranch loop and several paired regions, as shown in Figure 1. The loop is open, since no interactions between the flanking rRNAs (SSU and 5.8S rRNA) are developed, and its G content is low compared with the paired regions. The size of the loop ranges broadly from 26 to 62 unpaired bases. The GC content differs from 46 to 58%. Proportions of G–U pairs in the helices are usually low (9–20%).

At least two paired regions (helix I and helix II) can be identified in ITS1 (Figs 1 and 3A–H). Helix I is rather uniform in length in the species investigated (18–23 bp); however, in some cryptoperidiniopsoid dinoflagellates (Fig. 3F), it is longer (up to 30 bp), and in Symbiodinium species (Fig. 3H) it is shorter (11 bp). At the base of hairpin I, the conserved paired motif 5′-UGAG versus CUCR-3′ is almost ubiquitous in dinoflagellates. The two single-stranded nucleotides between helix I and II are purines in all investigated cases. Helix II also shows distinct size classes (Table 2): in Calciodinelloideae and other dinoflagellates (such as Pfiesteria and Thoracosphaera), it is 18–24 bp long, while it is elongated in species of Heterocapsa [H.triquetra: 29 bp (Fig. 3G)]. At the base of hairpin II, the conserved paired motif 5′-GGCGGC versus GUCGYC-3′ is found in most of the species investigated.

The region between hairpin II and the 5′ end of the 5.8S rRNA is very variant, even at species level. However, another helix III near the 3′ end of the ITS1 transcript can be identified and shows extreme variation in length (only 6 bp in, for example, Scrippsiella cf. trochoidea, but up to 18 bp in Symbiodinium species). Additionally, mFOLD found several local interactions in this divergent region as well as at the 5′ end of the ITS1 transcript, indicated with lower case letters in Figure 3A–H. These helices have no common position in the ITS1 transcript and are not universally present in dinoflagellates, but ‘helix a’ found in Pfiesteria and cryptoperidiniopsoid dinoflagellates can also be formed at exactly the same position in species of Heterocapsa. It is identical to the other two, except for a 1 bp short stem. It also occupies more or less the same position, relative to helix I.

### ITS2 structure

Secondary structure predictions of the ITS2 transcript are proposed in Figure 3A–H, from which a general secondary structure model consisting of a multibranch loop with several paired regions has been reconstructed (Fig. 1). The loop is closed, based on the 5.8S–LSU interaction (A′ in Fig. 2), and its G content is low compared with the paired regions. Numerical and statistical descriptions of the deviations are given in Table 3. The size of the loop lies within a rather narrow frame of 39–47 unpaired bases. The GC content in the sequences ranges from 43 to 64%, while higher values are found in, for example, Pfiesteria and Thoracosphaera. Proportions of G–U pairs in the helices are usually low (5–13%).

Four paired regions (helices I–IV) can be identified in ITS2. These helices show distinct size classes. Generally, helix I is 11–13 bp long, but it is elongated in Pfiesteria and cryptoperidiniopsoid dinoflagellates (up to 21 bp). Helix II varies from 9 bp (Symbiodinium sp.) to 20 bp (cryptoperidiniopsoid dinoflagellates) and has a more or less conserved motif 5′-GUGYGU versus ACGYRY-3′ at its base. Helix III is the longest of the four paired regions, ranging from 18 bp (Symbiodinium sp., in which it is probably divided into two hairpins IIIa and IIIb, Fig. 3H) to 33 bp (cryptoperidiniopsoid dinoflagellates). Finally, the generally short helix IV is only ~5 bp long in many species (e.g. Scrippsiella cf. trochoidea), but is extremely elongated in cryptoperidiniopsoid dinoflagellates (up to 16 bp).
Table 2. Numerical and statistical values of the secondary structures (ITS1) proposed in this study

<table>
<thead>
<tr>
<th>Species</th>
<th>Sequence length (in bases)</th>
<th>Loop length (unpaired bases)</th>
<th>GC content</th>
<th>Paired G–U in helices I–III</th>
<th>Length of helices (in paired bases) I</th>
<th>II</th>
<th>III</th>
<th>ΔG (25°C, kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Scripsiella cf. trochoidea</em></td>
<td>209</td>
<td>49</td>
<td>49%</td>
<td>9%</td>
<td>18</td>
<td>19</td>
<td>6</td>
<td>-48.2</td>
</tr>
<tr>
<td><em>Calciodinellum operosum</em></td>
<td>209</td>
<td>51</td>
<td>46%</td>
<td>10%</td>
<td>20</td>
<td>20</td>
<td>9</td>
<td>-41.3</td>
</tr>
<tr>
<td><em>Ensiculifera cf. imariensis</em></td>
<td>210</td>
<td>41</td>
<td>50%</td>
<td>20%</td>
<td>19</td>
<td>19</td>
<td>7</td>
<td>-71.3</td>
</tr>
<tr>
<td><em>Thoracosphaera heimi</em></td>
<td>218</td>
<td>38</td>
<td>57%</td>
<td>11%</td>
<td>20</td>
<td>18</td>
<td>7</td>
<td>-45.2</td>
</tr>
<tr>
<td><em>Pfiesteria piscida</em></td>
<td>251</td>
<td>60</td>
<td>58%</td>
<td>20%</td>
<td>23</td>
<td>22</td>
<td>14</td>
<td>-95.0</td>
</tr>
<tr>
<td>Cryptoperidiniopsoid dinoflagellate</td>
<td>261</td>
<td>62</td>
<td>50%</td>
<td>9%</td>
<td>30</td>
<td>19</td>
<td>16</td>
<td>-57.5</td>
</tr>
<tr>
<td><em>Heterocapsa triquetra</em></td>
<td>233</td>
<td>62</td>
<td>56%</td>
<td>13%</td>
<td>19</td>
<td>29</td>
<td>6</td>
<td>-53.7</td>
</tr>
<tr>
<td><em>Symbiodinium</em> sp.</td>
<td>223</td>
<td>26</td>
<td>52%</td>
<td>15%</td>
<td>11</td>
<td>24</td>
<td>18</td>
<td>-94.3</td>
</tr>
</tbody>
</table>

Table 3. Numerical and statistical values of the secondary structures (ITS2) proposed in this study

<table>
<thead>
<tr>
<th>Species</th>
<th>Sequence length (in bases)</th>
<th>Loop length (unpaired bases)</th>
<th>GC content</th>
<th>Paired G–U in helices I–III</th>
<th>Length of helices (in paired bases) I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>ΔG (25°C, kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Scripsiella cf. trochoidea</em></td>
<td>195</td>
<td>39</td>
<td>49%</td>
<td>12%</td>
<td>12</td>
<td>13</td>
<td>28</td>
<td>5</td>
<td>-43.2</td>
</tr>
<tr>
<td><em>Calciodinellum operosum</em></td>
<td>199</td>
<td>45</td>
<td>48%</td>
<td>10%</td>
<td>10</td>
<td>13</td>
<td>27</td>
<td>5</td>
<td>-31.5</td>
</tr>
<tr>
<td><em>Ensiculifera cf. imariensis</em></td>
<td>196</td>
<td>(24) 46</td>
<td>45%</td>
<td>6%</td>
<td>12</td>
<td>12</td>
<td>20</td>
<td>4</td>
<td>-34.1</td>
</tr>
<tr>
<td><em>Thoracosphaera heimi</em></td>
<td>219</td>
<td>41</td>
<td>64%</td>
<td>5%</td>
<td>13</td>
<td>16</td>
<td>27</td>
<td>7</td>
<td>-62.4</td>
</tr>
<tr>
<td><em>Pfiesteria piscida</em></td>
<td>244</td>
<td>46</td>
<td>63%</td>
<td>12%</td>
<td>21</td>
<td>16</td>
<td>30</td>
<td>9</td>
<td>-62.0</td>
</tr>
<tr>
<td>Cryptoperidiniopsoid dinoflagellate</td>
<td>297</td>
<td>43</td>
<td>53%</td>
<td>12%</td>
<td>19</td>
<td>20</td>
<td>33</td>
<td>16</td>
<td>-63.1</td>
</tr>
<tr>
<td><em>Heterocapsa triquetra</em></td>
<td>203</td>
<td>47</td>
<td>43%</td>
<td>10%</td>
<td>13</td>
<td>15</td>
<td>26</td>
<td>5</td>
<td>-41.6</td>
</tr>
<tr>
<td><em>Symbiodinium</em> sp.</td>
<td>195</td>
<td>47</td>
<td>50%</td>
<td>13%</td>
<td>11</td>
<td>9</td>
<td>18</td>
<td>10</td>
<td>-39.3</td>
</tr>
</tbody>
</table>

Complementary base substitutions

Several cases of complementary base substitutions can be identified. (i) In many dinoflagellates (e.g. Calciodinelloideae, *Thoracosphaera*, *Heterocapsa* and *Prorocentrum*), helix F of the 5.8S rRNA has the motif 5’-GCUUCCGGG versus CCUGAAAGC-3’ (Fig. 2). This motif has been changed into 5’-CCUUCCGGG versus CCUGAAAGG-3’ in *Pfiesteria* and 5’-ACUUCCGGG versus CCUGAAAGU-3’ in cryptoperidiniopsid dinoflagellates. (ii) The highly conserved motif of the 5.8S–LSU interaction is 5’-UGCCUGUCCUA versus UGAAGUYAGGUCA-3’ (Fig. 2) in many dinoflagellates, while it is modified to 5’-UGCCUGCUCAA versus UGAAGUGYAGGUA-3’ (Fig. 2) in some dinoflagellates. (iii) In *Heterocapsa illeifera*, the conserved motif 5’-GCCGGC versus GUUCYG-3’ at the base of helix II of the ITS1 transcript (Fig. 3B–G) has evolved into 5’-GGUACG versus GUUAAYC-3’. (iv) In *Calciodinellidae*, the helix I base of ITS2 has the motif 5’-GYAAY versus GUUGC-3’, which shows a substitution with 5’-GCAGC versus GCUGU-3’ in *T. heimii* (Fig. 3A–D).

**DISCUSSION**

**Deducing secondary structures from sequences**

Computer programs such as mFOLD (26,27) find highly divergent secondary structures (even from the same sequence), independently of the settings (e.g. temperature). However, the thermodynamically optimal structure does not necessarily reflect the *in vivo* structure, since (largely unknown) interacting factors such as other molecules (e.g. proteins) or tertiary structure constraints may occur. Consequently, common structural elements of rRNA transcripts should not be exclusively reconstructed by energy optimization (provided by mFOLD), but also by homologizing internal regions of sequences [‘phylogenetic comparative method’ (13)] by mutual comparison (i.e. plausibility of thermodynamically optimal and suboptimal hypotheses). Such investigations should include many sequences of both closely and distantly related species (14), both of which were investigated in this study. Paired regions (considered as homologous) can be proven by complementary base substitutions (6,12–14,16). Several such substitutions can be identified for paired regions of dinoflagellates (see Results).

The ITS helices are GC rich, while the G content of non-paired regions is low (in dinoflagellates often T rich) corresponding to the situation found in other groups of organisms (4,12–14). Thus, paired regions are generally composed of bases with strong bonds, while unpaired regions are rich in bases with lower bond pairing potential [e.g. A-rich in Boraginales (16)]. The amount of G–U pairings in the presumptive helices is mostly low (Tables 2 and 3), indicating a high stability of the helices. These findings also help identification of paired regions by mutual comparison of sequences.

Figure 3. ITS1 and ITS2 transcript secondary structures from (A) *Scripsiella cf. trochoidea*, (B) *Calciodinellum operosum*, (C) *Ensiculifera cf. imariensis*, (D) *Thoracosphaera heimi*, (E) *Pfiesteria piscida*, (F) Cryptoperidiniopsid dinoflagellate, (G) *Heterocapsa triquetra* and (H) *Symbiodinium* sp. Structures are numbered every 50 nucleotides with tick marks at 10 nucleotide intervals. Major helices are labeled I–III (ITS1) and I–IV (ITS2), respectively. a–c indicate less well supported helices.
Structure of the 5.8S rRNA
The 5.8S rRNA sequences investigated are more or less conserved. Their presumptive secondary structures correspond to the universal model of Vaughn et al. (6), which is supported by the results of nuclease digestion studies (28,29). However, of the paired regions, only helix F has been universally accepted (30,31).

Vaughn et al. (6) have proposed a very short ‘region E’ near helix F. Considering complementary base substitutions at the base of helix F in *Pfisteria* and cryptophycean species (see Results), the formation of a ‘region E’ seems impossible at least for dinoflagellates. Thus, the ubiquitous presence of ‘region E’ remains questionable.

Intermolecular bonds between the 5.8S rRNA and the LSU have been identified (3,6,32). These close rRNA–rRNA interactions support the hypothesis that the 5.8S rRNA of eukaryotes is derived from (parts of) the LSU of prokaryotes (2,3,33). The predominant bonding between the 3′ end of the 5.8S rRNA and the 5′ end of the LSU (Fig. 2) is supported by complementary base substitutions (see Results). In eukaryotes, it encloses ITS2 (Fig. 1), which is absent in prokaryotes (2,6). Additionally, another contact site near the 5′ end of the 5.8S rRNA, interacting with a complement in the LSU, has been assumed (32,34), but the precise complementary region in the LSU remains to be determined. In *Proorocentrum micans*, Lenaers et al. (3) located this contact site between the LSU domains D1 and D2. However, such a region cannot be identified in the sequences analyzed in this study, mainly based on missing data of the complete LSU (and, therefore, complementary base substitutions).

Structure of ITS1
Several authors have emphasized that the secondary structure of ITS1 is conserved at higher systematic levels (13,14,16). In most eukaryotes investigated so far, it consists of an open multibranch loop with several helices, although data for the ITS1 secondary structure are still very rare. Furthermore, the 3′ end, but not the 5′ end, of ITS1 is conserved regarding both the nucleotide sequence itself and its secondary structure (32,33). In contrast, the 3′ region in dinoflagellates (with a divergent, often short helix) is much more variant than the 5′ region.

Two major helices are found in dinoflagellates, together with a third, often short, and divergent, helix III near the 3′ end (Figs 1 and 3A–H; comparative data not available). ITS1 secondary structure regions with high nucleotide sequence stabilities (such as the basal base pairings in helices I and II) could also be identified for helices in Volvocales (14) and Boraginaceae (16).

Among eukaryotes, different numbers of helices and structural details occur in ITS1. Four helices are found in Chlorobionta (14,16) and yeast (35), and seven helices have been proposed for *Digenea* [Platyhelminthes (36)]. The homology of single helices among all eukaryotes remains to be determined, since taxon sampling (i.e. of known secondary structures) is still not representative (mainly due to insufficient case studies). Conserved motifs of the loop such as the angiosperm motif 5′-AAGGAA-3′ (11,16), the repeated motif 5′-CCAA-3′ in Volvocales (14) and the ‘Block D’ motif in Coccinellidae [Coleoptera (37)] cannot be identified for the sequences analyzed here. Such motifs have been discussed as key factors in recognition by the processing endonuclease during the maturation of rRNAs (11), which may be different in the ITS1 of dinoflagellates.

In conclusion, ITS1 secondary structures are conserved, even at high systematic levels, but may differ in some respects (number of helices, consistency of the loop) among eukaryotes. Homology of certain helices in different groups of organisms should be more rigorously established in future studies. One single, often short, helix near the 3′ end of the ITS1 transcript appears to be a highly conserved structure in eukaryotes.

Structure of ITS2
The general architecture of the ITS2 secondary structure is similar to that found in ITS1, consisting of a multibranch loop with several major helices (Fig. 3A–H). Similarities between structural features of ITS1 and ITS2 have been emphasized for, for example, species of *Chlamydomonas* (17) and *Schizosaccharomyces pombe* (38). However, the most striking difference between both spacers is that the loop is closed in the ITS2 transcript due to the 5.8S–LSU interaction (Fig. 1). Based on the criteria ‘position’ and ‘complexity’, the formation of the four helices can be regarded as homologous in the species investigated.

Four independent domains are frequently found in other taxa such as Vertebrata (12), Chlorobionta (13), Diptera (39) and Trematoda (40). In the basic secondary structure model generated by this study, helix III is the longest of the paired regions and joined to a short helix IV at the 3′ end. A model for Chlorobionta, proposed by Mai and Coleman (13) (four helices, helix III generally the longest, helix IV short and divergent), particularly resembles the secondary structure developed in this study (Fig. 1). Furthermore, a giant D1 domain with an additional short helix (D2) at the 3′ end has been identified for several other taxa such as yeast and vertebrates (12,15) and leaf beetles (41), supporting the homology of the corresponding paired regions among all eukaryotes.

Côté et al. (42) proposed a dynamically conformational model of the ITS2 secondary structure, arranging a ‘hairpin model’ with a ‘ring model’. In the ‘hairpin model’ of yeast (43), the ITS2 regions adjacent to the flanking 5.8S and LSU ends have an extensive base-paired structure, which is free (or shows only local interactions) in the ‘ring model’ (15). However, those base-paired structures of the ‘hairpin model’ cannot be identified either for dinoflagellates or for other taxa except yeast (14,40). mFOLD only occasionally found base pairings of the loop (e.g. in *Ensiliculifera* cf. *imariensis*, Fig. 3C) that are not universally present and are not homologous with those of the ‘hairpin model’ in the sense of Yeh and Lee (43) since they do not share a common structural alignment, the sequence of *E.imariensis* can also be adapted into the secondary structure model proposed here.

Another difference between the secondary structure models of both ITS1 and ITS2 is the size of the multibranch loop. The number of unpaired loop bases appears to be rather conserved in the ITS2 transcript (approximately 40–45), while it varies in ITS1 (26–62). Probably a principal difference in the architecture of both spacers allows different degrees of freedom for
putative conformational transformations. Only the ITS2 secondary structure is closed, based on the 5.8S rRNA–28S rRNA interaction, which may result in higher evolutionary pressure on this molecule than on the ITS1 transcript.

The identification of homologous base pairings is somewhat ambiguous based on the divergence of the ITS2 transcript. However, complementary base substitutions are present at least in helix I comparing sequences of Calciodinelloideae and *Thoracosphaera* (see Results), which support the existence of this helix. Similar (and probably homologous) motifs occur at least in helix I comparing sequences of Calciodinelloideae and *H.triquetra* (5’-GCCA versus UGGC-3’), *Pfisteria* (5’-GGAGU versus GCUCU-3’), cryptoperidinioid dinoflagellates (5’-GCCGU versus ACGGU-3’) and even *Symbiodinium* (5’-GCAAG versus CUUGC-3’). The 12–14 nucleotide distance from these motifs to the more or less conserved motif at the base of helix II (see Results) shows only minor variation, which is a strong argument for the possibility of structural alignments of at least parts of analyzed ITS2 sequences. Further conserved sequence motifs, whose presence would facilitate structural alignment of the ITS2 sequences, might also be present in helix III, but they are rather indistinct compared with those in helices I and II.

**Conclusion**

A complex network of interactions take place to assemble the pre-rRNA structural features directly involved in processing steps into a relatively compact structure. Not only do the rRNAs show highly conserved secondary structures, but also for the separate regions, ITS1 and ITS2, such common structural elements are plausible. The general organization of both the spacers, consisting of a multibranch loop with several paired regions, is very similar, but differs from the complex structures suggested for the mature rRNAs.

The identification of common structural elements indicates equivalent functionality of the corresponding molecules (44). This is a strong argument against non-functionality even of highly divergent regions such as the ITSs and implies similar processing requirements. Subtle secondary structural motifs may participate in the ITS excision process, spatially positioned by the conserved framework of helices.

Common structural elements, even of highly divergent markers, can be regarded as homologous (12,13,16,45) if criteria such as ‘position’ and ‘complexity’ are fulfilled (the criterion ‘specific quality’ remains to be determined since precise functions of helices are largely unknown). Identifying homologous regions, and reconstructing their evolution, increases the number of character traits available for phylogenetic analyses, as has been shown for the Volvocales (14,17), the Heliotrophiaceae from the Boraginales (16,45), and the Ixodidae from the Chelicera (46). Such transformations are also plausible for sequences investigated in this study. For example, all *Heterocapsa* species deposited in GeneBank show a uniquely long helix II in the ITS1 transcript (Fig. 3G). Hence, this (molecular) character can be regarded as a (structural) apomorphy (‘elongation of hairpin II’) for the corresponding group.

Structural constraints dictated by secondary structure models help to optimize alignments (11,13,16,41) by, for example, recognizing highly conserved paired motifs such as the base of the ITS1 helix I and identifying complementary base substitutions. An example of a structural alignment of the sequences investigated in this study is additionally provided as on-line material (Structural_Alignment.nex). It can be demonstrated that this application deepens the phylogenetic signal and can reduce the bias of divergent markers (16,17,47). In this context, secondary structure models have been shown to be useful for future phylogenetic studies. With respect to Calciodinelloideae, they may help to identify their closest relatives among dinoflagellates expanding on the preliminary work of D’Onofrio *et al.* (18) and Montresor *et al.* (19).

**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at NAR Online.

**ACKNOWLEDGEMENTS**

We would like to express our gratitude to Marina Montresor (Naples) for providing us with material, Monika Kirsch (Bremen), Ronny Knop and Robert Schreiber (both Berlin) for technical work, David Unwin (Berlin) for improving the written English, and two anonymous reviewers for helpful comments on first drafts of the manuscript. We are grateful to the Deutsche Forschungsgemeinschaft (grants Ke 322/30 and Pi 213/4) for financial support.

**REFERENCES**