Comparative analysis of tertiary structure elements in signal recognition particle RNA

ARTICLE *in* FOLDING AND DESIGN · FEBRUARY 1996

DOI: 10.1016/S1359-0278(96)00044-2 · Source: PubMed

3 AUTHORS, INCLUDING:

Christian Zwieb
University of Texas Health Science Center at …

160 PUBLICATIONS **4,099** CITATIONS

SEE PROFILE

Comparative analysis of tertiary structure elements in signal recognition particle RNA

Christian Zwieb[1], Florian Müller[2] and Niels Larsen[3]

Background: The signal recognition particle (SRP) is a ribonucleoprotein complex that associates with ribosomes to promote co-translational translocation of proteins across biological membranes. We have used comparative analysis of a large number of bacterial, archaeal, and eukaryotic SRP RNA sequences to derive shared tertiary SRP RNA structure elements.

Results: A representative three-dimensional model of the human SRP RNA is shown that includes single-stranded intrahelical and interhelical RNA loops and incorporates data from enzymatic and chemical modification, electron microscopy, and site-directed mutagenesis. Properties of the SRP RNA model are an overall extended dumbbell-shaped structure (260 Å × 70 Å) with a pseudoknot in the small SRP domain (a pairing of 12-UGGC-15 with 33-GCUA-36), and a tertiary interaction in the large SRP domain (198-GA-199 with 232-GU-233).

Conclusions: The RNA 'knuckle' formed in helix 8 of SRP RNA appears to constitute the binding site for protein SRP54 or its bacterial equivalent, protein P48. A dynamic property of this feature may explain the hierarchical assembly of proteins SRP19 and SRP54 in the large SRP domain. Furthermore, the human SRP RNA model serves as a framework to understand details of the structure and function of SRP in all organisms and is presented to stimulate further experimentation in this area.

Addresses: [1]Department of Molecular Biology, The University of Texas Health Science Center, PO Box 2003, Tyler, Texas 75710, USA; e-mail: zwieb@jason.uthct.edu. [2]Max Planck Institute for Molecular Genetics, Ihnestraβe 73, D14195, Berlin, Germany; e-mail: mueller_f@mpimg-berlindahlem.mpg.de. [3]Department of Microbiology, Giltner Hall, Room 180, Michigan State University, East Lansing, Missouri 48824, USA; e-mail: niels@truth.mph.msu.edu.

Correspondence: Christian Zwieb
e-mail: zwieb@jason.uthct.edu

Key words: comparative sequence analysis, pseudoknot, RNA folding, RNA–protein interactions, RNA tetraloop, tertiary interaction, translation

Received: 17 May 1996
Revisions requested: 10 Jun 1996
Revisions received: 14 Jun 1996
Accepted: 19 Jun 1996

Published: 22 Jul 1996
Electronic identifier: 1359-0278-001-00315

Folding & Design 22 Jul 1996, 1:315–324

Introduction

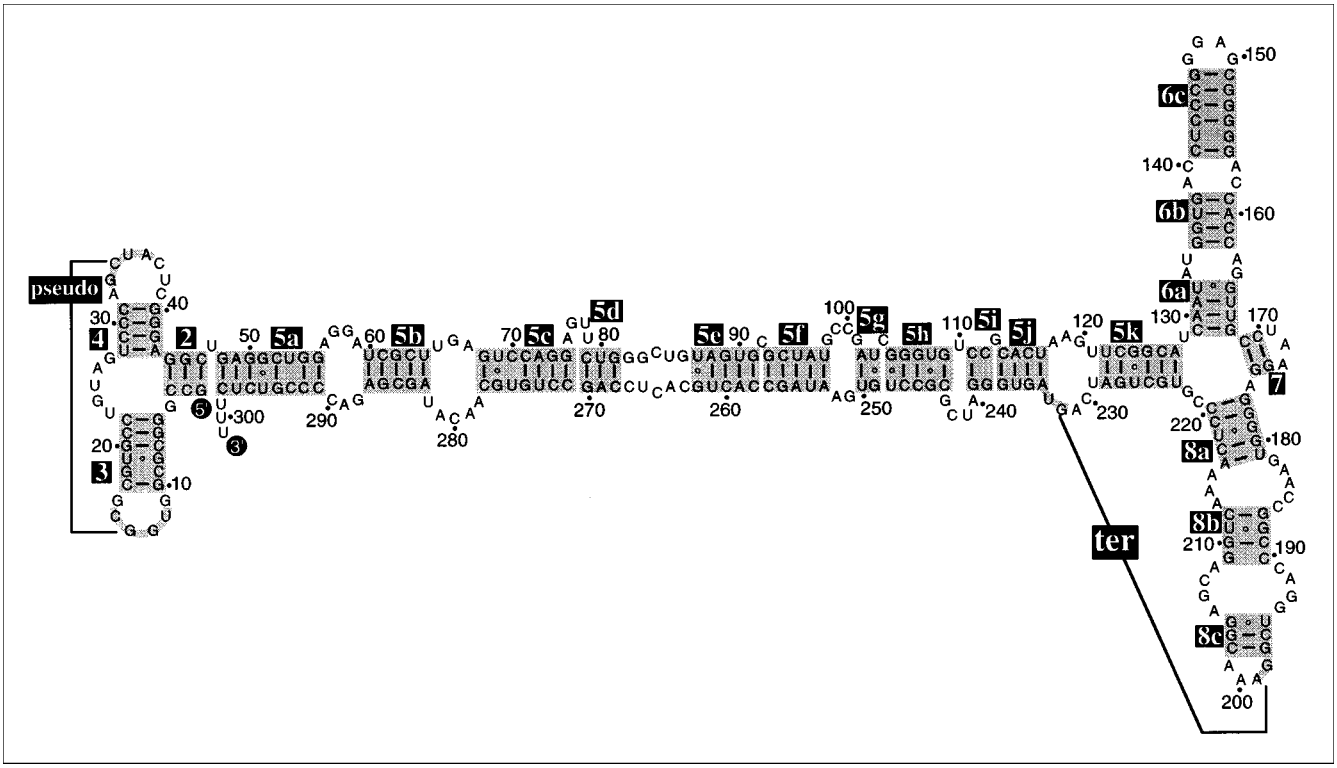
The signal recognition particle (SRP) is a cytosolic ribonucleoprotein complex that facilitates the co-translational translocation of proteins across lipid bilayers. SRP cycles through at least three functional stages where it either exists free in the cytosol, bound to ribosomes, or associated with the membrane (reviewed in [1]).

SRP, in its simplest form, contains one polypeptide, called SRP54, P48 or ffh [2–4], that is bound to the SRP RNA. The size of the RNA molecule varies from 519 nucleotides in *Saccharomyces cerevisiae* to 77 nucleotides in *Mycoplasma mycoides*. The SRP has been characterized best in the mammalian system where it is typically composed of a 300-nucleotide RNA and six proteins (SRP9, SRP14, SRP19, SRP54, SRP68, and SRP72) [2]. Whereas most SRP proteins interact directly with the RNA, the association of SRP54 is dependent on the binding of protein SRP19 to SRP RNA. It has been suggested that the sequential binding of SRP19 and SRP54 is mediated through a conformational change in the RNA, thereby assigning the SRP RNA a functional role in the assembly of the SRP.

The electron microscope reveals SRP to be an elongated particle with a small and a large domain that are separated by a flexible adapter [5]. The shape of the SRP RNA is similar to that of the assembled particle, showing that the RNA is present throughout the SRP [6]. Enzymatic and chemical modification experiments and site-directed mutagenesis approaches helped to assign protein-binding sites to portions of the RNA. The small domain (see Fig. 1) contains the protein SRP9/SRP14 heterodimer, whereas the large domain includes proteins SRP19, SRP54, and SRP68/SRP72. The inter-domain adapter was predicted to correspond to a variable region of helix 5 and appears to lack association with proteins. Preliminary model building attempts [7] did not include single-stranded interhelical or intrahelical regions, but nevertheless indicated that some RNA helices of the small and large domain must be in close proximity to fit within the dimensions (220–240 Å × 50–60 Å) of the particle as seen in the electron microscope [5,6].

Earlier comparative analysis of 39 SRP RNA sequences enabled us to determine common secondary structure elements of the SRP RNAs at the level of individual base pairs [8]. An increasing number of aligned SRP RNAs, reviewed in the SRP database (SRPDB) [9], helped to extract possible tertiary structure elements directly from the sequence alignment [10]. Here, we discuss two such

Figure 1



Secondary structure of human SRP RNA. Base pairings are shown as supported by comparative sequence analysis of SRP RNA sequences in the SRP database [40,41]. The 5′ and 3′ ends of the RNA molecule are labeled; helices are marked 2–8 according to the nomenclature of Larsen & Zwieb [8]; residues are numbered in 10-nucleotide increments; base-paired sections (highlighted in gray) of helices 5, 6, and 8 are suffixed a–k in helix 5, and a–c in helices 6 and 8; pairing between the loops of helices 3 and 4 (left part of the figure) indicates the proposed pseudoknot (pseudo) in the small SRP domain; and the postulated tertiary interaction (ter) between the tip of helix 8 and an internal loop between sections 5j and 5k in the large domain of the SRP is shown at the right.

elements that restrict the folding of the SRP RNA in three dimensions: a pseudoknot structure in the small domain, and a tertiary interaction in the large domain; the latter may serve as a dynamic mechanism for the binding of protein SRP54 in the eukaryotic SRPs. In developing a three-dimensional model of the human SRP RNA as a representative example, we also critically review the available experimental data and collectively incorporate and explore possible foldings of the RNA with the computer program ERNA-3D [11], a powerful RNA modeling tool.

Results and discussion

Comparative analysis of SRP RNA

The SRP RNA secondary structures, derived by comparison of more than 70 sequences (see Materials and methods), served as the basis for our investigation of the three-dimensional structure. Figure 1 shows as an example the secondary structure of the human RNA with seven helices, numbered 2–8 [8]. Where applicable, helix sections are named with suffixed letters. Helices 2–4 and sections 5a and 5b are part of the small SRP domain. The large SRP domain is composed of sections 5g–5k and helices 6–8.

Although the structural arrangement of the mammalian SRP is similar to the SRP RNAs from the lower eukaryotes, such as the yeasts, these organisms have a greatly reduced small domain. In the Archaea, the deduced secondary structure is very much like that of the eukaryotes, but there is an additional helix (helix 1) close to the RNA termini. Most bacterial SRP RNAs possess only helix 8 and a phylogenetically conserved portion of helix 5 that appears to correspond to helices 5g–5k (Fig. 1). An exceptional case are the SRP RNAs of the geni *Bacillus* and *Clostridium* that, as in all bacterial secondary structures, are missing helices 6 and 7 but are equipped with a small domain that includes helices 1–5. These structures, although not shown here, will be discussed below.

Folding restraints beyond the interactions of the secondary structure are necessary to accommodate the mammalian SRP RNA within the 220–240 Å × 50–60 Å dimensions of the particle [5–7,12]. This expectation is supported, among other findings, by investigations of the structure of the SRP with RNases, chemical modification, and systematic site-directed mutagenesis. Results from

Table 1

Data used in the modeling of the SRP RNA structure.

No	N	2°	C	V	R	9/14
1	G	h2				
2	C	h2				I
3	C	h2	M			I
4	G	I	M	S		I
5	G	h3		S		P
6	G	h3	L	S		P
7	C	h3		S		P
8	G	h3		S		P
9	C	h3		S		P
10	G	h3				P
11	G	I				
12	U	knt				
13	G	knt			W	
14	G	knt	R		W	
15	C	knt	M			I
16	G	I				I
17	C	h3				P
18	G	h3				P
19	U	h3				P
20	G	h3				P
21	C	h3		S		P
22	C	h3		S		P
23	U	I	H	S		P
24	G	I	H			P
25	U	I	H		S	P
26	A	I	M		S	P
27	G	I	L		W	P
28	U	h4				P
29	C	h4				P
30	C	h4	L			
31	C	h4				
32	A	I				
33	G	knt				
34	C	knt				
35	U	knt				
36	A	knt			S	
37	C	I			S	
38	U	I			S	
39	C	I			S	
40	G	h4			W	
41	G	h4	M			
42	G	h4				
43	A	h4				
44	G	h2	H	W		
45	G	h2		W		
46	C	h2	L			
47	U	I				
48	G	h5a		S		P
49	A	h5a		S		P
50	G	h5a				P

No	N	2°	C	V	R	9/14
51	G	h5a				P
52	C	h5a		S		
53	U	h5a		S		
54	G	h5a		S		
55	G	h5a	K	S		
56	A	I			S	
57	G	I	M		W	
58	G	I			W	P
59	A	I			S	P
60	U	h5b				P
61	C	h5b				P
62	G	h5b				P
63	C	h5b				
64	U	h5b	HS			
65	U	I	HS			
66	G	I				
67	A	I				
68	G	h5c		S		
69	U	h5c		S		
70	C	h5c		S		
71	C	h5c		S		
72	A	h5c		S		
73	G	h5c		S		
74	G	h5c				
75	A	I	HS		S	
76	G	I	HS		W	
77	U	I			S	
78	U	I				
79	C	h5d				
80	U	h5d				
81	G	h5d		S		
82	G	I		S		
83	G	I	M	S		
84	C	I		S		
85	U	I		S		
86	G	I	H	S		
87	U	h5e				
88	A	h5e				
89	G	h5e				
90	U	h5e				
91	G	h5e				
92	C	I				
93	G	h5f		S		
94	C	h5f		S		
95	U	h5f	L			
96	A	h5f				
97	U	h5f				
98	G	I	M		W	
99	C	I				
100	C	I	HS	W		

No	N	2°	C	V	R	19	68/72
101	G	I	HS		W		S
102	A	h5g			S		S
103	U	h5g					
104	C	I		S			
105	G	h5h	M	S			S
106	G	h5h					S
107	G	h5h		S			
108	U	h5h	Y				
109	G	h5h	H				
110	U	I	Y				
111	C	h5i	M				
112	C	h5i	H				
113	G	I					
114	C	h5j		S			
115	A	h5j		S			
116	C	h5j	M	S			
117	U	h5j		S	S		
118	A	I	M	S	S		
119	A	I	M		S		
120	G	I	L		W	W	S
121	U	I	Y			W	
122	U	h5k				W	
123	C	h5k				W	
124	G	h5k				W	
125	G	h5k				W	
126	C	h5k				W	
127	A	h5k				W	
128	U	I			S	W	
129	C	h6a			S	M	S
130	A	h6a			S	M	
131	A	h6a				M	S
132	U	h6a				M	S
133	A	I				M	S
134	U	I				M	
135	G	h6b	M			M	S
136	G	h6b				M	
137	U	h6b	K			M	
138	G	h6b				M	S
139	A	I				M	
140	C	I				M	
141	C	h6c				M	
142	U	h6c				M	
143	C	h6c		S		M	
144	C	h6c		S		M	
145	C	h6c				M	
146	G	h6c				M	
147	G	I	H			W	W
148	G	I			W		W
149	A	I	H		S	S	W
150	G	I	R		W	M	W

these experiments were compiled (Table 1) and, when viewed collectively and critically, led to a model for the three-dimensional structure of the human SRP RNA (Fig. 2) serving as a representative example for all SRP RNAs. The modeling process was assisted by the ERNA-3D software, a program that generates A-form RNA from the base-paired regions. The helical sections are connected by single strands derived from reiterated rotations along the sugar-phosphate backbone [11]. ERNA-3D allowed us to manipulate the resulting preliminary three-dimensional model in space and real time, as is possible with a physical model but with greater accuracy and ease of use (see Materials and methods).

For additional constraint of the model, we have expanded the comparative sequence analysis approach to identify loops that are apart in the primary and secondary structure but close in three dimensions. Two such interactions were considered and are discussed in detail below. They are, in the small SRP domain, a pseudoknot formed between the loops of helices 3 and 4 (12-UGGC-15 with 33-GCUA-36 in the human SRP RNA), and, in the large domain, a tertiary interaction between the terminal loop of helix 8 (198-GA-199) and the internal loop of helix 5 at 232-GU-233 (Fig. 3).

Helix 5 sections of the adapter were placed adjacent to each other at distances that tend to preserve the A-form helical character of RNA in the single-stranded intersectional regions. The resulting extended character of the adapter is in agreement with the positions of the four sites that are hypersensitive to digestion by micrococcal nucle-

Table 1 (continued)

Data used in the modeling of the SRP RNA structure.

No	N	2°	C	V	R	19	54	68/72
151	C	h6c				M		W
152	G	h6c				M		W
153	G	h6c				M		W
154	G	h6c				M		
155	G	h6c				M		
156	G	h6c				M		
157	A	l				W		
158	C	l		S		W		
159	C	h6b		S		W		
160	A	h6b		S		W		
161	C	h6b		S		W		
162	C	h6b		S		W		
163	A	l						S
164	G	l						S
165	G	h6a						S
166	U	h6a						S
167	U	h6a						S
168	G	h6a						S
169	C	h7						
170	C	h7						
171	U	l						
172	A	l	H					
173	A	l	M					S
174	G	h7			W			S
175	G	h7	L		W			S
176	A	l						S
177	G	h8a	M				+	S
178	G	h8a	L				+	S
179	G	h8a					+	S
180	G	h8a					+	S
181	U	h8a					+	S
182	G	l	H		W		+	S
183	A	l					+	S
184	A	l	H				+	
185	C	l					+	
186	C	l					+	
187	G	h8b	H				+	
188	G	h8b					+	
189	C	h8b					+	
190	C	h8b					+	
191	C	l					+	
192	A	l	H			M	+	S
193	G	l	H				+	S
194	G	l	H		W		+	W
195	U	h8c			S	M	+	
196	C	h8c			S	M	+	W
197	G	h8c	M		W	M	+	W
198	G	tr	H		W		+	W
199	A	tr			S		+	W
200	A	l			S		+	
201	A	l	H		S		+	
202	C	h8c	M			M	+	
203	G	h8c				M	+	S
204	G	h8c	L			M	+	S
205	A	l	H				+	S
206	G	l	H				+	S
207	C	l	H				+	
208	A	l	H				+	S
209	G	h8b	H				+	S
210	G	h8b					+	S
211	U	h8b			S		+	
212	C	h8b			S		+	
213	A	l			S		+	
214	A	l			S		+	
215	A	l	L		S		+	S
216	A	h8a			S		+	S
217	C	h8a					+	
218	U	h8a					+	
219	C	h8a	L				+	
220	C	h8a	L				+	
221	C	l	M					
222	G	l	M					
223	U	h5k						
224	G	h5k						
225	C	h5k						
226	U	h5k						
227	G	h5k			W			S
228	A	h5k			S			S
229	U	l			S			
230	C	l	M		S			
231	A	l			S			
232	G	tr	H		W			
233	U	tr	H		S			
234	A	h5j			S			S
235	G	h5j	M		W			S
236	U	h5j						S
237	G	h5j						
238	G	h5i	H					S
239	G	h5i	M					S
240	A	l	H					S
241	U	l	H					
242	C	l						
243	G	l	R		W			S
244	C	h5h	M					
245	G	h5h	R					S
246	C	h5h		W				
247	C	h5h		S				
248	U	h5h		W				
249	G	h5g						S
250	U	h5g						
251	G	l	M		W			S
252	A	l	HS		S			S
253	A	h5f	HS		S			S
254	U	h5f			S			
255	A	h5f			S			
256	G	h5f						
257	C	h5f						
258	C	h5e						
259	A	h5e						
260	C	h5e						
261	U	h5e						
262	G	h5e						
263	C	l						
264	A	l	W					
265	C	l		W				
266	U	l		S				
267	C	l		W				
268	C	h5d						
269	A	h5d						
270	G	h5d						
271	C	h5c						
272	C	h5c						
273	U	h5c		S				
274	G	h5c		S				
275	U	h5c		S				
276	G	h5c		S				
277	C	h5c		S				
278	A	l		S				
279	A	l			S			
280	C	l			S			
281	A	l			S			
282	U	l						
283	A	h5b						
284	G	h5b						
285	C	h5b						
286	G	h5b						
287	A	h5b						
288	G	l						
289	A	l	H					
290	C	l	H					
291	C	h5a						
292	C	h5a						
293	C	h5a						
294	G	h5a						
295	U	h5a						
296	C	h5a						
297	U	h5a						
298	C	h5a						
299	U	l						
300	U	l						
301	U	l						

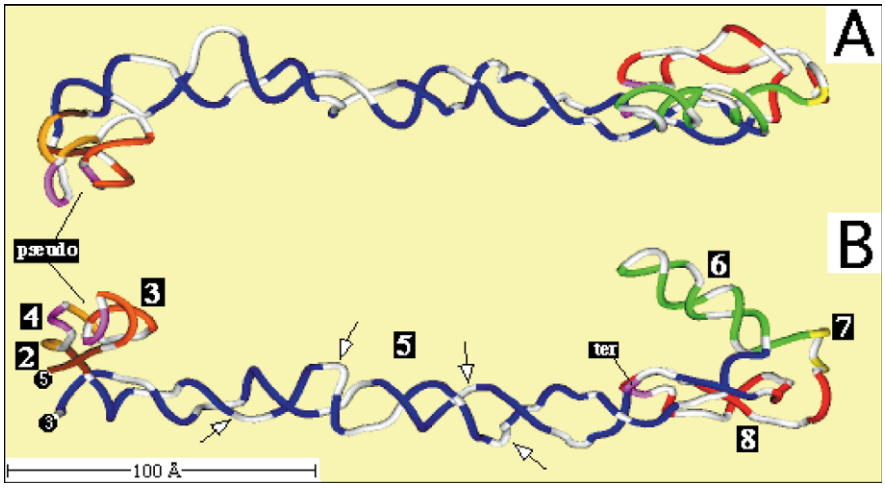
No, nucleotide position. N, nucleotides A, C, G, or U. 2°, RNA secondary structure feature: h, helices numbered according to the nomenclature of Larsen & Zwieb [8] with letter-suffixed sections; knt, a base that is part of the pseudoknot in the small domain of the SRP; tr, a base that participates in the tertiary interaction of the large domain. C, degree of conservation in a representative set of aligned eukaryotic SRP RNA sequences (see Materials and methods): H, conserved in more than 90% of the sequences; M, conserved in 80–90% of the sequences; L, conserved in 70–80% of the sequences; R indicates the conservation of a purine; Y, conserved pyrimidine; K, conserved G or U; W, conserved A or T; HS marks a site that is hypersensitive to digestion by micrococcal nuclease [15]. V, accessibility of the human SRP RNA to the double-stranded RNA-specific nuclease V1 [15]: S, highly accessible site; W, weakly accessible site. R, accessibility of the human SRP RNA to single-strand-specific RNases [15]: S, highly accessible site; W, weakly accessible site. 9/14, nucleotides that are protected (P) from hydroxyl radical cleavage by the SRP9/SRP14 protein heterodimer [16] or bases that are inaccessible (I) to an Fe-EDTA reagent in the naked RNA. 19, bases that affect the binding of protein SRP19 when altered by site-directed mutagenesis [27,28]: S, strong; M, intermediate; W, weak effect. 54, binding site of protein SRP54 as derived by the RNase protection analysis of the *Mycoplasma* SRP [30] are indicated by the + signs. 68/72, bases protected from cleavage by α-sarcin in the presence of the SRP68/SRP72 protein heterodimer [31].

ase ([13]; Fig. 2b). The human SRP RNA appears in the three-dimensional model as an elongated molecule with an maximum length of 260 Å (the distance between U300 and A172) and a maximum width of 70 Å (the distance between C116 and G148).

The small SRP domain

The small domain includes, in the human SRP RNA, the ends of the SRP RNA molecule with helices 2–4 and sections 5a and 5b. The boundary between the small domain and the adapter is defined by ribonuclease-sensi-

Figure 2



Three-dimensional model of human SRP RNA. A 100 Å ruler indicates the approximate dimensions of the model (260 Å × 70 Å); two different views (a) and (b) are derived by rotation by 90° around the long (helix 5) axis. The model was constructed on a Silicon Graphics Indigo 2 Extreme workstation with the program ERNA-3D [11] with considerations as described in Results and discussion. The sugar-phosphate backbone is shown as a narrow tube with base-paired regions of each helix colored differently: helix 2, dark brown; helix 3, orange; helix 4, light brown; helix 5, dark blue; helix 6, green; helix 7, yellow; helix 8, red; the pseudoknot interaction (pseudo) and the tertiary interaction (ter) are shown in pink. The 5′ and 3′ ends and helices 2–8 are labeled in the bottom view. Arrows show the sites of hypersensitivity to micrococcal nuclease digestions [15]. These four sites demarcate the small domain (left), the large domain (right), and the adapter (center). Small and large domains of the SRP RNA are described in detail in Figs 4,5.

tive sites at positions 64/65 and 280 ([13]; Table 1; Fig. 1).

Comparative sequence analysis strongly supports the formation of a pseudoknot between the loops of helices 3 and 4 in the SRP RNAs of the Archaea, *Clostridium*, and the Bacilli [8]. Figure 3a shows that a similar interaction is present in the SRP RNAs of eukaryotes, although the assignment is supported less stringently. The number of base pairs is particularly low in the plant sequences, and this may be because these organisms operate at comparatively lower average temperatures at which there is a reduced tendency to break hydrogen bonds. As shown in Figure 3, but not included in the three-dimensional model, the interaction may involve A–G and A–C non-Watson–Crick 'pairings', which occur commonly elsewhere in otherwise well supported regions [8,9].

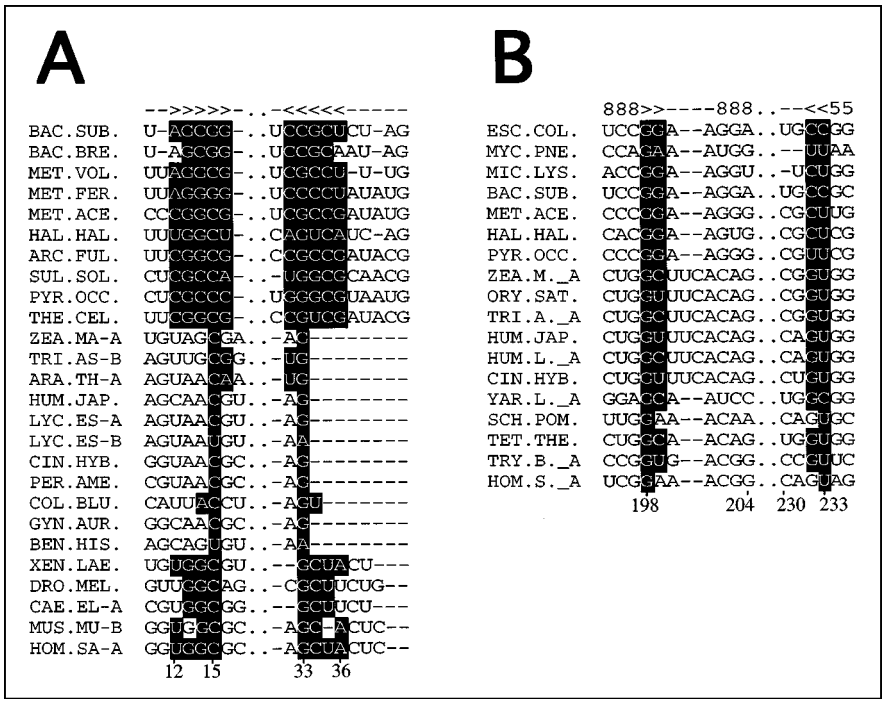
The suggested pseudoknot introduces considerable constraint upon the folding of the small SRP domain (Fig. 4). There is further restriction caused by a proposed coaxial stacking [14] between helices 2 and 4. The need to preserve helix stacking is supported by the immediate transition from helix 4 to helix 2 between A43 and G44 as seen in the alignment of the archaeal and eukaryotic SRP RNA sequences [9]. Furthermore, G44 and G45 are recognized by the double-strand-specific RNase V1 [15], demonstrating that they are accessible and located within a larger helical region. The transitional C3–G44 base pair is conserved which emphasizes its importance, potentially for the overall stability of the stack (Table 1).

The proposed structure of the mammalian SRP RNA in the small domain is mutually supported by findings listed in Table 1. Of those, we note that the bases at positions 2–4 and positions 15 and 16 are hidden within the folded RNA (marked dark blue in Fig. 4a) and are protected from hydroxyl radical cleavage [16]. In further agreement with the proposed structure, regions that are attacked by RNases [15] are located on the outside. The nuclease accessibilities of the 5′ portion of helical sections 5a and 5b are also consistent with the model.

Nucleotides at positions 5–10 and 17–29 are protected by the SRP9/SRP14 heterodimer from chemical modifications and cluster on one side of the pseudoknot-forming helices (Fig. 4). Two additional protected patches are located in helices 5a and 5b at positions 48–51 and 58–62 [16]. From these data it seems likely that the proteins are sandwiched between the knot and helix 5. Using another set of chemical reagents the areas that were protected in the SRP, but not the SRP RNA, included the 5′ portion of the pseudoknot and the 3′ portion of helices 4 and 2. It was suggested that these may be due to indirect effects of the proteins [17].

The binding of proteins SRP9/SRP14 is expected to limit movement around U47 considerably (Figs 1,2,4). Of course, the precise relative orientation of the protein-binding sites remains to be determined and must take into account the size and shape of the SRP9/SRP14 heterodimer. The interhelically located nucleotides U23, G24, and U25 are in an accessible loop that is likely to constitute an important protein-binding site [16]. For proximity reasons (Fig. 4), we considered an interaction between this region and an internal loop in helix 5 (nucleotides A289 and C290) but, because of the conserved nature of the involved nucleotides, this is irres-

Figure 3



Alignment of SRP RNA sequences. (a) Alignment in the pseudoknot region of the small SRP domain. (b) Alignment in the regions that are involved in an interaction between the tip of helix 8 and an internal loop in helix 5. Sequences are ordered phylogenetically with the bacteria on top; species IDs (left column of each panel) are as in the SRP database (SRPDB) [9]. Only one representative sequence is shown when it is identical to the sequence of a close relative. BAC.SUB., *Bacillus subtilis*; BAC.BRE., *Bacillus brevis*; MET.VOL., *Methanococcus voltae*; MET.FER., *Methanothermus fervidus*; MET.ACE., *Methanosarcina acetivorans*; HAL.HAL., *Halobacterium halobium*; ARC.FUL., *Archaeoglobus fulgidus*; SUL.SOL., *Sulfolobus solfataricus*; PYR.OCC., *Pyrodictium occultum*; THE.CEL., *Thermococcus celer*; ZEA.MA., *Zea mays*; TRI.AS., *Triticum aestivum*; ARA.TH., *Arabidopsis thaliana*; HUM.JAP., *Humulus japonicus*; LYC.ES., *Lycopersicon esculentu*; CIN.HYB., *Cineraria hybrida*; PER.AME., *Persea americana*; COL.BLU., *Coleus blumei*; GYN.AUR., *Gynura aurantiaca*; BEN.HIS., *Benincasa hispida*; XEN.LAE., *Xenopus laevis*; DRO.MEL., *Drosophila melanogaster*; CAE.EL., *Caenorhabditis elegans*; MUS.MU., *Mus musculus*; HOM.SA., *Homo sapiens*; ESC.COL., *Escherichia coli*; MYC.PNE., *Mycoplasma pneumoniae*; MIC.LYS., *Micrococcus lysodeikticus*; ORY.SAT., *Oryza sativa*; HUM.LU., *Humulus lupulus*; YAR.LI., *Yarrowia lipolytica*; SCH.POM., *Schizosaccharomyces pombe*; TET.THE., *Tetrahymena thermophila*; TRY.BR., *Trypanosoma brucei*. Base pairs of the pseudoknot and the tertiary interaction are shown in reverse print and indicated by >> <<; portions of helices 8 and 5 are indicated by numbers in the top row of (b), double dots between the interacting regions represent a variable number of nucleotides, and numbers in the bottom row are nucleotide positions in the human SRP RNA. The current alignment of all SRP RNA sequences is available from the SRPDB at the Internet address http://pegasus.uthct.edu/SRPDB/SRPDB.html or by request from the corresponding author.

olute. The ends of the RNA were placed relatively closely to each other to allow the attachment of helix 1 as found in the Archaea and the Bacilli. As the corresponding interhelical loops consist of at least four nucleotides [9], however, this information provides only slight constraints and is therefore of limited value.

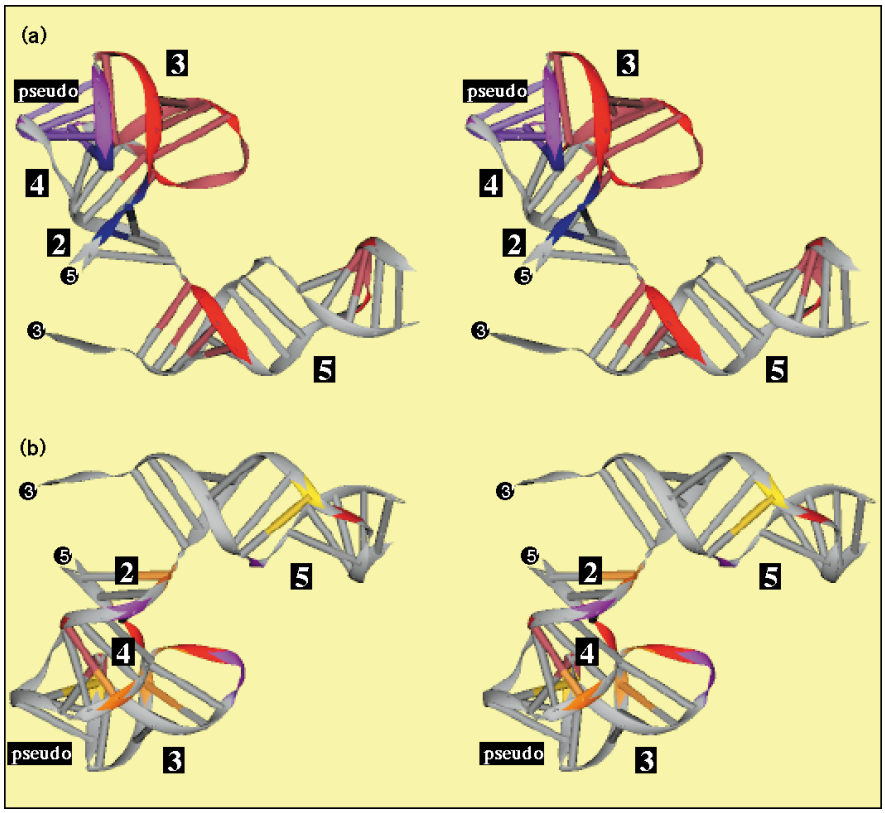
The large SRP domain

In humans, the large SRP RNA domain consists of a continuous 152-nucleotide RNA strand that includes in its secondary structure sections 5g–5k and helices 6, 7, and 8 (Fig. 1). Its boundaries are defined by micrococcal nuclease hypersensitive sites at positions 101 and 252 ([13]; see also Table 1 and Fig. 2b). Sections 5g–5k and helix 8 align with the bacterial 4.5S RNA and contain a cluster of conserved nucleotides [8,9].

We have constrained the three-dimensional structure of the large domain by a new tertiary interaction between the terminal loop of helix 8 (positions 198 and 199) and an internal loop located in helix 5 (positions 232 and 233; see Figs 1,5). These sites were identified first in an subalignment of representative SRP RNA sequences by the computer program Consensus Matrix ([18]; see Materials and methods) and are present in all SRP RNAs. Because of the conserved nature of G198, its pairing with U233 cannot be confirmed or disproved by compensatory base changes, but there are three nonconflicting U→C transitions in *Escherichia coli*, *Bacillus subtilis*, and the yeast *Yarrowia lipolytica*. The support for the pairing between positions 199 and 232 is robust, with commonly occurring A–G 'pairs' in the human and *Schizosaccharomyces pombe* sequences (Fig. 3b).

The terminal loop of helix 8, postulated to be engaged in the tertiary interaction, is a tetranucleotide loop (tetraloop) with the consensus sequence GNRA (where N can be any nucleotide and R any purine) in all but the plant SRP RNAs. Tetraloops occur frequently in many large RNA molecules [19] where, because of their peculiar structures [20,21], they increase the stability of the corresponding helix. It has been noticed that the GNRA-motif is structurally similar to a U-turn (or UNR-motif) [22], ideally suited for RNA tertiary interactions [23]. Interest-

Figure 4



Relaxed stereo views of the structure of SRP RNA in the small SRP domain. The drawings were made on a Silicon Graphics Indigo 2 Extreme with the program DRAWNA [42]. The 5′ and 3′ ends, helices 2–5, and the pseudoknot (pseudo) are labeled. (a) The loops of helices 3 and 4 are shown in purple; sites that are protected from hydroxyl radical cleavage by the SRP9/SRP14 protein heterodimer [16] are marked on the sugar-phosphate backbone in red and on the bases in orange; sites that are protected by the reagent in the naked RNA are in dark blue. (b) Shows the same structure turned approximately 180° around the helix 5 axis and highlights the degree of conservation for each base. Positions that are conserved or invariant in more than 90% of the representative sequences (Materials and methods) are shown in purple; bases with an intermediate degree of conservation (80–90%) are shown in red; bases that are conserved in 70–80% of the sequences are colored orange; and two sites (G14 and K55) are colored yellow, indicating their bias towards R (purine) and K (G or U), respectively.

ingly, plant SRP RNAs replace the tetraloop with a hexaloop that contains the sequence UCA [8,9], thereby conforming to a UNR-motif and the potential to perform the same function as the GNRA-tetraloop.

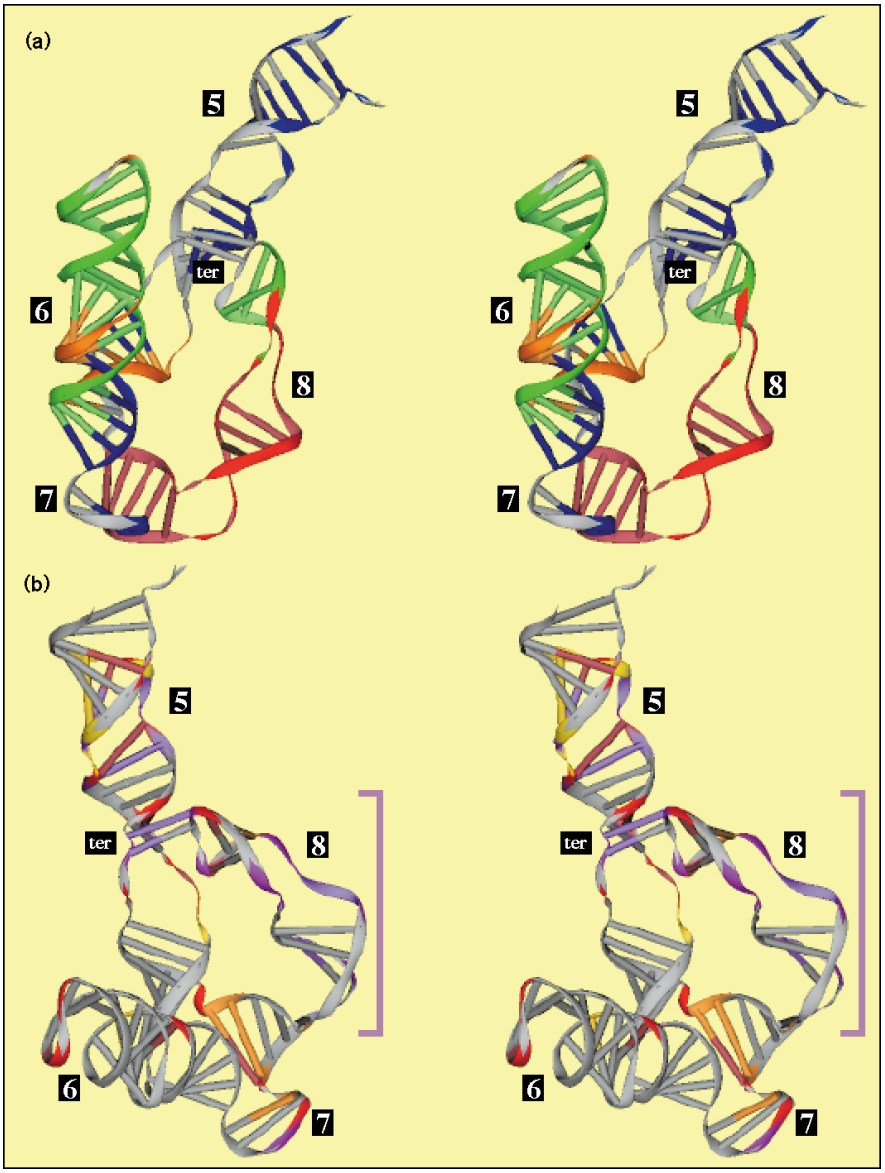
The proposed tertiary contact requires motion in the axes of sections 5k and helix 8 such that the RNA in helix 8 forms a 'knuckle'. This design is favored by the presence of two pronounced internal loops in helix 8 of all SRP RNAs, by the degree of freedom that is present at the junction of helices 5, 6, 7, and 8 (Fig. 1), and by the evidence for a bend in the equivalent portion of the bacterial SRP RNA [24].

The precise spatial placement of the various helix sections of the large domain is necessarily somewhat arbitrary. We have, however, considered nuclease and chemical susceptibility data ([15,25]; Table 1) to bring accessible sites to the periphery in the model of the human SRP RNA. For example, the RNase V1 sensitive sites C114–C116 and C246–U248 are located externally, as are the nuclease-accessible clusters at the junction of 5k with 6a (U128–A130), the internal loop in helix 8 (U213–A215), and the loop that connects sections 5k and 5j (U229–U233).

Because helix 6 lacks conspicuous internal loops, it appears to form a single more-or-less continuous stack with inconsequential deviation of the axial directions. This view is supported by the RNase V1 cuts at C143/C144 and C158–C162, and by the fact that base pairings, but not individual nucleotides, are conserved in helix 6. Two exceptions to this rule are the moderately conserved G135 and the tetraloop (positions 147–150; Table 1) which has the consensus sequence GNAR. We placed helix 6, with the RNase V1 sensitive sites at the periphery, at a protruding angle such that protein SRP19 is accommodated in the core of the large domain and, in relation to helix 5, is located on the same side as helix 8 (Fig. 5). In agreement with this aspect of our model, systematic site-directed mutagenesis showed effects on the binding of SRP19 when helix 6 was altered, but also when the conserved A192 or bases in section 8c were changed [26–28].

The domain of SRP54 that interacts with SRP RNA (M-domain) is large enough to cover a significant portion of helix 8 [29]. This view is justified by the fact that helix 8, a portion of helix 5, and the SRP54 homolog (ffh or P48) are the sole components of the bacterial SRP. Furthermore, ffh protected only helix 8 from RNase digestion in the *Mycoplasma* SRP [30]. Nevertheless, helix 8 is also in

Figure 5



Relaxed stereo views of the structure of SRP RNA in the large SRP domain. The drawings were made on a Silicon Graphics Indigo 2 Extreme with the program DRAWNA [42]. Labeled are helices 5–8 and the tertiary interaction (ter) involving the tip of helix 8 and an internal loop in helix 5. (a) Binding sites for proteins SRP19 (green, pronounced mutational effects; orange, mild effects), SRP54 (red), and SRP68/SRP72 (blue). (b) Degree of conservation for each base as in Fig. 4: bases conserved or invariant in more than 90% of the representative sequences are shown in purple; bases with an intermediate degree of conservation (80–90%) are shown in red; bases that are conserved in 70–80% of the sequences are colored orange. Seven sites in yellow represent conserved pyrimidines (U108, U110, and U121), conserved purines (G150, G243, and G225), and U137 (conserved as a G or U). The pink bracket indicates the conserved RNA knuckle, exposed to bind protein SRP54.

contact with SRP19 and possibly with SRP68/SRP72, or involved in tertiary interactions.

As could be expected from the large size of proteins SRP68/SRP72, their RNA-binding area is complex and diffuse, with sites that are located throughout the large domain. Several regions needed to bind SRP19 and SRP54 (e.g. sections 6a, 6b, 8b, and 8c; see Table 1) are also protected from digestion by α-sarcin when SRP68/SRP72 is present [31]. Conformational changes that occur in the SRP [17] may cause complications in the current interpretation of the experimental results. Despite this impediment, we wish to point out the SRP68/SRP72 interactions with the 3′-proximal part of helix 6 (A163–G168), helix 7 (A173–A176), and all SRP68/SRP72-binding sites in helix 5 (Table 1; Fig. 5) as compatible with our model of the SRP RNA.

Hierarchical assembly of SRP19 and SRP54

The assembly of the large SRP domain occurs in a stepwise fashion whereby protein SRP19 binds first, followed by SRP54 [2,32–34]. In the absence of SRP RNA there are no discernible interactions between SRP19 and SRP54, so the RNA appears to function as a mediator of SRP assembly, most likely through changes in the RNA conformation that occur upon binding of SRP19.

Conformational changes in the RNA are supported by the finding that there are two electrophoretic variants of the SRP RNA, called A-form and B-form [35]. Furthermore, SRP19 enhances the RNase accessibility of certain sites (e.g. U121/122, G135/136, G174/175, and C190) and binds to the high-mobility B-conformer about 3.5 times better than to the A-conformer [25]. As decreased flexibility and/or increased compactness are among the factors that cause RNA to migrate faster during electrophoresis in native polyacrylamide gels [36,37], we hypothesize that the SRP RNA exists either as a closed form that contains the proposed tertiary interaction between helix 8 and helix 5, or as an open form in which this interaction is absent. The rigid RNA molecule (the B-form) may bind SRP19 more avidly because it closely resembles the SRP19-binding site. As the SRP19–RNA complex exists as a single electrophoretic species [25], the interaction with the loosely structured A-conformer appears to cause a shift to a B-like conformer. This interpretation resolves an apparent discrepancy between data from site-directed mutagenesis experiments of the helix 8 tetraloop (198-GAAA-201) that indicated no effects on the binding of SRP19 [27] and earlier findings that suggested an interaction between the helix 8 tetraloop and SRP19 [31]. As G197 and G198 were less accessible to RNase T1 in the presence of SRP19 [25], our model suggests that the protection of these residues is due to a tertiary RNA–RNA contact with helix 5 (Fig. 5).

As a result of the binding of SRP19 to SRP RNA, an RNA 'knuckle' is formed in helix 8 to generate a binding site for SRP54. Helix 8 includes, among others, nucleotides 192-AGG-194, and 205-AGCAG-209 (Table 1) which are part of a conserved core in all SRP RNAs (Fig. 5). It is currently unclear whether SRP19 simply selects RNA molecules that fold properly thereby pushing the conformer equilibrium towards the compact B-like form [38] and an SRP54-favored helix 8 conformation, or whether it plays an more active role in creating the SRP54-binding site.

Nucleotides U117 and A118, which are located across from the tertiary interaction at 232-GU-233 (Fig. 1), are accessible to double-strand-specific and single-strand-specific nucleases ([15]; Table 1). This apparent contradiction is explained by the formation of a phylogenetically supported A118–U233 base pair in the A-conformer, thereby extending helical section 5j, which would be present in only a subset of SRP RNA molecules.

Implications for the structure of SRP

The dimensions of the human SRP RNA model approximate what has been determined in the electron microscope [5,6] with the model being slightly larger. The differences, in the order of 20–40 Å, may reflect ambiguities or dehydration in the electron microscopic determinations and/or our model building. For example, helical sections 5c–5f may zigzag and may not be oriented coaxially as shown in the model, thereby providing room for variation in the length of the long axis. Our limited knowledge about the size and shape of the SRP68/SRP72 heterodimer (discussed above) will almost certainly make it necessary to improve the model gradually when these data become available. Studies may have to take into account a substantial flexibility of the SRP including possibly major changes in SRP RNA conformation [17].

We view our model as a first step in the detailed molecular analysis of the SRP. Furthermore, the model will provide a framework to understand SRP structure and function in context with the other players in the SRP cycle, such as the ribosome and membrane substituents.

Materials and methods

Comparative sequence analysis

Alignment procedures have been described in detail previously [8]. Briefly, sequences of close relatives were grouped first on the basis of their primary structure similarity. Each group of aligned sequences was then aligned against other groups. Sets of conserved nucleotides were identified to align the more variable regions. Lastly, common secondary structural elements were used as markers in the more variable regions. Alignment columns were examined visually for compensating base changes of the Watson–Crick-type (including G–U pairs) that would support the existence of a particular base pair. Only base pairs for which there was at least twice as much positive evidence as negative were considered valid.

The RNA–RNA interaction in the large SRP domain (198-GA-199 with 232-GU-233) was identified with the program Consensus Matrix, version 1.86 [18]. Because of matrix size limitations of the program, a representative subalignment was generated with SRP RNA sequences from the following organisms, listed according to their ID in the SRPDB [9]: THE.THE., LEG.PNE., PSE.AER., ESC.COL., MYC.PNE., MYC.MYC., MIC.LYS., BAC.SUB., BAC.STE., BAC.CER., BAC.THU., BAC.MEG., BAC.BRE., MET.VOL., MET.FER., MET.THE., MET.ACE., HAL.HAL., ARC.FUL., SUL.SOL., PYR.OCC., THE.CEL., ZEA.M.-A, ORY.SAT., TRI.A.-A, ARA.T.-A, HUM.JAP., HUM.L.-A, LYC.E.-A, CIN.HYB., YAR.L.-A, SCH.POM., TET.THE., TRY.B.-A, XEN.LAE., DRO.MEL., CAE.E.-A, RAT.RAT., CAN.SPE., and HOM.S.-A. The thresholds in the analysis of covariation were set to one base pair and to a display filter value of 0.37; G–U interactions were allowed to form.

Nucleotide conservations of the eukaryotic SRP RNAs were calculated from a representative alignment that included the sequences from ZEA.M.-A, TRI.A.-A, ARA.T.-A, HUM.JAP., LYC.E.-A, CIN.HYB., YAR.L.-A, SCH.POM., TET.THE., TRY.B.-A, XEN.LAE., DRO.MEL., CAE.E.-A, RAT.RAT., CAN.SPE., and HOM.S.-A (listed with their SRPDB ID [9]).

SRP RNA model building

At an early stage, a physical model (scale 1 Å ≈2 mm), made of wire for the sugar-phosphate backbone and cylinders for the helix sections, was used to test basic ideas about the folding of the SRP RNA in three dimensions. Although still useful for that purpose, greater accuracy is provided by models in electronic form. We have used a pre-release version of the program ERNA-3D on an Silicon Graphics Indigo 2 Extreme workstation running IRIX version 5.2. As input for ERNA-3D we used the secondary structure of human SRP RNA determined by comparative sequence analysis [8], including a pseudoknot in the small domain. The three-dimensional model generated by the ERNA-3D was viewed with wireless LCD stereo glasses (Stereographics). Individual helices or helix-sections were treated as clusters and manipulated with the mouse and the on-screen cursor box. The conformations of the

single-stranded regions or loops were accepted as calculated by the program, unless collisions occurred, in which case the strands were untangled manually.

Experimental data were compiled and incorporated into the model by color-coding of the bases. Repetitively, the structure was modified until it satisfied the experimental results. Structural details without theoretical or experimental support (such as in the adapter and, possibly, in the large SRP domain) were not included, resulting in a minimal three-dimensional model. The structure is available in PDB file format [39] from the SRPDB at the Internet address http://pegasus.uthct.edu/SRPDB/SRPDB.html or by request from the corresponding author.

Acknowledgements

We thank Shaun D Black and Richard Brimacombe for critical reading of the manuscript. This work was supported by NIH grant GM-49034 to C Zwieb.

References

1. Lütcke, H. (1995). Signal recognition particle (SRP), a ubiquitous initiator of protein translocation. *Eur. J. Biochem.* 228, 531–550.
2. Walter, P. & Blobel, G. (1983). Disassembly and reconstitution of the signal recognition particle. *Cell* 34, 525–533.
3. Römisch, K., *et al.*, & Dobberstein, B. (1989). Homology of 54K protein of signal-recognition particle, docking protein and two *E. coli* proteins with putative GTP-binding domains. *Nature* 340, 478–482.
4. Bernstein, H., Poritz, M., Strub, K., Hoben, P.J., Brenner, S. & Walter, P. (1989). Model for signal sequence recognition from amino-acid sequence of 54K subunit of signal recognition particle. *Nature* 340, 482–486.
5. Andrews, D., Walter, P. & Ottensmeyer, F. (1985). Structure of signal recognition particle by electron microscopy. *Proc. Natl. Acad. Sci. USA* 82, 785–789.
6. Andrews, D., Walter P. & Ottensmeyer, F. (1987). Evidence for an extended 7SL RNA structure in the signal recognition particle. *EMBO J.* 6, 3471–3477.
7. Zwieb, C. & Schüler, D. (1989). Low resolution three-dimensional models of the 7SL RNA of the signal recognition particle, based on an intramolecular cross-link introduced by mild irradiation with ultraviolet light. *Biochem. Cell Biol.* 67, 434–442.
8. Larsen, N. & Zwieb, C. (1991). SRP–RNA sequence alignment and secondary structure. *Nucleic Acids Res.* 19, 209–215.
9. Larsen, N. & Zwieb, C. (1996). The signal recognition particle database (SRPDB). *Nucleic Acids Res.* 24, 80–81.
10. Larsen, N. (1992). Higher order interactions in 23s rRNA. *Proc. Natl. Acad. Sci. USA* 89, 5044–5048.
11. Müller, F., *et al.*, & Brimacombe, R. (1995). Getting closer to an understanding of the three-dimensional structure of ribosomal RNA. *Biochem. Cell Biol.* 73, 767–773.
12. Zwieb, C. (1989). Structure and function of signal recognition particle RNA. *Prog. Nucleic Acid Res. Mol. Biol.* 37, 207–234.
13. Gundelfinger, E.D., Krause, E., Melli, M. & Dobberstein, B. (1983). The organization of the 7SL RNA in the signal recognition particle. *Nucleic Acids Res.* 11, 7363–7374.
14. Walter, A., *et al*., & Zuker, M. (1994). Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl. Acad. Sci. USA* 91, 9218–9222.
15. Gundelfinger, E.D., DiCarlo, M., Zopf, D. & Melli, M. (1984). Structure and evolution of the 7SL RNA component of the signal recognition particle. *EMBO J.* 3, 2325–2332.
16. Strub, K., Moss, J. & Walter, P. (1991). Binding sites of the 9- and 14-kilodalton heterodimeric protein subunit of the signal recognition particle (SRP) are contained exclusively in the Alu domain of SRP RNA and contain a sequence motif that is conserved in evolution. *Mol. Cell. Biol.* 11, 3949–3959.
17. Andreazzoli, M. & Gerbi, S.A. (1991). Changes in 7SL RNA conformation during the signal recognition particle cycle. *EMBO J.* 10, 767–777.
18. Davis, J.P., Janjic, N., Pribnow, D. & Zichi, D.A. (1995). Alignment editing and identification of consensus secondary structures for nucleic acid sequences: interactive use of dot matrix representations. *Nucleic Acids Res.* 23, 4471–4479.
19. Woese, C.R., Winker, S. & Gutell, R.R. (1990). Architecture of ribosomal RNA: constraints on the sequence of 'tetra-loops'. *Proc. Natl. Acad. Sci. USA* 87, 8467–8471.
20. Cheong, C., Varani, G. & Tinoco, I.J. (1990). Solution structure of an unusually stable RNA hairpin, 5′GGAC(UUCG)GUCC. *Nature* 346, 680–682.
21. Heus, H.A. & Pardi, A. (1991). Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* 253, 191–194.
22. Jucker, F. & Pardi, A. (1995). GNRA tetraloops make a U-turn. *RNA* 1, 219–222.
23. Quigley, G. & Rich, A. (1976). Structural domains of transfer RNA molecules. The ribose 2 hydroxyl which distinguishes RNA from DNA plays a key role in stabilizing tRNA structure. *Science* 194, 796–806.
24. Lentzen, G., Dobberstein, B. & Wintermeyer, W. (1994). Formation of SRP-like particles induces a conformational change in *E. coli* 4.5S RNA. *FEBS Lett.* 348, 233–238.
25. Walker, P., Black, S. & Zwieb, C. (1995). Cooperative assembly of signal recognition particle RNA with protein SRP19. *Biochemistry* 34, 11989–11997.
26. Zwieb, C. (1991). Interaction of protein SRP19 with signal recognition particle RNA lacking individual RNA helices. *Nucleic Acids Res.* 19, 2955–2960.
27. Zwieb, C. (1992). Recognition of a tetranucleotide loop of signal recognition particle RNA by protein SRP19. *J. Biol. Chem.* 267, 15650–15656.
28. Zwieb, C. (1994). Site-directed mutagenesis of signal recognition particle RNA: identification of the nucleotides in helix 8 required for interaction with protein SRP19. *Eur. J. Biochem.* 222, 885–890.
29. Czarnota, G., Andrews, D., Farrow, N. & Ottensmeyer, F. (1994). A structure for the signal sequence binding protein SRP54: 3D reconstruction from STEM images of single molecules. *J. Struct. Biol.* 113, 35–46.
30. Samuelsson, T. (1992). A *Mycoplasma* protein homologous to mammalian SRP54 recognizes a highly conserved domain of SRP RNA. *Nucleic Acids Res.* 20, 5763–5770.
31. Siegel, V. & Walter, P. (1988). Binding sites of the 19-kDa and 68/72-kDa signal recognition particle (SRP) proteins on SRP RNA as determined in protein–RNA `footprinting'. *Proc. Natl. Acad. Sci. USA* 85, 1801–1805.
32. Römisch, K., Webb, J., Lingelbach, K., Gausepohl, H. & Dobberstein, B. (1990). The 54-kD protein of signal recognition particle contains a methionine-rich RNA binding domain. *J. Cell Biol.* 111, 1793–1802.
33. Zopf, D., Bernstein, H.D., Johnson, A.E. & Walter, P. (1990). The methionine-rich domain of the 54 kd protein subunit of the signal recognition particle contains an RNA binding site and can be crosslinked to a signal sequence. *EMBO J.* 9, 4511–4517.
34. Krolkiewicz, S., Sänger, H. & Niesbach-Klösgen, U. (1994). Structural and functional characterisation of the signal recognition particle-specific 54-kD protein (SRP54) of tomato. *Mol. Gen. Genet.* 245, 565–576.
35. Zwieb, C. & Ullu, E. (1986). Identification of dynamic sequences in the central domain of 7SL RNA. *Nucleic Acids Res.* 14, 4639–4657.
36. Gast, F. & Hagerman, P. (1991). Electrophoretic and hydrodynamic properties of duplex ribonucleic acid molecules transcribed *in vitro*: evidence that A-tracts do not generate curvature in RNA. *Biochemistry* 30, 4268–4277.
37. Tang, R. & Draper, D. (1994). Bend and helical twist associated with a symmetric internal loop from 5S ribosomal RNA. *Biochemistry* 33, 10089–10093.
38. Weeks, K. & Cech, T. (1996). Assembly of a ribonucleoprotein catalyst by tertiary structure capture. *Science* 271, 345–348.
39. Bernstein, F., *et al.*, & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112, 535–542.
40. Levitt, M. (1969). Detailed molecular model for transfer ribonucleic acid. *Nature* 224, 759–763.
41. Fox, G. & Woese, C. (1975). 5S RNA secondary structure. *Nature* 256, 505–507.
42. Massire, C., Gaspin, C. & Westhof, E. (1994). DRAWNA: a program for drawing schematic views of nucleic acids. *J. Mol. Graphics* 12, 201–206.