ELSEVIER

Construction and analysis of base-paired regions of the 16S rRNA in the 30S ribosomal subunit determined by constraint satisfaction molecular modelling

Michael A. Dolan,[1] Patricia Babin, and Paul Wollenzien

Department of Biochemistry, North Carolina State University, Raleigh, North Carolina, USA

Structure models for each of the secondary structure regions from the Escherichia coli 16S rRNA (58 separate elements) were constructed using a constraint satisfaction modelling program to determine which helices deviated from classic A-form geometry. Constraints for each rRNA element included the comparative secondary structure, H-bonding conformations predicted from patterns of base-pair covariation, tertiary interactions predicted from covariation analysis, chemical probing data, rRNA–rRNA crosslinking information, and coordinates from solved structures. Models for each element were built using the MC-SYM modelling algorithm and subsequently were subjected to energy minimization to correct unfavorable geometry. Approximately two-thirds of the structures that result from the input data are very similar to A-form geometry. In the remaining instances, the presence of internal loops and bulges, some sequences (and sequence covariants) and accessory information require deviation from A-form geometry. The structures of regions containing more complex base-pairing arrangements including the central pseudoknot, the 530 region, and the pseudoknot involving base-pairing between G570-U571/A865-C866 and G861-C862/G867-C868 were predicted by this approach. These molecular models provide insight into the connection between patterns of H-bonding, the presence of unpaired nucleotides, and the overall geometry of each element. © 2001 by Elsevier Science Inc.

[1]Present address: Department of Biophysics, The Johns Hopkins University, Baltimore, Maryland 21218, USA

Corresponding author: P. Wollenzien, Department of Biochemistry, North Carolina State University, Raleigh, North Carolina 27695-762, USA. Tel.: 919-515-5703; Fax: 919-515-2047.

Keywords: molecular modelling; RNA structure; ribosomal RNA; comparative sequence analysis

INTRODUCTION

Computer-assisted modelling of RNA structures using biochemical and comparative sequence information has resulted in the accurate prediction and reproduction of many structures including tRNA,[1] the group I intron,[2] and the hepatitis delta virus ribozyme.[3] For the tRNA, the root-mean-square (rms) deviation between the predicted model and the structure determined by X-ray diffraction was 3.1–3.8 Å. This level of agreement was reached by including constraints for the secondary structure, tertiary interactions, and helical stacking arrangements. In this instance, the RNA backbone fold is correctly predicted.[1] Other RNAs that have been modelled where an atomic resolution structure is not yet known include RNase P[4–7] and individual rRNA elements.[8,9] The fold of larger RNAs such as the 5S rRNA[10] from the *Escherichia coli* large ribosomal subunit and the 16S rRNA from the *E. coli* small ribosomal subunit have been modelled in the past using a variety of computer algorithms.[11–16]

Recently, a structure of the *E. coli* 70S ribosome has been derived from cryo-electron microscopy (cryo-EM) experiments to 11.5 Å resolution.[17] The dimensions and overall organization of both subunits can be discerned and some of the ribosomal components and nature of the connections between ribosomal subunits were identified with the aid of X-ray crystallography data. However, the internal fold of the RNA was not easily distinguished from the protein component of the particle. The structure of the 30S ribosomal subunit from *T. thermophilus* has been determined at 5.5 Å[18] and 4.5 Å[19] by

Journal of Molecular Graphics and Modelling 19, 495–513, 2001

655 Avenue of the Americas, New York, NY 10010

1093-3263/01/$–see front matter
PII S1093-3263(00)00097-8

X-ray crystallography. At these resolutions a portion of the RNA phosphate backbone was resolved and seven crystal structures of small-subunit ribosomal proteins were positioned in the electron density map. A crystal structure of the *T. thermophilus* 70S ribosome complex containing mRNA and tRNA was resolved at up to 7.8 Å resolution.[20] In this instance rRNA elements, particularly in the vicinity of the tRNAs, were successfully identified and described. Recently, an S15, S6, S18-rRNA complex containing a 104 nucleotide RNA fragment from *T. thermophilus* was solved by X-ray crystallography to a resolution where hydrogen bonds within rRNA elements and between ribosomal proteins and rRNA are observed.[21]

Previous computer-assisted models of 16S rRNA were useful in suggesting RNA arrangements within the ribosome and allowed for the design of experiments to investigate ribosomal tertiary structure,[11–16] but little focus has been placed on using experimental information to determine the types of base-pairing and nucleotide arrangements within helical elements. One approach to modelling 16S rRNA included only the phosphates and generated three-dimensional solutions while attempting to meet several hundred experimental constraints.[13] Another approach used manual manipulation of the RNA elements to fit them into an electron density map and succeeded in modelling all-atom rRNA structures.[15,16] Both of these approaches used type-A geometry for each rRNA element and, in the latter case, an energy optimization routine was not used to remove incorrect bond lengths and angles. No computer-assisted modelling approach to date has produced all-atom, energy optimized structures of 16S rRNA that combines secondary and tertiary structure information.

Stereochemical rules and structural data can be expressed in terms of a set of distance constraints from which macromolecular modelling can be very useful in predicting such nucleotide arrangements. Here, we describe the three-dimensional construction of the secondary structure regions of the *E. coli* 16S rRNA using constraint satisfaction modelling. There are several reasons for using this approach. First, since it is possible to systematically investigate how geometry differences at a particular nucleotide affect the whole RNA element, the procedure gives insight into the connection between H-bonding patterns, the significance of unpaired nucleotides, and the overall geometry of each element. This aspect of the approach takes advantage of the very large database of rRNA sequences from which sequence covariation statistics can be calculated.[22] Second, it will be possible to test the modelling procedure eventually, since the ribosome structure is being solved by X-ray analysis at resolutions that should indicate shapes of each region. Third, the predictive procedure is an absolutely general one that can be used even when sequence structure motifs[23] are not present.

There have been many attempts to infer tertiary structures of proteins or nucleic acids from primary sequence. One way to approximate three-dimensional structures is to use constraint satisfaction algorithms that combine stereochemical rules with structural information. For this RNA study, the Macromolecular Conformations by SYMbolic (MC-SYM) program was chosen.[24] An ASCII script containing the sequence, base-pairing arrangements, and geometric constraints of an RNA molecule serve as input into MC-SYM. Nucleotide and atomic coordinates are computed dynamically using the information provided in the script. The order in which the nucleotides are entered into the script indicates the order in which each nucleotide is added to the polymer during construction. If a nucleotide conformation is encountered that conflicts with a set of user-specified constraints, MC-SYM backtracks up the conformational search tree to a previously sampled solution, "prunes" this branch of the search tree, and no longer explores this path. This substantially reduces the number of conformations searched and effectively reduces computational time. A second way that computational time is reduced is by a "look-ahead" feature incorporated into MC-SYM that functions when a loop is detected during construction. The gap to be filled between a nucleotide at the beginning of a loop and a nucleotide at the end of a loop is approximated by a distance equal to the sum of the length of all the nucleotides in the gap, multiplied by a user-specified phosphoester distance between nucleotides. If it is determined that a gap will not possibly be closed, that branch of the search tree is abandoned.

The conformational space to be investigated is determined by a set of pre-computed nucleotide conformations and transformations found in NMR and X-ray structures or obtained from theoretical calculations. Specific sets of conformations and transformations used in the MC-SYM algorithm can be selected by the user. An upper limit of 10,000 constructions was selected as this amount was the maximum number of solutions that were manageable for any given element in this study. In addition, solved structures such as those determined by NMR or crystallography can be imported into MC-SYM and used as a template on which to build.

Structures were based on the comparative secondary structure model of 16S rRNA (Figure 1).[25] Comparative sequence analysis established the presence of all base-pairs in the 16S rRNA and also provided information about the pattern of concerted sequence changes[26] from which H-bonding patterns can be obtained.[27] Chemical probing data to determine base reactivity[28] and N7 reactivities,[29] information from intramolecular rRNA-rRNA UV-crosslinking experiments,[30] structure information by NMR,[31–34] and from isolated rRNA regions determined by X-ray crystallography[35] are also explicitly or implicitly included in the modelling. The scheme used in modelling the regions is: (i) summarize secondary and tertiary structural information and incorporate this data into scripts that serve as input into the MC-SYM program; (ii) generate models consistent with the structural constraints using a progressive increase in conformational freedom in the nucleotide geometries; and (iii) refinement of the models with an energy optimization routine (Figure 2). The results point to those base-paired regions that must be significantly different from A-form helical geometry and provide a resource of structures for future modelling studies.

METHODS

Algorithms Used in the Construction of 16S rRNA Helices

The 16S rRNA secondary structure model[25] was divided into 58 base-paired elements (Figure 1) and numbered according to the scheme of Brimacombe[36] with many of the helices further divided into subsegments and indicated as a, b, etc. Most regions contain a run of base-paired residues interrupted by internal bulges or internal loops, or contain terminal loops. However, several regions contain more complex base-pairing arrangements such as the central pseudoknot (helices 1/2), the

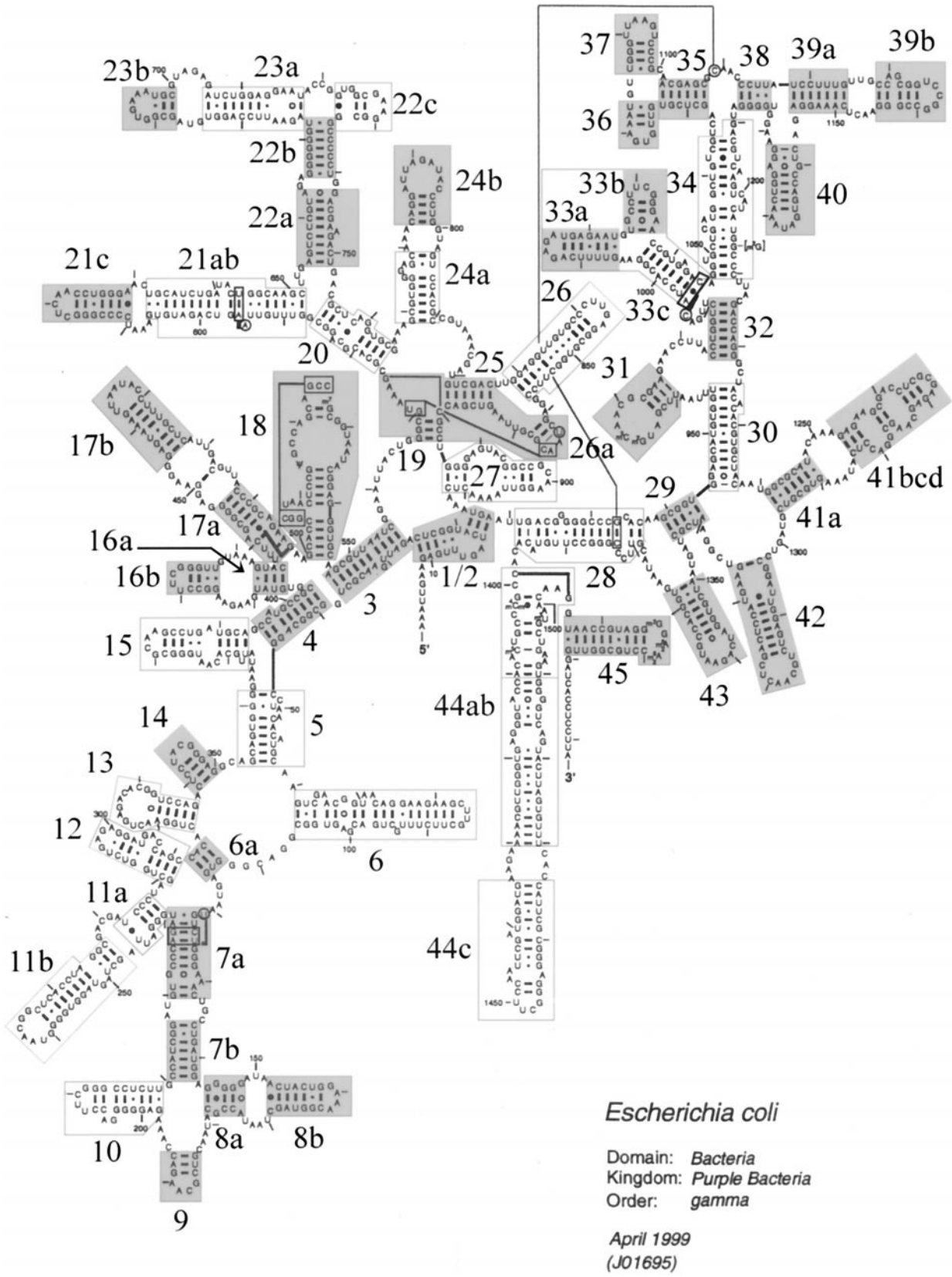

Figure 1. Secondary structure diagram of 16S rRNA.[25] The numbering has been adopted from the work of Brimacombe.[36] Each boxed region indicates a separate model constructed by the MC-SYM program, which sometimes included several base-paired elements. All nucleotides within a given box were included in the modelling. Helices that are A-form or near A-form geometry (deviating by <3 Å) are indicated by a shaded box.

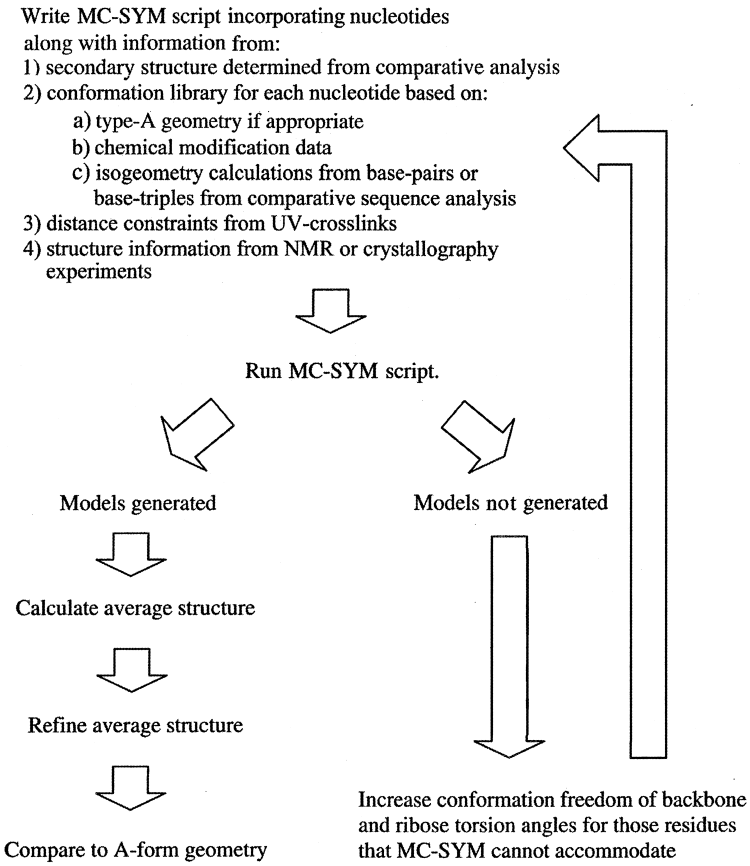

Figure 2. Flow diagram showing the step-wise construction of a 16S rRNA helical region. Secondary structure information such as chemical modification data, UV-crosslinks, comparative analysis to determine base-pairing arrangements, and NMR structures serve as input for the MC-SYM program.[24] MC-SYM use nucleotides with A-form geometries as a first choice to generate structures that incorporate all experimental constraints. If structures are not generated, MC-SYM is allowed to sample nucleotide conformations that deviate from A-form geometry. Structures are subsequently averaged and energy optimized.

central junction (helices 19, 25, and 26a), and the region containing the 530 loop (helix 18) (Figure 1).

Each helical region was modelled first using the Macromolecular Conformations by SYMbolic programming (MC-SYM) tool,[24] a modelling program that uses real nucleotide conformations and a constraint satisfaction algorithm to generate models of RNA structures. The MC-SYM algorithm is used to search the conformational space of RNA such that all generated models are consistent with a given set of constraints, and all possible structures are returned. Atomic coordinates for each nucleotide are computed using the information provided in an ASCII script that contains the RNA description and geometric constraints.

For a given helix, a conservative approach was taken by writing scripts that generated models that were as close to A-form RNA as possible, while attempting to meet all the experimental constraints. Frequently, little constraint information is available for nucleotides contained within internal loops and bulges (see helix 16 in Figure 1 as an example); these nucleotides were stacked into a helix unless involved in a secondary or tertiary arrangement, such as a crosslink. If no structures were produced, conformational freedom was added to a script, resulting in a broader range of torsion angles for the ribose and/or backbone of nucleotides to be sampled. Increased conformational freedom within a script increases the likelihood that resultant models will deviate from A-form geometry. Additional conformational freedom was incorporated in this way until models were produced that met experimental constraints. Generally, this approach had the effect of reducing the number of conformations searched while still meeting experimental data and resulted in the generation of only a small group of models that differed slightly based on their rms deviation. For most helical regions, <30 models resulted that were very similar based on their average rms deviation (<1 Å). All structures were used to calculate an average structure, which was subjected to energy minimization using the Insight II modelling suite (Molecular Simulations, Inc.). Occasionally, the addition of a small amount of conformational freedom to a script would result in a large number of models produced ($\sim$1,000–10,000). In the case of helix 18, $>10{,}000$ models were generated due to a large number of alternative conformations adopted within regions containing unpaired nucleotides caused by a lack of constraint data for these regions (Figure 1). Here, a single model was chosen arbitrarily because: (i) all models produced met a set of input parameters; (ii) all agreed with experimental information; and (iii) the rms deviation for the helical portions between all models was low (1.0 Å for the 500–504/541–545 and 511–515/536–540 helices, measured as a unit). In the case of helix 16, a sizable internal loop caused two groups of models to be produced from a single script. Models were separated into two groups based on differences in rms deviation and an average structure was determined and energy optimized for each group. The average structure of helix 16 presented here was chosen on the basis that it would be better accommodated in a tertiary structure of 16S rRNA, although both averaged structures meet the parameters input into MC-SYM. For helix 27, two alternative base-pairings were investigated separately each in a different script, resulting in two groups of models, both known to be present in the 30S

subunit. An average structure for each group was calculated and both averaged structures were energy optimized.

Experimental data such as those from chemical modification experiments[28,29] provided information as to whether the Watson-Crick side of a base or the N7 position of a base is available for hydrogen bonding. This information was used to select possible base-pairings and conformational arrangements. If many different base-pairings for a given position could be accommodated based on chemical reactivity data for those positions, comparative sequence analysis was performed to determine nucleotide covariance and the results served as input into the ISOPAIR program[27]. The ISOPAIR algorithm is used to find base-pair or base-triple conformations that are common to a set of base-pair or base-triple sequences determined from covariation analyses. For example, the G145●G177 base-pair in helix 8 is known to covary with U●G (21.8%) and C—G (24.1%). The G●G, U●G, and C—G base-pairings serve as input into the ISOPAIR program along with a requested number of H-bonds occurring between bases in the base-pairs. Probable base-pairings for the G145●G177 base-pair that are isogeometric to all three input pairings are output. The program generates a postscript file that graphically represents isomorphic conformations for the G145●G177 base-pair along with an "isomorphism value" (I value). A G●G conformation with an I value closer to zero is more isomorphic to the set of input structures than a conformation with a higher I value. The conformation with the lowest I value was always chosen for utilization with MC-SYM.

Where applicable, UV-crosslinks that occur close in the secondary structure were incorporated into a model, such as the U1052 × C1200 UV-crosslink and the U1064 × G1190 UV-laser crosslink in helix 34. A maximal distance of 4 Å was specified between the C5 atoms of nucleotides participating in a crosslink. This conservative value, approximately twice a covalent bond length, was chosen because the atoms comprising a given crosslink are as yet unknown and because it provides enough conformational flexibility to place two nucleotides in proximity.

Sixteen helical elements in 16S are terminated by four-membered loops referred to as tetraloops.[37] The most frequently occurring tetraloop in 16S rRNA contains the GNRA consensus sequence.[37] The solution structures of the GCAA and GAAA hairpins have been determined by nuclear magnetic spectroscopy and were incorporated into our models where possible or used as a template to model other GNRA tetraloops.[31] Before incorporation into a model, the average solution structure for the tetraloop was calculated. The tetraloop and applicable closing base-pairs were imported into the MC-SYM program and used as a starting point for modelling the remainder of the helix. Other tetraloops, such as the second most common UNCG motif,[38] were incorporated in the same way. Tetraloops with chemical reactivities that differed from that of the solved structures, or tetraloops having no corresponding solved structure, were modelled successfully using MC-SYM, in some cases using a solved structure as a template.[39]

MC-SYM scripts allow the user to specify a maximum distance between the 3′ hydroxyl of one nucleotide to the 5′ phosphate of an adjacent nucleotide during model construction. In general, lower phosphoester distances (~3 Å) were used in building structures thought to contain noncanonical base-pairings and unusual chemical reactivites (less A-form-like), whereas a larger adjacency specification (~ 3 Å) was used for structures having an abundance of Watson-Crick pairings and no unusual chemical reactivity (more A-form-like). Lower phosphoester constraints restrict the number of conformations that are sampled for a given space and reduces solutions that are sterically hindered. For instance, an adjacency value of 6.5 Å for helix 23a produced 104 models, all of which have major steric hindrances consisting of van der Waal's overlaps between several residues. Changing the adjacency constraint to 4.5 Å produces 40 models having minor van der Waal's overlaps and few models having major steric problems. Further reduction of the adjacency distance to 3 Å results in 2 models being generated with only minor steric overlaps. Although adjusting the adjacency constraint minimizes major steric collisions between nucleotides, it does not always prevent the spatial proximity of nucleotides. In addition, most models contain larger than acceptable phosphoester bond lengths and angles because of a specified adjacency constraint. To correct any improper geometries and undesirable steric interactions, each model was energy minimized.

Structural Refinement

Molecular models were energy minimized using the Amber forcefield associated with the Discover module from the Insight II *v*.98.0 modelling suite (Molecular Simulations, Inc.). For each model, steepest descent gradient minimization was performed until the maximal derivative was <100 kcal · $mol^{-1} \cdot Å^{-1}$ followed by conjugate gradient minimization until the maximal derivative reached < 10 kcal · $mol^{-1} \cdot Å^{-1}$. In addition, a "distance-dependent" dielectric value (4*r) that simulates water solvation was added to the electrostatic portion of the forcefield equation to avoid artifacts produced by a "cutoff" value that does not take into account charge–charge interactions after a certain distance. A cutoff distance of 20 Å for each atom was used, after which nonbonded interactions are not included in the potential energy function. Ignoring Coulombic interactions beyond a cutoff distance can produce conformational disparity. The introduction of a distance-dependent dielectric value is based on the assumption that electrostatic forces are effectively "screened" in real systems. This screening effect is approximated by introducing a dielectric term that increases with distance. In this case, the charge–charge interaction term in the potential energy function decreases as the square of the distance between the charges, instead of just the distance as in the standard Coulombic potential equation. MC-SYM scripts and coordinates for the 58 rRNA regions can be obtained at URL: http://www4.ncsu.edu/~plwollen/model.html.

Comparison of Modelled 16S rRNA Elements to A-form Geometry

Helices were compared with ideal A-form RNA geometry by superimposing the phosphates of a given helical model onto those of a generic A-form helix of similar sequence and the rms deviation between phosphates measured. Generic A-form helices did not include bulged nucleotides and internal loops but did span segments that contained these features when bulges or loops were ~3 nucleotides in length as in the case of helix 15. In helices having large internal bulges or loops (~4 nucleotides) such as helix 8, a comparison to A-form geometry was

made for each portion separately. A loose criterion was chosen to evaluate A-form geometry (rms deviation of >2.5 Å). At this value or less, a region has right-handed helical geometry with approximately normal major and minor grooves and no bends in the helical axis. Values for all comparisons are listed in Table 1. In helices where a deviation from A-form geometry was detected (rms deviation >2.5 Å), a second measurement was taken to determine the extent to which bulged nucleotides, internal loops, or sequence composition spanned by the generic A-form helix in the first measurement contributed to deviation from A-form geometry. This was achieved by creating A-form helices corresponding to each base-paired region of a modelled helix and superimposing the phosphates of each A-form helix with those of the modelled helix. As expected, in most cases internal bulges and loops contributed the most to A-form deviation. The causes for helices deviating from A-form geometry are listed in Table 1.

RESULTS

Elements and Regions Containing Unusual Features

Some helices have unusual conformations and arrangements due to their involvement in pseudoknot interactions, the presence of interactions that can be detected by UV-crosslinks within the domain of the structure, or other constraints such as long-range base-pairings to other parts of the 16S rRNA. The considerations that were made for fifteen helices or regions containing several helices (grouped because they contain long-range base-pair interactions) are specifically described here.

Central Pseudoknot (Helices 1/2)

The central pseudoknot is a functionally important, universally conserved structural element joining the three major domains of the small subunit rRNA. Mutational analysis has implicated the central pseudoknot in subunit assembly and function.[40–43] This region consists of a stem-loop structure composed of bases G9–C25/U13–G21 (helix 1) and a terminal loop (bases U14–U20) in which residues U17–A19 are involved in base-pairings (helix 2) to nucleotides A916–U918 from the central domain (Figure 1). The model that was constructed contains all of the base-pairs including the U17–A918 and C18–G917 and Watson-Crick base-pairs and a noncanonical A19–U916 base-pair (determined by N3 reactivity of U916). The arrangement of U916–A918 anticipates the connections that need to be made in the ribosome so that the adjacent regions (helix 27 and 28) are in the correct positions. Further, the arrangement of U14–A19 is consistent with crosslinks made by s^4U at position U14, U17, and U20 in the 30S subunit.[44]

Helix 7a: A base triple interaction in helix 7a was first proposed from comparative sequence studies (Figure 1).[45] Molecular modelling of two potential candidates for a base triple interaction in helix7a were attempted: U121●(C124–G237) and U121●(U125–A236). It was determined using MC-SYM and base-pair selections from the ISOPAIR program that U121 interacts with C124 in the U121●(C124–G237) base triple configuration and that only the U121●(C124–G237) base triple interaction can be optimized successfully.[46] The same isomorphic configuration was modelled for the three most frequent sequence motifs and is consistent with chemical reactivity data.

Helix 8: Helix 8 contains a GNRA tetraloop, a G●G pair (of the two G●G pairs that occur in *E. coli* 16S rRNA) and an A●C pair (of four that occur in *E. coli* 16S rRNA)(Figure 1). Initial models of this helix were found consistent with chemical reactivity that indicated position G177:N1 available for modification (not used in H-bonding). However the initial conformation of the G145●G177 base-pair in the structure disrupted helical stacking in such a way that a different G●G pairing was sought based on phylogenetic data. The G●G base-pair, present in 40.8% of eubacteria species, varies with U●G (21.8%) and C—G (24.1%). (Other pairings occur at significantly lower frequencies and are not considered here.) The G●G H-bonding pattern was selected using the ISOPAIR program comparing G●G, U●G and C—G pairings with at least one pair having three hydrogen-bonds (in an effort to generate Watson-Crick-like pairings). The resulting G●G base-pair had a single H-bond between the G145:N2 hydrogen to the O6 of G177 and the incorporation of this pairing into helix 8 results in a regular helical structure with a wide minor groove. A comparison to A-form geometry was taken for both helix 8a and 8b (0.85 and 1.31 Å rms deviation) although the entire helix was modelled as a unit (Table 1).

Helix 12: The GAGA tetraloop and U●G closing base-pair in helix 12 occur in more than 98% of the bacterial sequences.[37,47] This GAGA tetraloop differs from the NMR-derived structure of the GNRA tetraloop[31] by having the first and third residues in the tetraloop reactive to chemical probing (G297 WC side and G299 N7 position). Thus, in this case, the NMR-derived GNRA tetraloop could not be used as a reference for the construction. In addition, this tetraloop is the only GNRA tetraloop in 16S closed by a U●G base-pair. This U●G pair allows only a small number of possible base-pairings between G297 and A300 (residue 1 and 4 of the tetraloop) that agreed with chemical probing data. We decided to model helix 12 with A298-A300 (tetraloop residues 2–4) stacking on one another as in the NMR structure, but with residue G297 in a position to form a hydrogen bond with G299 (4.35 Å distance between G297:O2′ and G299:N7). This was accomplished by specifying the base-pairing between G297 and A300. In the refined version of this structure, G297 is not stacked on U296.

A second feature of helix 12 involves nucleotides G305–C307, which comprise an internal bulge (Figure 1). All three nucleotides are sensitive to modification reagents with C307 being the most reactive.[28] All structures for helix 12 produced using MC-SYM have nucleotide C307 turned out of the helical stack creating a wider and more accessible major groove compared with that of an A-form helix. No additional constraints during model construction other than a specified stacking arrangement was necessary to position these three nucleotides in such a way as to be accessible by solvent. Construction of this helix using MC-SYM beginning with the terminal loop automatically places nucleotides G305-C307 away from the helix center.

Helix 17a: Comparative sequence analysis was performed as previously described[46] to determine nucleotide positions in *E. coli* 16S rRNA that displayed patterns of covariation consistent with the presence of a base-triple. The analysis of likely base-triple candidates in 16S rRNA was performed using the latest release of the Ribosomal Database Project (RDP).[26] Approximately 6,200 prokaryotic 16S rRNA sequences—in-

Table 1. Data used in modeling each helical region of 16S rRNA. Information listed in this table served as constraints and was incorporated by the MC-SYM program reducing the number of possible solutions

Helix	Nucleotide	Data incorporated[a]	Root-Mean-Square Deviation from A-form geometry (Ä)	Notes	Reference
1	9-25/13-21		0.79		
2	17-918/19-916	●U916	3.28	Odd A19-U916 base-pairing causes deviation from A-form geometry	28
3	27-556/37-547		2.09		
4	39-397/46-405		2.19		
5	47-361/58-354	C54 × A353 UV -crosslink	6.87	Deviation from A-form geometry caused by several internal bulges	30
6a	113-115/312-314		0.22		
6	61-106	G82-C87 NMR CUUG tetraloop including closing base-pair	7.43	Deviation from A-form geometry caused by large number of internal bulges	32
7a	122-132/230-239	U121●(U124-A237) base-triple	2.78		
7b	136-142/221-227		0.61		
8a	144-148/174-178	ISOPAIR data for G145●G177 base-pair	0.85		
8b	152-169	G159●A162 NMR GAAA tetraloop	1.31		31
9	184-193	C186-G191 NMR GCAA tetraloop including closing base-pair	1.39		31
10	198-219	C207-G212 NMR UUCG tetraloop including closing base-pair	3.63	Deviation from A-form geometry caused by internal bulge	39
11a	240-245/283-286		3.56	Deviation from A-form geometry caused by internal bulge	
11b	247-277		3.59	Deviation from A-form geometry caused by internal bulge	
12	289-311	● G297, ● A298, ●G299, ●G299 (N7)	6.39	Deviation from A-form geometry caused by internal bulge	28,29
13	316-337	●G318, ●G319	5.46	Base-pairings involving 318, 319 cause deviation from A-form geometry	28,29
14	339-350	C342-U343 and C345-G347 NMR UUCG tetraloop including closing base-pair	1.48		39
15	367-393	G377-C386 NMR GCAA tetraloop including three base-pairs	7.94	Internal bulge and loop cause deviation from A-form geometry	31
16a	406-409/433-436		0.47		
16b	416-427	C419-G424 NMR UUCG tetraloop including closing base-pair	1.76		39
17a	438-446/488-496	G497●(C440-G494) base-triple, A441 × G494 UV-crosslink	0.56	Compared only nucleotides in helical region to A-form geometry (442-446/488-492)	27,30

(*Continued*)

Table 1. (*Continued*)

Helix	Nucleotide	Data incorporated[a]	Root-Mean-Square Deviation from A-form geometry (Å)	Notes	Reference
17b	455-477		1.79		
18	500-504/541-545	●●G527, coaxially stacked helices (G500-C504/G541-C545 and C511-G515/C536-G540), long-range base-pairs (G505-C507/G524-C526)	1.91	Both G500-C504/G541-C545 and C511-G515/C536-G540 helices compared to A-form geometry as a unit	29,50,52
	511-515/536-540				
	505-507/524-526		0.85		
	521-522/527-528		2.80	Deviation from A-form geometry for the 521-522/527-528 base-pairs caused by reactive G527	
19	567-569/881-883		0.68		
20	577-586/755-764	●●U762, C582 × G760 UV-crosslink, ISOPAIR data for C582-C760 base-pair	6.42	Deviation from A-form geometry caused by internal loop, internal bulge and a C582-C760 base-pair	27,29,30
21a/b	580-606/632-651	NMR structure of S8 binding site	8.44	Deviation from A-form geometry caused by two base-triples, a bulged loop and the sequence 588-593/646-651	34
21c	611-629		1.69		31,65,80
22a	655-663/741-751		1.88		
22b	666-672/734-740		0.67		
22c	722-733	G725-A729 and G731-C732 NMR GAAA tetraloop including two closing base-pairs	4.00	Deviation from A-form geometry caused by a bulged nucleotide and a G722-G733 base-pairing	31
23a	673-684/706-717		5.55	Deviation from A-form geometry caused by an internal loop	
23b	688-699		0.82		
24a	769-774/803-810		4.89	Deviation from A-form geometry caused by an internal loop	
24b	783-799		0.44		
25	821-827/874-879		0.95		
26a	861-862/867-868		1.39		
	570-571/865-866		0.83		
26	829-857	● G833 (N7), ● G836 (N7), ● G838 (N7), ● G847 (N7), ● G849 (N7)	4.05	Deviation from A-form geometry caused by large number of G·U base-pairings	28,29

(*Continued*)

Table 1. (*Continued*)

Helix	Nucleotide	Data incorporated[a]	Root-Mean-Square Deviation from A-form geometry (Å)	Notes	Reference
27	885-912	C897-G902 NMR GCAA tetraloop including closing base-pair, A889-A892●A907-A909 loop E motif from 5S rRNA crystal structure	5.61	The model for helix 27 represents the "ram" or "912-885" conformation. Deviation from A-form geometry caused by loop E motif and internal bulge	31,65,80
28	921-936/1379-1396	●●●●G926, A1394●(C924-G1392) base-triple	6.59	Deviation from A-form geometry caused by base-triple, internal loop and bulged nucleotides	28,64
29	939-943/1340-1344		1.62		
30	944-955/1225-1237	●U952, ●U955	7.42	Base-pairings involving 952,955 cause deviation from A-form geometry along with the sequence 1230-1237/944-951	29
31	960-976	●U960, ●U961, ●C972	2.28	Base-pairings involving 960, 961 cause deviation from A-form geometry	29
32	984-990/1215-1221		0.25		
33a	996-1003/1037-1045		5.19	Deviation from A-form geometry caused by bulged nucleotide and by overall sequence	
33b	1006-1023	G1013-A1016 NMR GAGA tetraloop	1.33		31
33c	1025-1036	C1028-G1033 NMR UUCG tetraloop including closing base-pair	1.05		39
34	1047-1064/1192-1210	●C1192, U1052 × C1200 UV-crosslink, G1064 × G1190 UV-laser crosslink	9.07	Deviation from A-form geometry caused by two UV-crosslinks, a base-pair involving 1192, and internal bulges	29,30,85–87
35	1068-1073/1102-1107		0.22		
36	1074-1083	● A1082, G1077-A1080 base-pair from NMR GAAA tetraloop	2.20		29,31
37	1086-1099		0.56		
38	1113-1116/1184-1187	●● U1116	2.28		29
39a	1118-1124/1149-1155		0.77		
39b	1128-1144		2.09		
40	1158-1178		1.33		
41a	1241-1247/1290-1296		1.10		
41bcd	1253-1284	C1265-C1267 and A1269-G1270 NMR GCAA tetraloop including closing base-pair	0.61(41b)		

(*Continued*)

Table 1. (*Continued*)

Helix	Nucleotide	Data incorporated[a]	Root-Mean-Square Deviation from A-form geometry (Å)	Notes	Reference
			0.46 (41c) 0.66 (41d)		31
42	1303-1334		2.04		
43	1350-1372		0.85		
44a	1399-1410/1490-1504	● C1501, C1402 × C1501 UV-crosslink	6.19	Four internal loops cause deviation from A-form geometry. Only nucleotides 1404-1410/1490-1497 were measured because of a bend occurring at 1404-1497 base-pair. The sequence 1399-1403/1498-1504 differs significantly from A-form geometry.	28–30
44b	1411-1430/1470-1489	●● A1413, ● U1414, ●● U1420	6.91	Deviation from A-form geometry caused by large number of G●U pairings, an internal loop, and base-pairs involving 1413, 1414 and 1420	28,29
44c	1435-1466	C1449-G1454 UUCG tetraloop NMR structure including closing base-pair	4.87	Deviation from A-form geometry caused by internal loop, internal loop, internal bulge and large number of G●U pairings	93
45	1506-1529	G1516-A1519 from methylated NMR GGAA tetraloop	0.81		33

[a] Results of chemical probing experiments are indicated as ●●●● = hyper, ●●● = strong, ●● = moderate, ● = slight. Chemical reactivities are included for nucleotides occurring in base-paired regions and only when it directly affects the resulting model of an element.

cluding complete and partial sequences—were used in the alignment. One potential base-triple in particular was identified at positions G494●(C440–G497) in helix 17a (Table 2). The suggested presence of this base-triple is supported by experimental data including a UV-crosslink between nucleotides A441 and G494[30] and an unusual organization of base-pairs, U438●A496[48] and C440●G497[22] that indicates a brief parallel arrangement of the backbones.

Because it could not be resolved statistically whether nucleotide G494 was hydrogen bonded to either of the base-paired nucleotides C440 or G497, the ISOPAIR program was used to predict the base-triple hydrogen bonding pattern that was isomorphic across the major covariation patterns (Table 2). The second most frequent base-pairing variant at positions 440–497 is a U●U pair. A lengthy effort was made to find a structure isogeometric with the G●(U●U) triple, but no satisfactory structures could be generated. One explanation is that a base-triple is present in these organisms at a different position to maintain the overall tertiary arrangement in this region. This is shown to occur with base-triples in different tRNAs.[49] Molecular modelling using the MC-SYM program was employed to test whether the proposed hydrogen bonding pattern predicted by ISOPAIR was consistent with experimental data. Three sets of structures were modelled to determine which was consistent with the A441 × G494 crosslink: (i) helix 17a without a base-triple, (ii) helix 17a containing a hydrogen bond between C440 and G497, and (iii) helix 17a containing a hydrogen bond between G494 and G497. Only the conformation that included the base-triple with a hydrogen bond between G494 and G497 resulted in a suitable arrangement for nucleosides A441 and G494 in which they were stacked, the arrangement expected for UV crosslink formation. The resulting refined structure in shown in Figure 3.

Helix 18: A pseudoknot structure in helix 18 involving base-pairing between residues G505–C507 and G524–C526 (Figure 1) was first proposed by covariance analysis[50] and subsequently verified with site-directed mutagenesis that showed that the pseudoknot interaction is essential for the ribosome to function.[51] Phylogenetic data also indicate that the two base-paired helices (G500–C504/G541–C545 and C511–G515/C536–G540) are coaxially stacked.[50,52] We have produced models that contain the pseudoknot structure with coaxially stacked helices. The arrangement results in the terminal loop (residues G524–C526) sharply bending toward the middle of the stacked helices to base-pair with residues G505–C507 (Figure 4). Although the unpaired nucleotide A510 is not sensitive to chemical modification,[28] MC-SYM predicts an arrangement for this nucleotide that is turned out of the helix. One explanation for this inconsistency is that A510 is involved in a tertiary hydrogen-bonding arrangement or other configuration that prevents modification of this nucleotide within the 30S subunit. Residues G530–A532 are positioned away from the coaxially stacked helices in such a manner that when incorporated into a global model, they are easily accessible to mRNA[53–56] and the A- and P-site. The rms deviation from A-form RNA for the coaxially stacked helices G500–C504/G541–C545 and C511–G515/C536–G40 is 1.91 Å (Table 1). A large number of models were produced for this helix due to the lack of constraint data for the sizable internal loop near the terminus of the helix (Ψ516–A520 and G529–A535). All models produced were nearly identical in the base-paired regions (<0.5 Å rms deviation) with differences in the single-stranded regions. Nucleotides contained within the unpaired terminal loop are predicted by MC-SYM to be splayed out of the structure, agreeing with modification information for these residues and consistent with the idea that they interact with mRNA[55,56] and tRNA.[57]

There has been long-standing debate as to the location and function of helix 18 within the 30S subunit. This helix is involved in a number of translational processes indicated by footprints of A- and P- site tRNA[57] on helix 18 and crosslinks from mRNA substituted with photoreactive compounds to the terminal loop of this helix.[55,56] This information along with crosslinking data from arylazide-containing antisense oligonucleotides,[58] crosslinking data from psoralen placed at the 530 region,[59] and ribosomal protein footprinting data[60] suggest a position of helix 18 at or near the decoding site. The molecular model for helix 18 indicates that it could undergo a conformational change especially with regard to the arrangement of the terminal loop, which would allow it to reach closer toward the decoding region. This unfolded conformation for helix 18 may function to assist in the movement of tRNA from the A-site into the P-site. Support for such a conformational change is shown by the accessibility of the 530 loop in 30S subunits to complementary oligonucleotides, suggesting that this region is at least for a time not in a pseudoknotted arrangement having the G524–C526/G505–C507 base-pairs absent.[61] Models were also produced for this arrangement and result in a more elongated and A-form helix relative to models where these base-pairs are present (results not shown). At the same time, the structure of the two helical regions would remain relatively unchanged throughout translation as indicated from phylogenetic studies supporting a coaxially stacked arrangement for these helices.[50,52]

Helix 19/25/26a junction: The helix 19, 25, and 26a junction region was modelled as a unit to include the pseudoknot base-pairings between the hairpin residues A865–C866 in helix 26a and U571–G570 at the 3′ end of helix 19[62,63] and the lone base-pair between G575–C880 (Figure 1). The resulting structure (Figure 5) is accommodated well into a global model and agrees with hydroxyl radical probing data from the ribosomal protein S8[60] (the binding site for S8 is in helix 21). In addition, psoralen crosslinks from position U788/U799 to C866 and A560 are consistent with the arrangement shown here.[64] The placement of these elements relative to one another agrees very well with a recently proposed model for this region based on

Table 2. Nucleotide identities at positions 440, 494 and 497 of 6205 prokaryotes

	A	C	G	U	494
CG	117		2079	177	
UU			1285		
UA		423	199		
AU			144		
UG			125		
440:497					

Note: Combinations which account for less than 1.5% of all covariations are omitted.

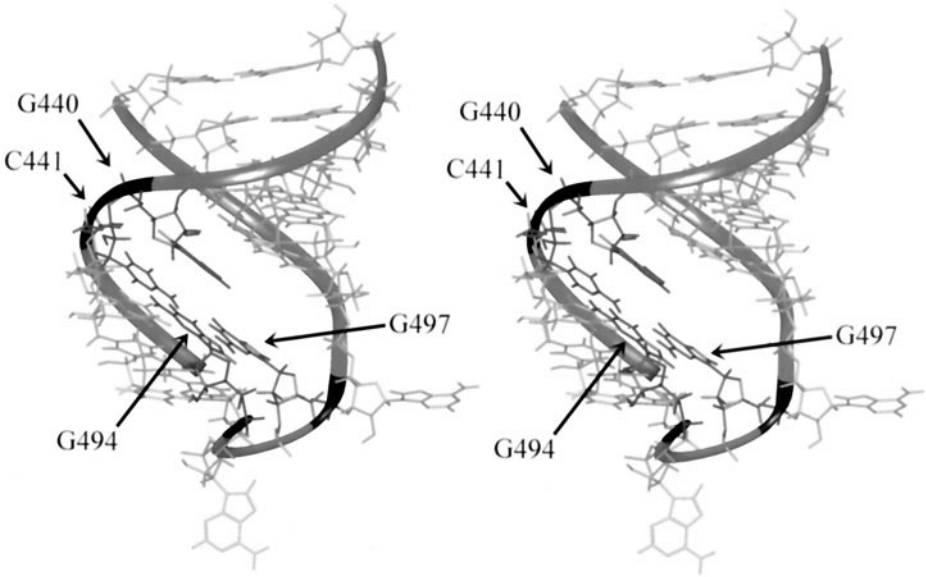

Figure 3. Stereo image of helix 17a showing the G494●(C440–G497) base triple interaction. The location of nucleotide A441 involved in a crosslink with G494 is also shown. Nucleotides involved in the base-triple and the crosslink are colored in black. Helix 17a differs little from A-form geometry when comparing residues G442–G446/C488–C492 but deviates substantially when comparing the region in and around the nucleotides involved in forming the base-triple.

the ribosomal protein map derived from neutron scattering experiments, and phylogenetic, biochemical, and crosslinking data.[65] This region is also included in the partial RNA model derived from X-ray analysis[18] and the computer-derived model has the same relative arrangement for each of the elements that comprise this region.

Helix 20: The constraints for helix 20 include: a UV-induced crosslink between residues C582 and G760[30] and a proposed C●C base-pairing between C582 and C758. Comparative sequence analysis reveals that a C●C pair is found in ~11% of sequences, a C—G pair in ~8% of the sequences and the remainder (>80%) having a U●G pair at these positions.[66] This covariance information was used as input into the ISOPAIR program and the most probable base-pairing arrangement for this base-pair is highly asymmetric.

It was proposed that a bacterial loop E submotif occurs between residues 581–583 and 758–760.[67] Incorporation of the most probable C582 and C758 base-pairing and a UV-induced crosslink at C582 × G760[30] results in a structure for helix 20 that is not consistent with the presence of the loop E submotif. In addition, experiments done in the organisms *Bacillus subtilis* and *Thermus aquaticus* indicate UV-crosslinks at 586 × 756 and 581 × 759 (*E. coli* numbering) respectively.[68] The presence of these crosslinks indicate the possibility of alternate structures within helix 20 in *E. coli* that differ from the proposed loop E submotif and also some differences in the structure of this element in different organisms.

Helix 21a/b: Helix 21a/b contains the binding site for ribosomal protein S8.[69] We used the S8 binding site structure as determined by NMR spectroscopy as a template on which the remaining nucleotides of this helix were added.[34] The NMR structure contains the base-triple A595●(A596–U644) and a metal binding motif. This base-triple was also identified by covariation analysis and the primary combinations of nucleotides for this triple in *Bacteria* and *Archaea* are A●(A-U).[66] Another base-triple nearby at positions U641●(A597–U643) was suggested by Moine et al.[70] from a Selex experiment. Although this base-triple does not covary, the primary comparative compositions are A●(A–U) and G●(C–G). This triple was not ruled out in the NMR determination and we therefore incorporated it into the helix 21a/b model along with the triple suggested by NMR. In fact, the model containing the

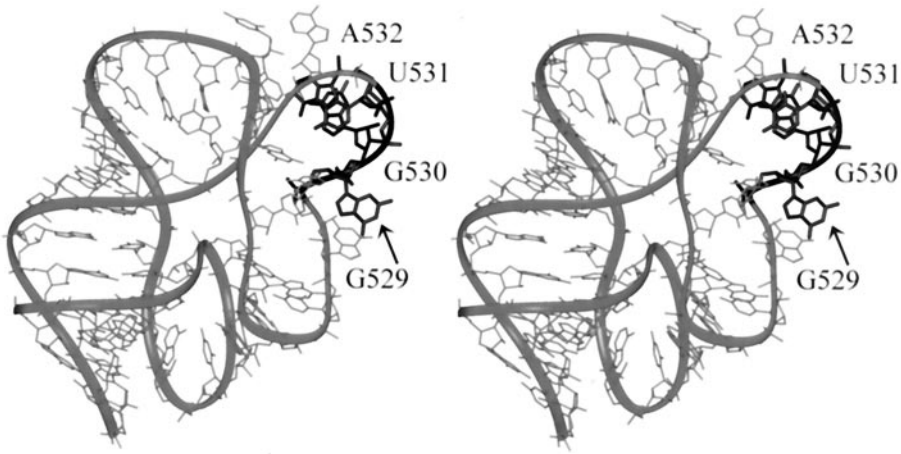

Figure 4. Stereo diagram of helix 18 with residues footprinted by A- and P-site tRNA (G529–A532) represented by thick lines. Helices G500–C504/G541–C545 and C511–G515/C536–G540 are in a coaxial stacking arrangement. The helix bends near the top of the C511–G515/C536–G540 helix to accommodate the G505–C507/G524–C526 base-pairings moving G529–A532 into an arrangement consistent with these nucleotides being at or near the decoding region.

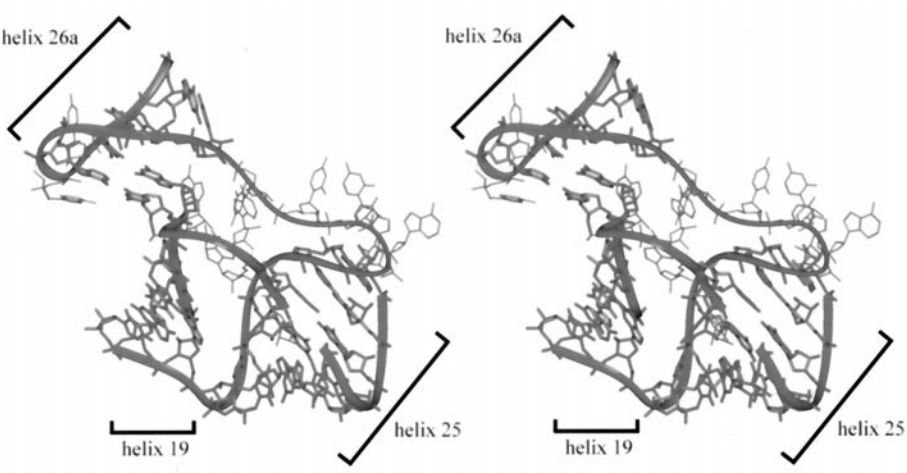

Figure 5. Stereo image of helices 19/25/26a involved in forming the pseudoknot at the juncture between the three major domains. This model includes all residues for this region shown boxed in Figure 1. Base-paired residues within each helix are represented as thick lines. The base-pairings between G570–U571/A865–C866 in helix 26a cause a turn in the structure involving helix 19, 25, and 26a. This results in the terminus of helix 26a being proximal to the 5′ end of helix 19 consistent with a psoralen crosslink from U788/U789 to C866[65] and with the positioning of helices 20 and 26 in the context of a global arrangement (Figure 1).

U641●(A597–U643) triple differs only slightly in backbone conformation and general placement of the bases involved in the triple relative to the structure determined by NMR.

The model for helix 21a/b contains two tandem G●U wobbles (G645●U594 and G646●U593) which along with the base-triple A595●(A596–U644) cause a sharp bend in the helical axis near these residues. A second sharp turn occurs near the G606●U632 terminus caused by an internal bulge several nucleotides away (U641–A642) and propagates along the helix terminating at two G●U wobbles (U605●G633 and G606●U632).

Helix 23b: Helix 23b terminates in a loop that contains several highly conserved residues including G693. G693 must be very accessible as it has been crosslinked by P- and E-site bound tRNA,[71] crosslinked to s^4U-substituted mRNA,[55] is footprinted by tRNA bound to the P-site[61,72] and is footprinted by a number of antibiotics.[73,74] In addition, G693 has been crosslinked by psoralen tethered to the apical loop of helix 24.[64]

Models were generated for this structure that had G693 protruding from the top of the helix agreeing with this residue being modified at its Watson-Crick face and putting it in position to form a mRNA crosslink in the 30S subunit. Differences between the 18 models generated by MC-SYM occur at or near residue G693 suggesting some conformational flexibility in the terminal loop of helix 23b. The dynamic nature of G693 and nucleotides immediately adjacent to this residue may aid in the recognition and binding of mRNA by the 30S subunit.

Helix 24a/b: The 790 loop contains a highly conserved motif present in all small subunit rRNAs.[22,66] This loop is thought to be involved in many important translational processes including tRNA binding,[57,72] subunit association,[75–77] and IF3 binding.[78,79] The terminal loop of this helix has hydrogen bonding and stacking within the loop, agreeing with observed nucleotide covariations.[80] The models for helix 24a contain a sharp bend caused by the internal bulge (residues G775–A777) near the terminus of the seven base-pair helix. This configuration for helix 24a results in an overall bend in helix 24 consistent with the proposal that in the context of a global model, the location of the base of the helix is near the cytoplasmic side of the platform and the 790 loop is placed near the decoding region, specifically at or near the P-site.[71,72] This arrangement was seen in a recent X-ray crystal structure of the 30S subunit with both the 690 and 790 loops protruding into the intersubunit interface.[18] The model for helix 24b shows nucleotides comprising the terminal loop stacked because of a lack of experimental constraints for this region. However, it is likely that the 790 loop is a dynamic structure that undergoes conformational changes during translation and its overall shape and location within the subunit may reflect its function in creating part of the P-site and in translocating deacylated-tRNA from the P-site into the E-site. Crosslinks from psoralen placed at U788 to four other helices[64] suggest either changes in the relative position of helix 24 to meet all the crosslinking distances or conformational changes in the terminal loop.

Helix 26: Helix 26 contains five G●U wobble base-pairs with four of these occurring consecutively. The presence of tandem G●U wobble base-pairs has been shown to create deviations from A-form geometry in RNA.[81] Additionally, there are five N7 reactive positions in helix 26[29] (Table1). It is not possible to find models with MC-SYM for regions containing N7 reactivities if residues are restricted to A-form conformations. Therefore, the construction of a model for helix 26 that could be refined required additional conformational freedom in the ribose and phosphate backbone torsion angles for those nucleotides with N7 reactivities. The final energy optimized model for helix 26 deviated from A-form geometry by 4 Å (Table 1).

Helix 27: Helix 27 is involved in maintaining translational accuracy and has alternative conformations that affect the tRNA ribosome interaction.[82] Mutations favoring an alternative "912–885" or *ram* base-pairing motif had higher rates of reading through an in-frame stop codon while mutations favoring a second "912–888" or "restrictive" motif had lower read-through rates, but an elevated rate of frameshifting. Both the *ram* (912–885) and restrictive (912–888) conformations were modelled. For the *ram* conformation, the crystal structure of the

5S rRNA loop E motif[67] was incorporated (agreeing with the base-pairing scheme in Figure 1) to produce an "S-turn" structure in the internal loop.[83] This results in a straight helical axis overall compared with the restrictive conformation, which produced models with helices of either bent of straight character. Thus the MC-SYM algorithm is able to accurately predict the *ram* conformation containing a loop E motif as it agrees with chemical modification data and the structure observed by X-ray crystallography.[20] This helix differs from A-form geometry, having an rms deviation of 5.61 Å caused in large part by the loop E motif (Table 1).

Recently, it has been proposed that helix 27 modulates the conformation of the decoding region by direct interaction with helices 24 and 44.[18] The two different configurations for helix 27 resulting from the modelling of the *ram* and restrictive conformations may represent two functionally important states for this helix as it interacts with the decoding region. One conformation may alter the structure of a portion of helix 24 or helix 44, creating a less stable binding site for tRNA in the P-site and changing the way the anticodon interacts with mRNA. Both helix 24 and helix 44 are footprinted by tRNA in the P-site.[57] The two models for helix 27 will be of use in determining the mechanism by which helix 27 maintains translational accuracy and the dynamics of the decoding region during translation.

Helix 28: Helix 28 is integral in forming the tRNA P-site. It is footprinted at residue G926 by P-site bound tRNA,[57] indicating the proximity of tRNA to G926. The model for this helix contains a A1394●(C924–G1392) base triple supported by mutational data.[8] The incorporation of this base-triple into the model causes a significant bend of the helical axis and the orientation of A1394 requires an unstacked and accessible arrangement for G926. At the 3′ end of helix 28, C934, which is present as a single nucleotide bulge in the current secondary structure model, is arranged near residue U1345 meeting a UV-crosslink constraint.[30] This is consistent with the tentative base-pair C934●A938 identified by comparative sequence analysis.[66] A second base-triple interaction between C1109●(G933–C1384) is proposed in the newest comparative secondary structure,[66] but not included in the present modelling. However, the G933–C1384 base-pair is accessible due to the four unpaired nucleotides flanking the 3′ end of this base-pair.

Helix 30: Helix 30 contains two A–U base-pairs at positions U952–A1229 and U955–A1225 that have reactive uridines and unreactive adenines.[29] Since the N3 position is the only hydrogen donor in uracil, the unreactive N1 atom of the adenine must be protected by another means such as a hydrogen bond to a ribose or another nucleotide. Hydrogen bonding between the N1 of adenine and the O2′ from a neighboring base has recently been shown in the minor groove RNA triple helix within the catalytic core of the *Tetrahymena* group I intron using nucleotide analog interference mapping (NAIM)[84] and in the crystal structure of the hammerhead ribozyme.[85] We therefore added a distance constraint corresponding to a hydrogen bond length between the N1 of the adenine and the O2′ of the uridine residue. The resulting model for helix 30 is an elongated structure, deviating from an A-form geometry by >7 Å. This is caused in part by the odd A●U base-pairings and in part by the particular sequence of base-pairings found in helix 30.

Helix 34: The helix 34 sequence is highly conserved among small ribosomal subunit rRNAs.[86] Chemical probing studies suggest this helix is a dynamic structure changing configuration between two functional states during the translation cycle resulting in efficient UGA-dependent termination.[87] Alternative conformations for this region have been proposed based on mutational data that suggest two different base-pairings between either one of two UCA motifs (1199–1201 or "upper" conformer and the 1202–1204 or "lower" conformer) and the complementary UGA stop codon of mRNA.[88] The model for helix 34 was constructed according to the proposed secondary structure (Figure 1) which has a bulged loop at positions C1200–U1202 and represents the proposed 1199–1201 conformer. A G1064 × G1190 UV-laser crosslink[89] beyond the terminal base-pair is also included (Figure 1). Although both the 1199–1201 and 1202–1204 conformers can be modelled, the model of helix 34 containing the 1199–1201 conformer is better able to accommodate the U1052 × C1200 UV-induced crosslink,[29] the UV-laser crosslink (G1064 × G1190), and chemical reactivity data for this region. Both conformers for helix 34 are compatible with a tentative base-triple interaction between G1053●(G1057–C1203) identified with comparative sequence analysis,[66] although it was not included as a constraint during modelling. Incorporation of the crosslinks, especially the U1052 × C1200 crosslink, contributes to a gradual, but significant bend of the helical axis (Figure 6). This positions U1052 and A1196 in proximity to the proposed mRNA tract through the subunit and is consistent with crosslinks from s^4U-containing mRNA analogues to these residues.[55,56]

Helix 44abc: The 3′ minor domain of 16S rRNA includes residues C1399–A1542, encompassing some highly conserved regions within the penultimate helix and the 3′ terminal helix. The top portion of the penultimate helix comprises part of the P-site as the proximity of tRNA to this portion of the helix is indirectly supported by the protection of bases C1399, C1400, and G1401 by tRNA bound to the P-site.[57,72] This is strengthened by the description of a short-range UV-crosslink from C1400 to the 5′anticodon of $tRNA_1^{Val}$.[90] Footprinting of A1492 and A1493 occurs for tRNA in the A-site; however the arrangement of A-site tRNA in the 70S ribosome now shows this to be an indirect effect.[20]

We have modelled the penultimate helix as three separate elements: the first encompasses (C1399–A1410)●(U1490–G1504)(helix 44a), the second includes (C1411–A1430)●(U1470–G1489) (helix 44b), and the third comprises G1435–C1466 making up the terminal portion of the penultimate helix (helix 44c). The model of helix 44a includes a C1402 × C1501 UV-crosslink,[28] which allows for the placement of C1400 at the P-site and for nucleotide A1492 to be in such a position as to accommodate a crosslink from residue U14 in the central pseudoknot when incorporated into the global model.[44,91] The axes of helix 44b and 44c remain linear overall agreeing with models proposed from recent crystal structures of the 70S ribosome.[18,20] A description of the details of the helix1/2, helices 44a, 44b, and 44c and helix 28 arrangement will be given elsewhere.[92]

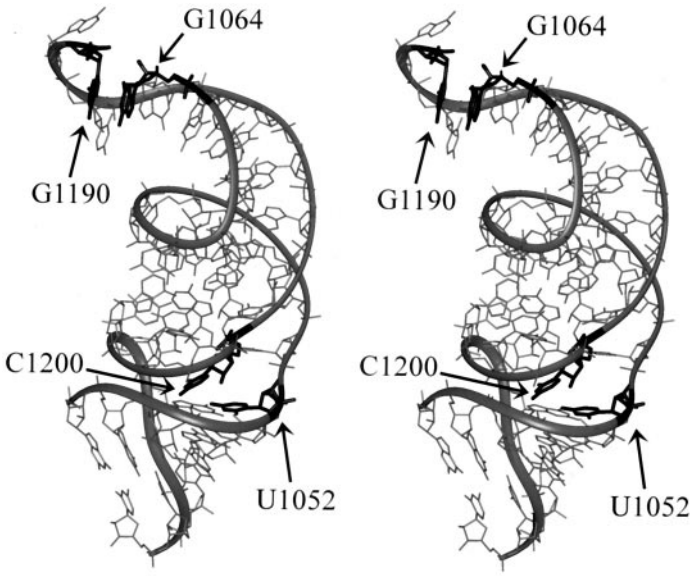

Figure 6. Stereo view of helix 34 containing the proposed 1199–1201 conformer.[86] The U1052 × C1200 UV-crosslink and the G1190 × G1064 UV-laser crosslink are represented as thick lines. The crosslinked residues in each pair are ~4 Å apart. Helix 34 has a relatively large deviation from A-form geometry (>9 Å) due to the two crosslinks and three internal bulges. The 1199–1201 conformer for helix 34 shown here is consistent with U1199–A1201 being available for interaction with the UGA stop codon of mRNA resulting in translation termination.

DISCUSSION

Usefulness of Computer-Based Models

Computer-modelled or solved rRNA structures can aid in the process of determining a relatively large structure such as the ribosome. Several attempts to fit solved structures or computer-based models of RNA to lower resolution structures of the 30S ribosomal subunit have been made. These include the GNRA tetraloop structure determined by NMR fitted to several regions of the electron-density map of the 50S ribosomal subunit crystal structure,[93] and the entire 16S, 23S, and 5S rRNAs into three-dimensional cryo-EM reconstructions.[17] Given the pending high resolution crystal structures, it will be important to compare the structure models predicted here with experimentally derived structures. It will be possible to determine how well MC-SYM, comparative sequence analysis, crosslinking data, and other information support static crystal structures or the dynamic movement of 16S rRNA.

One feature in any constraint satisfaction process as opposed to a motif identification process or threading process is that in absence of sufficient information, sections of models will be arbitrary. In the present case, any nucleotides that are not base-paired are arranged only in a general way to be consistent with chemical reactivity positions or crosslinking arrangements if such information exists. Thus the present models can only approximate biologically relevant structures. However, we already know that the structures presented here can be readily incorporated into a folded structure for the whole 16S rRNA that has a compact shape and incorporates many additional distance constraints that occur in the 30S subunit between these individual elements.[94] Moreover, the prediction process brings forth some surprising conclusions about the connection between sequence, secondary structure, and three-dimensional structure. There are many irregularities (interior loops and bulges) within the base-paired elements of 16S rRNA. The modelling presented here demonstrates that sometimes these necessitate non-A-form helices and sometimes they can be accommodated into A-form helices. On the other hand, in other instances the nucleotide sequence itself requires non-A-form geometry. These conclusions are not only for *E. coli* since the modelling has extensively used the Ribosomal Database Project (RDP)[26] for information.

Comparison can be made to the recently published structure of the *T. thermophilus* 30S ribosome central domain.[95,96] We found that our modelled structures for helixes 22a, 22b, and 20 differed little from the crystal structure having rms deviations of 1.2, 0.67, 1.8 Å, respectively between phosphate backbones (Table 3). It is difficult to compare individual base-pairings of modelled structures to those derived from crystallography experiments because of sequence differences between *E. coli* and *T. thermophilus.* However, the model for helix 22c agreed with the overall nucleotide placement in the crystal structure, including nucleotide 723, which is turned out of the helix. The deviation between the model and the solved structure for this helix was 2.8 Å due to the difference in stacking arrangement of the 722–733 base-pair. In *E. coli*, this base-pair is G●G and in *T. thermophilus* it is an A●A. Although, MC-SYM did not stack the 722–733 base-pair as seen in the crystal structure, the conformation for this base-pair was isogeometric with that seen in *T. thermophilus.* This arrangement was the only one that allowed MC-SYM to produce models. The similarity between

Table 3. Comparison of modeled structures to structures determined by X-ray crystallography.[93,94] All rms deviation measurements were made by superimposition of phosphates

Modeled helix	rms deviation from X-ray crystal structure (Å)	Notes
20	1.8	residues 583-586/755-758, *E. coli* numbering
22a	1.2	
22b	0.67	
22c	2.8	

the modelled structures and those solved by X-ray crystallography support the ability of MC-SYM to accurately predict three-dimensional structures based on experimental data.

Deviation from A-form Geometry

Almost 2/3 of our modelled helices were A-form or near A-form in geometry (<3 Å rms deviation), while 1/3 of the helices were less A-form in structure (>3 Å rms deviation). In general, elements that contain a lower number of internal loops or bulges (or none at all) and no unusual base-pairs or wobble base-pairs have lower rms deviations from A-form geometry. Data used in the modelling of each RNA region, comparison to A-form geometry, and the causes of A-form deviation for certain helices are noted in Table 1. Helices with the greatest deviation from A-form geometry are: (i) helix 5 containing a single nucleotide bulge and a UV-crosslink, (ii) helix 6 having several internal bulges, (iii) helix 15 containing tandem U•G base-pairs, an internal loop, and a bulged nucleotide, (iv) helix 21a/b containing two sets of tandem U•G base-pairs, two bulged loops, and the NMR-determined structure of the S8 binding site,[34] (v) helix 28 containing a base triple, several bulged nucleotides, and an internal loop, (vi) helix 30 having two unusual A-U base-pairings determined by chemical modification data, and (vii) helix 34 containing four nucleotide bulges, two UV-induced rRNA-rRNA crosslinks, and a U•U base-pair (Table 1).

ACKNOWLEDGEMENTS

This work was supported by National Institutes of Health Grant GM43237 to P.W. and by a computing grant from the North Carolina Supercomputing Center to M.A.D. We are grateful to Dr. Robin Gutell for comments on this manuscript and for providing information from comparative sequence studies before publication. We also thank Dr. Francois Major for the use and guidance of the MC-SYM program and Dr. Daniel Gautheret for kindly providing the ISOPAIR program.

REFERENCES

1 Major, F., Gautheret, D., and Cedergren, R. Reproducing the three-dimensional structure of a tRNA molecule from structural constraints. *Proc. Natl. Acad. Sci. U.S.A.* 1993, **90**, 9408–9412
2 Michel, F., and Westhof, E. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* 1990, **216**, 585–610
3 Tanner, N.K., Schaff, S., Thill, G., Petit-Koskas, E., Crain-Denoyelle, A.M., and Westhof, E. A three-dimensional model of hepatitis delta virus ribozyme based on biochemical and mutational analyses. *Curr. Biol.* 1994, **4**, 488–498
4 Massire, C., Jaeger, L., and Westhof, E. Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. *J. Mol. Biol.* 1998, **279**, 773–793
5 Harris, M.E., Kazantsev, A.V., Chen, J.-L., and Pace, N. Analysis of the tertiary structure of the ribonuclease P ribozyme–substrate complex by site-specific photoaffinity crosslinking. *RNA* 1997, **3**, 561–576
6 Westhof, E., and Altman, S. Three dimensional working model of M1 RNA, the catalytic RNA subunit of ribonuclease P from *Escherichia coli. Proc. Nat. Acad. Sci.* 1994, **91**, 5133–5137
7 Harris, M.E., Nolan, J.M., Malhotra, A., Brown, J.W., Harvey, S.C., and Pace, N.R. Use of photoaffinity crosslinking and molecular modeling to analyze the global structure of ribonuclease P RNA. *EMBO J.* 1994, **13**, 3953–3963
8 Ericson, G., Minchew, P., and Wollenzien, P. Structural changes in base-paired region 28 in 16S rRNA close to the decoding region of the 30S ribosomal subunit are correlated to changes in tRNA binding. *J. Mol. Biol.* 1995, **250**, 407–419
9 Philippe, C., Benard, L., Portier, C., Westhof, E., Ehresmann, B., and Ehresmann, C. Molecular dissection of the pseudoknot governing the translational regulation of *Escherichia coli* ribosomal protein S15. *Nucl. Acids Res.* 1995, **23**, 18–28
10 Brunel, C., Romby, P., Westhof, E., Ehresmann, C., and Ehresmann, B. Three-dimensional model of *Escherichia coli* ribosomal 5S RNA as deduced by structure probing in solution and computer modeling. *J. Mol. Biol.* 1991, **221**, 293–308
11 Stern, S., Weiser, B., and Noller, H.F. Model for the three-dimensional folding of 16S ribosomal RNA. *J. Mol. Biol.* 1988, **204**, 447–481
12 Hubbard, J., and Hearst, J.E. Computer modeling 16S ribosomal RNA. *J. Mol. Biol.* 1991, **221**, 889–907
13 Malhotra, A., and Harvey, S.C. A quantitative model of the *Escherichia coli* 16S RNA in the 30S ribosomal subunit.*J. Mol. Biol.* 1994, **240**, 308–340
14 Fink, D., Chen, R.O., Noller, H.F., and Altman, R.B. Computational methods for defining the allowed conformational space of 16S rRNA based on chemical footprinting data. *RNA* 1996, **2**, 851–866
15 Mueller, F., and Brimacombe, R. A new model for the three-dimensional folding of *Escherichia coli* 16S ribosomal RNA. I. Fitting the RNA to a 3D electron microscopic map at 20 Å. *J. Mol. Biol.* 1997, **271**, 524–544
16 Mueller, F., Stark, H., van Heel, M., Rinke-Appel, J., and Brimacombe, R. A new model for the three-dimensional folding of *Escherichia coli* 16S ribosomal RNA. III. The topography of the functional centre. *J. Mol. Biol.* 1997, **271**, 566–587
17 Gabashvili, I.S., Agrawal, R.K., Spahn, C.M., Grassucci, R.A., Svergun, D.I., Frank, J., and Penczek, P. Solution structure of the *E. coli* 70S ribosome at 11.5 Å resolution. *Cell* 2000, **100**, 537–549
18 Clemons, W.M.J., May, J.L.C., Wimberly, B.T., McCutcheon, J.P., Capel, M.S., and Ramakrishnan, V. Structure of a bacterial 30S ribosomal subunit at 5.5 Angstroms resolution. *Nature* 1999, **400**, 833–840
19 Tocilj, A., Schlunzen, F., Janell, D., Gluhmann, M., Hansen, H.A.S., Harms, J., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., and Yonath, A. The small ribosomal subunit from *Thermus thermophilus* at 4.5 Å resolution: Pattern fittings and the identification of a functional site. *Proc. Natl. Acad. Sci. U.S.A.* 1999, **96**, 14252–14257
20 Cate, J.H., Yusupov, M.M., Yusupova, G.Z., Earnest, T.N., and Noller, H.F. X-ray crystal structures of 70S ribosome functional complexes. *Science* 1999, **285**, 2095–2104
21 Agalarov, S.C., Sridhar Prasad, G., Funke, P.M., Stout,

C.D., and Williamson, J.R. Structure of the S15, S6, S18-rRNA complex: assembly of the 30S ribosome central domain. *Science* 2000, **288**, 107–113
22 Gutell, R.R. Collection of small subunit (16S and 16S-like) ribosomal RNA structures: 1994. *Nucl. Acids Res.* 1994, **22**, 3502–357
23 Massire, C., and Westhof, E. MANIP: an interactive tool for modelling RNA. *J. Mol. Graph. Model.* 1998, **16**, 197–205, 255–257
24 Major, F., Turcotte, M., Gautheret, D., Lapalme, G., Fillion, E., and Cedergren, R. The combination of symbolic and numerical computation for three-dimensional modeling of RNA. *Science* 1991, **253**, 1255–1260
25 Unpublished results, Gutell R.R., Subashchandran, S., Schnare, M., Du, Y., Lin, N., Madabusi, L., Muller, K., Pande, N., Yu, N., Shang, Z., Date, S., Konings, D., Schweiker, V., Weiser, B., and Cannone, J.J., 2000. see http://www.rna.icmb.utexas.edu/
26 Maidak, B.L., Cole, J.R., Lilburn, T.G., Parker, C.T., Saxman, P.R., Stredwick, J.M., Garrity, G.M., Li, B., Olsen, G.J., Pramanik, S., Schmidt, T.M., and Tiedje, J.M. The RDP (Ribosomal Database Project) continues. *Nucl. Acids Res.* 2000, **28**,173–174
27 Gautheret, D., and Gutell, R.R. Inferring the conformation of RNA base pairs and triples from patterns of sequence variation. *Nucl. Acids Res.* 1997, **25**, 1559–1564
28 Moazed, D., Stern, S., and Noller, H.F. Rapid chemical probing of conformation in 16S ribosomal RNA and 30S ribosomal subunits using primer extension. *J. Mol. Biol.* 1986, **187**, 399–416
29 Unpublished results, Wollenzien, P., 2000
30 Wilms, C., Noah, J.W., Zhong, D., and Wollenzien, P. Exact determination of UV-induced crosslinks in 16S ribosomal RNA in 30S ribosomal subunits. *RNA* 1997, **3**, 602–612
31 Heus, H.A., and Pardi, A. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* 1991, **253**, 191–194
32 Jucker, F.M., and Pardi, A. Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif. *Biochemistry* 1995, **34**, 14416–14427
33 Rife, J., and Moore, P. The structure of a methylated tetraloop in 16S ribosomal RNA. *Structure* 1998, **6**, 747–756
34 Kalurachchi, K., and Nikonowicz, E. NMR structure determination of the binding site for ribosomal protein S8 from *Escherichia coli* 16 S rRNA. *J. Mol. Biol.* 1998, **280**, 639–654
35 Correll, C.C., Freeborn, B., Moore, P.B., and Steitz, T.A. Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell* 1997, **91**, 705–12
36 Brimacombe, R. RNA-protein interactions in the *Escherichia coli* ribosome. *Biochimie* 1991, **73**, 927–936
37 Woese, C.R., Winker, S., and Gutell, R.R. Architecture of ribosomal RNA: Constraints on the sequence of "tetra-loops." *Proc. Natl. Acad. Sci. U.S.A.* 1990, **87**, 8467–8471
38 Cheong, C., Varani, G., and Tinoco, I. Jr. Solution structure of an unusually stable RNA hairpin, 5′GGAC(UUCG)GUCC. *Nature* 1990, **346**, 680–682
39 Butcher, S.E., Allain, F.H., and Feigon, J. Solution structure of the loop B domain from the hairpin ribozyme. *Nat. Struct. Biol.* 1999, **6**, 212–216
40 Pinard, R., Cote, M., Payant, C., and Brakier-Gingras, L. Positions 13 and 914 in *Escherichia coli* 16S ribosomal RNA are involved in the control of translational accuracy. *Nucl. Acids Res.* 1994, **22**, 619–624
41 Dammel, C.S., and Noller, H.F. Suppression of a cold-sensitive mutation in 16S rRNA by overexpression of a novel ribosome-binding factor, RbfA. *Genes Dev.* 1995, **9**, 626–637
42 Pinard, R., Payant, C., and Brakier-Gingras, L. Mutations at postions 13 and/or 914 in *Escherichia coli* 16S ribosomal RNA interfere with the initiation of protein synthesis. *Biochemistry* 1995, **34**, 9611–9616
43 Poot, R., Pleij, C., and van Duin, J. The central pseudoknot in 16S ribosomal RNA is needed for ribosome stability but is not essential for 30S initiation complex formation. *Nucl. Acids Res.* 1996, **24**, 3670–3676
44 Juzumiene, D.I., and Wollenzien, P. Arrangement of the central pseudoknot region of 16S rRNA in the 30S ribosomal subunit determined by site-directed 4-thiouridine crosslinking. *RNA* 2001, **7,** 71–84
45 Gutell, R.R., *Comparative sequence analysis and the structure of 16S and 23S rRNA.* in *Ribosomal RNA. Structure, Evolution, Processing, and Function in Protein Biosynthesis* Dahlberg, A., and Zimmerman, B., Eds. CRC Press, Boca Raton, 1996, pp. 111–128
46 Babin, P., Dolan, M., Wollenzien, P., and Gutell, R.R. Identity and geometry of a base triple in 16S rRNA determined by comparative sequence analysis and molecular modeling. *RNA* 1999, **5**, 1430–1439
47 Unpublished results, Woese, C., 2000
48 Gutell, R.R., and Woese, C.R. Higher order structural elements in ribosomal RNAs: Pseudo-knots and the use of noncanonical pairs. *Biochemistry* 1990, **87**, 663–667
49 Gautheret, D., Damberger, S.H., and Gutell, R.R. Identification of Base-triples in RNA using Comparative Sequence Analysis. *J. Mol. Biol.* 1995, **248**, 27–43
50 Woese, C.R. and Gutell, R.R. Evidence for several higher order structural elements in ribosomal RNA. *Proc. Natl. Acad. Sci. U.S.A.* 1989, **86**, 3119–3122
51 Powers, T., and Noller, H.F. A functional pseudoknot in 16S ribosomal RNA. *EMBO J.* 1991, **10**, 2203–2214
52 Winker, S., and Woese,C.R. A definition of the domains Archaea, Bacteria and Eucarya in terms of small subunit ribosomal RNA characteristics. *System. Appl. Microbiol.* 1991, **14**, 305–310
53 Rinke-Appel, J., Junke, N., Stade, K., Junke, N., and Brimacombe, R. The path of mRNA through the *Escherichia coli* ribosome; site-directed cross-linking of mRNA analogues carrying a photoreactive label at various points 3′ to the decoding site. *EMBO. J.* 1991, **10**, 2195–2202
54 Wollenzien, P., Expert-Bezancon, A., and Favre, A. Sites of contact of mRNA with 16S rRNA and 23S rRNA in the *Escherichia coli* ribosome. *Biochemistry* 1991, **30**, 1788–1795
55 Juzumiene, D.I., Shapkina, T.G., and Wollenzien, P. Distribution of crosslinks between mRNA analogues and 16S rRNA in *Escherichia coli* 70S ribosomes made under equilibrium conditions and their response to tRNA binding. *J. Biol. Chem.* 1995, **270**, 12794–12800
56 Dontsova, O., Dokudovskaya, S., Kopylov, A., Bogdonov, A., Rinke-Appel, J., Junke, N., and Brimacombe, R.

Three widely separated positions in the 16S RNA lie in or close to the ribosomal decoding region; a site directed cross-linking study with mRNA analogues. *EMBO J.* 1992a, **11**, 3105–3116
57 Moazed, D., and Noller, H.F. Transfer RNA shields specific nucleotides in 16S ribosomal RNA from attack by chemical probes. *Cell* 1986, **47**, 985–994
58 Wang, R., Alexander, R.W., VanLoock, M., Vladimirov, S., Bukhtiyarov, Y., Harvey, S.C., and Cooperman, B.S. Three-dimensional placement of the conserved 530 loop of 16S rRNA and of its neighboring components in the 30 S subunit. *J. Mol. Biol.* 1999, **285**, 521–540
59 Unpublished data, Mundus, D. and Wollenzien, P
60 Powers, T., and Noller, H.F. Hydroxyl radical footprinting of ribosomal proteins on 16S rRNA. *RNA* 1995, **1**, 194–209
61 van Waes, M.A., Smith, C.E., and Hill, W.E. Probing of the 530 loop of *the Escherichia coli* small ribosomal-subunit with short oligodeoxynucleotides. *FASEB J.* 1997, **7**, A1093
62 Gutell, R.R., Noller, H.F., and Woese, C.R. Higher order structure in ribosomal RNA. *EMBO J.* 1986, **5**, 1111–1113
63 Vila, A., Viril-Farley, J., and Tapprich, W.E. Pseudoknot in the central domain of small subunit ribosomal RNA is essential for translation. *Proc. Natl. Acad. Sci. U.S.A.* 1994, **91**, 11148–11152
64 Mundus, D., and Wollenzien, P. Neighborhood of 16S rRNA nucleotides U788/U789 in the 30S ribosomal subunit determined by site-directed crosslinking. *RNA* 1998, **4**, 1373–1385
65 Masquida, B., Felden, B., and Westhof, E. Context dependent RNA-RNA recognition in a three-dimensional model of the 16S rRNA core. *Bioorg. Med. Chem.* 1997, **5**, 1021–1035
66 Personal communication, Gutell, R. (University of Texas, Austin), 2000.
67 Leontis, N.B., and Westhof, E. The 5S rRNA loop E: chemical probing and phylogenetic data versus crystal structure. *RNA* 1998, **4**, 1134–1153
68 Noah, J., Shapkina, T., and Wollenzien, P. UV-induced crosslinks in the 16S rRNAs of *Escherichia coli, Bacillus subtilis* and *Thermus aquaticus* and their implications for ribosome structure and photochemistry. *Nucl. Acids Res.* 2000, **28**, 3785–3792
69 Mougel, M., Allmang, C., Eyermann, F., Cachia, C., Ehresmann, B., and Ehresmann, C. Minimal 16S rRNA binding site and role of conserved nucleotides in *Escherichia coli* ribosomal protein S8 recognition. *Eur. J. Biochem.* 1993, **215**, 787–792
70 Moine, H., Nurse, K., Ehresmann, B., Ehresmann, C., and Ofengand, J. Conformational analysis of *Eschericia coli* 30S ribosomes containing the single-base mutations G530U, U1498G, G1401C, C1501G and the double-base mutation G1401C/C1501G. *Biochemistry* 1997, **44**, 13700–13709
71 Doring, T., Mitchell, P., Osswald, M., Bochkariov, D., and Brimacombe, R. The decoding region of 16S RNA; a cross-linking study of the ribosomal A, P, and E sites using tRNA derivatized at position 32 in the anticodon loop. *EMBO J.* 1994, **13**, 2677–2685
72 Moazed, D., and Noller, H.F. Binding of tRNA to the ribosomal A and P sites protects two distinct sets of nucleotides in 16S rRNA. *J. Mol. Biol.* 1990, **211**, 135–145
73 Egebjerg, J., and Garrett, R.A. Binding sites of the antibiotics pactamycin and celesticetin on ribosomal RNAs. *Biochimie* 1991, **73**, 1145–1149
74 Woodcock, J., Moazed, D., Cannon, M., Davies, J., and Noller, H.F. Interaction of antibiotics with A- and P-site-specific bases in 16S ribosomal RNA. *EMBO J.* 1991, **10**, 3099–3103
75 Herr, W., Chapman, N., and Noller, H. Mechanism of ribosomal subunit association: discrimination of specific sites in 16S RNA essential for association activity. *J. Mol. Biol.* 1979, **130**, 433–439
76 Santer, M., Bennett-Guerrero, E, Byahatti, S, Czarnecki, S, O'Connell, D, Meyer, M, Khoury, J., Cheng, X., Schwartz, I., and McLaughlin, J. Base changes at position 792 of *Escherichia coli* 16S rRNA affect assembly of 70S ribosomes. *Proc. Natl. Acad. Sci. U.S.A.* 1990, **87**, 3700–3704
77 Tapprich, W.E., and Hill, W.E. Involvement of bases 787–795 of *Escherichia coli* 16S ribosomal RNA in ribosomal subunit association. *Proc. Natl. Acad. Sci. U.S.A.* 1986, **83**, 556–560
78 Tapprich, W.E., Goss, D.J., and Dahlberg, A.E. Mutation at position 791 in *Escherichia coli* 16S ribosomal RNA affects processes involved in the initiation of protein synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 1989, **86**, 4927–4931
79 Moazed, D., Samaha, R., Gualerzi, C., and Noller, H. Specific protection of 16S rRNA by translational initiation factors. *J. Mol. Biol.* 1995, **248**, 207–210
80 Lee, K.S., Varma, S., SantaLucia, J., Jr., and Cunningham, P. *In vivo* determination of RNA structure–function relationships: analysis of the 790 loop in ribosomal RNA. *J. Mol. Biol.* 1997, **269**, 732–743
81 Trikha, J., Filman, D.J., and Hogle, J.M. Crystal structure of a 14 bp RNA duplex with non-symmetrical tandem GxU wobble base pairs. *Nucl. Acids Res.* 1999, **27**, 1728–1739
82 Lodmell, J., and Dahlberg, A. A conformational switch in *Escherichia coli* 16S ribosomal RNA during decoding of messenger RNA. *Science* 1997, **277**, 1262–1267
83 Correll, C.C., Munishkin, A., Chan, Y.L., Ren, Z., Wool, I.G., and Steitz, T.A. Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *Proc. Natl. Acad. Sci. U.S.A.* 1998, **95**, 13436–1341
84 Szewczak, A.A., Ortoleva-Donnelly, L., Ryder, S.P., Moncoeur, E., and Strobel, S.A. A minor groove RNA triple helix within the catalytic core of a group I intron. *Nat. Struct. Biol.* 1998, **5**, 1037–1042
85 Pley, H.W., Flaherty, K.M., and McKay, D.B. Three-dimensional structure of a hammerhead ribozyme. *Nature* 1994, **372**, 68–74
86 Gutell, R., Weiser, B., Woese, C., and Noller, H. Comparative anatomy of 16S-like ribosomal RNA. *Prog. Nucl. Acid Res. Mol. Biol.* 1985, **32**, 155–216
87 Baudin, F., Mougel, M., Romby, P., Eyermann, F., Ebel, J.P., Ehresmann, B., and Ehresmann, C. Probing the phosphates of the *Escherichia coli* ribosomal 16S RNA in its naked form, in the 30S subunit, and in the 70S ribosome. *Biochemistry* 1989, **28**, 5847–5855
88 Prescott, C.D., Kleuvers, B., and Goringer, H.U. A rRNA-mRNA base pairing model for UGA-dependent termination. *Biochimie* 1991, **73**, 1121–1129

89 Unpublished results, Shapkina, T., and Wollenzien, P., 2000
90 Prince, J.B., Taylor, B.H., Thurlow, D.L., Ofengand, J., and Zimmermann, R.A. Covalent cross-linking of $tRNA_1^{Val}$ to 16S RNA at the ribosomal P site: Identification of cross-linked residues. *Proc. Natl. Acad. Sci. U.S.A.* 1982, **79**, 5450–5454
91 Juzumiene, D.I., and Wollenzien, P. Organization of the 16S rRNA around its 5′ terminus determined by photochemical crosslinking in the 30S ribosomal subunit. *RNA* 2000, **6**, 26–40
92 Unpublished results, Babin, P., Juzumiene, D.I., Dolan, M.A., and Wollenzien, P., 2000
93 Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P.B. and Steitz, T. A. Placement of protein and RNA structures into a 5Å-resolution map of the 50S ribosomal subunit. *Nature* 1999, **400**, 841–847
94 Unpublished results, Dolan, M., Babin, P., and Wollenzien, P. Three-dimensional arrangement of the *Escherichia coli* 16S rRNA in the 30S ribosomal subunit using constraint satisfaction algorithms
95 PDB ID: 1EKC. Agalarov, S.C., Sridhar Prasad, G., Funke, P.M., Stout, C.D., and Williamson, J.R. Structure of the S15, S6, S18-rRNA complex: assembly of the 30S ribosome central domain. *Science* 2000, **288**, 107–113
96 Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. The Protein Data Bank. *Nucl. Acids Res.* 2000, **28**, 235–242