RNA (1996), 2:851-866. Cambridge University Press. Printed in the USA.

Computational methods for defining the allowed conformational space of 16S rRNA based on chemical footprinting data

DORAN L. FINK,[1] RICHARD O. CHEN,[1] HARRY F. NOLLER,[2] and RUSS B. ALTMAN[1]

[1] Section on Medical Informatics, Stanford University, MSOB X215, Stanford, California 94305-5479, USA
[2] Department of Biology and Center for Molecular Biology of RNA, Sinsheimer Laboratories, University of California–Santa Cruz, Santa Cruz, California 95064, USA

ABSTRACT

Structural models for 16S ribosomal RNA have been proposed based on combinations of crosslinking, chemical protection, shape, and phylogenetic evidence. These models have been based for the most part on independent data sets and different sets of modeling assumptions. In order to evaluate such models meaningfully, methods are required to explicitly model the spatial certainty with which individual structural components are positioned by specific data sets. In this report, we use a constraint satisfaction algorithm to explicitly assess the location of the secondary structural elements of the 16S RNA, as well as the certainty with which these elements can be positioned. The algorithm initially assumes that these helical elements can occupy any position and orientation and then systematically eliminates those positions and orientations that do not satisfy formally parameterized interpretations of structural constraints. Using a conservative interpretation of the hydroxyl radical footprinting data, the positions of the ribosomal proteins as defined by neutron diffraction studies, and the secondary structure of 16S rRNA, the location of the RNA secondary structural elements can be defined with an average precision of 25 Å (ranging from 12.8 to 56.3 Å). The uncertainty in individual helix positions is both heterogeneous and dependent upon the number of constraints imposed on the helix. The topology of the resulting model is consistent with previous models based on independent approaches. The result of our computation is a conservative upper bound on the possible positions of the RNA secondary structural elements allowed by this data set, and provides a suitable starting point for refinement with other sources of data or different sets of modeling assumptions.

Keywords: computational methods; molecular modeling; ribosome active sites; ribosome structure; tRNA binding

INTRODUCTION

In the bacterium *Escherichia coli*, the 70S ribosome is comprised of two separate subunits, the 30S subunit (the site of codon–anticodon interaction) and the 50S subunit (the site of peptide bond formation). Each of these subunits consists of ribosomal RNA (rRNA) tightly associated with a specific set of ribosomal proteins (21 for the 30S, and 31 for the 50S). The three-dimensional structures of these ribosomal subunits have been the subject of intense investigation, in order to understand the mechanisms of protein synthesis. The structure of the 30S subunit has been pursued because it is the smaller of the two subunits, and there are many more experimental constraints available to help deduce its structure.

Unfortunately, the 30S subunit has not yielded to direct methods of structure determination, such as X-ray crystallography (Yonath et al., 1990). Nevertheless, a large amount of experimental data has been compiled, and several models based on subsets of this data (using various modeling approaches) have been presented (Brimacombe et al., 1988; Stern et al., 1988; Hubbard & Hearst, 1991; Malhotra & Harvey, 1994; Brimacombe, 1995; Chen et al., 1995). The *E. coli* 30S subunit has a 1,542-nt 16S rRNA (Brosius et al., 1981). Regions of base pairing and helix formation have been determined by comparative sequence analysis using the 16S rRNA sequences of different prokaryotic species (Noller &

Reprint requests to: Russ B. Altman, Section on Medical Informatics, Stanford University, MSOB X215, Stanford, California 94305-5479, USA; e-mail: altman@smi.stanford.edu.

Woese 1981; Woese et al., 1983). The locations of the 21 30S proteins have been determined by neutron scattering (Capel et al., 1987). For three of these proteins, S5, S6, and S17, a high-resolution structure is available (Ramakrishnan & White, 1992; Golden et al., 1993; Lindahl et al., 1994). A critical question is, then, how do the proteins and RNA associate to form the active ensemble?

There are several techniques for measuring physical proximity of protein and RNA within the 30S structure, including protection of the RNA from chemical probes by interaction with proteins (Moazed et al., 1986; Stern et al., 1989), and by chemical or photochemical crosslinking (Greuer et al., 1987; Osswald et al., 1987; Brimacombe, 1991). In addition, there are now low-resolution structural envelopes available from three-dimensional reconstruction of electron microscopy images (Frank et al., 1995; Stark et al., 1995).

Previously proposed models of the 30S subunit have relied primarily on protein protection data and RNA–RNA tertiary interactions as the primary basis for predicting structure. The reliability of protein protection data has been questioned recently because of varying experimental conditions, the possibility of heterogeneous conformational changes induced by the experimental methods, and possible errors in the reported protein locations from neutron diffraction (Brimacombe, 1995). Some recently reported protein protection data, in which hydroxyl radicals generated from Fe(II)-EDTA are used as the chemical probing agent, provide an opportunity to address some of these concerns (Powers & Noller, 1995). Hydroxyl radicals attack the ribose moiety in the backbone of RNA, providing an extensive set of reproducible protection patterns gathered under standard conditions. These data are more abundant and are relatively indifferent to whether the target ribose lies in single- or double-stranded structure (Latham & Cech, 1989). For this reason, hydroxyl radical footprinting is less likely to be complicated by the effects of protein-induced conformational changes than those collected using base-specific probes, which are typically specific for unpaired bases (Powers & Noller, 1995). Eventually, model building studies should incorporate reliable information from all these data sources. In this study, however, we seek to establish the independent structural information content of the hydroxyl radical data in the context of the published protein positions. With this information in hand, we will be in a position to subsequently combine data sets and better understand any inconsistencies that may arise. The questions we ask here are: (1) Do the hydroxyl radical data allow the definition of structural models that are relatively well defined? (2) Are the resulting structures largely compatible with previous modeling efforts? Because the hydroxyl radical data may not be sufficient to define a unique conformation for 16S rRNA, we sought to identify the allowed ranges of helix positions (and conformations) that are compatible with the data.

Previously, we have described a computational method for determining structure from sparse and noisy constraint information. PROTEAN I, designed originally for interpreting distance constraints from protein NMR experiments, uses a constraint satisfaction algorithm to determine an upper bound to the number of structures that satisfy a parametric interpretation of the constraint (Brinkley et al., 1987). It has been tested in both model systems of known structure (Brinkley et al., 1987; Lichtarge et al., 1987), as well as used to solve the structure of unknown molecules based on the distances derived from NMR (Arrowsmith et al., 1991). It has been shown to produce a reliable upper bound on the set of locations for helical components, given distance constraints. It is made freely available at the Quantum Chemistry Program Exchange (Brinkley et al., 1990), or by request from the authors.

The method begins with the assumption that each helix can be located at any position, in any orientation. By sequentially introducing structural constraints (such as distances to the proteins or between RNA helices), the list of allowed positions and orientations is pruned until a minimal set is found. This set provides, in a sense, an upper bound on the range of conformations that can be adopted within the given constraints and thus provides a measure of the net information content of the structural constraints. We have used our constraint satisfaction algorithms using the hydroxyl radical footprinting data to define the range of positions for the 16S rRNA secondary structural elements that are compatible with the data. Unlike previous model-building efforts, we conduct a systematic search through conformation space and retain all structures (within our sampling limit) that are allowed. We show that a conservative interpretation of the hydroxyl radical data is sufficient to define a gross topology of the 16S RNA. We also show that the uncertainty in individual helix positions is very heterogeneous, and depends on the number of protein protection constraints, as well as the number of single-stranded RNA tethering constraints to other helices. The result of our computation is a conservative upper bound on the possible positions of the RNA secondary structural elements and is a suitable starting point for refinement using other techniques (based on energetics or maximum likelihood, for example).

The other methods for deriving structure from constraints include a modified version of molecular mechanics (Malhotra & Harvey, 1994), interactive computer graphical model building (Stern et al., 1988; H.F. Noller, T. Powers, G.M. Heilek, S. Mian, & B. Weiser, in prep.), and variations of distance geometry (Brimacombe et al., 1990; Hubbard & Hearst, 1991). In addition, several qualitative models have been built (Expert-Benzancon & Wollenzien, 1985; Nagano et al., 1988; Oakes et al., 1990). Each of these has been successful at combining

multiple data sources to arrive at a single (or a few closely related) alternative structure. Our method is distinguished from these in that it is designed specifically to explore the full conformational space compatible with the data, and does not produce any typical or consensus model. (For the purposes of comparison, we do make available the average positions and spatial uncertainty of the phosphate atoms in the helical backbones, as discussed in the Materials and Methods.) We have shown the feasibility of applying the PROTEAN program to ribosomal structure in a preliminary report of the structure of the central domain of 16S rRNA (Altman et al., 1994). The data set used for the modeling reported here has also been interpreted using interactive manual model building. Thus, there is an opportunity to evaluate the relative benefits of automatic versus manual interpretation of structural data.

RESULTS AND DISCUSSION

Our method of representing and interpreting the hydroxyl protection data was sufficient to position 93% of the 16S rRNA double-helical elements (61 of 64 helices) within well-defined volumes of uncertainty. In fact, the locations of the individual helices are consistent in morphology with the recent 23-Å resolution electron micrograph reconstruction of the 30S subunit (Stark et al., 1995) and with the model published by Malhotra and Harvey (1994). In particular, our coherent instances show a "head" at the top of the subunit, a "body" at the bottom, a large groove and protruding notch on the left (using the view of Fig. 7 in Malhotra & Harvey, 1994), and a smaller groove on the lower right (see Fig. 5). Another interesting feature of our model is that there is a three-dimensional partitioning of helices that are grouped together in the secondary structure. Helices in the 5′ domain occupy the lower portion of the subunit, whereas helices in the central domain occupy the middle portions, near the presumptive tRNA binding cleft. Helices in the 3′ domain are localized to the upper portions. This partitioning has been reported previously (Stern et al., 1988; Brimacombe et al., 1990; Malhotra & Harvey, 1994).

The uncertainty of any helix in our model can be best determined by considering the volume of its cloud, which is the total space occupied by the helix in all of its allowed locations. The mean cloud volume using only footprinting data (after the anchoring step) was 350,000 $Å^3$, (corresponding to an uncertainty radius of 44 Å, assuming spherical clouds). The final cloud volumes (after high-order yoking is completed) vary considerably, from a minimum of 8,800 $Å^3$ (helix 13) to a maximum of 750,000 $Å^3$ (helix 40). The average reduction in cloud size was almost 60% from start to finish (see Table 4).

The mean helix cloud size at the final stage of calculation was ~64,000 $Å^3$ (see Table 4), which corresponds to a 22-Å radius of uncertainty. The diameter of an A-form double helix is approximately 19 Å, and so a positional uncertainty of 22 Å is only slightly more than one helix diameter from the center of the cloud. More than 60% of the helix clouds occupy a volume of less than 30,000 $Å^3$ (see Table 4), which is indicative of an uncertainty radius less than 19.2 Å. This uncertainty is comparable to the resolution of available electron microscopy measurements of the 30S subunit, which describes its topography at a resolution of 23 Å (Stark et al., 1995).

The uncertainties in our helix positions are somewhat higher than those reported by Malhotra and Harvey (1994)—75% of the RNA helices in their models were positioned to within 15 Å. Our approach differs from theirs in at least two respects. First, Malhotra and Harvey used experimental data gathered from various sources, including several types of footprinting and crosslinking studies of protein-RNA interactions, as well as crosslinking data for RNA-RNA interactions. Our calculation is based only on a single data set provided by protection of 16S rRNA from hydroxyl radical attack by specific proteins (Powers & Noller, 1995). Second, we used a conservative estimate for all distance ranges (both the experimental and RNA tether distances), in order to maintain a conservative "upper bound" on the conformation space. Malhotra and Harvey used a very different method based on molecular mechanics and pseudoenergy constraints derived from experimental data, which sought models with the smallest errors possible. Even though we invoked fewer constraints than Malhotra and Harvey, the general topology and level of structural uncertainty within our model are quite similar. We have experimented with further pruning of our location clouds with RNA-RNA tertiary constraints and these can reduce cloud sizes significantly.

The heterogeneity that exists among final cloud volumes also exists among the final number of locations in each cloud. The number of final locations varied from 1 (helix 22, helix 21.1) to 29,000 (helix 40), where the reduction in number of locations from initial anchoring to *N*-yoking ranged from 10% to greater than 99.9% (Fig. 4). (It is important to note that one location at our sampling interval may still be compatible with multiple locations at resolutions below 4–7 Å.). As expected, clouds with fewer locations have smaller volumes of uncertainty. A correlation coefficient of 0.89 exists between the number of locations in a cloud and the physical dimensions of the cloud. Interesting exceptions exist, however, in cases where a helix is well localized to a small volume by strong protein protection constraints, but is free to assume many orientations within the volume due to weakness in tether constraints to other helices (for example, helix 31). Exceptions also exist when strong tether constraints allow a helix to occupy only a few locations that are not close to each other (such as helix 4).

As would be expected, the heterogeneity in both final cloud volume and final number of locations correlates well with the number of constraints imposed on a given helix. For example, helices 4, 6.3, and 22 all have multiple constraints to proteins and to other helices (see constraint network, Fig. 3), and their clouds diminished steadily throughout the calculation (see Fig. 4). Helix 17.2, on the other hand, has only one tether constraint to another helix and no constraints to the ribosomal proteins. Not surprisingly, the cloud for helix 17.2 remained large in both volume and number of locations throughout the calculation (shown in Table 4).

Just as the final cloud volumes provide a conservative upper bound to the uncertainties in helix location attributable to the hydroxyl footprinting data, they also provide an upper bound to the number of conformations of the RNA that simultaneously satisfy all constraints imposed by the data. Based on our cloud sizes, there are $\sim 10^{66}$ coherent instances possible after using our boundary sampling procedure. Although this is a seemingly gargantuan number of possible conformations, the difference between one coherent instance and another might be as small as moving a single helix 7 Å in one direction or rotating it slightly about its vertical axis. Furthermore, removing the most poorly constrained helices from our model decreases the number of coherent instances by a factor of 100 for each helix removed. Perhaps most importantly, coherent instances that vary greatly in individual helix locations can look quite similar in terms of general topography (see Fig. 5).

The hydroxyl radical footprinting data provide a set of qualitative proximity relationships between the ribosomal proteins and bases in the rRNA. In calculating a structure from the data, it was therefore important to systematically interpret the qualitative relationships into quantitative values, i.e., maximum distances between bases and proteins. A parameterization of the data in terms of uncertainty in protein shape (μ = morphology factor) and protein position (σ = number of standard deviations from the mean position) made such an interpretation possible, as is shown in Figure 2. The calculation assumed that ribosomal proteins were roughly globular, and that their positions in the subunit were confined to the error ellipsoids determined by neutron diffraction. Higher values of μ represent situations in which proteins could be elongated rather than spherical. Higher values of σ allow for greater uncertainty in the accuracy of the neutron diffraction positions by increasing the lengths of the error ellipsoid axes. Our parameterization of the hydroxyl data determines only the upper bound of the constraint, i.e., the maximum distance between a protein and a base. The minimum bound of protein–RNA interactions was uniformly set at zero, because the structures of most ribosomal proteins are not known, and an RNA strand could theoretically run through a deep binding cleft within a protein. In fact, those 30S proteins whose structures are known (S5, S6, and S17) contain potential RNA-binding clefts (Liljas & Garber, 1995). We used values of $\mu = 1$ and $\sigma = 1.5$ for most computations reported here. The increase in cloud size caused by increasing the value of either μ or σ above 2.0 results in a loss of structural information such that many of the helices cannot be well localized. Setting either parameter below its default value generally over-constrained the helices, which resulted in no allowed locations for all of the helices.

A comparison of helix clouds after initial anchoring to those at the end of the 21-way yoking shows that the clouds not only decrease in volume, but also move their centers of mass, by an average of 16.0 Å (see Table 4). This movement is a significant effect of tether constraints between helices, and demonstrates that reducing cloud size by simply decreasing the values of μ and σ would not result in an accurate representation of the data. Thus, it seems that our parameter values are consistent with a conservative interpretation of the data.

As expected, the model built by interactive model building, using the same protein protection data, is quite similar to ours, with only a few differences. This interactive model provides the phosphate positions for 43 of the 64 helical regions modeled in this study (H.F. Noller, T. Powers, G.M. Heilek, S. Mian, & B. Weiser, in prep.). Of these 43 helices, 46% (20 helices) are positioned within 19 Å (one RNA helix diameter) of the cloud center for our helix, and 86% (37 helices) differ in position by less than 38 Å. The remaining six helices are all less than 53 Å from our helix cloud centers. On average, the distance between our cloud centers and the interactive model positions was 22 Å, only slightly more than one helix diameter (see Table 4 for summary). Of the 43 helices represented in the interactive model, 15 do not satisfy our parametric interpretation of the footprinting data, and these helix locations would not be allowed in our model. A correlation exists between the ability of the interactively positioned helices to satisfy our parametric constraints and the difference in position between those helices and the corresponding helices in our model. Interactive model building is able to introduce additional knowledge about how to "strain" distance constraints acceptably in order to find good locations. Our method, on the other hand, is strict in its application of the distance ranges defined parametrically. The similarity between the models is gratifying, however, because the automatic method employed here is fast and can be run many times.

It is also reassuring to see the level of agreement with the Malhotra and Harvey model, even though the agreement with this model was somewhat less than with the interactively built model. On average, their

helix positions differed from ours by 30.8 Å, or about 1.5 helix diameters, with 74% of comparable helices less than 38 Å from our cloud centers and 96% of comparable helices differing in position by less than 57 Å (see Table 4 for summary). In this case, both the data set and the computational method used for interpreting this data were different, so it was expected that disagreements might be more widespread and larger in magnitude. In addition, Malhotra and Harvey represented the secondary structure as 86 double-helical elements, as opposed to 64 in our representation, allowing for greater flexibility of certain helical components. One such instance is our helix 22, which Malhotra and Harvey modeled as three separate helices. One of the three helices in their model agrees moderately well with our position for helix 22, (35.5 Å from the cloud center), but the other two agree much less well (51.9 Å and 62.7 Å – see Table 4).

An important factor to be noted in comparing these two models is that, although our helices are unconditionally consistent with our interpretation of the distance constraints, the Malhotra and Harvey model, like the interactive model, allowed some leniency in satisfying constraints, which again adds flexibility to the range of possible positions. Of the 86 helices represented in the Malhotra and Harvey model, 33 fail to satisfy our parametric interpretation of the footprinting data. In fact, only 57 of the 188 protein–RNA constraints that we use are satisfied in the Malhotra and Harvey model. As with the interactively built model, helices in the Malhotra and Harvey model whose positions differ the most from ours tend not to satisfy our protein–RNA constraints. Notable exceptions are helices H17.2, H28, and H33.2, all of which have extremely weak protein protection constraints and large cloud volumes in our model. Applying the protein–RNA protection data and interpretation of that data used by Malhotra and Harvey to our clouds reveals that 15 of our 61 helix clouds do not contain at least one location that satisfies these constraints. Again, helices in our model that do not satisfy the Malhotra and Harvey protein–RNA constraints tend to be those which, in the Malhotra and Harvey model, do not satisfy our protein–RNA constraints. This clearly shows how differences between constraint sets used in computational model building can lead to differences between the resultant models. The general concordance of the models, however, can be taken as evidence that the positions of the helices are now known to an average resolution of about 30 Å. Given the high-quality electron microscopic envelopes that are now available as yet another set of structural constraints, and the increasing availability of high-resolution protein structures, we can soon look forward to glimpses of the 30S subunit structure at even higher resolution.

Even though we considered only the hydroxyl radical footprinting data in our calculation, our model fares well when put to the test of satisfying additional constraint information. One such source of constraint information is the eight RNA–RNA contacts derived from crosslinking data (Stiege et al., 1986). Using a parameterized interpretation of this data as reported in Brimacombe et al. (1990), our model satisfies seven of the eight RNA–RNA constraints. The eighth constraint can be satisfied by adding 10 Å to the maximum distance allowed, which is not much larger than our sampling error of 7 Å. Another source of constraint information is the topographical boundaries imposed by the EM envelope of the 30S subunit. Using the EM reconstruction reported in Verschoor et al. (1984), we found that 58 of our 61 helix clouds are contained well within the EM envelope. Both the RNA–RNA crosslinking data and the EM envelope could be used to prune our clouds in future refinements of our model.

As noted above, we did not include data involving S20, because of uncertainty about its location. In this case, unlike the other proteins, there is clearly conflicting evidence about its location. The neutron map places S20 near the midsection of the 30S subunit, as shown in the figures. Immunoelectron microscopy data places it at the bottom of the subunit (Schwedler et al., 1993). However, we were able to position many of the RNA structural elements protected by S20 using other constraints. We could thus use the locations of these helices (helix 8.1, helix 9, helix 13) to compute a plausible location for S20. This computationally derived position for the center of mass of S20 is at (39.0, 5.0, 44.2), which moves S20 from the base of the 30S subunit head (as predicted by neutron scattering) toward the bottom of the subunit. This new position is in between the neutron map location and the position at the bottom of the subunit suggested by immunoelectron microscopy (Schwedler et al., 1993). Future calculations may allow estimates of the protein positions and orientations, based not only upon neutron diffraction distances, but also upon protein–RNA distance constraints.

Our results in identifying the binding sites for important 30S ligands are useful for assessing the degree to which these sites are currently understood. Although the locations of these ligands can be defined reliably, the orientation is still not clear at this resolution. Our attempt to dock P-site tRNA to our model, for example, shows that the anticodon stem loop is well localized to the bottom of the major cleft, but that the entire tRNA molecule is still relatively unrestricted in its orientation (see Fig. 7). We experienced an interesting problem in docking P-site tRNA in that helix 31, with which P-site tRNA has been shown to interact at nt 966 (vonAhsen & Noller, 1995), is located more than 50 Å away from the tRNA-binding cleft and the rest of the P-site tRNA-associated helices (see Fig. 6). Not only does our location for H31 make docking P-site tRNA difficult, but it also differs by 40 Å with the locations

for the corresponding helix in both Malhotra and Harvey's computed model and Noller's interactively built model (see Table 4). In any case, it seems that some distance constraints must be strained in order to place helix 31 near its P-site neighbors. One possible cause of this disagreement is that the protein–RNA constraints that localized H31 to the area of S19 (see Table 1) might require different parameterizations than others, or they might even be the result of secondary conformational effects rather than true protein–RNA contacts. Alternatively, helix 31 might be dynamic, and the constraints between helix 31 and the proteins might arise from a different ribosomal conformation than do the constraints between helix 31 and P-site tRNA. Further corroboration with emerging high-resolution data (Rinke-Appel et al., 1995) will undoubtedly aid the positioning of P-site tRNA in our model. For example, it appears that the position of P-site tRNA represented in (Agrawal et al., 1996) is contained within the cloud defined in Figure 7.

MATERIALS AND METHODS

Modeling of the components of the 30S subunit

The 30S subunit can be subdivided into its constituent 21 proteins and 64 double-helical RNA segments. Helices are represented as ideal A-form double helices (Saenger, 1984). Each base within a helix is represented by a single pseudo atom, placed at the position of the associated phosphorous atom. Single-stranded regions are represented as flexible tether constraints between helices and their conformations are not explicitly modeled. In partitioning the secondary structure of the 16S rRNA into helices, we assumed that any stretch of base pairs (including G-U pairs) not containing a helical bulge of more than three bases constituted a single helix. The 64 helices are labeled as shown in Figure 1. They roughly follow the numbering system used by Brimacombe and coworkers (Brimacombe, 1991).

For the purposes of our computation, the proteins are represented as points, located at the mean position determined by neutron scattering (Capel et al., 1987), and therefore establish a global coordinate system in which to introduce the RNA helices. Each protein position was associated with ellipsoid axes representing the standard error in the x, y, and z directions, as reported (Capel et al., 1987). Because the precise structures and orientations of the proteins are not known (high-resolution structures are available for S5, S6, and S17, but their relative orientations in the 30S subunit are not known), the RNA helices were allowed to overlap the protein center of mass, in order to allow for the possibility of tight association with some elements of the RNA. The position of protein S20, and all constraints involving it, were not included in this data set in light of present uncertainty concerning its position. S20 is mapped by neutron diffraction to the upper region of the subunit, near protein S3. Immunoelectron microscopy, however, has placed it at the bottom of the subunit (Schwedler et al., 1993), and there is currently some debate about its true location (Brimacombe, 1995). For this reason, protein protection constraints involving S20 were ex-

TABLE 1. Fe(II)-EDTA protection data, grouped by helix (Powers & Noller, 1995).[a]

Helix	Base-protein constraints
H1	21-S16, 22-S16, 23-S16, 24-S16
H2	20-S16*
H3	32-S16, 33-S16, 34-S12, 35-S12, 36-S12, 560-S16*
H4	50-S16*
H5	51-S16*, 53-S16
H7.1	124-S16, 125-S16, 126-S16, 230-S16, 232-S17, 233-S17, 234-S17, 235-S17
H7.2	137-S16, 138-S16, 228-S16*
H11.1	245-S17, 280-S17*
H11.2	246-S17*, 248-S17, 253-S17, 254-S17, 255-S17, 262-S17*, 266-S17*, 275-S17, 276-S17, 280-S17*
H12	294-S16, 295-S16, 301-S16, 302-S16, 303-S16, 304-S16
H15	375-S16, 376-S16, 377-S16
H16.1	406-S4, 407-S4, 408-S4, 409-S4, 429-S4*
H16.2	411-S4*, 426-S4, 427-S4
H17.1	438-S4, 439-S4, 440-S4, 449-S16*
H17.2	453-S16*
H18.1	500-S4, 501-S4, 502-S4, 503-S4, 542-S4, 543-S4, 544-S4
H18.2	511-S4, 512-S4, 517-S12, 524-S12*, 536-S12, 537-S12, 539-S4
H19	563-S16*
H20.1	758-S8*
H20.2	754-S15, 755-S8, 756-S8, 757-S8
H21.1	596-S16, 598-S8, 598-S8, 643-S8, 644-S8, 653-S8*
H21.2	599-S8, 608-S16*, 632-S16
H21.3	612-S16, 631-S16*
H22	655-S15, 653-S8*, 656-S15, 657-S15, 658-S15, 670-S6S18, 670-S21, 671-S6S18, 671-S21, 672-S6S18, 672-S21, 736-S6S18, 737-S6S18, 738-S6S18, 749-S15, 750-S15, 751-S15
H23.1	673-S6S18, 674-S6S18, 675-S6S18, 677-S11, 678-S11, 679-S11, 684-S11, 706-S11, 707-S11, 708-S11
H23.2	696-S11*, 697-S11, 698-S11, 705-S11*
H24.1	804-S11, 820-S8*
H24.2	794-S11*, 796-S11, 803-S11*
H25	821-S8, 822-S8, 823-S8, 824-S8, 825-S8, 826-S8, 827-S8, 873-S8, 874-S8, 875-S8, 876-S8, 877-S8, 878-S8
H26.1	844-S6S18*, 845-S6S18*
H28	1377-S7*
H29	939-S7, 940-S7, 1342-S7, 1343-S7, 1349-S7*
H30	947-S7, 956-S19*, 1233-S7, 1234-S7, 1235-S7
H31	960-S19, 961-S19, 969-S3*, 971-S3*, 976-S19*
H32	982-S19*, 985-S19, 987-S19, 990-S19
H33.1	994-S14*
H33.2	1013-S19*, 1014-S19*
H34.2	1055-S3, 1056-S3, 1057-S3
H34.3	1046-S19, 1047-S3, 1047-S19, 1048-S3, 1050-S3, 1051-S3, 1052-S3
H34.1	1189-S3, 1190-S3, 1191-S3, 1192-S3
H35	1072-S2, 1073-S2, 1102-S2, 1103-S2, 1104-S2
H36	1079-S3*, 1081-S3, 1082-S3, 1083-S3
H37	1086-S3, 1093-S3*, 1096-S3, 1097-S3
H38	1115-S9, 1116-S9, 1188-S3*
H39	1118-S9, 1123-S10, 1124-S10, 1125-S9, 1126-S9, 1127-S9, 1146-S9, 1147-S9, 1148-S9, 1149-S9, 1150-S9, 1150-S10, 1151-S10, 1152-S10
H41	1241-S7, 1248-S9, 1249-S9, 1250-S9, 1251-S9, 1252-S9, 1281-S9*, 1289-S7, 1290-S7, 1291-S7, 1301-S13*
H42	1302-S13*, 1307-S13, 1308-S13, 1309-S13, 1310-S13, 1312-S19, 1313-S19, 1314-S19, 1323-S7, 1321-S19*, 1330-S13, 1333-S7, 1334-S7
H43	1350-S7, 1351-S7, 1352-S7, 1353-S9, 1353-S7, 1354-S7, 1355-S7, 1356-S7, 1364-S7*, 1370-S7, 1371-S7, 1373-S7*

[a] Protein S6S18 represents a disjunctive constraint involving protein S6 or S18; the base in question is protected by one of these proteins (which cannot be distinguished for experimental reasons). A * denotes a constraint from a protein to a base outside of a helix, in which the total distance is the directional distance (Dd) from the protein to the base (see Fig. 2) plus the tether distance from the base to the closest helix (see Table 2).

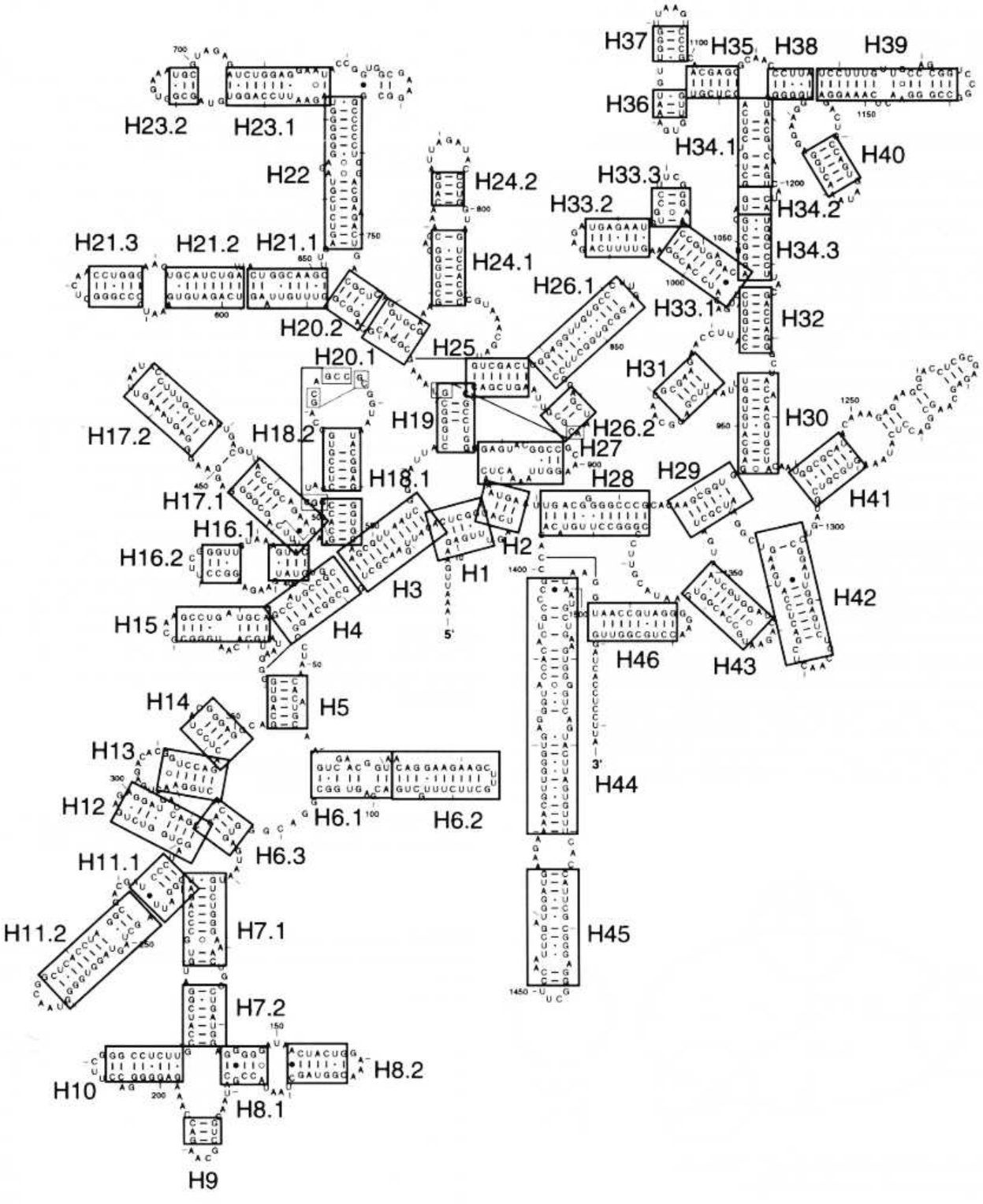

FIGURE 1. Secondary structure of the 16S rRNA, as predicted by comparative sequence analysis (Noller & Woese, 1981; Woese et al., 1983). Numbered boxes show our partitioning of helical elements, with numbering roughly following the system used by Brimacombe et al. (1991).

cluded from the calculation. In addition, protein S19 may be closer to proteins S3 and S7 than is predicted by the neutron diffraction map presented by Capel et al. (Brimacombe, 1995). Preliminary calculations showed that moving S19 36 Å (about 3 standard deviations in its uncertainty, according to the neutron diffraction data) toward S7 and S13 greatly augments the ability of helices protected by S19 to simultaneously satisfy constraints to other proteins and helices. These new coordi-

nates were therefore used for the location of S19 in the constraint satisfaction calculation. Otherwise, the positions of the centers of mass of the 30S proteins used were the same as reported by Capel et al. (1987).

Parametric representation of structural data

The hydroxyl radical footprinting data provide distance constraints between proteins and nucleotide positions within individual helices (see Table 1). However, the allowed range of distances corresponding to these data must be computed using principled approximations. Distance ranges between proteins and nucleotides were given a minimum value of zero and a maximum value calculated by summing the anhydrous radius of the protein (estimated by its molecular weight and multiplied by a morphology factor, μ, to account for possible nonspherical morphology) and the length of the directional error axis between the base and the protein's mean position. The distance along the directional error axis that was considered allowable is specified as the number of standard deviations (σ) from the protein's mean position. A graphical depiction of this parameterization of the distance associated with the hydroxyl footprinting data is shown in Figure 2. Hydroxyl radical footprinting data were reported in Powers and Noller (1995) as being either strong, medium, or weak. Because the weak data were dispersed more widely along the secondary structure and could represent tertiary structural protection at a distance, we used only the strong and medium interactions, and these were assigned equal weight in the calculation. Parameterization of the footprinting data was nearly uniform, with almost all of the constraints being assigned a value of 1.5 for σ and 1.0 for μ. In the single case for which these parameters were hand-optimized (helix 7.1), σ was set at 2.0, and μ was set at 1.5. The default values of 1.5 and 1.0 for σ and μ were determined to be acceptable through prior calculations (Altman et al., 1994).

The distance constraints between helices arise from the single-stranded RNA connecting them, and are referred to as *tether* constraints. These represent the range of distances allowed between the bases at either end of the single-stranded RNA connecting the helices. For a tether of length N bases, the minimum allowed value was set to the smallest distance observed in tRNA between phosphates separated by N bases. The maximum value was set to the length of a fully extended chain of RNA N bases long ($n \times 7.0$ Å) for $n = 1$–3, and a slightly less extended chain ($n \times 5.79$ Å) for $n > 3$ (see Table 2).

Taken together, the tether constraints (Table 3) and the protein protection distance constraints form a network of distances that connect the components of the 30S subunit (see Fig. 3, a depiction of this constraint network). The network forms the basis for our constraint satisfaction computation. Each node in the network is an object: the central node is the set of proteins in a single, fixed relative orientation. The other nodes are the RNA helices. It is apparent from Figure 3 that some helices are tightly constrained by many constraints to both proteins and RNA elements, whereas other helices are relatively poorly "tied in" to the constraint network.

Three helices (numbers 44, 45, and 46, comprising the 3′ end of the RNA) could not be positioned meaningfully. These are protected only by protein S20, which was not used in the calculation, and they are connected to the rest of the 16S rRNA by a 12-base-long tether, which does not sufficiently constrain their location. Thus, their positions were not modeled.

Calculation of structure by sampled constraint satisfaction

The method of constraint satisfaction has been described and validated in detail elsewhere (Brinkley et al., 1987, 1990;

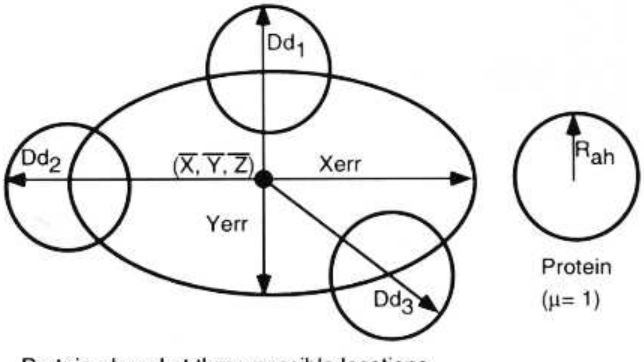

FIGURE 2. Schematic diagram of the parameterization of protein protection constraints and calculation of maximal directional distance (Dd) between a base and a protein, in three different directions. The Dd is the sum of the scaled directional error and a scaled anhydrous radius.

TABLE 2. Lengths of tether constraints used.[a]

Base separation	Dmin (Å)	Dmax (Å)
1	4.90	7.00
2	5.11	14.00
3	8.41	21.00
4	8.90	23.80
5	9.20	28.95
6	9.50	34.74
7	7.32	40.53
8	9.71	46.32
9	10.6	52.11
10	9.77	57.90
11	7.65	63.69
12	7.75	69.48
13	7.12	75.27
14	7.36	81.06
15	7.49	86.85
16	9.96	92.64

[a] Dmin = minimum observed in tRNA; Dmax = fully extended: ($P - P = n \times 7.0$ Å) for $n = 1$–3, ($P - P = n \times 5.7$ Å) for $n > 3$.

TABLE 3. Tether constraints between helices.[a]

From helix	To helix	Bases involved	From helix	To helix	Bases involved
H1	H2	13–17, 21–19	H20.2	H22	754–751
H1	H3	25–27	H21.1	H21.2	598–599, 642–639
H2	H27	916–912	H21.1	H22	651–655
H2	H28	918–923	H21.2	H21.3	606–612, 632–628
H3	H4	37–39	H22	H23.1	672–673
H3	H18.1	547–545	H23.1	H23.2	684–688, 706–699
H3	H19	556–564	H24.1	H24.2	775–783, 804–799
H4	H5	47–52	H24.1	H25	810–821
H4	H15	394–393	H25	H26.1	827–829
H4	H16.1	403–406	H25	H26.2	873–868
H5	H6.1	58–61	H26.1	H26.2	857–861
H5	H14	354–350	H28	H29	933–938
H5	H15	359–368	H28	H43	1384–1372
H6.1	H6.2	70–73, 98–97	H29	H30	943–946
H6.1	H6.3	106–113	H29	H42	1340–1334
H6.3	H7.1	115–122	H29	H43	1345–1350
H6.3	H12	312–311	H30	H31	955–960
H6.3	H13	314–316	H30	H32	1225–1221
H7.1	H7.2	132–136, 230–227	H30	H41	1235–1241
H7.1	H11.1	239–240	H31	H32	975–984
H7.2	H8.1	142–144	H32	H33.1	990–997
H7.2	H10	221–219	H32	H34.3	1215–1211
H8.1	H8.2	150–153, 172–168	H33.1	H33.2	1003–1006
H8.1	H9	178–184	H33.1	H33.3	1037–1036
H9	H10	193–198	H33.1	H34.3	1044–1046
H11.1	H11.2	245–247, 281–277	H33.2	H33.3	1023–1025
H11.1	H12	286–289	H34.1	H34.2	1058–1057, 1199–1203
H13	H14	337–339	H34.1	H35	1067–1068
H16.1	H16.2	409–416, 433–427	H34.1	H38	1189–1187
H16.1	H17.1	436–437	H34.2	H34.3	1055–1052, 1205–1206
H17.1	H17.2	446–455, 488–477	H35	H36	1073–1074
H17.1	H18.1	497–500	H35	H37	1102–1099
H18.1	H18.2	504–511, 541–540	H35	H38	1107–1113
H19	H20.1	569–576	H36	H37	1083–1086
H19	H25	881–879	H38	H39	1116–1118
H19	H27	886–888	H38	H40	1184–1175
H20.1	H20.2	580–584, 761–757	H39	H40	1155–1161
H20.1	H24.1	765–769	H41	H42	1296–1303
H20.2	H21.1	587–588			

[a] The identity of each helix involved in a single-stranded RNA tether is listed in the first two columns. The bases forming the tether are listed in the final column. These constraints were used to limit the distance between connected helices.

Lichtarge et al., 1987), but can be summarized here. At the beginning of a constraint satisfaction calculation, each helix can assume any location (comprising both position and orientation) in three-dimensional space. Helical locations are specified with six numbers: the x, y, and z positions of the helix center of mass, and the three Eulerian angles, ϕ, θ, and ω, (rotations around the initial object-based z, x, and z axes, respectively) describing the orientation of the helix. Our constraint satisfaction algorithm defines an initial set of locations that are compatible with some of the constraints (in this case, those to the proteins), and then systematically prunes out locations that are inconsistent with both the distance constraints between pairs of helices and distance constraints between proteins and helices. Conformations in which components overlap and thus violate van der Waals radii are also eliminated. The basic operations used in constraint satisfaction are (in the sequence they are applied): anchoring, appending, 2-way yoking, and N-way yoking, as described below.

Anchoring

Using the mean positions of the ribosomal proteins as a global coordinate system for the 30S subunit, an initial set of allowed locations is determined for each helix. In order to maintain a computationally tractable number of locations, we sample a grid with positional increments of 7 Å and angular increments of 30°. (This sample interval guarantees that the sampling error is roughly comparable to the radius of an RNA double helix.) At each grid point (or location), we check that all the constraints between the RNA helix and the proteins can be satisfied (that is, all the distances fall within the specified ranges). Locations that do not satisfy all the constraints

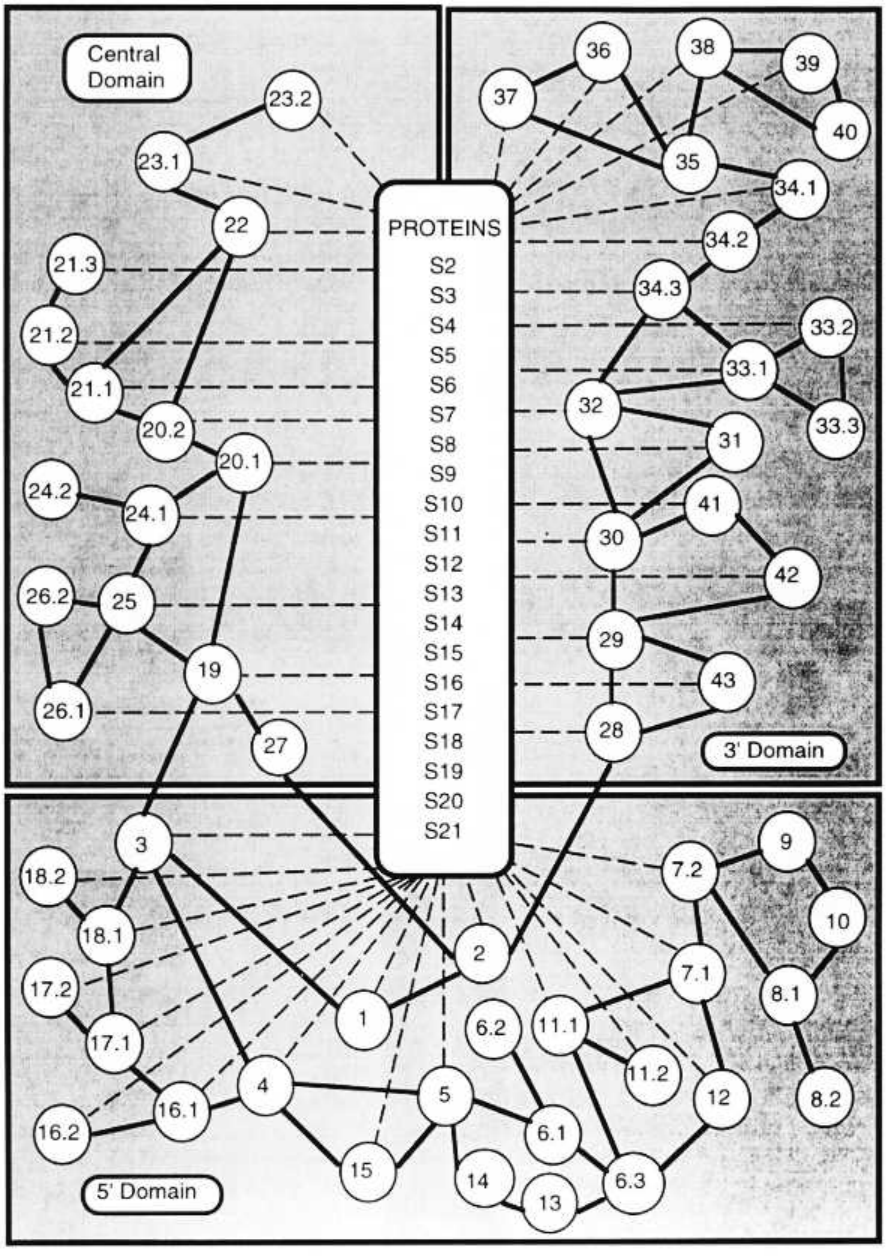

FIGURE 3. Constraint network for the helices in the 16S rRNA. Helices are represented as circled numbers. Solid lines represent tether constraints between helices, based on the connecting single-stranded regions. Dashed lines represent protection constraints between helices and proteins. The network of dependencies shown here influences the precision to which each helix position can be defined, as shown in Table 4.

are removed from the list of allowed locations. In some instances, constraints between a helix and the proteins are sufficiently constraining that a finer sampling level of 4 Å and 15° can be used. To graphically display the results of anchoring, we can render each allowed location as a set of dots in the protein coordinate system. When all such locations are drawn, a *location cloud* is formed that gives a qualitative impression of the allowed range of locations of the helix. We extended the PROTEAN system for this problem by allowing distance constraints to depend on the directional error of the protein center-of-mass position. For example, the uncertainty of protein S13 along the x-axis is 9.0 Å, whereas along the z-axis it is 17.7 Å. Thus, the allowed distance of a nucleotide that is protected by protein S13 (as illustrated in Fig. 2) is different in the x direction and the z direction.

Appending

Some helices contain no direct constraints to the proteins, and so they cannot be anchored to them as described above. These helices, however, do have constraints to other helices, and so they can be anchored to one of them (the secondary anchor) using the tether constraints between the two helices. The set of valid locations in the global coordinate system can then be generated for these helices as the cross-product of locations relative to the secondary anchor, and the locations of the secondary anchor relative to the global coordinate system.

2-way yoking

Given the initial clouds, as determined by anchoring and appending, the tether constraints between helices can be used

to further reduce the size of the location clouds. Consider the case of two clouds for helices 4 and 5, and a short tether constraint between them (bases 47–51). If there is some location of helix 4 for which there are no locations of helix 5 that allow the tether constraint to be satisfied, then that location of helix 4 can be removed, because it cannot possibly be part of a valid conformation. By systematically checking all pairs of helices in the constraint network, we can drastically reduce the size of the initial, anchored (or appended) clouds using this simple test. Helix locations are also checked to ensure that volume overlap does not occur. Once the clouds for helices 4 and 5 are 2-yoked, they may need to be re-yoked if one of them is further pruned by yoking with helix 15. Thus, the 2-yoking operation is performed iteratively until location clouds can be reduced no further.

N-way yoking

N-way yoking is a generalization of the 2-way yoke described above. In this case, however, helices are considered in groups of 3, 4, 5, etc. When *N* helices are considered together, then the location cloud for each helix is reduced to only those locations capable of participating in a group that simultaneously satisfies all distance and volume overlap (van der Waals) constraints. Ideally, we seek $N = 64$-way consistency in order to ensure that every location is part of at least one allowed conformation of the complete ensemble. However, the computational cost of computing 64-way consistency is prohibitive (it is related to the product of the size of the location clouds for each of the 64 helices). Fortunately, we have shown that, for problems featuring short-range distances within biological structures, the clouds computed above approximately 10-way consistency are very similar to one another (Altman et al., 1994). In this computation, we achieved 21-way consistency, just to ensure that the clouds were pruned maximally (see Fig. 4). Thus, the final set of helix locations is comprised of those that are able to participate in at least one coherent conformation (termed a *coherent instance*) that satisfies all constraints simultaneously. Three such coherent instances, each providing a slightly different position for the helices, but satisfying all the provided constraints, are shown in Figure 5. The ability to test high-order consistency efficiently is a second enhancement of the PROTEAN system created by this modeling effort.

Boundary sampling for efficiency

When manipulating the lists of helix locations, it becomes apparent that many locations are equivalent, differing in small translational or orientational increments. In general, we are most interested in the range of locations that are distinctly different. In order to make our computations more efficient, we use a method of sampling that reduces the number of locations in a cloud while preserving the cloud's volume and spatial characteristics (Chen et al., 1995). Reduction is accomplished by a clustering algorithm that chooses the subset of locations that is most representative of the cloud. We used this sampling procedure to limit cloud size (for individual helices) to 100,000 locations. The effect of sampling is to reduce the number of locations in a cloud significantly while reducing the precision with which the cloud volume can be measured only slightly.

Comparison with other models

The task of comparing models with nonatomic representations of structure is nontrivial because of the different modeling choices that can be made. Nonetheless, in order to assess the agreement between models, we examined the de-

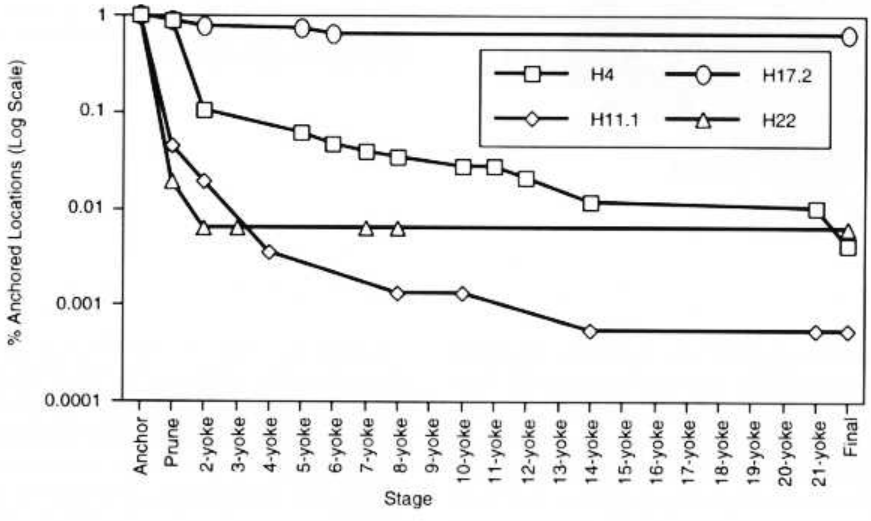

FIGURE 4. Progress of computations. Representative tracings of the number of locations (as a percent of the number of initial, anchored locations) are displayed for four specific helices. Helices with strong protein protection constraints, such as helix 22 (triangle), decreased radically in number of locations from the outset. Those with less strong protein protection constraints but strong tether constraints to other helices, such as helix 11.1 (diamond) and helix 4 (square), decreased steadily in number of locations throughout the calculation. Those with weak or few constraints, such as helix 17.2 (circle), remained high in percentage of anchored locations.

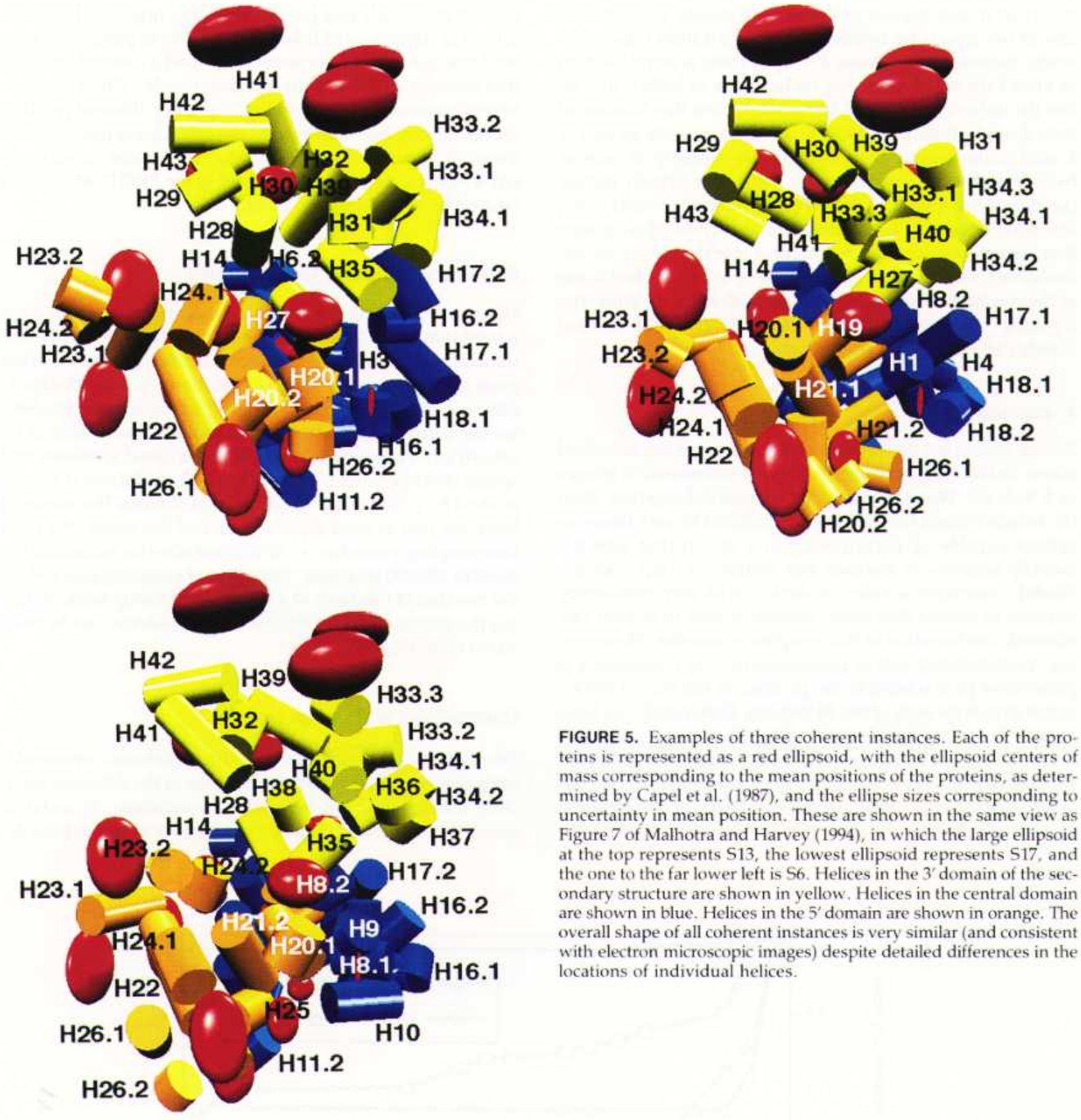

FIGURE 5. Examples of three coherent instances. Each of the proteins is represented as a red ellipsoid, with the ellipsoid centers of mass corresponding to the mean positions of the proteins, as determined by Capel et al. (1987), and the ellipse sizes corresponding to uncertainty in mean position. These are shown in the same view as Figure 7 of Malhotra and Harvey (1994), in which the large ellipsoid at the top represents S13, the lowest ellipsoid represents S17, and the one to the far lower left is S6. Helices in the 3′ domain of the secondary structure are shown in yellow. Helices in the central domain are shown in blue. Helices in the 5′ domain are shown in orange. The overall shape of all coherent instances is very similar (and consistent with electron microscopic images) despite detailed differences in the locations of individual helices.

gree to which the clouds reported here "contain" the helix positions reported in two models: a recent interactively built model based on the same data set (H.F. Noller, T. Powers, G.M. Heilek, S. Mian, & B. Weiser, in prep.), and the model of Malhotra and Harvey (1994). We superimposed each of these models using the common protein positions. For each helix, we computed its location (or locations if more than one was provided) in each of the comparison models. We then evaluated whether the location was contained by the corresponding cloud (within our sampling error). The results are shown in Table 4. For the model proposed by Brimacombe et al. (1988), the coordinates of the helix locations are provided in the text, but not the protein coordinates. Without the protein coordinates, we are only able to compute a simple RMS distance between the reported helix location and the center of our helix clouds. The computed distance is 35.2 Å, which is quite similar to the average distances observed between the other models. The model of Hubbard and Hearst (1991) is interesting because it did not use the protein position map, but relied solely on RNA-based constraints in the model building. The resulting structure is significantly different from the models proposed by Brimacombe, Noller, Malhotra and Harvey, and us (a computed RMS distance of 59 Å from our model, for example). The differences, how-

TABLE 4. Statistics for each helix on cloud sizes, number of locations, cloud centroid movement, and differences between cloud centroids and corresponding helix positions in two previous modeling efforts (Malhotra & Harvey, 1994; H.F. Noller, T. Powers, G.M. Heilek, S. Mian, & B. Weiser, in prep.).[a]

Helix	Initial # locations	Final volume (Å^3)	Final # locations	Final volume (Å^3)	Δ Center of mass (Å)	Cloud radius (Å)	Δ Malhotra model (Å)	Δ Noller model (Å)
H1	7,775	71,372	512	43,753	8.8	21.9	11.6	25.6
H2	10,903	303,660	593	54,465	22.9	23.5	8.4	34.8
H3	2,752	58,304	42	24,614	13.7	18.0	10.6, 13.2	19.0
H4	9,498	501,432	40	53,325	20.3	23.3	22.5, 30.1	6.6
H5	9,745	120,750	155	9,520	26.9	13.1	35.4, 37.1	N/M
H6.1	53,272	762,436	620	29,650	34.8	19.2	25.9, 29.6	N/M
H6.2	36,185	1,655,454	998	84,269	40.8	27.2	45.1, 48.9	N/M
H6.3	25,112	481,636	231	10,102	33.1	13.4	29.4	47.0
H7.1	154	25,796	2	12,639	7.2	14.5	29.5	N/M
H7.2	23,819	104,141	26	13,702	14.6	14.8	54.3	N/M
H8.1	39,932	596,390	588	20,107	21.9	16.9	46.6	N/M
H8.2	83,713	1,704,772	1,081	30,601	22.3	19.4	55.0	N/M
H9	159,545	2,379,541	2,105	35,638	23.3	20.4	70.5	N/M
H10	29,671	536,837	402	50,741	23.5	23.0	48.5	N/M
H11.1	34,654	157,043	19	10,833	38.0	13.7	27.7	4.9
H11.2	7,878	67,366	304	46,443	10.7	22.3	22.5, 40.3	17.6
H12	7,177	58,670	59	16,504	10.6	15.8	25.1, 27.2	27.4
H13	72,869	1,202,158	482	8,831	56.2	12.8	12.0	N/M
H14	60,876	772,973	925	18,128	51.3	16.3	30.4	N/M
H15	9,377	116,372	429	42,636	11.8	21.7	34.1, 35.0	18.2
H16.1	7,769	44,128	388	20,916	6.4	17.1	11.2	N/M
H16.2	8,550	64,792	4,677	56,984	1.3	23.9	26.8	N/M
H17.1	4,444	56,393	384	51,622	4.9	23.1	24.1	N/M
H17.2	9,573	441,934	6,075	41,5536	8.7	46.3	64.5	N/M
H18.1	3,976	24,012	322	17,010	1.8	16.0	23.3	8.9
H18.2	115	17,107	13	12,425	4.3	14.4	44.3	34.2
H19	10,935	329,113	166	48,712	25.3	22.7	23.1	13.0
H20.1	9,099	401,384	374	49,155	32.1	22.7	7.0	22.2
H20.2	7,460	50,864	133	14,744	10.0	15.2	32.2	17.2
H21.1	92	42,030	1	11,819	14.3	14.1	24.6	24.6
H21.2	315	38,230	5	24,163	7.2	17.9	14.9	19.8
H21.3	14,637	127,821	6,562	66,732	10.9	25.2	26.5	14.2
H22	157	44,452	1	19,070	8.0	16.6	35.5, 51.9, 62.7	16.9
H23.1	1,793	81,863	5	20,309	20.7	16.9	32.7	15.9
H23.2	10,050	168,682	1,463	44,366	13.5	22.0	12.3	39.5
H24.1	6,286	140,498	779	60,806	13.6	24.4	12.7	16.9
H24.2	6,212	198,478	3,002	124,051	3.1	30.9	19.7	17.9
H25	6,295	34,418	930	25,990	4.1	18.4	28.8	19.3
H26.1	14,626	780,829	132	138,151	40.8	32.1	27.8	21.5
H26.2	58,013	988,636	5,308	194,264	18.3	35.9	17.2	38.6
H27	45,097	814,901	1,199	71,782	25.7	25.8	46.3	9.7
H28	34,977	1,405,004	1,275	148,036	28.1	32.8	33.1, 37.8	31.1
H29	15,263	82,046	2,538	61,660	1.8	24.5	31.5	16.3
H30	9,619	82,907	807	58,466	8.5	24.1	30.3	14.5
H31	6,588	52,874	1,382	24,405	9.9	18.0	39.6	45.7
H32	10,194	60,666	1,217	31,911	8.5	19.7	10.8	19.3
H33.1	23,169	620,157	2,908	65,631	37.1	25.0	29.2	N/M
H33.2	9,283	417,262	1,523	198,282	12.8	36.2	47.9	N/M
H33.3	31,930	511,121	3,738	155,824	7.8	33.4	39.5	N/M
H34.1	4,109	43,565	69	34,768	7.8	20.2	29.9, 36.1	14.1
H34.2	5,634	38,405	614	25,397	4.7	18.2	41.2	29.3
H34.3	178	21,134	27	19,199	4.6	16.6	30.7, 32.9	26.1
H35	9,404	103,450	32	18,289	24.1	16.3	25.5	40.1
H36	5,653	38,043	176	18,582	17	16.4	12.4	40.4
H37	3,010	23,687	301	16,279	9.4	15.7	25.0	19.2
H38	2,032	33,388	1,231	30,176	3.2	19.3	32.1	13.1
H39	166	48,059	79	44,738	1.8	22.0	9.9, 14.5, 30.1	11.4
H40	44,017	721,221	29,396	748,835	10.2	56.3	29.6	20.6
H41	3,218	65,068	1,182	55,160	1.2	23.6	28.9, 38.4, 39.2	22.0
H42	135	33,186	3	18,942	7.3	16.5	40.0	20.7
H43	5,180	38,321	4,099	37,823	1.5	20.8	32.0	12.7
Mean	18,593	344,381	1,543	64,221	16.0	21.7	30.8	22.0

[a] The volume of a cloud was determined by calculating the spatial distribution of helix phosphate groups, across all possible helix locations, at 2 standard deviations. The radius of a cloud was determined by assuming that the cloud was spherical. In comparing helix positions between structures, multiple error distances indicate that our helix was modeled as several separate helices in the other structure, and the errors are for each helix component in the other model. N/M indicates that the helix was not modeled in the other structure.

TABLE 5. Sites of interaction between the 16S rRNA and various cofactors and antibiotics.

Cofactor	Function	Helices involved	Ref[a]	Fig. 6 color
A-site tRNA	Continue translation	H18.2, H44	1	Rose
P-site tRNA	Begin translation	H23.2, H24.2, H28, H31, H44	1	Olive
mRNA	Specify aa sequence	H16.2, H18.2, H23.2, H26.1, H39, H42, H28	2	N/S[b]
50S subunit	Part of ribosome	H24.2, H46	1	Olive
IF-1	Initiation factor	H18.2, H44	1	Rose
IF-3	Initiation factor	H23.2, H24.2,	1	Olive
Tetracycline	A-site interference	H27	3	Burnt sienna
Streptomycin	Initiation inhibition, misreading	H2, H27	3	Burnt sienna
Neomycin	A-site interference	H44	3	N/S
Edeine	P-site interference	H23.2, H24.2, H28	3	Olive
Hygromycin	Misreading, translocation inhibition	H44	3	N/S
Spectinomycin	Translocation inhibition	H34.1	3	Lavender

[a] References: 1, Noller (1991); 2, Wollenzien et al. (1991); 3, Moazed and Noller (1987).
[b] N/S, not shown.

ever, are not homogeneous, and it appears that the 5′ domain of this model (as noted previously by Malhotra & Harvey, 1994) is generally consistent with the other models.

Mapping important functional sites

With our helical positions and uncertainties mapped, we are able to examine the location of bases that are known to interact with important functional ligands. Using reported data on the positions of tRNA binding (see Table 5), the A-site (site of aminoacyl-tRNA interaction) and P-site (site of nascent peptidyl-tRNA interaction), as well as data on the binding sites of several antibiotics, we evaluated the degree of structural precision with which these sites can be defined. The results are shown in Figure 6. Additionally, we used the P-site tRNA footprinting data (Moazed et al., 1986; Moazed & Noller, 1990) to determine the range of positions and orientations that P-site tRNA could adopt in our model. In completing this step, we had to extend the allowed distance from the tRNA anticodon stem loop to helix 31 from 15 Å to 30 Å in order to satisfy this constraint while simultaneously satisfying constraints to other ribosomal helices. The results of docking P-site tRNA are shown in Figure 7.

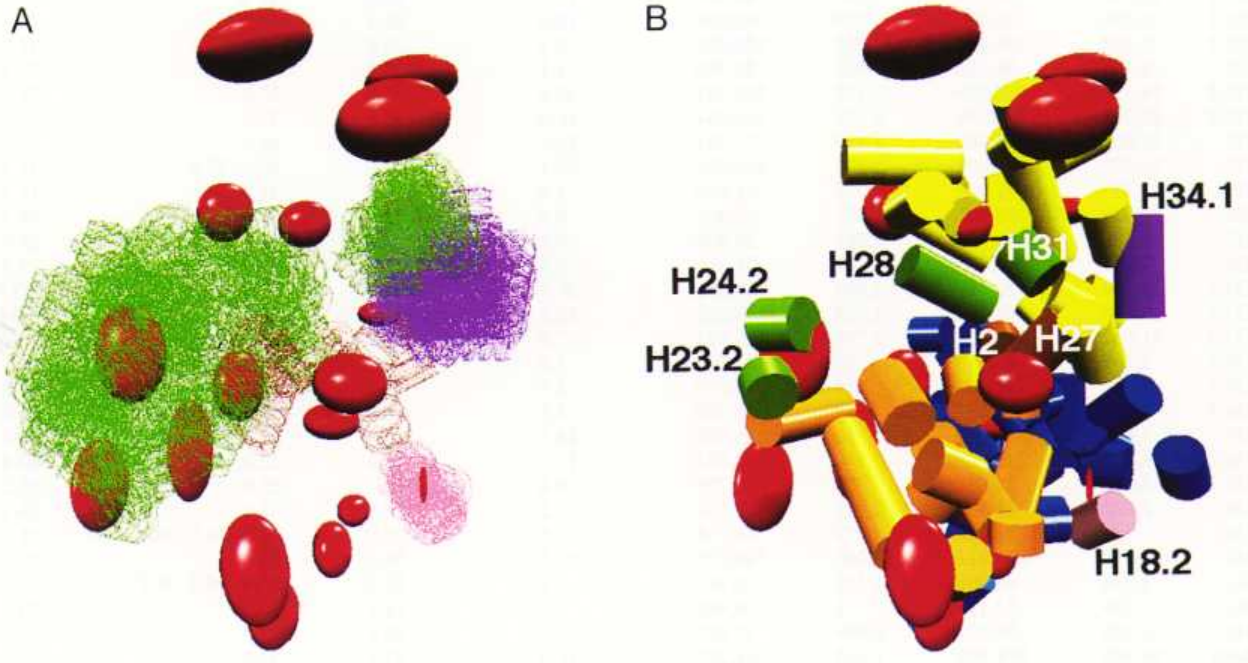

FIGURE 6. Identification of functional sites on our model of folded 16S rRNA. **A:** Spatial extent of the allowed location clouds for helices involved in various functional sites. Rose, A site tRNA; olive green, P-site tRNA, 50S subunit, edeine, and IF-3; burnt sienna, tetracycline and streptomycin; lavender, spectinomycin. **B:** Specific locations of these helices (using the same color system) in a coherent instance. Helices not involved in functional sites are colored according to the domain-specific scheme presented in Figure 5.

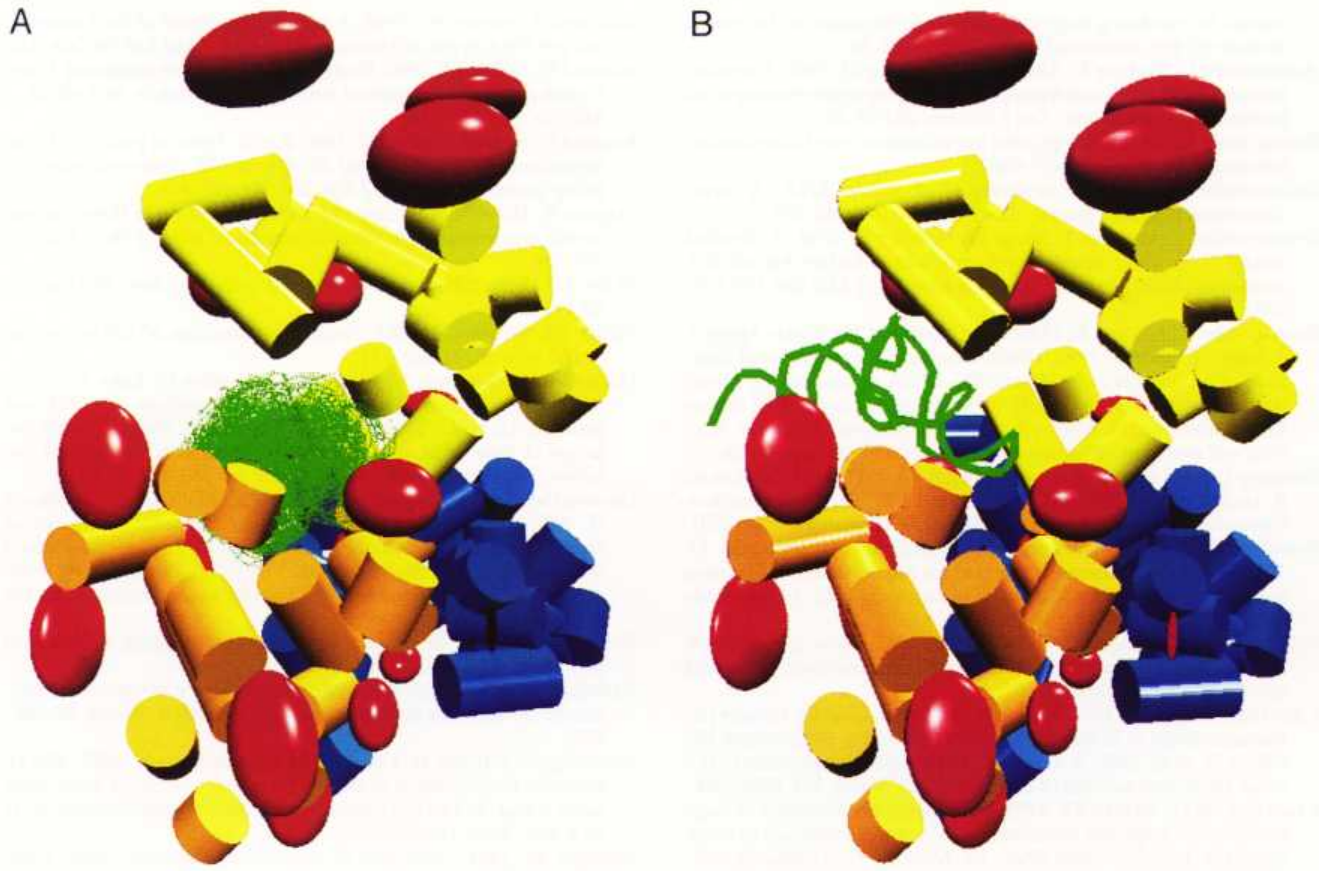

FIGURE 7. Docking P-site tRNA to a coherent instance of 16S rRNA. **A:** Cloud of allowed locations for the anticodon stem loop in green. **B:** One particular allowed position and orientation of the tRNA, based on the cloud of allowed locations for the anticodon stem loop and other distance constraints, as discussed in the text.

Reporting the coordinates

The primary product of our computation is a list of allowed locations (translations and orientations) for each helix. In order to facilitate comparison of our model with others, we have made available on an internet WWW site (http://www-smi.stanford.edu/projects/helix/pubs/finketal/) four summaries of our results: (1) the average center of mass location for each helix, along with its three-dimensional uncertainty (variance in each of three Cartesian directions); (2) the average position of all the modeled phosphates, along with their three-dimensional uncertainty; (3) the phosphate positions for the three instances shown in Figure 5; and (4) the protein positions used (or computed, in the case of S20).

It is important to note that the phosphate average positions reported do not satisfy strict stereochemistry and van der Waals packing constraints, because the models were built with a simplified representation whose precision does not allow detailed checks on these constraints (although rough checks for packing violations are included in the method). Because the precision of our resulting models are no better than 15 Å, there is little need to strictly enforce constraints that are relevant only at much finer levels of detail. The average uncertainty for each structural component is sufficient to allow the refinements necessary, as data becomes available, to ensure good packing and stereochemistry.

ACKNOWLEDGMENTS

R.B.A. is a Culpeper Medical Scholar, and is funded by NIH grant #LM-05652 and #LM-06442. H.F.N. is funded by NIH grant GM-17129, and by a grant to the Center for Molecular Biology of RNA from the Lucille P. Markey Charitable Trust. Computing resources are provided partially by the CAMIS resource, NIH grant #LM-05305. We thank Bryn Weiser for his contributions to modeling studies conducted at UCSC and A. Malhotra and J. Hubbard for their model coordinates. Coordinates for our helical clouds are available at http://www-smi.stanford.edu/projects/helix/pubs/finketal/. The PROTEAN program is available through the QCPE at file://qcpe6.chem.indiana.edu/, or by request from the authors. The input files to PROTEAN for the 16S structure can also be provided.

Received May 13, 1996; returned for revision May 28, 1996; revised manuscript received June 17, 1996

REFERENCES

Agrawal RK, Penczek P, Grassucci RA, Li Y, Leith A, Nierhaus KH, Frank J. 1996. Direct visualization of A-, P-, and E-site transfer RNAs in the *Escherichia coli* ribosome. *Science 271*:1000–1002.

Altman RB, Weiser B, Noller HF. 1994. Constraint satisfaction tech-

niques for modeling large complexes: Application to the central domain of 16S ribosomal RNA. *ISMB 2*:10–18.
Arrowsmith C, Pachter R, Altman R, Jardetzky O. 1991. The solution structures of *E. coli* trp repressor and trp aporepressor at an intermediate resolution. *Eur J Biochem 202*:53–66.
Brimacombe R. 1991. RNA–protein interactions in the *Escherichia coli* ribosome. *Biochimie 73*:927–936.
Brimacombe R. 1995. The structure of ribosomal RNA: A three-dimensional jigsaw puzzle. *Eur J Biochem 230*:365–833.
Brimacombe R, Atmadja J, Stiege W, Schuler D. 1988. A detailed model of the three-dimensional structure of *Escherichia coli* 16 S ribosomal RNA in situ in the 30 S subunit. *J Mol Biol 199*:115–136.
Brimacombe R, Greuer B, Mitchell P, Osswald M, Rinke-Appel J, Schüler D, Stade K. 1990. Three-dimensional structure and function of *Escherichia coli* 16S and 23S rRNA as studied by cross-linking techniques. In: Hill WE, Dahlberg A, Garrett RA, Moore PB, Schlessinger D, Warner JR, eds. *The ribosome: Structure, function, and evolution*. Washington, DC: ASM Press. pp 93–106.
Brinkley J, Cornelius C, Altman R, Brugge J, Pachter R, Buchanan B, Jardetzky O. 1990. PROTEAN/I and PROTEAN/II. *Quantum Chemistry Program Exchange Newsletter X*:596. (Catalog section VIII).
Brinkley JF, Altman RB, Duncan BS, Buchanan BG, Jardetzky O. 1987. Heuristic refinement method for the derivation of protein solution structures: Validation on cytochrome *b*562. *J Chem Information Comput Sci 28*:194–210.
Brosius J, Dull TJ, Sleeter DD, Noller HF. 1981. Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli*. *J Mol Biol 148*:107–127.
Capel MS, Engelman DM, Freeborn BR, Kjeldgaard M, Langer JA, Ramakrishnan V, Schindler DG, Schneider DK, Schoenborn BP, Sillers IY, et al. 1987. A complete mapping of the proteins in the small ribosomal subunit of *Escherichia coli*. *Science 238*:1403–1406.
Chen R, Fink D, Altman RB. 1995. Computing the structure of large complexes: Applying constraint satisfaction techniques to modeling the 16S ribosomal RNA. In: Markley JL, Opella SJ, eds. *Biomolecular NMR spectroscopy*. Oxford: Oxford University Press. pp 279–299.
Expert-Bezancon A, Wollenzien PL. 1985. Three-dimensional arrangement of the *Escherichia coli* 16s ribosomal RNA. *J Mol Biol 184*:53–66.
Frank J, Zhu J, Penczek P, Li Y, Srivastava S, Verschoor A, Radermacher M, Grassucci R, Lata RK, Agrawal RK. 1995. A model of protein synthesis based on cryo-electron microscopy of the *E. coli* ribosome [see comments]. *Nature 376*:441–444.
Golden BL, Hoffman DW, Ramakrishnan V, White SW. 1993. Ribosomal protein S17: Characterization of the three-dimensional structure by 1H and 15N NMR. *Biochemistry 32*:12812–12820.
Greuer B, Osswald M, Brimacombe R, Stöffler G. 1987. RNA–protein cross-linking in *Escherichia coli* ribosomal subunits; determination of sites on 16S RNA that are cross-linked to proteins S3, S4, S7, S9, S10, S11, S17, S18 and S21 by treatment with *bis*-(2-chloroethyl)-methylamine. *Nucleic Acids Res 15*:3241–3255.
Hubbard JM, Hearst JE. 1991. Computer modeling 16 S ribosomal RNA. *J Mol Biol 221*:889–907.
Latham JA, Cech TR. 1989. Defining the inside and outside of a catalytic RNA molecule. *Science 245*:276–282.
Lichtarge O, Cornelius C, Buchanan B, Jardetzky O. 1987. Validation of the first step of the heuristic refinement method for the derivation of solution structures of proteins from NMR data. *Proteins Struct Funct Genet 2*:340–358.
Liljas A, Garber M. 1995. Ribosomal proteins and elongation factors. *Curr Opin Struct Biol 5*:721–727.
Lindahl M, Svensson LA, Liljas A, Sedelnikova SE, Eliseikina IA, Fomenkova NP, Nevskaya N, Nikonov SV, Garber MB, Muranova TA, et al. 1994. Crystal structure of the ribosomal protein S6 from *Thermus thermophilus*. *EMBO J 13*:1249–1254.
Malhotra A, Harvey SC. 1994. A quantitative model of the *Escherichia coli* 16S RNA in the 30S ribosomal subunit. *J Mol Biol 240*:308–340.
Moazed D, Noller HF. 1990. Binding of tRNA to the ribosomal A and P sites protects two distinct sets of nucleotides in 16 S rRNA. *J Mol Biol 211*:135–145.
Moazed D, Stern S, Noller HF. 1986. Rapid chemical probing of conformation in 16S ribosomal RNA and 30S ribosomal subunits using primer extension. *J Mol Biol 187*:399–416.
Nagano K, Harel M, Takezawa M. 1988. Prediction of three-dimensional structure of *Escherichia coli* ribosomal RNA. *J Theor Biol 134*:199–256.
Noller HF. 1991. Ribosomal RNA and translation. *Annu Rev Biochem 60*:191–227.
Noller HF, Woese CR. 1981. Secondary structure of 16S ribosomal RNA. *Science 212*:403–411.
Oakes MI, Scheinman A, Atha T, Shankweiler G, Lake JA. 1990. Ribosome structure: Three-dimensional locations of rRNA and proteins. In: Hill WE, Dahlberg A, Garrett RA, Moore PB, Schlessinger D, Warner JR, eds. *The ribosome: Structure, function and evolution*. Washington, DC: ASM Press. pp 180–193.
Osswald M, Greuer B, Brimacombe R, Stöffler G, Bäumert H, Fasold H. 1987. RNA–protein cross-linking in *Escherichia coli* ribosomal subunits; determination of sites on 16S RNA that are cross-linked to proteins S3, S4, S7, S9, S8, S9, S11, S13, S19 and S21 by treatment with methyl *p*-azidophenyl acetimidate. *Nucleic Acids Res 15*:3221–3240.
Powers T, Noller H. 1995. Hydroxyl radical footprinting of ribosomal proteins on 16S rRNA. *RNA 1*:194–209.
Ramakrishnan V, White SW. 1992. The structure of ribosomal protein S5 reveals sites of interaction with 16S rRNA. *Nature 358*:768–771.
Rinke-Appel J, Jünke N, Osswald M, Brimacombe R. 1995. The ribosomal environment of tRNA: Crosslinks to rRNA from positions 8 and 20:1 in the central fold of tRNA located at the A, P, or E site. *RNA 1*:1018–1028.
Saenger W. 1984. *Principles of nucleic acid structure*. New York: Springer-Verlag.
Schwedler G, Albrecht-Ehrlich R, Rak KH. 1993. Immunoelectron microscopic localization of ribosomal proteins BS8, BS9, BS20, BL3 and BL21 on the surface of 30S and 50S subunits from *Bacillus stearothermophilus*. *Eur J Biochem 217*:361–369.
Stark H, Mueller F, Orlova EV, Schatz M, Dube P, Erdemir T, Zemlin F, Brimacombe R, van Heel M. 1995. The 70S *Escherichia coli* ribosome at 23 Å resolution: Fitting the ribosomal RNA. *Structure 3*:815–821.
Stern S, Powers T, Changchien LM, Noller HF. 1989. RNA–protein interactions in 30S ribosomal subunits: Folding and function of 16S rRNA. *Science 244*:783–790.
Stern S, Weiser B, Noller HF. 1988. Model for the three-dimensional folding of 16 S ribosomal RNA. *J Mol Biol 204*:447–481.
Stiege W, Atmadja J, Zobawa M, Brimacombe R. 1986. Investigation of the tertiary folding of *Escherichia coli* ribosomal RNA by intra-RNA cross-linking in vivo. *J Mol Biol 191*:135–138.
Verschoor A, Frank J, Radermacher M, Wagenknecht T, Boublik M. 1984. Three-dimensional reconstruction of the 30 S ribosomal subunit from randomly oriented particles. *J Mol Biol 178*:677–698.
vonAhsen U, Noller HF. 1995. Identification of bases in 6S rRNA essential for tRNA binding at the 30S ribosomal P site. *Science 267*:234–237.
Woese CR, Gutell R, Gupta R, Noller HF. 1983. Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol Rev 47*:621–669.
Yonath A, Bennett W, Weinstein S, Wittmann HG. 1990. Crystallography and image reconstruction of ribosomes. In: Hill WE, Dahlberg A, Garrett RA, Moore PB, Schlessinger D, Warner JR, eds. *The ribosome: Structure, function and evolution*. Washington, DC: ASM Press. pp 134–147.