RNA (2002), *8*:279–289. Cambridge University Press. Printed in the USA.
Copyright © 2002 RNA Society.
DOI: 10.1017.S135583820202407X

BIOINFORMATICS

Mining biochemical information: Lessons taught by the ribosome

MICHELLE WHIRL-CARRILLO, IRENE S. GABASHVILI, MICHAEL BADA, D. REY BANATAO, and RUSS B. ALTMAN

Stanford Medical Informatics, Stanford University, Stanford, California 94305-5479, USA

ABSTRACT

The publication of the crystal structures of the ribosome offers an opportunity to retrospectively evaluate the information content of hundreds of qualitative biochemical and biophysical studies of these structures. We assessed the correspondence between more than 2,500 experimental proximity measurements and the distances observed in the ribosomal crystals. Although detailed experimental procedures and protocols are unique in almost each analyzed paper, the data can be grouped into subsets with similar patterns and analyzed in an integrative fashion. We found that, for crosslinking, footprinting, and cleavage data, the corresponding distances observed in crystal structures generally did not exceed the maximum values expected (from the estimated length of the agent and maximal anticipated deviations from the conformations found in crystals). However, the distribution of distances had heavier tails than those typically assumed when building three-dimensional models, and the fraction of incompatible distances was greater than expected. Some of these incompatibilities can be attributed to the experimental methods used. In addition, the accuracy of these procedures appears to be sensitive to the different reactivities, flexibilities, and interactions among the components. These findings demonstrate the necessity of a very careful analysis of data used for structural modeling and consideration of all possible parameters that could potentially influence the quality of measurements. We conclude that experimental proximity measurements can provide useful distance information for structural modeling, but with a broad distribution of inferred distance ranges. We also conclude that development of automated modeling approaches would benefit from better annotations of experimental data for detection and interpretation of their significance.

Keywords: bacterial ribosome; cleavage; crosslinking; crystal structure; footprinting; proximity measures; ribosomal proteins; ribosomal RNA

INTRODUCTION

The ribosome is responsible for translating mRNA into protein, and its structure has been the subject of intense study for more than 30 years. The bacterial 70S ribosome (with its two subunits—30S and 50S) measures approximately 210 Å in each direction and is composed of more than 50 polypeptides and three RNA molecules. The recent advances in the X-ray crystallography of the ribosome (e.g., Ban et al., 2000; Schluenzen et al., 2000; Wimberly et al., 2000; Yusupov et al., 2001) provide the first detailed look at a very large macromolecular complex and promise to increase our understanding of the mechanism of translation.

Reprint requests to: Russ B. Altman, Stanford Medical Informatics, Stanford University, 251 Campus Drive, MSOB x-215, Stanford, California 94305-5479, USA; e-mail: altman@smi.stanford.edu.

Before the crystal structures were published, a number of groups mapped experimental proximity measures to specific distance ranges and built ribosomal models. With the crystal structures, we have an unprecedented opportunity to review these qualitative proximity measures, which were determined by a variety of biochemical and biophysical experiments and used in the construction of models before the availability of crystal structures. RiboWeb is an online resource that draws together such experimental observations relevant to the structure of the 70S bacterial ribosome (Chen et al., 1997; Bada & Altman, 2000). In total, it contains about 10,000 experimental observations from biochemical, biophysical, and phylogenetic studies (http://riboweb.stanford.edu/riboweb/login-frozen.html).

The experimental techniques used in the ribosomal studies remain important sources of information on molecular neighborhoods. Information regarding the rela-

tionships between different types of data may be useful for future experimental designs and structural modeling approaches. It may help to devise better structured data repositories for researchers working on three-dimensional structures and biomolecular interactions in the future.

INFORMATION AVAILABLE FOR MODELING AND COMPARISONS

For comparisons with interatomic distances in crystal structures, we selected results of crosslinking, footprinting, and cleavage experiments, representing important sources of proximity information for building three-dimensional models. We selected more than 2,600 observations of crosslinking, footprinting, and cleavage experiments from 84 articles in the RiboWeb knowledge base to examine (see the Appendix), including only proximities that could be compared with those derived from the available crystal structures (Protein Data Bank identifications listed in Materials and Methods). We analyzed the data based on general and specific similarities between the experiments, such as individual steps in procedures, objects of study, probing agents, and methods, taking into account possible structural deviations between the different species studied and conformational states.

Altogether, the observations considered involved 19 small and 20 large ribosomal subunit proteins. Counting only two bases of rRNA for each crosslink (those closest to each other in the reported stretches) and summing this information with nucleotides determining the footprinting and cleavage patterns, the observations processed in this work dealt with more than 650 of the 1,542 bases in *Escherichia coli* 16S rRNA, more than 1,000 of the 2,917 bases in the 23S rRNA and 24 from the 120 bases of the 5S rRNA. All of these nucleotides were relatively evenly distributed in the ribosomal RNAs. Some of the bases were involved in up to 10 (large subunit) or even more than 30 (small subunit) independent experiments. However, not all of this information was equally valuable.

We found that the qualitative observations frequently indicate close distances (usually less that 20–30 Å); however, the distribution of distances has a long tail that stretches out to over 100 Å. Among crosslinking agents, we found no correlation between the molecular size of the agent and the mean distance between residues that were crosslinked by them. We did, however, see some correlation between the type of experimental analysis used and the mean distance for crosslinks. Some footprinting agents have tighter distance ranges than others, but there is no clear correlation to the size of the agent. There are, however, differing distance ranges depending on the subunit involved. Cleavage-based distances can extend to over 60 Å and vary by protein or rRNA fragment with the attached probe.

CROSSLINKS

We define crosslinking experiments as those experiments in which two or more entities within the molecular ensemble become covalently linked through their reaction with chemical and/or radiative agents. The results are usually interpreted to indicate proximity of the joined entities. This common type of experiment is performed using a wide variety of agents and methods of identification of the participating elements, which either are not known a priori, or are targeted with the preliminarily incorporated agents. Methods of a posteriori identification include decomposition of the structure by specific enzymes, separation and purification of crosslinked from nonlinked components, and various approaches to determine exact nucleotides and amino acids or their stretches involved (Sergiev et al., 2001).

We examined instances of crosslinking within and between the 16S, 23S, and 5S rRNA molecules, crosslinks between proteins and rRNA, and crosslinks between proteins. We evaluated 411 crosslinking observations taken from 59 articles, computing the span of each crosslink as the distance in the crystal structure between (1) the phosphorus atoms of the linked rRNA nucleotides, (2) the phosphorus atom of the rRNA nucleotide to the alpha carbon of the closest amino acid in the linked protein, or (3) the alpha-carbons of the amino acids closest to each other in the linked proteins. We note that various chemical agents and UV irradiation bridge atomic groups that might be located as far as 5 to 15 Å from the closest phosphorus and alpha-carbon atoms. However, such precision is of minimal use in the modeling of low-resolution structural data, especially if no information is provided regarding the exact nucleotides or amino acids linked together.

About half of all rRNA crosslink distances are less than 20Å, and the remainder are generally less than 40 Å. However, there are some outliers as high as 130 Å and higher. rRNA/rRNA and rRNA/protein distances have modes at 13 Å and 4 Å, respectively. It is difficult, however, to draw conclusions based on these observations because of the uncertainty in the experimental results, as the closest amino acid or a nucleotide in a stretch was taken to calculate the distance.

Sergiev et al. (2001) thoroughly analyzed experimental procedures used for crosslinking and pointed to the steps possibly responsible for errors. Our data (Fig. 1) agree with their analysis, which was performed for a smaller representative data set describing only the rRNA/rRNA interactions. As noticed by these authors, both small and large subunit inter-rRNA crosslinks produce very large distances when reverse transcription is involved in RNA identification (data displayed with the “primer extension” label in Fig. 1). We determined an “acceptable distance range” for each data point using the length of the agent, an estimated distance of reacting atomic groups from the phosphorus or alpha-carbon

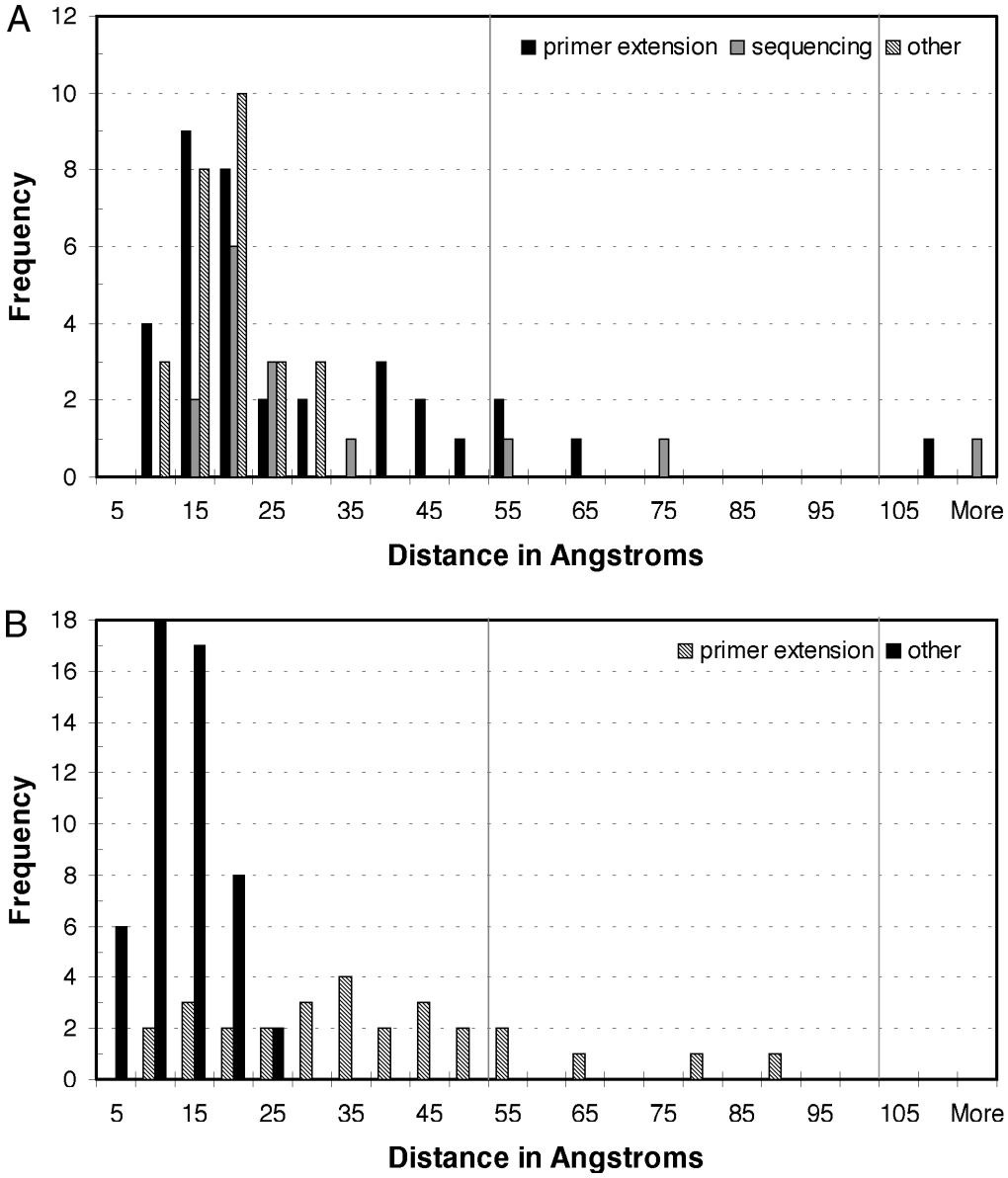

FIGURE 1. Histograms of rRNA/rRNA crosslinking distances, plotted based on method of analysis of experimental results. **A:** 30S subunit crosslinks. **B:** 50S subunit crosslinks. The crosslinks analyzed with primer analysis generate the largest distances and appear to be the least reliable method.

atom of the corresponding nucleotide or amino acid, and the maximal conformational variations expected. A distance greater than this range was termed a "gross mismatch." We found that earlier crosslinking results interpreted on the basis of direct sequencing of the involved fragments had a greater probability of being a gross mismatch, although most of the results were in the acceptable range (Fig. 1A).

Distances calculated from RNA/protein experiments displayed a more complex pattern (Fig. 2). The 50S ribosomal subunit distances ranged to 45 Å, whereas the distances from the 30S subunit ranged up to 133 Å. This dichotomy is most probably due to the different experimental approaches used to probe the two subunits. The best methods used in the studies of both subunits were oligonucleotide analysis and matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS). The latter yielded only one result that did not match the *Thermus thermophulis* structure, although it was comparable with the distance found in *Haloarcula marismortui*. This could be due to the current resolution for the 50S subunit of the thermophilic ribosome (Yusupov et al., 2001), which is not high enough to visualize individual amino acids. Two measurements detected by oligonucleotide analysis corresponded to distances that were larger than expected. These could be explained by the greater flexibility of the proteins involved in the crosslinks—L9 (Lillemoen

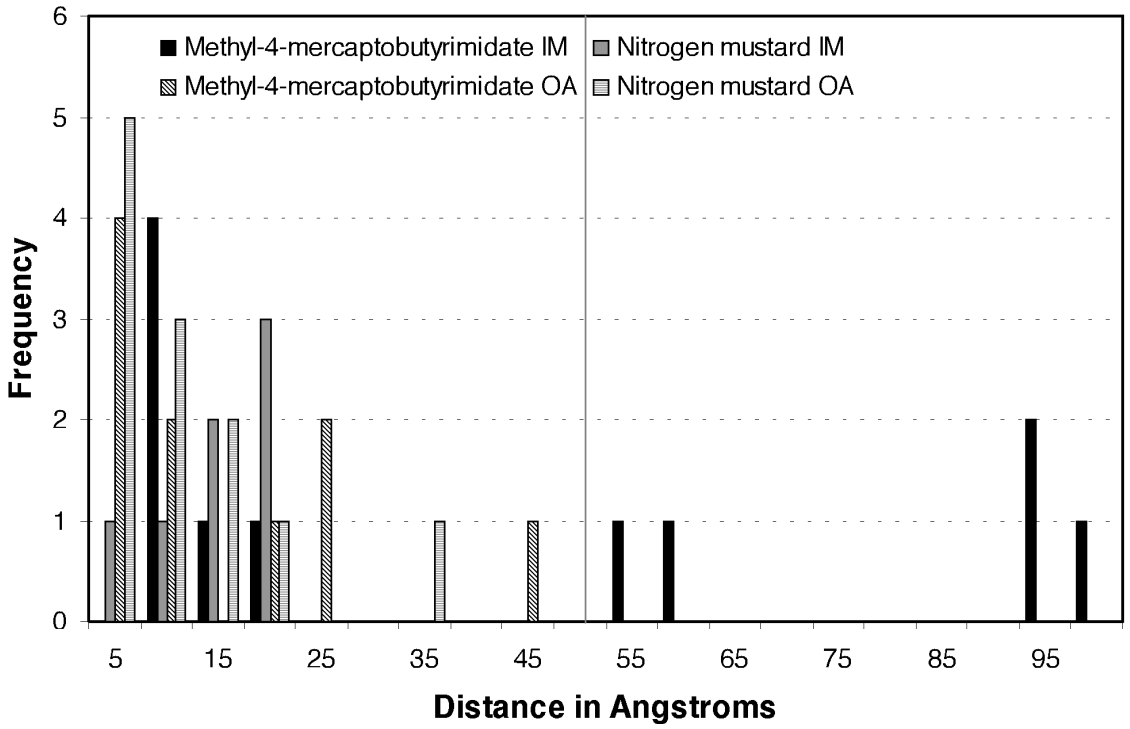

FIGURE 2. Histogram of rRNA/protein crosslink distances, plotted based on both agent and method of analysis. Methods of analysis are abbreviated OA for oligonucleotide analysis and IM for immunological method and fingerprinting. Agents performed differently based on method of analysis, but nitrogen mustard generated smaller distances for each method as compared to methyl-4-mercaptobutyrimidate.

& Hoffman, 1998) and S7. In the case of S7, the agent used for site-specific protein labeling had a greater length, and so it is not surprising that the experiment would have lower resolution.

Protein/protein crosslinks displayed another pattern. Classification of these results based on the exact experimental procedures did not allow separation into more and less reliable approaches. Immunological techniques are often supposed to be more accurate than gel electrophoresis-based identification (Stoffler et al., 1988), but we observed practically the same probability of gross mismatch in the experiments.

However, classification of the protein/protein crosslinking data based on the object studied displayed very strong distinctions between the data sets (Fig. 3). To be more precise in this assessment, we also plotted those crosslinking experiments performed with a common agent in a similar fashion (Fig. 3, inset). The 30S subunit protein experiments yielded more reliable results, as the probability of gross mismatch was only about 15% with distances up to 88 Å. Comparatively, the gross mismatch was 50% for the large subunit (even though some of the experiments were performed with immunoaffinity chromatography, as opposed to electrophoretic analysis) and 70% for the intersubunit protein contacts, both groups having distances outliers over 125 Å. We conclude that the different nature of the large subunit proteins (Moore, 1971) and the nature of bridges between the subunits were the probable causes. A dependence of the results on the details of the object under study was also observed in the cleavage experiments, as described below.

It is a common assumption in modeling that the crosslink distance and the size of the crosslinking agent are generally linearly associated. However, we did not see this correlation. This may be partially attributed to possible experimental errors we discussed previously and uncertainties in the exact atomic groups that are linked together. It can be concluded, however, that length of the agent probably should not be treated as a significant parameter in low-resolution structural modeling. This observation is in accordance with recent results of molecular dynamics experiments (Green et al., 2001) showing that longer agents might become compact in some environments, resulting in shorter distances between the reactive groups.

Although the agents seemed to play a less important role as compared to the methods of analysis, we noted that nitrogen mustard and UV were among the best agents in all types of crosslinking experiments, that is, they yielded consistently small distances. Sergiev et al. (2001) noted this in the analysis of RNA/RNA interactions. Other agents performed well, though there were very few data points available for their analysis, making generalization difficult. Among these other agents were: tatryl di(glycylazide) (estimated length: 13 Å; num-

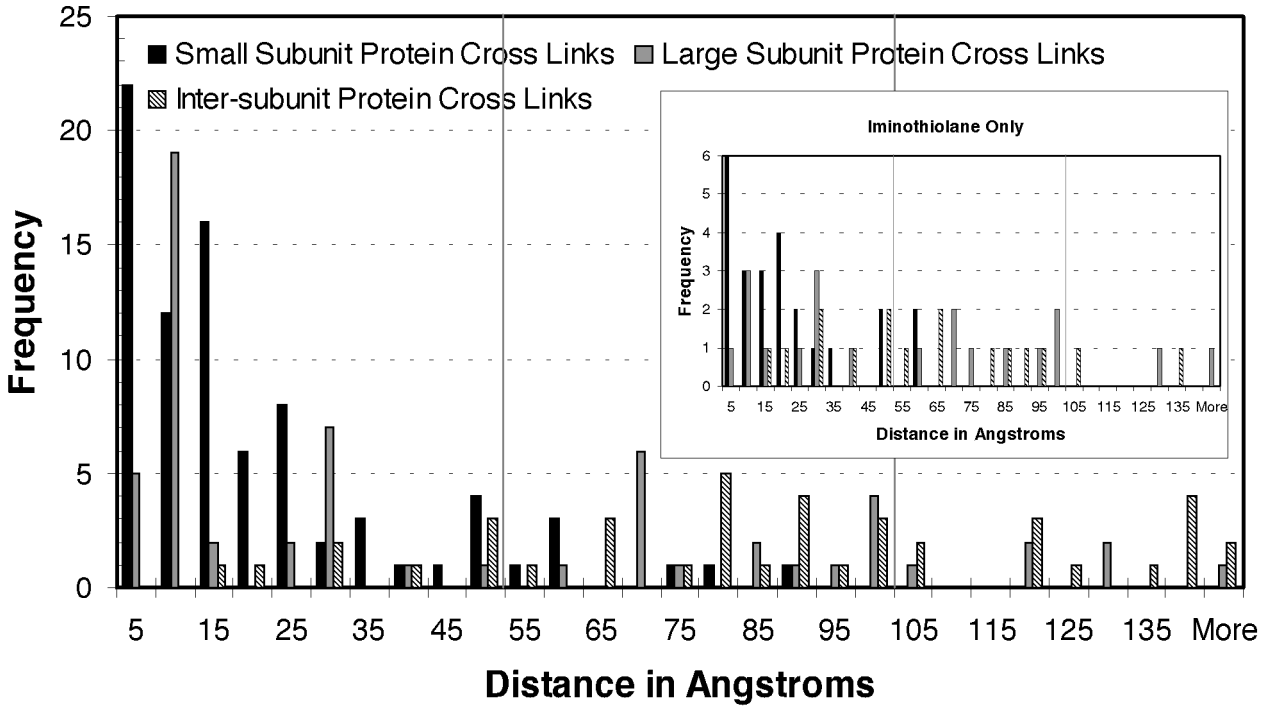

FIGURE 3. Histogram of protein/protein crosslinking distances, plotted based on the ribosomal subunit(s) involved. 30S crosslinks display the tightest distances. Intersubunit protein crosslinks represent very large distances that are likely erroneous. Inset: Distances are plotted for protein crosslinks resulting from iminothiolane only, and the same dependence on ribosomal subunits can be seen.

ber of data points: 1), dimethylapidimidate (length: 8.8 Å; points: 3), 3,3′-thiobispropioniomidate (length: 12 Å; points: 3), tatryl diazide (length: 6 Å; points: 7) and aminomethyltrimethylpsoralen (length: 23 Å; points: 1). *O*- and *p*-phenylenedimaleimide were some of the best agents used for protein/protein crosslinking studies. Although these agents contained only five data points, they were among the very few reliable points obtained for the 50S ribosomal subunit protein/protein crosslinks. Conversely, methyl-4-mercaptobutyrimidate was one of the worst agents. It is possible that this chemical causes destabilization and destruction of ribosomal particles.

FOOTPRINTS

We define footprinting experiments as those experiments in which (1) a specific structure-digesting agent (substance) is introduced into the medium, and (2) the data are expressed as protections from attack. The procedure is often based on a preliminary guess of the protected region. This experimental technique is relevant if the molecular structure can be reconstituted in vitro from its components with and without the studied part, or if that part can be removed from the original complex so that the overall structure is not affected. The protection pattern from the footprinting agent can be examined at the resolution of individual nucleotides. A wide variety of agents can be used, some of which have been recognized as more effective than others based on nonribosomal RNA and DNA structural studies. One of the advantages of footprinting is that it allows simultaneous detection of multiple protected elements—and thus assignment of multiple proximities, yielding more information for modeling.

Primer extension is the method commonly used for identification of affected RNA nucleotides in these types of experiments. Many different primers can be used. They detect breaks in the RNA to allow reading of the surrounding region by priming the action of reverse transcriptase. There are several steps in this procedure that include end labeling of the probe nucleic acid (primer), incorporation of the probe, formation of heteroduplexes between the probe and test samples, chemical reactions applied to these heteroduplexes, cleavage of sites of mismatch, and separation of the reaction products by polyacrylamide gel electrophoresis.

We evaluated 837 footprinting observations, involving all small ribosomal (S-) proteins except *E. coli* S1, S10, and S21, and three large subunit (L-) proteins, L9, L11, and L15. These observations were taken from 14 articles, representing distances between proteins and the protected rRNA base(s). We calculated the distance between the alpha-carbon of each amino acid in the protecting protein and the phosphorus atom of the protected rRNA nucleotide and used the shortest dis-

tance in creating the data set. Like crosslinking data, a little more than half of the footprinting observations have crystal distances less than 20 Å, with the remainder predominantly less than 50 Å (Fig. 4). A few outliers represent distances as high as 98 Å.

We expected that larger proteins might protect more RNA, but we did not see this pattern within the observed distances. A variety of agents were used for footprinting experiments, ranging from small inorganic molecules to large enzymes. Most significantly, we noted that hydroxyl radical footprints are always associated with distances less than 45 Å. There were no large distance outliers for this footprinting agent. This is consistent with the fact that the radius of diffusion of free hydroxyl radicals able to cleave internucleotide bonds has been estimated to be about 40 Å (Sergiev et al., 2001). Iodine was used in fewer footprints, but maintained a range of distances less than 25 Å. Dimethyl sulfate was the agent involved in most of the gross-mismatch distances for footprinting.

Sometimes, footprinting data are associated with strengths (strong, medium, or weak) based on gel band density. Strong footprints are assumed to originate from shorter distances from the probe (0–22 Å), medium strength is associated with distances in the range between 12–36 Å, and weak cleavages are attributed to 20 to 44 Å distances (Joseph et al., 1997). We did not find this correlation across all footprinting data, although sometimes the trend can be seen within individual experiments included in large data sets published in the same articles.

CLEAVAGES

We define cleavage experiments as those experiments in which (1) an agent that attacks some portion of the structure is introduced into the medium, and (2) the data are expressed in terms of which entities are attacked (rather than protected from attack). Cleavage experiments include many steps similar to those described above for the footprinting approach, and require cloning and site-directed mutagenesis techniques for the incorporation of the probe and its further activation. For example, in a set of experiments on the 30S ribosomal subunit, an Fe(II) was attached to a base or an amino acid via an EDTA linker. Chemicals were then added to the medium that caused hydroxyl radicals to be generated from the Fe(II) site; these hydroxyl radicals then cleaved nearby nucleic acid bonds.

We evaluated 1,307 cleavage observations taken from nine articles. We computed the distance between phosphorus atoms of the rRNA nucleotide with an attached probe and a cleaved rRNA nucleotide, or the closest alpha-carbon atom of a protein with an attached probe and the phosphorus atom of the cleaved rRNA nucleotide.

Given that the linker is approximately 12 Å long and the hydroxyl radicals are effective for about 10 Å, and considering possible conformational variations, we expected reactions to be localized to regions within about 30 or 40 Å from the linker attachment site. However, we observed a larger range of distances. All cleavage data for the 30S subunit (both rRNA and protein tethers)

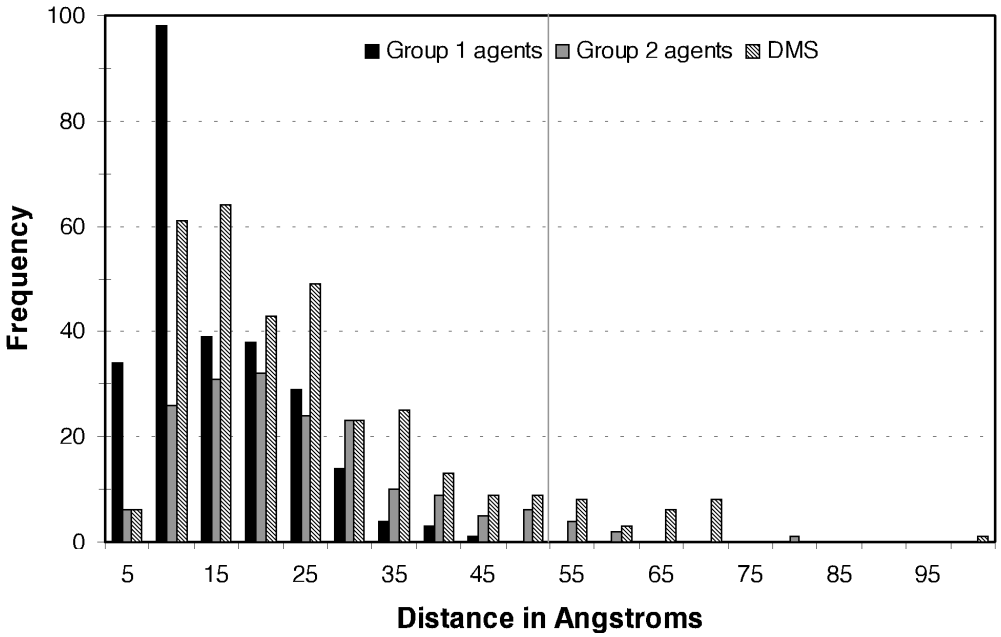

FIGURE 4. Histogram of footprinting distances, plotted based on the footprinting agent. Group 1 agents include hydroxyl radicals and iodine and yield the smallest distances. Group 2 agents include Fe[2+]-EDTA and kethoxal, which represented somewhat larger distances. Footprints generated with dimethyl sulfate (DMS) yielded a wide range of distances, including several that were extremely large.

were consistently associated with distances less than 60 Å, the majority of which mapped to distances less than 30 Å. For the 50S subunit, there were outlying distances up to 83 Å. We saw a clear dependence of distance on the object with the attached probe (Fig. 5A,B). Protein L9 yielded most of the very large distances observed (Fig. 5B). We noted that the same protein was associated with large distances in the RNA/protein crosslinking data. Clearly, this highly flexible protein can produce cleavages in many different regions after reconstitution with other components of the 50S subunit. The tightest group of cleavage distances was observed for 16S rRNA base G922, extending out to 30 Å (Fig. 5C).

As with footprinting data, cleavage data are often associated with strengths (strong, medium, or weak). In this case, we found the expected result that the stronger the cleavage, the shorter the distance. As was noted earlier, assignments of strength for crosslinking and footprinting data did not correlate with distinct ranges. We conclude that this parameter may be unreliable for molecular modeling in cases other than cleavage data.

COMPARISON OF MODELS

Given the large range of distances that map to the experimental proximity measures, it would not be surprising if the published models based on these data differed significantly from the crystal structures. We evaluated four models of the 30S ribosomal subunit—three published (Brimacombe et al., 1988; Malhotra & Harvey, 1994; Fink et al., 1996) and one unpublished (H.F. Noller et al.). We found that the root mean square deviation (RMSD) of the model phosphate locations to the crystal phosphate locations ranged from 26 to 46 Å. Given the overall size of the ribosomal subunit, and the demonstrable noise in the qualitative proximity measurements, these model errors seem reasonable.

CONCLUSION

All measurements have errors and uncertainties. Careful experimental design and interpretation of results, the development of controls, and evaluation of potential sources of errors, both pre- and postanalytical, are constant themes in the laboratory. Some results regarding proximities within the ribosome have been revised in the past as better techniques emerged. Less effective fractionation procedures may have contributed to incongruities in some protein/protein crosslink data (Lambert et al., 1983), exposure to some chemical agents in the media may have destabilized and partially destroyed ribosomal particles (Lutter & Kurland, 1975), and differences in reactivity between the 30S and 50S ribosomal subunit proteins were recognized very early (Moore, 1971). In addition, underlying assumptions about the ability of cleavage agents to diffuse into relatively protected regions may have been too conservative (Travers & Buckle, 2000). Protection data is difficult to interpret because nucleotides may be shielded by a tertiary fold or other features of the complex—including conformational changes. In general, reproducibility is known to not always be associated with accuracy of the results.

We know experimental techniques are complex and fallible. In this study, we were interested in identifying trends in the reliability of classes of experiments. In their paper, Sergiev et al. (2001) performed a careful analysis of rRNA/rRNA crosslink studies applied to the bacterial ribosome. Their main conclusion was that the simpler the methods, the more reliable the results. We saw the same trend in our study. The authors argued that there are no good methods for probing long RNA molecules (at least by incorporation of nucleotide analogs). They also formulated recommendations for experiments to be designed in the future.

In our work, we were able to analyze a larger group of data, including other types of crosslinks and footprinting and cleavage data. We observed that experimental results vary when probing different regions of the ribosome, probably because of flexibility. We also noted that certain agents may be more reliable for footprinting and crosslinking, though there is no perfect correlation between the size of the agent and the length of the crosslink.

The distances we observed in experimental proximity measures were generally less than 30 Å. However, the tail of the distance distribution extended much farther than we expected. This is informative from a modeling perspective. Model builders must be aware that distance distributions for crosslinking data can be very large. Footprinting observations may also vary in distance depending on the agent used. For example, DMS generated distances with a larger distribution than iodine or hydroxyl radicals. On the other hand, cleavage experiments can generate almost normally distributed distance data, as seen with the 30S data. Strength labels associated with this type of data, in particular, can be informative. Our results suggest that, with appropriate care, relatively low-resolution structural information can be used to probe the structure of large RNA/protein complexes, and can provide useful information for the construction of models of these molecular ensembles.

METHODS

Distance calculations

Ribosomal distances are calculated using the Cartesian coordinates from the Protein Data Bank files (accession numbers: 1fjf, 1fjc, 1hnw, 1hnx, 1hnz, 1hro, 1ibk, 1ibl, 1ibm, 1fka, 1i94, 1i95, 1i96, 1i97, 1gix, 1fgo, 1ffz, 1ffk, 1jj2, 1giy, 1c2w, 1eg0). Results for 1fjf, 1gix,

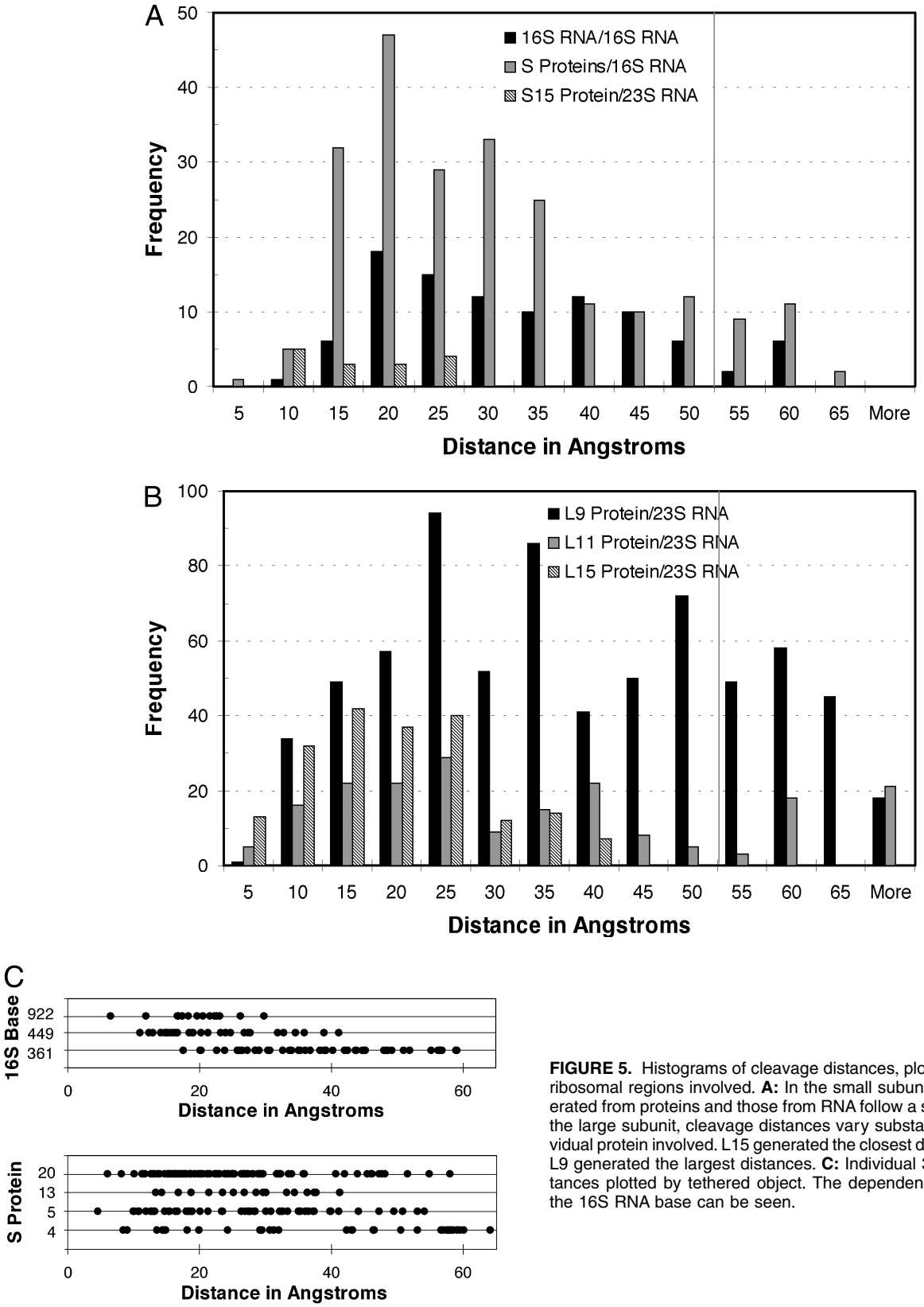

FIGURE 5. Histograms of cleavage distances, plotted based on the ribosomal regions involved. **A:** In the small subunit, cleavages generated from proteins and those from RNA follow a similar trend. **B:** In the large subunit, cleavage distances vary substantially by the individual protein involved. L15 generated the closest distances, whereas L9 generated the largest distances. **C:** Individual 30S cleavage distances plotted by tethered object. The dependence of distance on the 16S RNA base can be seen.

and 1giy are used for plotting the histograms shown in figures. Phosphorus atoms are used to represent the location of ribonucleotides, and alpha-carbon atoms are used to represent the location of amino acids. The positions of crosslinked proteins or RNA subsequences (i.e., continuous segments of the nucleic acid) are

represented by the position of the alpha-carbon or phosphorus atom of the closest amino acids or ribonucleotides, respectively. We used RiboWeb and other in-house automatic programs to calculate all distances.

To be consistent, we treated all experiments equally and calculated the distances exactly the same for all types of data, that is, between the phosphorus and alpha-carbon atoms (closest from the stretches given in the articles). It is known that different atomic groups and bonds are targeted by different agents, but our intention was to evaluate the general types of data from the perspective of a modeler working with low-resolution information.

Low-resolution macromolecular models are often built using "pseudo-atoms" to represent entire nucleotides. Because of the very large scale, representing each atom is impractical in the initial stages of modeling. Modelers are typically working with comparatively sparse data sets of spatial information from these types of experiments. Also, when modeling on such a large scale with such vague pieces of information as "two things are linked together," approximate positions are the only reasonable option. Later, the model may be filled in to include all atoms. Mostly, pseudo-atoms are not actual atoms at all, but rather large spheres used to represent the spatial positioning of nucleotides. In some cases, one atom, such as the phosphorus in the nucleic acid backbone, is used to represent the approximate position of a nucleotide. Because the kinds of experimental information studied here are those often used for low-resolution modeling of macromolecular structures, we adopted that representation in order to utilize atoms in the crystal structure with reported spatial coordinates. We extended that paradigm to proteins by using the alpha-carbon atom to represent the approximate positions of amino acids.

Data selection

We selected all crosslinking, footprinting and cleavage observations contained in the RiboWeb knowledge base, with the exception tRNA, mRNA, and protein cofactors. Results with bigger uncertainties (e.g., crosslinks reported to be between the entire 23S rRNA and an L-protein) were not included in the analysis. We used RiboWeb programs to automatically select and sort all appropriate observations.

ACKNOWLEDGMENTS

This work is supported by National Institutes of Health (NIH) Grants LM-05652 and LM-06422, National Science Foundation DBI-9600637, and grants from IBM and Sun Microsystems (to R.B.A.). M.W.-C. is also supported by a predoctoral fellowship GM-08294 from NIH. M.B. was a trainee on Genome Training Grant NIH T32 HG 00044. D.R.B. is supported by NIH 2 R25 GM056847-04. We thank Dr. Harry Noller for providing many useful conversations.

Received June 29, 2001; returned for revision July 20, 2001; revised manuscript received December 11, 2001

REFERENCES

Bada M, Altman RB. 2000. Computational modeling of structural experimental data. *Methods Enzymol 31*:470–491.

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science 289*:905–920.

Brimacombe R, Atmadja J, Stiege W, Schuler D. 1988. A detailed model of the three-dimensional structure of *Escherichia coli* 16S ribosomal RNA *in situ* in the 30S subunit. *J Mol Biol 199*:115–136.

Chen RO, Felciano R, Altman RB. 1997. RIBOWEB: Linking structural computations to a knowledge base of published experimental data. *Proc Int Conf Intell Syst Mol Biol 5*:84–87.

Fink DL, Chen RO, Noller HF, Altman RB. 1996. Computational methods for defining the allowed conformational space of 16S rRNA based on chemical footprinting data. *RNA 2*:851–866.

Green N, Reisler E, Houk KN. 2001. Quantitative evaluation of the lengths of homobifunctional protein cross-linking used as molecular rulers. *Protein Sci 10*:1293–1304.

Joseph S, Weiser B, Noller HF. 1997. Mapping the inside of the ribosome with an RNA helical ruler. *Science 278*:1093–1098.

Lambert JM, Boileau G, Cover JA, Traut RR. 1983. Cross-links between ribosomal proteins of 30S subunits in 70S tight couples and in 30S subunits. *Biochemistry 22*:3913–3920.

Lillemoen J, Hoffman DW. 1998. An investigation of the dynamics of ribosomal protein L9 using heteronuclear NMR relaxation measurements. *J Mol Biol 281*:539–551.

Lutter LC, Kurland CG. 1975. Chemical determination of protein neighborhoods in a cellular organelle. *Mol Cell Biochem 7*:105–116.

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.

Moore PB. 1971. Reaction of n-ethyl maleimide with ribosomes of *Escherichia coli*. *J Mol Biol 60*:169–184.

Schluenzen F, Tocilj A, Zarivach R, Harms J, Gluehmann M, Janell D, Bashan A, Bartels H, Agmon I, Franceschi F, Yonath A. 2000. Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell 102*:615–623.

Sergiev PV, Dontsova OA, Bogdanov AA. 2001. Chemical methods for the structural study of the ribosome: Judgment day. *Mol Biol 35*:472–495.

Stoffler G, Redl B, Walleczek J, Stoffler-Meilicke M. 1988. Identification of protein–protein cross-links within the *Escherichia coli* ribosome by immunoblotting techniques. *Methods Enzymol 164*:64–76.

Travers EA, Buckle M. 2000. *DNA protein interactions*. New York: Oxford University Press.

Wimberly BT, Brodersen DE, Clemons WM Jr, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V. 2000. Structure of the 30S ribosomal subunit. *Nature 407*:327–339.

Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JHD, Noller HF. 2001. Crystal structure of the ribosome at 5.5 Å resolution. *Science 292*:883–896.

APPENDIX

Alexander RW, Muralikrishna P, Cooperman BS. 1994. Ribosomal components neighboring the conserved 518–533 loop of 16S rRNA in 30S subunits. *Biochemistry 33*:12109–12118.

Atmadja J, Brimacombe R, Blocker H, Frank R. 1985. Investigation of the tertiary folding of *Escherichia coli* 16S RNA by *in situ* intra-RNA cross-linking within 30S ribosomal subunits. *Nucleic Acids Res 13*:6919–6936.

Atmadja J, Stiege W, Zobawa M, Greuer B, Osswald M, Brimacombe

R. 1986. The tertiary folding of *Escherichia coli* 16S RNA, as studied by *in situ* intra-RNA cross-linking of 30S ribosomal subunits with bis-2-chloroethyl-methylamine. *Nucleic Acids Res 14*: 659–673.

Baranov PV, Gurvich OL, Bogdanov AA, Brimacombe R, Dontsova OA. 1998. New features of 23S ribosomal RNA folding: The long helix 41–42 makes a "U-turn" inside the ribosome. *RNA 4*:658–668.

Batey RT, Williamson JR. 1996. Interaction of the *Bacillus stearothermophilus* ribosomal protein S15 with 16 S rRNA: II. Specificity determinants of RNA–protein recognition. *J Mol Biol 261*:550–567.

Bode U, Lutter C, Stoffler G. 1974. Proteins S14 and S19 are near-neighbors in the *E. coli* ribosome. *FEBS Lett 45*:232–236.

Chiaruttini C, Expert-Bezancon A, Hayes D, Ehresmann B. 1982. Protein–RNA crosslinking in *Escherichia coli* 30S ribosomal subunits. Identification of a 16S rRNA fragment crosslinked to protein S12 by the use of the chemical crosslinking reagent 1-ethyl-3-dimethyl-aminopropylcarbodiimide. *Nucleic Acids Res 10*:7657–7676.

Chiaruttini C, Milet M, Hayes DH, Expert-Bezancon A. 1986. Multiple crosslinks of proteins S7 and S9 to domains 3 and 4 of 16S ribosomal RNA in the *Escherichia coli* 30S particle. *Eur J Biochem 160*:363–370.

Chiaruttini C, Milet M, Hayes DH, Expert-Bezancon A. 1989. Crosslinking of ribosomal proteins S4, S5, S7, S8, S11, S12 and S18 to domains 1 and 2 of 16S rRNA in the *Escherichia coli* 30S particle. *Biochimie 71*:839–852.

Cover JA, Lambert JM, Norman CM, Traut RR. 1981. Identification of proteins at the subunit interface of the *Escherichia coli* ribosome by cross-linking with dimethyl 3,3′-dithiobis(propionimidate). *Biochemistry 12*:2843–2852.

Culver GM, Cate JH, Yusupova GZ, Yusupov MM, Noller HF. 1999. Identification of an RNA–protein bridge spanning the ribosomal subunit interface. *Science 285*:2133–2136.

Culver GM, Noller HF. 1998. Directed hydroxyl radical probing of 16S ribosomal RNA in ribosomes containing FeII tethered to ribosomal protein S20. *RNA 4*:1471–1480.

Dokudovskaya S, Dontsova O, Shpanchenko O, Bogdanov A, Brimacombe R. 1996. Loop IV of 5S ribosomal RNA has contacts both to domain II and to domain V of the 23S RNA. *RNA 2*:146–152.

Dontsova O, Tishkov V, Dokudovskaya S, Bogdanov A, Doring T, Rinke-Appel J, Thamm S, Greuer B, Brimacombe R. 1994. Stem-loop IV of 5S rRNA lies close to the peptidyltransferase center. *Proc Natl Acad Sci USA 91*:4125–4129.

Doring T, Greuer B, Brimacombe R. 1991. The three-dimensional folding of ribosomal RNA; localization of a series of intra-RNA cross-links in 23S RNA induced by treatment of *Escherichia coli* 50S ribosomal subunits with bis-(2-chloroethyl)-methylamine. *Nucleic Acids Res 19*:3517–3524.

Doring T, Greuer B, Brimacombe R. 1992. The topography of the 3′-terminal region of *Escherichia coli* 16S ribosomal RNA; an intra-RNA cross-linking study. *Nucleic Acids Res 20*:1593–1597.

Ehresmann B, Backendorf C, Ehresmann C, Millon R, Ebel JP. 1980. Effect of ultraviolet irradiation on 30-S ribosomal subunits. Identification of the RNA region crosslinked to protein S7. *Eur J Biochem 104*:255–262.

Expert-Bezancon A, Milet M, Carbon P. 1983. Precise localization of several covalent RNA–RNA cross-link in *Escherichia coli* 16S RNA. *Eur J Biochem 136*:267–274.

Gregory RJ, Zeller ML, Thurlow DL, Gourse RL, Stark MJ, Dahlberg AE, Zimmermann RA. 1984. Interaction of ribosomal proteins S6, S8, S15 and S18 with the central domain of 16 S ribosomal RNA from *Escherichia coli*. *J Mol Biol 178*:287–302.

Greuer B, Osswald M, Brimacombe R, Stoffler G. 1987. RNA–protein cross-linking in *Escherichia coli* 30S 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.

Gulle H, Hoppe E, Osswald M, Greuer B, Brimacombe R, Stoffler G. 1988. RNA–protein cross-linking in *Escherichia coli* 50S ribosomal subunits; determination of sites on 23S RNA that are cross-linked to proteins L2, L4, L24 and L27 by treatment with 2-iminothiolane. *Nucleic Acids Res 16*:815–832.

Heilek GM, Marusak R, Meares CF, Noller HF. 1995. Directed hydroxyl radical probing of 16S rRNA using FeII tethered to ribosomal protein S4. *Proc Natl Acad Sci USA 92*:1113–1116.

Heilek GM, Noller HF. 1996. Directed hydroxyl radical probing of the rRNA neighborhood of ribosomal protein S13 using tethered FeII. *RNA 2*:597–602.

Heilek GM, Noller HF. 1996. Site-directed hydroxyl radical probing of the rRNA neighborhood of ribosomal protein S5. *Science 272*: 1659–1662.

Holmberg L, Noller HF. 1999. Mapping the ribosomal RNA neighborhood of protein L11 by directed hydroxyl radical probing. *J Mol Biol 289*:223–233.

Hui CF, Cantor CR. 1985. Mapping the location of psoralen cross-links on RNA by mung bean nuclease sensitivity of RNA.DNA hybrids. *Proc Natl Acad Sci USA 82*:1381–1385.

Kenny JW, Traut RR. 1979. Identification of fifteen neighboring protein pairs in the *Escherichia coli* 50 S ribosomal subunit cross-linked with 2-iminothiolane. *J Mol Biol 127*:243–263.

Kyriatsoulis A, Maly P, Greuer B, Brimacombe R, Stoffler G, Frank R, Blocker H. 1986. RNA–protein cross-linking in *Escherichia coli* ribosomal subunits: Localization of sites on 16S RNA which are cross-linked to proteins S17 and S21 by treatment with 2-iminothiolane. *Nucleic Acids Res 14*:1171–1186.

Lambert JM, Boileau G, Cover JA, Traut RR. 1983. Cross-links between ribosomal proteins of 30S subunits in 70S tight couples and in 30S subunits. *Biochemistry 22*:3913–3920.

Lambert JM, Traut RR. 1981. The subunit interface of the *Escherichia coli* ribosome. Identification of proteins at the interface between the 30 S and 50 S subunits by crosslinking with 2-iminothiolane. *J Mol Biol 149*:451–476.

Lieberman KR, Firpo MA, Herr AJ, Nguyenle T, Atkins JF, Gesteland RF, Noller HF. 2000. The 23 S rRNA environment of ribosomal protein L9 in the 50 S ribosomal subunit. *J Mol Biol 297*:1129–1143.

Lieberman KR, Noller HF. 1998. Ribosomal protein L15 as a probe of 50 S ribosomal subunit structure. *J Mol Biol 284*:1367–1378.

Lutter LC, Bode U, Kurland CG, Stoffler G. 1974. Ribosomal protein neighborhoods. 3. Cooperativity of assembly. *Mol Gen Genet 129*:167–176.

Lutter LC, Kurland CG, Stoffler G. 1975. Protein neighborhoods in the 30S ribosomal subunit of *Escherichia coli*. *FEBS Lett 54*:144–150.

Lutter LC, Zeichhardt H, Kurland CG. 1972. Ribosomal protein neighborhoods. I. S18 and S21 as well as S5 and S8 are neighbors. *Mol Gen Genet 119*:357–366.

Maassen JA, Schop EN, Moller W. 1981. Structural analysis of ribosomal protein L7/L12 by the heterobifunctional cross-linker 4-(6-formyl-3-azidophenoxy)butyrimidate. *Biochemistry 20*:1020–1025.

Maly P, Rinke J, Ulmer E, Zwieb C, Brimacombe R. 1980. Precise localization of the site of cross-linking between protein L4 and 23S ribonucleic acid induced by mild ultraviolet irradiation of *Escherichia coli* 50S ribosomal subunits. *Biochem 19*:4179–4188.

Mitchell P, Osswald M, Brimacombe R. 1992. Identification of intermolecular RNA cross-links at the subunit interface of the *Escherichia coli* ribosome. *Biochem 31*:3004–3011.

Mitchell P, Osswald M, Schueler D, Brimacombe R. 1990. Selective isolation and detailed analysis of intra-RNA cross-links induced in the large ribosomal subunit of *E. coli*: A model for the tertiary structure of the tRNA binding domain in 23S RNA. *Nucleic Acids Res 18*:4325–4333.

Moine H, Bienaime C, Mougel M, Reinbolt J, Ebel JP, Ehresmann C, Ehresmann B. 1988. Crosslinking of ribosomal protein S18 to 16 S RNA in *E. coli* ribosomal 30 S subunits by the use of a reversible crosslinking agent: Trans-diamminedichloroplatinumII. *FEBS Lett 228*:1–6.

Montpetit A, Payant C, Nolan JM, Brakier-Gingras L. 1998. Analysis of the conformation of the 3′ major domain of *Escherichia coli* 16S ribosomal RNA using site-directed photoaffinity crosslinking. *RNA 4*:1455–1466.

Mundus D, Wollenzien P. 1998. Neighborhood of 16S rRNA nucleotides U788/U789 in the 30S ribosomal subunit determined by site-directed crosslinking. *RNA 4*:1373–1385.

Newcomb LF, Noller HF. 1999. Directed hydroxyl radical probing of 16S ribosomal RNA in 70S ribosomes from internal positions of the RNA. *Biochemistry 38*:945–951.

Osswald M, Brimacombe, R. 1999. The environment of 5S rRNA in the ribosome: Cross-links to 23S rRNA from sites within helices II and III of the 5S molecule. *Nucleic Acids Res 27*:2283–2290.

Osswald M, Greuer B, Brimacombe R. 1990. Localization of a series of RNA–protein cross-link sites in the 23S and 5S ribosomal RNA from *Escherichia coli*, induced by treatment of 50S subunits with three different bifunctional reagents. *Nucleic Acids Res 18*:6755–6760.

Osswald M, Greuer B, Brimacombe R, Stoffler G, Baumert H, Fasold H. 1987. RNA–protein cross-linking in *Escherichia coli* 30S ribosomal subunits; determination of sites on 16S RNA that are cross-linked to proteins S3, S4, S5, S7, S8, S9, S11, S13, S19 and S21 by treatment with methyl p-azidophenyl acetimidate. *Nucleic Acids Res 15*:3221–3240.

Peretz H, Towbin H, Elson D. 1976. The use of a cleavable cross-linking reagent to identify neighboring proteins in the 30-S ribosomal subunit of *Escherichia coli*. *Eur J Biochem 63*:83–92.

Powers T, Changchien LM, Craven GR, Noller HF. 1988. Probing the assembly of the 3′ major domain of 16 S ribosomal RNA. Quaternary interactions involving ribosomal proteins S7, S9 and S19. *J Mol Biol 200*:309–319.

Powers T, Noller HF. 1995. Hydroxyl radical footprinting of ribosomal proteins on 16S rRNA. *RNA 1*:194–209.

Powers T, Stern S, Changchien LM, Noller HF. 1988. Probing the assembly of the 3′ major domain of 16 S rRNA. Interactions involving ribosomal proteins S2, S3, S10, S13 and S14. *J Mol Biol 201*:697–716.

Redl B, Walleczek J, Stoffler-Meilicke M, Stoffler G. 1989. Immunoblotting analysis of protein–protein crosslinks within the 50S ribosomal subunit of *Escherichia coli*. A study using dimethylsuberimidate as crosslinking reagent. *Eur J Biochem 181*:351–356.

Rosendahl G, Douthwaite S. 1993. Ribosomal proteins L11 and L10.(L12)4 and the antibiotic thiostrepton interact with overlapping regions of the 23 S rRNA backbone in the ribosomal GTPase center. *J Mol Biol 234*:1013–1020.

Samaha RR, Joseph S, O'Brien B, O'Brien TW, Noller HF. 1999. Site-directed hydroxyl radical probing of 30S ribosomal subunits by using FeII tethered to an interruption in the 16S rRNA chain. *Proc Natl Acad Sci USA 96*:366–370.

Scheibe U, Wagner R. 1986. Identification of neighbouring proteins by cross-linking of intact 70 S ribosomes from *Escherichia coli*. *Biochim Biophys Acta 869*:1–7.

Sergiev P, Dokudovskaya S, Romanova E, Topin A, Bogdanov A. Brimacombe R, Dontsova O. 1998. The environment of 5S rRNA in the ribosome: Cross-links to the GTPase-associated area of 23S rRNA. *Nucleic Acids Res 26*:2519–2525.

Sommer A, Traut RR. 1975. Identification by diagonal gel electrophoresis of nine neighboring protein pairs in the *Escherichia coli* 30 S ribosome crosslinked with methyl-4-mercaptobutyrimidate. *J Mol Biol 97*:471–481.

Sommer A, Traut RR. 1976. Identification of neighboring protein pairs in the *Escherichia coli* 30 S ribosomal subunit by crosslinking with methyl-4-mercaptobutyrimidate. *J Mol Biol 106*:995–1015.

Stern S, Changchien LM, Craven GR, Noller HF. 1988. Interaction of proteins S16, S17 and S20 with 16 S ribosomal RNA. *J Mol Biol 200*:291–299.

Stern S, Powers T, Changchien LM, Noller HF. 1988. Interaction of ribosomal proteins S5, S6, S11, S12, S18 and S21 with 16 S rRNA. *J Mol Biol 201*:683–695.

Stern S, Wilson RC, Noller HF. 1986. Localization of the binding site for protein S4 on 16 S ribosomal RNA by chemical and enzymatic probing and primer extension. *J Mol Biol 192*:101–110.

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.

Stiege W, Glotz C, Brimacombe R. 1983. Localization of a series of intra-RNA cross-links in the secondary and tertiary structure of 23S RNA, induced by ultraviolet irradiation of *Escherichia coli* 50S ribosomal subunits. *Nucleic Acids Res 11*:1687–1706.

Stiege W, Kosack M, Stade K, Brimacombe R. 1988. Intra-RNA cross-linking in *Escherichia coli* 30S ribosomal subunits: Selective isolation of cross-linked products by hybridization to specific cDNA fragments. *Nucleic Acids Res 16*:4315–4329.

Stiege W, Zwieb C, Brimacombe R. 1982. Precise localization of three intra-RNA cross-links in 23S RNA and one in 5S RNA, induced by treatment of *Escherichia coli* 50S ribosomal subunits with bis-(2-chloroethyl)-methylamine. *Nucleic Acids Res 10*:7211–7229.

Stoffler G, Redl B, Walleczek J, Stoffler-Meilicke M. 1988. Identification of protein–protein cross-links within the *Escherichia coli* ribosome by immunoblotting techniques. *Methods Enzymol 164*:64–76.

Sun TT, Bollen A, Kahan L, Traut RR. 1974. Topography of ribosomal proteins of the *Escherichia coli* 30S subunit as studied with the reversible cross-linking reagent methyl 4-mercaptobutyrimidate. *Biochemistry 13*:2334–2340.

Svensson P, Changchien LM, Craven GR, Noller HF. 1988. Interaction of ribosomal proteins, S6, S8, S15 and S18 with the central domain of 16 S ribosomal RNA. *J Mol Biol 200*:301–308.

Szymkowiak C, Wagner R. 1985. Analysis of a sequence region of 5S RNA from *E. coli* cross-linked *in situ* to the ribosomal protein L25. *Nucleic Acids Res 13*:3953–3968.

Thiede B, Urlaub H, Neubauer H, Grelle G, Wittmann-Liebold B. 1998. Precise determination of RNA–protein contact sites in the 50 S ribosomal subunit of *Escherichia coli*. *Biochem J 334*:39–42.

Thompson JF, Hearst JE. 1983. Structure of *E. coli* 16S RNA elucidated by psoralen crosslinking. *Cell 32*:1355–1365.

Traut RR, Dey D, Bochkariov DE, Oleinikov AV, Jokhadze GG, Hamman B, Jameson D. 1995. Location and domain structure of *Escherichia coli* ribosomal protein L7/L12: Site specific cysteine crosslinking and attachment of fluorescent probes. *Biochem Cell Biol 73*:949–958.

Traut RR, Lambert JM, Kenny JW. 1983. Ribosomal protein L7/L12 cross-links to proteins in separate regions of the 50 S ribosomal subunit of *Escherichia coli*. *J Biol Chem 258*:14592–14598.

Turner S, Noller HF. 1983. Identification of sites of 4′-(hydroxymethyl)-4,5′,8-trimethylpsoralen cross-linking in *Escherichia coli* 23S ribosomal ribonucleic acid. *Biochemistry 16*:4159–4164.

Urlaub H, Kruft V, Bischof O, Muller EC, Wittmann-Liebold B. 1995. Protein–rRNA binding features and their structural and functional implications in ribosomes as determined by cross-linking studies. *EMBO J 14*:4578–4588.

Urlaub H, Thiede B, Muller EC, Brimacombe R, Wittmann-Liebold B. 1997. Identification and sequence analysis of contact sites between ribosomal proteins and rRNA in *Escherichia coli* 30 S subunits by a new approach using matrix-assisted laser desorption/ionization-mass spectrometry combined with N-terminal microsequencing. *J Biol Chem 272*:14547–14555.

Walleczek J, Martin T, Redl B, Stoffler-Meilicke M, Stoffler G. 1989. Comparative cross-linking study on the 50S ribosomal subunit from *Escherichia coli*. *Biochemistry 28*:4099–4105.

Wiener L, Brimacombe R. 1987. Protein binding sites on *Escherichia coli* 16S RNA; RNA regions that are protected by proteins S7, S14 and S19 in the presence or absence of protein S9. *Nucleic Acids Res 15*:3653–3670.

Wilms C, Noah JW, Zhong D, Wollenzien P. 1997. Exact determination of UV-induced crosslinks in 16S ribosomal RNA in 30S ribosomal subunits. *RNA 3*:602–612.

Wower I, Brimacombe R. 1983. The localization of multiple sites on 16S RNA which are cross-linked to proteins S7 and S8 in *Escherichia coli* 30S ribosomal subunits by treatment with 2-iminothiolane. *Nucleic Acids Res 11*:1419–1437.

Wower I, Wower J, Meinke M, Brimacombe R. 1981. The use of 2-iminothiolane as an RNA–protein cross-linking agent in *Escherichia coli* ribosomes, and the localization on 23S RNA of sites cross-linked to proteins L4, L6, L21, L23, L27 and L29. *Nucleic Acids Res 9*:4285–4302.

Zecherle GN, Oleinikov A, Traut RR. 1992. The proximity of the C-terminal domain of *Escherichia coli* ribosomal protein L7/L12 to L10 determined by cysteine site-directed mutagenesis and protein–protein cross-linking. *J Biol Chem 267*:5889–5896.

Zecherle GN, Oleinikov A, Traut RR. 1992. The C-terminal domain of *Escherichia coli* ribosomal protein L7/L12 can occupy a location near the factor-binding domain of the 50S subunit as shown by cross-linking with N-[4-(p-azidosalicylamido)butyl]-3-(2′-pyridyldithio)propionamide. *Biochemistry 31*:9526–9532.

Zimmermann RA, Mackie GA, Muto A, Garrett RA, Ungewickell E, Ehresmann C, Stiegler P, Ebel JP, Fellner P. 1975. Location and characteristics of ribosomal protein binding sites in the 16S RNA of *Escherichia coli*. *Nucleic Acids Res 2*:279–302.

Zwieb C, Brimacombe R. 1979. Max-Planck-Institut fur Molekulare Genetik, Abteilung Wittmann, Berlin-Dahlem, GFR. *Nucleic Acids Res 6*:1775–1790.