Title
STRUCTURE OF E. coli 16S RNA ELUCIDATED BY PSORALEN CROSSLINKING

Authors
Thompson, J.F.
Hearst, J.E.

Publication Date
1983

LBL-15508
Preprint c.2

Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

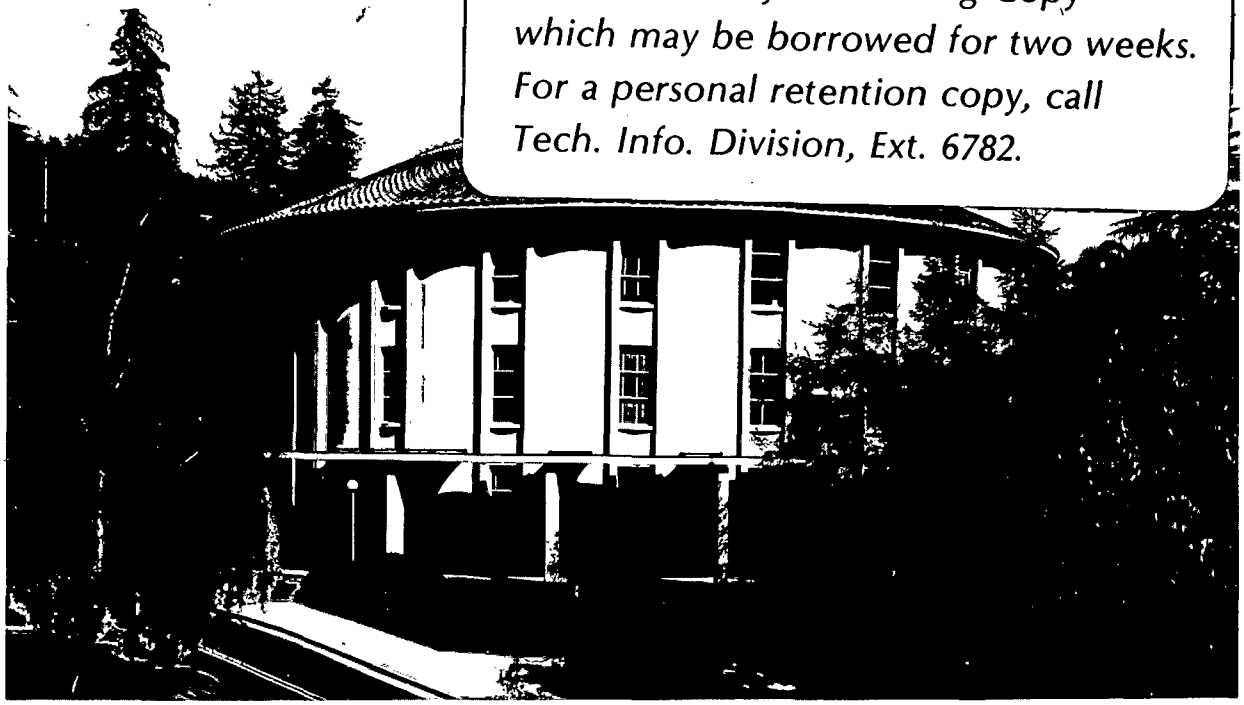
CHEMICAL BIODYNAMICS DIVISION

Submitted to Cell

STRUCTURE OF E.coli 16S RNA ELUCIDATED BY PSORALEN CROSSLINKING

John F. Thompson and John E. Hearst

January 1983



Prepared for the U.S. Department of Energy under Contract DE-AC03-76SF00098

DISCLAIMER

LBL-15508

STRUCTURE OF *E.coli* 16S RNA ELUCIDATED BY PSORALEN CROSSLINKING

John F. Thompson* and John E. Hearst

Laboratory of Chemical Biodynamics
Lawrence Berkeley Laboratory
and the
Department of Chemistry
University of California, Berkeley
Berkeley, California 94720

January 1983

*Present Address: Division of Biology & Medicine
Brown University
Providence, RI 02912

ABSTRACT

E. coli 16S RNA in solution was photoreacted with hydroxymethyltrimethyl-psoralen and long wave ultraviolet light. Positions of crosslinks were determined to high resolution by partially digesting the RNA with T_1 RNase, separating the crosslinked fragments by two-dimensional gel electrophoresis, reversing the crosslink, and sequencing the separated fragments. This method yielded the locations of crosslinks to $\pm$15 nucleotides. Even finer placement has been made on the basis of our knowledge of psoralen reactivity. Thirteen unique crosslinks were mapped. Seven crosslinks confirmed regions of secondary structure which had been predicted in published phylogenetic models, three crosslinks discriminated between phylogenetic models, and three proved the existence of new structures. The new structures were all long range interactions which appear to be in dynamic equilibrium with local secondary structure. Because this technique yields direct information about the secondary structure of large RNAs, it should prove invaluable in studying the structure of other RNAs of all sizes.

INTRODUCTION

The secondary structure of E. coli 16S RNA has been speculated on ever since the first, partial sequences were determined (Fellner et al., 1970). Real progress was not made until the complete sequence of the RNA was unequivocably determined (Brosius et al., 1978; Carbon et al., 1979). Once the sequence was established, chemical modification and enzymatic digestion data (reviewed by Noller & Woese, 1981) could be used with greater confidence. Availability of the small subunit RNA sequences of Z. mays chloroplasts (Schwarz & Kossel, 1980), P. vulgaris (Carbon et al., 1981), S. cerevisiae (Ruptsov et al., 1980), X. laevis (Salim & Maden, 1981) and various mitochondria (Eperon et al., 1980; Van Etten et al., 1980; Sor & Fukuhara, 1980) has made phylogenetic comparisons possible. On the basis of these data, Noller & Woese (1981) and Stiegler et al. (1981) have proposed a general secondary structure model for all small subunit RNAs. Zwieb et al. (1981) have also proposed a general model based on this evidence as well as UV crosslinking (Zwieb & Brimacombe, 1980), and denaturation studies (Ross & Brimacombe, 1979; Glotz & Brimacombe, 1980). All three models are very similar.

Despite this agreement, the models are far from complete. Psoralen crosslinking data generated by electron microscopy yield results that, in many instances, cannot be easily incorporated into the other models (Wollenzein et al., 1979; Wollenzein & Cantor, 1982). Because of the low resolution of the electron microscope, the unambiguous placement of crosslinks is not possible. It is clear that psoralen crosslinking is generating new information, so this study has been undertaken to obtain a more detailed localization of crosslinks and thus a more complete picture of the 16S RNA secondary structure.

A technique, based on that of Zwieb and Brimacombe (1980), has been developed which allows resolution to at least 15 bases. Because psoralen reacts with only a certain class of sites, and only a few secondary structures can be drawn for any given fragments, actual resolution is generally to the exact nucleotides crosslinked. Psoralen has been observed to react primarily with uridines in natural RNAs (Thompson et al., 1981; Bachellerie & Hearst, 1982; Youvan & Hearst, 1982), but crosslinking to cytidine and purines has been observed (Bachellerie et al., 1981) and cannot be ruled out. While crosslink assignment to fragments is certain, the nucleotide assignments are tentative.

RESULTS

A schematic diagram of the methodology used to form and analyze crosslinks is shown in Figure 1. Crosslinking was done with protein-free 16S RNA in reconstitution buffer at low levels of HMT reaction in order to attain the most biologically relevant conformation. HMT has been shown to have a minimal effect on RNA structure when added at these low doses, 3-5 HMTs per 16S (Thompson et al., 1981; Thompson et al., 1982). The crosslinked and uncrosslinked 16S melted at exactly the same temperature in reconstitution buffer. The psoralen sample had about 10% less hyperchromicity and also renatured much more quickly (results not shown). When examined at lower ionic strength, the crosslinked sample melted at lower temperatures, suggesting that a different structure was more stable. The crosslinked sample was locked in the reconstitution conformation and could not change as easily as the uncrosslinked sample.

Zwieb & Brimacombe (1980) have used a similar two dimensional gel technique to examine UV induced crosslinks. In the first dimension gel, any pre-existing secondary structure is stable because there are no denaturants and the gels are run at room temperature. The only fragments which will have structure are snap-back hairpins and covalent crosslinks. Prior to loading on the gel, all intermolecular secondary structure is destroyed by heating and rapid cooling. This produces a more uniform population of molecules. The second dimension contains molecules with no secondary structure because it contains 7 M urea and is run hot. Most molecules are unaffected because they contained no secondary structure in the first dimension and thus run as a diagonal. Uncrosslinked hairpin loops actually run slightly faster in the second dimension because their radius decreases somewhat. Covalently crosslinked molecules, however, are severely retarded in the second dimension. After melting, crosslinks have four ends and are forced to move through the gel in a spread-out, octopus-like conformation (Zwieb & Brimacombe, 1980).

As shown in Figure 2, the only bands which appear above the diagonal are caused by crosslinking. In 2D gels which contain total 16S RNA, the number of off-diagonal spots is simply too large to handle. Only a few crosslinks can be retrieved in pure form with enough ^{32}P to sequence. In order to circumvent both these problems, the T_1 digests of crosslinked RNA were first run through either a RPC-5 or BND cellulose column. Fractions from these columns were then labelled and analyzed by the 2D gel system. This protocol has the double advantage of reducing cross-contamination of crosslinked bands and also providing better incorporation of γ -^{32}P ATP.

The basis for RPC-5 separation of crosslinks is not entirely clear. When fractions from RPC-5 are run on the two-dimensional gel system, two diagonals are observed. The lower diagonal corresponds to the normal single-stranded

material while the upper, steeper diagonal consists entirely of crosslinked material. When a 0.1 to 2.0 M KCl gradient was used for elution, the most useful fractions for analysis came off the column at .45-.80 M KCl. Later fractions tended to contain material too large for analysis. These can be recombined, redigested, and run on the column again.

BND cellulose separates principally on the basis of secondary structure (Sedat et al., 1967). Before adding RNA to the column, it was heated and quick-cooled to minimize noncrosslinked secondary structure. As expected, the elution profile for OD and ^{3}H (from ^{3}H-HMT) do not coincide. Most RNA has little secondary structure and elutes early. The ^{3}H elutes more slowly because the crosslinks stabilize the secondary structure to the denaturing effects of DMSO. Later fractions contained a large number of crosslinked hairpin loops while the earlier fractions tended to contain more long range crosslinks with less complementarity. There were significant amounts of hairpin structures in these later fractions as well.

The crosslinks obtained vary tremendously with the level of T_1 digestion. If the samples are heavily digested as in Figure 2b (note the lack of material near the top of the diagonal), a small set of crosslinks is found. Light digestion results in more crosslinks which are less well-resolved. If the full range of crosslinks is to be found, several different digestions must be done.

In order for the two dimensional gel system described here to separate crosslinked and uncrosslinked material, the fragments must each have a minimum length of approximately 15 nucleotides. Complete digestion with any RNase would make the fragments too small to be of use. In order to produce oligonucleotides that are large enough to separate but small enough to be resolvable, partial digestion was done with T_1 RNase. This has the undesirable

property of producing many different-sized fragments which actually contain the same crosslink. The number of off-diagonal spots produced even when using a crosslinking agent as specific as psoralen becomes overwhelming. To compound this even further, it was observed that many spots can reverse to yield exactly the same pair of fragments. This appears to be primarily a problem in the running of the first dimension gel.

To reduce the number of off-diagonal bands, first dimension gels containing a short 4% polyacrylamide, 7 M urea stacking gel were employed. This ensured that no intermolecular base-pairing could be present. While this did reduce the number of bands, it also reduced the separation of the bands from the diagonal. The decision of which type of first dimension gel to use is thus dependent upon the seriousness of the multiple spot phenomenon.

After photoreversal of the crosslink and elution from the gel, fragments were separated on a 20% polyacrylamide, 7 M urea gel. In some cases, a single band was observed. If the mobility of this band changed distinctly after reversal (R_f in 2nd dimensional gel and separation gel were compared), the band was classified as a hairpin loop and sequenced. When two bands were found, both were sequenced. Frequently, more than two bands were found. Unfortunately, relative intensities of labelling cannot be used to match pairs of bands because the halves of crosslinks are not necessarily labelled equally. The 5' terminal bases are usually different and the ends are usually differentially offset in secondary structure. These ambiguous crosslinks were discarded.

Because the fragments are 5' end labelled, rapid sequencing techniques can be employed in determining their positions. Many fragments were sequenced simultaneously so that the less time-consuming method of partial enzymatic digestion was chosen over chemical cleavage. The sequence of 16S RNA is

already known (Brosius et al., 1978) so that ambiguities in enzymatic cuts are easily resolved. The positions of the 5' and 3' bases in the unambiguous crosslinks are shown in Table I. Frequently, different sets of termini were obtained from different off-diagonal spots for the same crosslink. These overlapping fragments allow an even finer resolution of the crosslink pattern.

DISCUSSION

Reactivity of Psoralen

In order to best assign the locations of crosslinks within the observed fragments, a knowledge of the specificity of reaction of HMT is necessary. Reaction of different psoralens with various RNAs (Bachellerie et al., 1981; Thompson et al., 1982) and DNAs (Straub et al., 1981; Kanne et al., 1982) has demonstrated that uridine and thymine are the preferred sites for photoreaction. There are only a few studies in which the position of psoralen has been mapped to high resolution (Rabin & Crothers, 1979; Thompson et al., 1981; Bachellerie & Hearst, 1982; Youvan & Hearst, 1982; Turner et al., 1982). These studies indicate that the most reactive positions occur in locations which are relatively unstable and facilitate intercalation. Particularly susceptible sites occur near the ends of helices, adjacent to G·U pairs, or at the end of runs of uridine. Crosslinks involving cytidine have also been found but at low yields. Certainly, there are non-Watson-Crick pairing schemes which could be crosslinked. Such structures may be inherently less reactive because, in tRNA, melting of the tertiary structure greatly enhanced psoralen reaction (Bachellerie & Hearst, 1982).

Another type of tertiary interaction has been observed in tRNA (Kim et al, 1974). Coaxial stacking of helices appears to be a major structural feature in tRNA and has also been proposed as an important feature of large

ribosomal RNAs (Noller et al., 1981). Such structures may be reactive if they can be unwound sufficiently to allow stacking of the psoralen between helices. The non-continuous phosphodiester backbone should allow unwinding to occur more easily.

Nomenclature

The nomenclature used to describe the crosslinks found is an extension of that used by Noller et al. (1981) and Wollenzein & Cantor (1982). Noller et al. use the symbol "/" to mean "is base paired to". Because psoralen crosslinked nucleotides are not base paired to each other but rather to adjacent residues, we shall use the symbol "x" to denote "is crosslinked to". Wollenzein & Cantor have adopted a prefix to signify that their psoralen crosslinks (Ps) have been localized by the electron microscope (E). The crosslinks described here have been resolved using gel techniques (G). Thus, if bases 625 and 1420 are crosslinked by psoralen, this interaction is written as GPs 625 x 1420. In the following discussion, the three most widely cited secondary structure models will be called, for simplicity's sake, the American model (Noller & Woese, 1981), the German model (Zwieb et al., 1981), and the French model (Stiegler et al., 1981).

Method of Crosslink Assignment:

Assignment of crosslinks to specific bases in the following section was done with a number of criteria in mind. First, and most obvious, was that the crosslink must occur within the isolated fragments. Possible secondary structures between the two fragments were then determined by hand. These possibilities were discriminated on the basis of phylogenetic conservation and the presence of suitable psoralen crosslinking sites. Crosslinking sites were

assumed to be between uridines only. While psoralen can certainly react with cytidine and even purines to some extent (Thompson, 1982), reaction with polymers has shown uridine to be the preferred site. Hot spots for monoaddition and crosslinking (Youvan & Hearst, 1982; Thompson et al., 1981), have all been at uridines. To some extent, looking for uridine-uridine crosslinks and then using these to prove the specificity of psoralen is a circular argument. However, the great ease in finding such sites argues in favor of this bias.

Phylogenetic comparisons were made using the best alignments available in the literature. The secondary structure alignments of Zwieb et al. (1981) and Stiegler et al. (1981) were found to be the most useful while the primary structure alignment of Kuntzel and Kochel (1981) was also helpful. The assignments below are separated into three categories: confirmatory crosslinks which are present in all models, discriminatory crosslinks present in some models, and new crosslinks.

Confirmatory Crosslinks:

GPs 594 x 644: This crosslink confirms the presence of 588-617/623-651. While the presence of the helical structure is definite, the resolution of the crosslink does not allow an unambiguous assignment because there are two other likely crosslinking sites in the secondary structure shown in Figure 3. Also possible are GPs 598 x 641 and GPs 603 x 636. The secondary structure drawn is present in all three models in the literature. A UV-induced crosslink was also found in this region (Zwieb & Brimacombe, 1981).

GPs 1351 x 1372: This crosslink could also occur as GPs 1348 x 1376. 1350-1356/1366-1372 is present in all three models while 1347-1349/1376-1378 is found only in the German model. This helical region has also been

crosslinked by UV light (Zwieb & Brimacombe, 1981).

GPs 252 x 273: This is the only likely site of crosslinking within the fragment found. It is near the end of a helix terminated by two A·U pairs so that psoralen can easily intercalate. The secondary structure proposed in all three models is identical in this region except for an additional G·U pair in the American model. The crosslink is isolated as a single fragment despite the presence of accessible guanines in the hairpin loop.

GPs 458 x 473: This fragment was also isolated as a single fragment. The three consecutive uridines opposite a G·U pair are ideal for crosslinking. This helix is not present in many species and has been shown to be near the bound mRNA (Wagner et al., 1976). This same psoralen crosslink has also been mapped by Turner et al. (1982) by a slightly different gel technique.

GPs 1007 x 1023: This crosslinking site is ideal for intercalation of psoralen. It contains a run of four uridines with a G·U pair at the end of the helix. This helix is present in all models and is supported by compensating base changes in other species.

GPs 1240 x 1298: This crosslinking site is at the base of an extended helix present in all three models. There are two potential psoralen sites within the proposed helix which are virtually identical and thus cannot be distinguished.

GPs 1308 x 1330: The hairpin stem proven by this crosslink is present in all models. The only uridine-uridine crosslink possible involves the terminal uridine on one strand with the first uridine beyond the helix on the opposite strand. This type of interaction might even be preferred over normal intercalation. There would be no unwinding of the helix necessary and nearly as much stabilization by stacking would be gained. Stacking from the terminal base pair would be normal and some would also be gained from the next unpaired

residues . Nucleotide 1308 was previously observed to be a hot spot for monoaddition by Youvan & Hearst (1982).

Discriminatory Crosslinks

GPs 1116 x 1183: The German model contains no sites of potential crosslinking that would generate the observed fragments, while the newest version of the American model (H. Noller, personal communication) and the French model have two each. The most likely site in the American model (Figure 4) involves the terminal uridine in a helix with a uridine in a G·U pair. Also possible is a crosslink between terminal uridines in a coaxial stack (GPs 1118 x 1183). This helical region is deleted in some mitochondria but conserved in other species. A very similar base-pairing scheme is present in the French model. In this model, GPs 1115 x 1183 would involve the terminal uridine in a helix with the first uridine beyond the helix. This type of crosslink has been proven for GPs 956 x 1506 and is likely for GPs 1308 x 1330. The other possible crosslink would arise if the helices formed by 1118-1124/1149-1155 and 1063-1067/1184-1193 stacked on each other. The terminal uridines would then be suitably positioned for crosslinking. These four crosslinks cannot be distinguished on the basis of crosslinking data available.

GPs 1189 x 1202: The helical region identified by this crosslink was present in two stretches in the original American model but has since been deleted (H. Noller, personal communication). It was not present at all in the French or German models. If GPs 1189 x 1202 is to occur simultaneously with GPs 1116 x 1183, the helices in the original American model need to be

shortened as shown in Figure 4.

GPs 14 x 921: This crosslink must be classified as tentative because of problems in identifying one strand. The sequence for 918-941 is definite; however, 7-46 could not be read clearly enough to be sure of its uniqueness. Since nucleotides 920 and 921 are the only uridines in the first sequence, one of them is most likely involved in the crosslink. If 17-20/915-918, present in the American model, is extended by four base pairs with a single-base bulge, a good crosslinking site would be present. This crosslink may have been observed in early electron microscopic work (Wollenzein et al., 1979) and erroneously classified as EPs 530 x 1540. The latter crosslink has been confirmed with a more rigorous polarity assignment recently (Wollenzein & Cantor, 1982); thus the presence of this crosslink does not contradict the original work.

New Crosslinks

GPs 358 x 1330: While there are two stretches of complimentarity between the crosslinked fragments found (330-340/1333-1343 and 357-362/1325-1330), only one of these is conserved in other species as well (Figure 5). In this region, there is only one likely crosslinking site. In the American and German models, both of these regions are involved in other interactions. In the French model, one of the regions is involved in another interaction. This crosslink had been observed previously in the electron microscope and mapped as EPs 353 x 1344 (Cantor et al., 1980).

In order to conserve this interaction in Z. mays chloroplasts, one strand must be offset by two bases and a one-base bulge introduced. One additional base pair can be made and more G·C pairs are present so that this structure is actually more stable than in E. coli. Similar structures can

be drawn for the homologous regions in eukaryotic RNAs. In mitochondria, however, only very weak interactions are found. Human and yeast mitochondriae are only stable by 5.5 and 5.6 kcal/mol while no reasonable structure at all is present in mouse mitochondria. Secondary structure calculations were done by the method of Tinoco et al. (1973) using the most recent values available (I. Tinoco, Jr., personal communication).

It does not seem likely that coaxial stacking would generate this interaction. 368-379/384-393 has a suitably placed terminal uridine but there is no structure between 1317 and 1343 with a terminal uridine except for 1301-1305/1335-1339. This is present only in the German model and would necessitate stacking within a bulged loop of an extended helix.

GPs 625 x 1420: Overlapping T_1 fragments have allowed the mapping of this crosslink to high resolution. There is only one good crosslinking site in the secondary structure shown (Figure 6). A similar structure is present in Z. mays chloroplasts. One base is bulged but additional G·C pairs more than make up for this. Deciding which structures are homologous is difficult in eukaryotes. Most of the additional sequences which have been inserted in 18S RNA are added between bases 600 and 640 of E. coli. Because of this, no region is strictly homologous to bases 620-626. There is no such difficulty with the other crosslinked fragment. When the inserted sequence is scanned for complementarity with the known homologous region, one stretch can be found that has good base pairing. This is shown in Figure 6. Finding the appropriate secondary structures for mitochondrial RNAs is much simpler. Homologous regions for both crosslinked strands have been deleted. This crosslink may correspond to Feature X found by electron microscopy and mapped at EPs 740 x 1370 (Wollenzein et al., 1979).

GPs 956 x 1506: This crosslink is the most accurately known of those

described. Not only were the fragments in Table I found, but, using a technique involving complete T_1 digestion, followed by running in a 2D gel system similar to that described by Turner et al. (1982), the exact T_1 fragments were found. Even with this knowledge, we cannot distinguish between two different structures. 946-955/1225-1235 could stack on 1506-1515/1520-1529 such that uridines 1506 and 955 could crosslink. This structure would not be a true coaxial stack because the orientations of the two helices are incorrect. Nonetheless, a crosslinkable conformation can certainly be envisioned. Alternatively, an interaction which is not present in any of the current models and would, in fact, necessitate the melting of one of the most highly conserved features of small sub-unit RNAs, the m_2^6A hairpin structure, can be drawn. The high degree of conservation of the interaction, shown in Figure 7, indicates that this helix is significant and is not an artifact caused by the crosslinking or the conditions used. One exception to the conservation of this feature has been found. The chloroplast RNA of C. reinhardii has three base substitutions which destroy much of the base pairing. Whether this absence is related to the unusual structure of the large subunit RNA is not known (Dron et al., 1982). The functional significance of this structure will be discussed elsewhere. This feature was mapped by electron microscopy and was the most prevalent crosslink found (Wollenzein et al., 1979). The locations of the observed crosslinks are summarized in Figure 8.

Dynamics of Long Range Crosslinks

The most surprising aspect of the long range interactions which have been observed is the presence of multiple structures for the same nucleotides. These structures are probably in equilibrium in the conditions used.

In the functioning ribosome, such conformational switches might be used to produce the physical movement necessary in translation. Evidence for switches has been found by Glotz et al. (1981). Psoralen may react particularly well with this type of structural feature because, by necessity, the helices must be relatively free to move and unwind. Since the helices are not tightly constrained by other parts of the RNA, psoralen is able to intercalate and crosslink.

Two of the crosslinks found, GPs 1189 x 1202 and GPs 1116 x 1183, are in a region in which there is evidence that large conformational changes occur (Glotz et al., 1981). The data from crosslinking and denaturation studies could not all be incorporated into a single secondary model. When our data are combined with that of Glotz et al. (1981), it is quite apparent that large conformational changes are required to explain the data. Possibly linked to this structural change is the crosslink GPs 358 x 1330. This crosslink is between regions similar to those in one interaction postulated by Glotz et al. (1981). The two interactions could easily co-exist and provide a stable link between the two regions of the RNA. Extensive conformational changes can be drawn which combine the data of all three models as well as the evidence presented here. GPs 1308 x 1330 must also be in dynamic equilibrium because it obviously could not co-exist with GPs 358 x 1330.

GPs 625 x 1420 also appears to be in a helix which undergoes conformational change. The formation of the helix shown in Figure 6 would require the opening of two other helices. The conformational change may be facilitated by the binding of protein S8. S8 has been shown to strongly protect 588-603/636-651 upon binding (Ungewickall et al., 1975); it also induces a large conformational change when bound (Nomura et al., 1969). This change could very well be the bringing together of the regions around

620 and 1420. S8 does not necessarily have to bind directly to the region around 1420; instead, this interaction could be mediated by another protein.

Additional evidence for this conformational switch can be deduced from the work of Stark et al. (1982). They have studied the effect of deletion mutations on the processing and function of E. coli 16S RNA. One of these mutations, a single base deletion at position 614, has a profound effect. Very little of this RNA is processed correctly or incorporated into 30S subunits. While this base is not directly involved in the long range interaction described, it would have an effect on the equilibrium of 612-617/623-628 620-626/1420-1426. The stability of the first helix and hairpin loop is 12.0 kcal/mol while the second helix is 11.8 kcal/mol. Omitting base 614 does not change the stability of the second helix, but the first structure becomes more stable because an additional base pair can form (14.1 kcal/mole). Instead of being slightly more stable, it is considerably stronger than the second helix after deletion. The same type of finely tuned equilibrium can be seen in Z. mays chloroplasts. The homologous helices have stabilities of 14.6 and 14.1 kcal/mol. Eukaryotes behave similarly with S. cerevisiae, having stabilities of 7.4 and 7.2 kcal/mol, and X. laevis, having stabilities of 8.7 and 7.8 kcal/mol. While the sequence and stabilities of these helices in the four species vary widely, the tendency of the short range interaction, to be slightly stronger than the long range interaction is uniform.

In the above analysis, the German secondary structure model has been used in all cases. The stabilities were calculated for the entire helix, including the hairpin loop, from where the disruption caused by the long range interaction would occur. Additional stability caused by base pairing of the other two sequences involved in the short range interactions has not

been included because there is no information on this. Disruption of the 1409-1430/1470-1491 helix has not been calculated either. This contribution is small because, of the seven base pairs needed to be broken for E. coli, four are G·U pairs and two are A·U pairs. Whether this structure is present as drawn in the German model is also subject to debate.

The helix crosslinked by GPs 956 x 1506 would also need to be in dynamic equilibrium. It would require the unpairing of the highly conserved $m_2^6Am_2^6A$ hairpin loop. That this helix should open is not surprising in view of the results of Van Charldorp et al. (1981). They found that the presence of the four methyl groups destabilized the helix because of the greater energy required to unstack them for placement in a hairpin loop compared to unmodified adenines. This destabilization would not be present in 950-956/1507-1513 because this helix would not introduce a loop, hence the two m_2^6As could effectively stack.

Other Methods of Psoralen Crosslink Mapping: Electron microscopy has been the most powerful tool for localizing psoralen crosslinks up until now. This technique is limited primarily by resolution. Short range crosslinks cannot be observed at all while long range crosslinks cannot be mapped as accurately as described here. Now that the type of site that psoralen prefers to crosslink is known better, mapping can be done with somewhat more confidence. Other gel techniques have also been used (Thompson et al., 1981; Turner et al., 1982). While providing valuable information, these techniques are simply not as versatile as those described here.

Complete digestion of fragments with T_1 provides higher resolution but short fragments are sometimes impossible to place in the sequence. Complete T_1 digestion followed by reversal could be employed in finding GPs 956 x 1506

but only because it was present in high yield (Wollenzein et al., 1979) and contained two relatively long T_1 fragments.

CONCLUSIONS

The large number of psoralen crosslinks which have been localized in this study clearly shows the power of this technique for determining the secondary structure of any large RNA. The technique is also simple enough that most laboratories can readily employ it on any RNA of interest.

The results presented here, while supporting the generally accepted models for 16S RNA structure, also clearly point out the dynamic nature of the molecule. Only when we can correlate conformational changes with specific events will we begin to understand the ribosomal machinery. The value of psoralen crosslinking in future studies is high. Psoralen crosslinks precisely those features which are most difficult to observe using other techniques.

EXPERIMENTAL PROCEDURES

E. coli MRE 600 cells were grown as by Traub et al. (1971). Frozen cells (1 g) were suspended in 10 ml of 50 mM sodium acetate (pH 5.0) .5% SDS and 10 mM VRC (vadanylribonucleoside complex; Berger & Berkenmeier, 1979). Cells were homogenized in a ground glass tissue homogenizer until the solution became viscous. DNase (Worthington, RNase free) was added to about 50 µg/ml and homogenization continued until the solution was no longer viscous. An equal volume of redistilled phenol was added. Phenol extraction was repeated at least three more times (until the water-phenol interface was clear). The solution was made .2 M NaCl and ethanol precipitated at -20° twice.

The precipitated RNA was dissolved in a minimum volume of .1 M LiCl, 10 mM EDTA, 0.5% SDS, 10 mM Tris base (LES). Approximately 5 mg was placed on top of a 15-30% sucrose gradient containing LES buffer. The RNA was centrifuged for 24 hr at 27,000 rpm in an SW27.1 rotor. Bands corresponding to 23S RNA and 16S RNA were resolved by pumping the solution through a Beckman Analytical Optical Unit (254 nm). 16S RNA fractions were combined according to purity and then precipitated. When necessary, the centrifugation was repeated in order to get pure 16S RNA. The 5S/tRNA band was sometimes obscured by residual VRC.

Purity and intactness of the RNA was examined by gel electrophoresis in 4% polyacrylamide (20:1, w/w, acrylamide to bis-acrylamide), 7 M urea gels buffered with 50 mM Tris-borate (pH 8.3), and 10 mM EDTA. Polymerization was catalyzed by 0.075 g ammonium persulfate and 50 µl TEMED per 100 ml. All gel materials were purchased from Bio Rad except urea which was from Schwarz-Mann.

Before crosslinking, 16S RNA was incubated for 30 min or more at 37° in TMA I buffer (10 mM $MgCl_2$, 100 mM NH_4Cl, 10 mM Tris-HCl, pH 7.2, 14 mMβ -mercaptoethanol). Crosslinking was done in the apparatus described previously (Thompson et al., 1981) at 10°. Three additions of HMT were made from a stock solution in DMSO (2 mg/ml). HMT was obtained from HRI Associates (Emeryville, CA), either unlabelled or labelled with ^3H at 4 x 10^7 cpm/µg. Each addition made the aqueous solution 20 µg/ml in HMT and additions were separated by 10 min. This protocol produced an average of 3-5 HMTs per 16S molecule. After phenol extraction and ethanol precipitation, the RNA was redissolved in 50 mM Tris-HCl (pH 8.5), 10 mM $MgCl_2$ at a concentration of 10 OD_{260}/ml. This was digested for 2 hr at 37° with 100 u/ml T_1 RNase

(Sigma). This was phenol extracted and ethanol precipitated.

The digested RNA was then separated by two different methods. Method one involved running the RNA through an RPC-5 column (0.4 x 20 cm; Astro Enterprises) and eluting with 80 ml of a 0.1 to 2 M KCl gradient. The KCl solutions also contained 2 mM sodium thiosulfate, 10 mM EDTA, .06% sodium azide, and 10 mM Tris-HCl (pH 6.8). One ml fractions were collected and ethanol precipitated. Later fractions required dialysis to remove excess KCl before precipitation. Method two involved running the RNA through a BND cellulose column (0.5 ml. Sigma) with a 15 ml 0% DMSO, 0.3 M NaCl to 25% DMSO, 0.65 M NaCl gradient for elution.

The purified, precipitated RNA was redissolved in 25 µl of 50 mM Tris-HCl (pH 8.5), 10 mM $MgCl_2$, 13 mM β-mercaptoethanol, 2 mM spermine and labelled wih 0.5 mCi of γ-^{32}P ATP (ICN, crude) and 2 units of polynucleotide kinase overnight at 37°. The sample was then made 1 M in urea, heated at 90-95° for 1 min, fast cooled on ice and loaded onto a .08 x 15 x 40 cm 12% polyacrylamide gel (made like the 4% gel described earlier except no urea was present). This gel was run at room temperature until the bromophenol blue had migrated off the gel (about 7 hr at 800 V). The upper glass plate was removed and the gel covered with plastic wrap. The lane(s) containing the RNA was cut using a razor blade and a straight-edge. The lower 30 cm was placed on a 34 x 43 cm plate. After removal of the plastic wrap, a 20% polyacrylamide (30:1,acrylamide to bis), 7 M urea gel was polymerized around the gel strip. The gel was buffered at twice the salt concentration of the first dimension gel. The second dimension was run warm until the xylene cyanol marker had migrated off the gel. The gel was covered

with plastic wrap and autoradiographed with Kodak XAR-5 film. Cross-links were reversed in the gel by exposure to a 40 W germicidal lamp at a distance of about 10 cm for 2 hr. After reversal, fragments of interest were cut from the gel and eluted in 200 µl of LES buffer for 6-16 hours twice. Ten micrograms of tRNA were added and the samples ethanol precipitated. The RNA was run on another 20% polyacrylamide, 7 M urea gel to separate the previously crosslinked fragments. After autoradiography, the fragments of interest were eluted from the gel as described above.

Fragments were sequenced enzymatically. Digestions were carried out at 55° for 15 min in 10 µl which contained 20 mM sodium citrate, 1 mM EDTA, and 2 µg carrier tRNA. In addition, the RNase T_1 (G specific), U_2 (A specific), and Phy M (A and U specific) contained 7 M urea while the B. cereus (U and C specific) did not. The U_2 reaction was buffered at pH 3.5 and the others at pH 5.0. Phy M and B. cereus were purchased from P. L. Biochemicals and used as described. T_1 and U_2 were purchased from Sigma and dissolved in stock solutions at 50 u/ml. One microliter of the stock solution was needed for each digestion. After reaction, the samples were run on 20% polyacrylamide gels and autoradiographed using Dupont Lightening Plus intensifying screens.

ACKNOWLEDGEMENTS

We would like to thank Dr. George Cimino for growing the E. coli cells used in this study, and for suggesting the use of BND cellulose,

Drs. Peter Butler and Robert Traut for their invaluable assistance in preparation of E. coli ribosomes used in the initial stages of these experiments, and Drs. Harry Noller and Stephen Beckendorf for useful comments about the manuscript. This work was supported by the National Institutes of Health Grant #GM-11180 and by the Office of Energy Research, Office of Health and Environmental Research, Health Effects Research Division of the U.S. Department of Energy under Contract Number DE-AC03-76SF00098.

REFERENCES

Bachellerie, J. -P., Thompson, J. F., Wegnez, M. R., and Hearst, J. E. (1980). Identification of the modified nucleotides produced by covalent photoaddition of hydroxymethyltrimethylpsoralen to RNA. Nucleic Acids Res. 9, 2207-2222.

Bachellerie, J. -P., and Hearst, J. E. (1982). Specificity of the photoreaction of 4'-(hydroxymethyl)-4,5',8-trimethylpsoralen with ribonucleic acid. Identification of reactive sites in Escherichia coli phenylalanine-accepting transfer ribonucleic acid. Biochemistry 21, 1357-1363.

Berger, S. L., and Birkenmeier, C. S. (1979). Inhibition of intractable nucleases with ribonucleoside-vanadyl complexes: Isolation of messenger ribonucleic acid from resting lymphocytes. Biochemistry 18, 5143-5149.

Brosius, J., Palmer, M. L., Kennedy, P. J., and Noller, H. F. (1978). Complete nucleotide sequence of a 16S ribosomal RNA gene from Escherichia coli. Proc. Nat. Acad. Sci. USA 75, 4801-4805.

Cantor, C. R., Wollenzein, P. L., and Hearst, J. E. (1980). Structure and topology of 16S RNA. An analysis of the pattern of psoralen cross-linking. Nucliec Acids Res. 8, 1855-1872.

Carbon, P. Ehresmann, C., Ehresmann, B., and Ebel, J. -P. (1979). The complete nucleotide sequence of the ribosomal 16S RNA from Escherchia coli. Eur. J. Biochem. 100, 399-410.

Carbon, P., Ebel, J. -P., and Ehresmann, C. (1981). The sequence of the ribosomal 16S RNA from Escherchia coli. Nucleic Acids Res. 9, 2325-2333.

Dron, M., Rahire, M., Rochaix, J.-D. (1982). Sequence of the chloroplast 16S rRNA gene and its surrounding regions of Chlamydomonas reinhardii. Nucleic Acids Res. 10, 7609-7620.

Edelmann, P., and Gallant, J. (1977). Mistranslation in E. coli. Cell 10, 131-138.

Eperon, I. C., Anderson, S., and Nierlich, D. P. (1980). Distinctive sequence of human mitochondrial ribosomal RNA genes. Nature 286, 460-464.

Fellner, P., Ehresmann, C., and Ebel, J. -P. (1970). Nucleotide sequences present within the 16S ribosomal RNA of Escherichia coli. Nature 225, 26-29.

Glotz, C., and Brimacombe, R. (1980). An experimentally-derived model for the secondary structure of the 16S ribosomal RNA from Escherichia coli. Nucl. Acid. Res. 8, 2377-2395.

Kanne, D., Straub, K., Rapoport, H., and Hearst, J. E. (1982). Psoralen-deoxyribonucleic acid photoreaction. Characterization of the monoaddition products from 8-methoxypsoralen and 4,5',8-trimethylpsoralen. Biochemistry 21, 861-871.

Kim, S. -H., Sussman, J. L., Suddath, F. L., Quigley, G. J., McPherson, A., Wang, A., Seeman, N. C., and Rich, A. (1974). The general structure of transfer RNA molecules. Proc. Nat. Acad. Sci. USA 71, 4970-4979.

Kuntzel, H., and Kochel, H. G. (1981). Evolution of rRNA and origin of mitochondria. Nature 293, 751-755.

Noller, H. F., and Woese, C. R. (1981). Secondary structure of 16S ribosomal RNA. Science 212, 403-411.

Noller, H. F., Kop, J., Wheaton, V., Brosius, J., Gutell, R. R., Kopylov, A. M., Dohme, F., and Herr, W. (1981). Secondary structure model for 23S

ribosomal RNA. Nucleic Acids. Res. 9, 6167-6189.

Nomura, M., Mizushima, S., Ozaki, M., Traub, P., and Lowry, C. V. (1969). Structure and function of ribosomes and their molecular components. Cold Spring Harbor Symp. Quant. Biol. 34, 49-61.

Rabin, D., and Crothers, D. M. (1979). Analysis of RNA secondary structure by photochemical reversal of psoralen crosslinks. Nucleic Acids Res. 7, 689-703.

Ross, A., and Brimacombe, R. (1979). Experimental determination of interacting sequences in ribosomal RNA. Nature 281, 271-276.

Rubtsov, P. M., Musakhanov, M. M., Zakharyev, V. M., Krayev, A. S., Skyrabin, K. F., and Bayev, A. A. (1980). The structure of the yeast ribosomal RNA genes. I. The complete nucleotide sequence of the 18S ribosomal RNA gene from saccharomyces cerevisiae. Nucleic Acids Res. 8, 5779-5795.

Salim, M., and Maden, B. E. H. (1981). Nucleotide sequence of Xenopus laevis 18S ribosomal RNA inferred from gene sequence. Nature 291, 205-208.

Schwarz, Z., and Kossel, H. (1980). The primary structure of 16S rDNA from Zea mays chloroplast is homologous to E. coli 16S rRNA. Nature 283, 739-742.

Sedat, J. W., Kelly, R. B., and Sinsheimer, R. L. (1967). Fractionation of nucleic acid on benzoylated-napthalated DEAE cellulose. J. Mol. Biol. 26, 537-540.

Sor, F., and Fukuhara, H. (1980). Sequence nucleotidique du gene de l'ARN ribosomique 15S mitochondrial de la levure. C. R. Acad. Sci. (Paris) Ser. D. 291, 933-936.

Stark, M. J. R., Ghourse, R. L., and Dahlberg, A. E. (1982). Site directed mutagenesis of ribosomal RNA. Analysis of ribosomal RNA deletion

mutants using maxicells. J. Mol. Biol. 159, 417-439.

Stiegler, P., Carbon, P., Ebel, J. -P., and Ehresmann, C. (1981). A general secondary structure model for procaryotic and eucaryotic RNAs of the small ribosomal subunits. Eur. J. Biochem. 120, 487-495.

Straub, K., Kanne, D., Hearst, J. E., and Rapoport, H. (1981). Isolation and characterization of pyrimidine-psoralen photoadducts from DNA. J. Am. Chem. Soc. 103, 2347-2355.

Thompson, J. F., Wegnez, M. R., and Hearst, J. E. (1981). Determination of the secondary structure of Drosophila melanogaster 5S RNA by hydroxymethyltrimethyl psoralen crosslinking. J. Mol. Biol. 147, 417-436.

Thompson, J. F., Bachellerie, J. -P., Hall, K., and Hearst, J. E. (1982). Dependence of 4'-(Hydroxymethyl)-4,5',8-trimethylpsoralen photoaddition on the conformation of ribonucleic acid. Biochemistry 21, 1363-1368.

Thompson, J. F. (1982). Reaction of psoralen with RNA: Specificity and use as a probe for secondary structure analysis. Ph.D. Thesis, University of California, Berkeley, California.

Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M., and Gralla, J. (1973). Improved estimation of secondary structure in ribonucleic acids. Nature New Biol. 246, 40-41.

Traub, P., Mizushima, S., Lowery, C. V., and Nomura, M. (1971). Reconstitution of ribosomes from subribosomal components. Meth. Enzymol. 20, 391-401.

Turner, S., Thompson, J. F., Hearst, J. E., and Noller, H. F. (1982). Identification of a site of psoralen crosslinking in E. coli 16S ribosomal RNA. Nucleic Acids Res. 10, 2839-2849.

Ungewickall, E., Garrett, R., Ehresmann, C., Stiegler, P., and Fellner, P. (1975). An investigation of the 16S RNA binding sites of ribosomal proteins S4, S8, S15, and S20 from Escherichia coli. Eur. J. Biochem. 51, 165-180.

Van Charldorp, R., Heus, H.A., Van Knippenberg, P.H., Joordens, J., DeBruin, S.H., & Hilbers, C.W. (1981). Destabilization of secondary structure in 16S Ribosomal RNA by dimethylation of two adjacent adenosines. Nucl. Acids. Res. 9, 4413-4422.

Van Etten, R. A., Walberg, M. W., and Clayton, D. A. (1980). Precise localization and nucleotide sequence of the two mouse mitochondrial rRNA genes and three immediately adjacent novel tRNA genes. Cell 22, 157-170.

Wagner, R., Gassen, H. G., Ehresmann, C., Stiegler, P., and Ebel, J. -P., (1976). Identification of a 16S RNA sequence located in the decoding site of 30S ribosomes. FEBS. Let. 67, 312-315.

Wollenzein, P. L., Hearst, J. E., Thammana, P., and Cantor, C. R. (1979). Base-pairing between distant regions of the Escherichia coli 16S ribosomal RNA in solution. J. Mol. Biol. 135, 255-269.

Wollenzein, P. L., and Cantor, C. R. (1982). Gel electrophoretic technique for separating crosslinked RNAs. Application to improved electron microscopic analysis of psoralen crosslinked 16S ribosomal RNA. J. Mol. Biol. 159, 151-166.

Youvan, D. C., and Hearst, J. E. (1982). Sequencing psoralen photochemically reactive sites in Escherichia coli 16S rRNA. Anal. Biochem. 119, 86-89.

Zwieb, C., and Brimacombe, R. (1980). Localization of a series of intra-RNA cross-links in 16S RNA, induced by ultraviolet irradiation of Escherichia coli 30S ribosomal subunits.

Zwieb, C., Glotz, C., and Brimacombe, R. (1981). Secondary structure comparisons between small subunit ribosomal RNA molecules from six different species. Nucleic Acids Res. 9, 3621-3640.

TABLE I

Locations of Crosslinked Oligonucleotides

Fragment 1		Fragment 2	
5' Ends	3' Ends	5' Ends	3' Ends
	Confirmatory Crosslinks		
238	289	—	—
435	481	—	—
576,592	616	617	645,650,655
994	1015	1016	1043
1317,1335	1361	1362	1379,1385
1280	1316	1317	1361
1234	1255	1280	1304
	Discriminatory Crosslinks		
1100	1127	1167,1179	1206,1215
1167,1179	1206,1215	---	---
7	46*	918	941
	New Crosslinks		
327	369	1317	1343
618	645	1405,1418	1432,1453
955	976	1498	1514

* This crosslinked oligonucleotide assignment is tentative. See Discussion

Multiple 5' and 3' ends are found because most crosslinks have been sequenced more that once with differing termani present in the different diagonal spots.

FIGURE LEGENDS

Figure 1. Schematic diagram of methods used to analyze crosslinks.

Figure 2. Autoradiograms of two-dimensional gels. A) 2D gel of uncrosslinked 16S RNA. Heavy T_1 digestion results in very little large material near the top of the gel. The "X" marks the position of xylene cyanole. B) 2D gel of crosslinked 16S RNA. This sample was treated exactly as that in A except it was crosslinked as described in Materials and Methods. Discrete off diagonal spots result from heavy T_1 digestion. C) 2D gel of fraction from RPC-5 column. Note the two diagonals, both of which are heavier near bottom of the gel. D) 2D gel of fraction from BND cellulose column The family of dark, off diagonal spots near the center are all GPs 1116 x 1183. Other fractions yield gels with a much different pattern of off diagonal spots.

Figure 3. Sequences of confirmatory crosslinks arranged in proposed secondary structures. The entire fragments of crosslinks present in published structure models are shown. Arrows mark the locations of observed T_1 cuts.

Figure 4. Sequences and secondary structures of discriminatory crosslinks. Arrows show the positions of T_1 cuts. Three nucleotides have been added to the observed fragments of GPs14 x 921 in order to show the entire helical stem.

Figure 5. Phylogenetic conservation of GPs 358 x 1330. The homologous secondary structures for 6 species of small subunit RNAs are shown. Base

changes in going from the species on the left to the one on the right are shown by arrows.

Figure 6. Phylogenetic conservation of GPs 625 x 1420. In human and yeast mitochondrial RNA this interaction is deleted.

Figure 7. Phylogenetic conservation of GPs 956 x 1506.

Figure 8. Positions of observed crosslinks. The locations of all thirteen crosslinks are superimposed over the skeleton of the secondary structure model presented by Noller and Woese (1981).

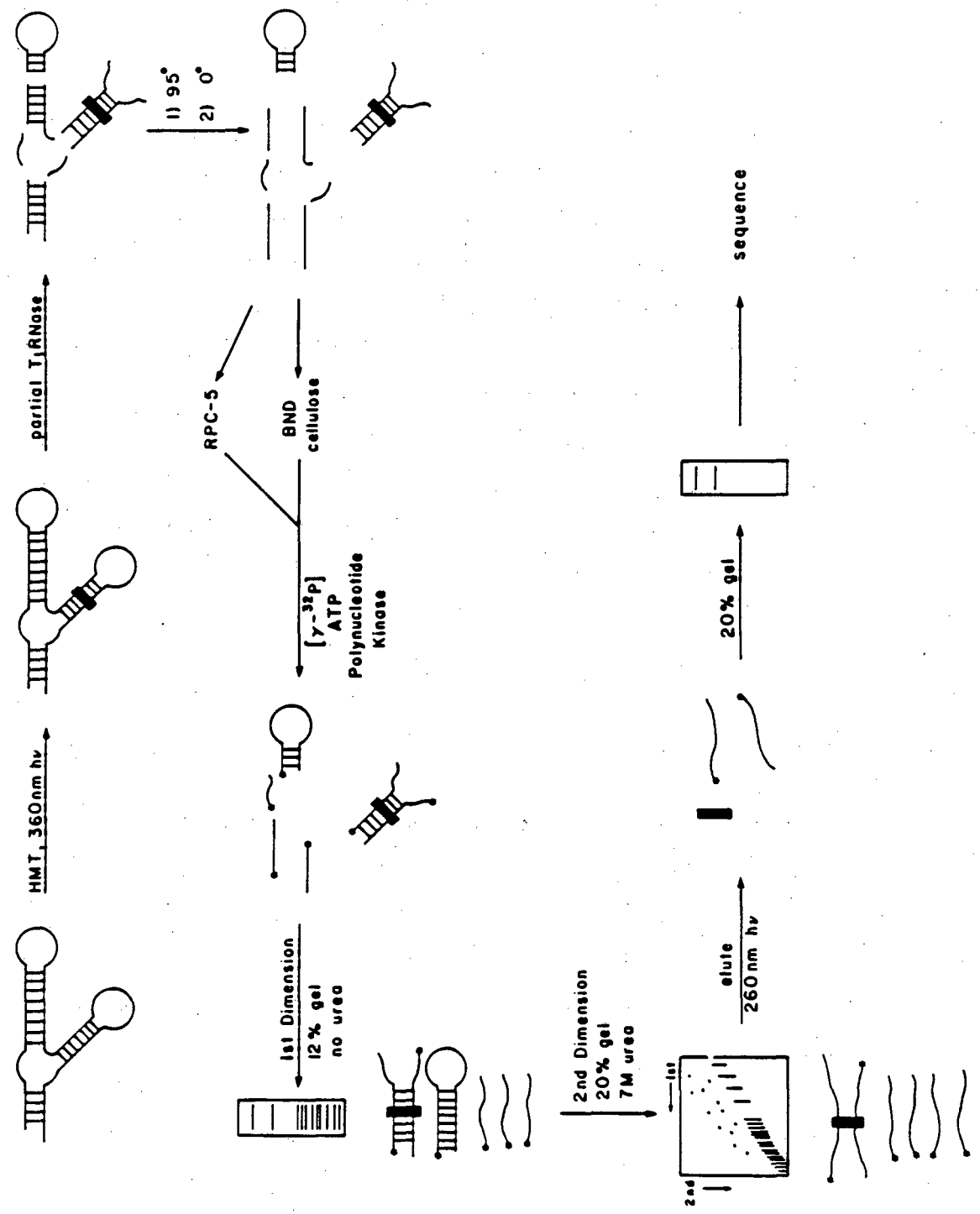

HMT, 360nm hv
partial T1RNase
1) 95°
2) 0°
RPC-5
BND cellulose
[γ-32P] ATP Polynucleotide Kinase
1st Dimension
12% gel
no urea
2nd Dimension
20% gel
7M urea
elute
260nm hv
20% gel
sequence

Fig. 1

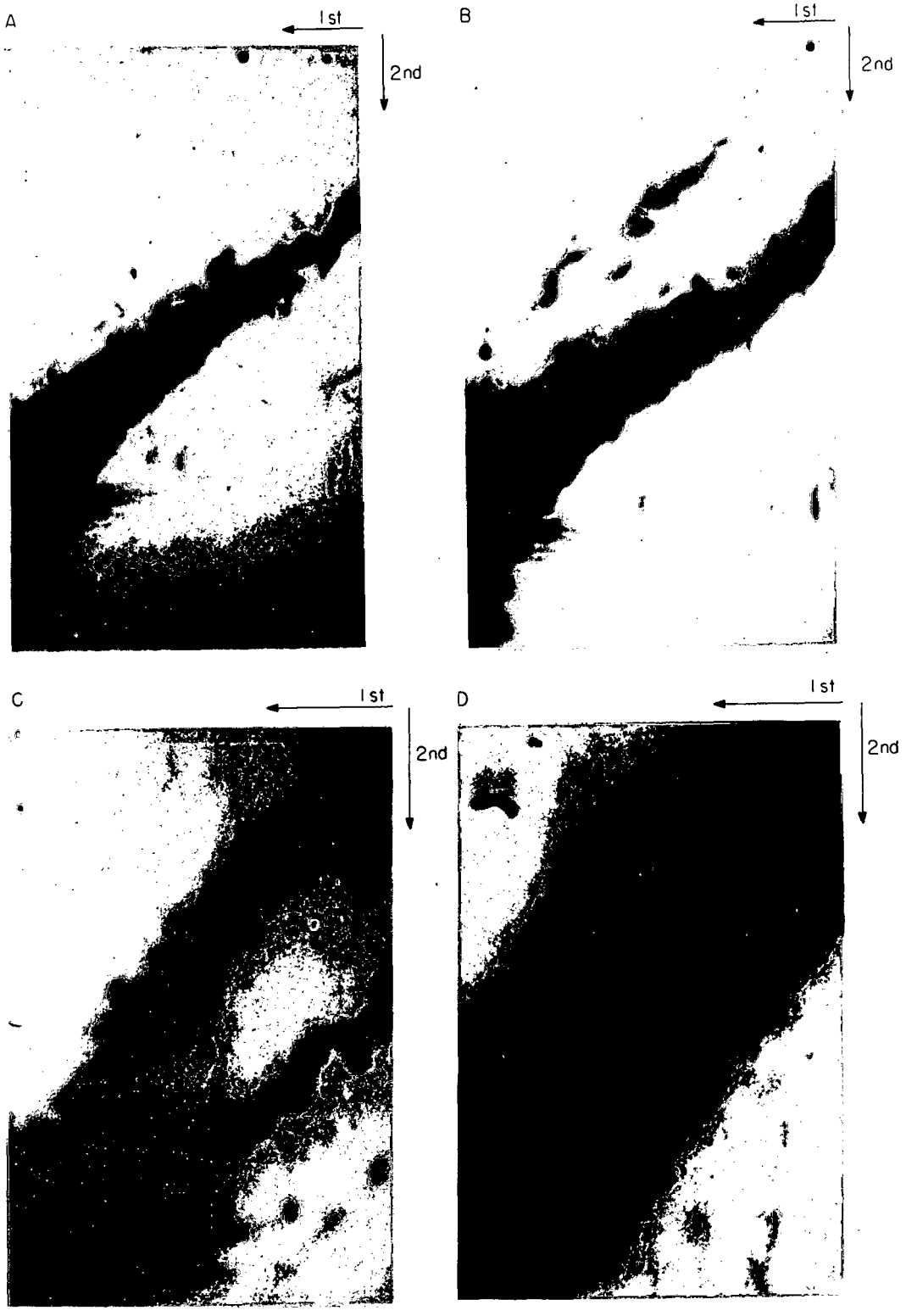


Fig. 2

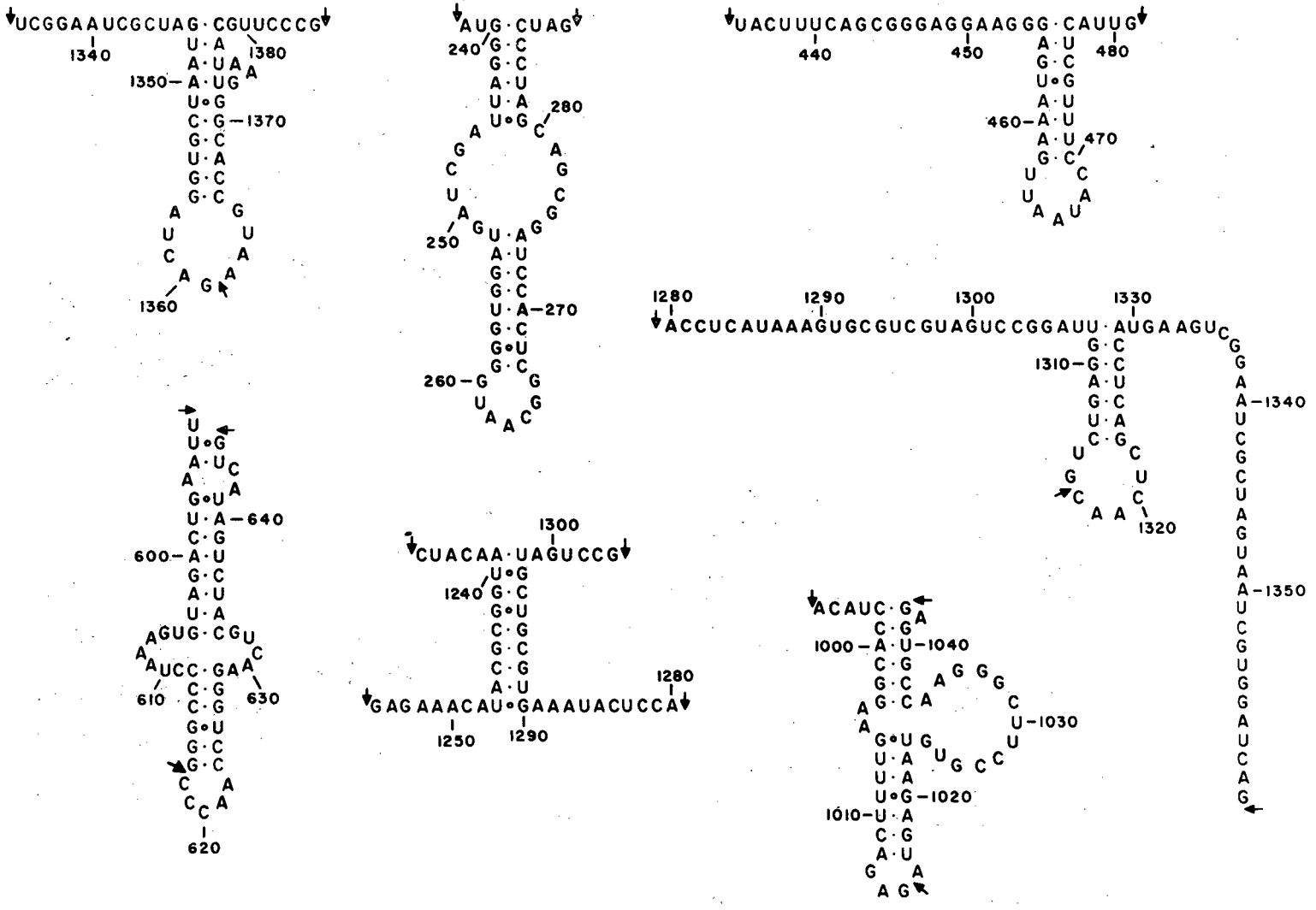


Fig. 3

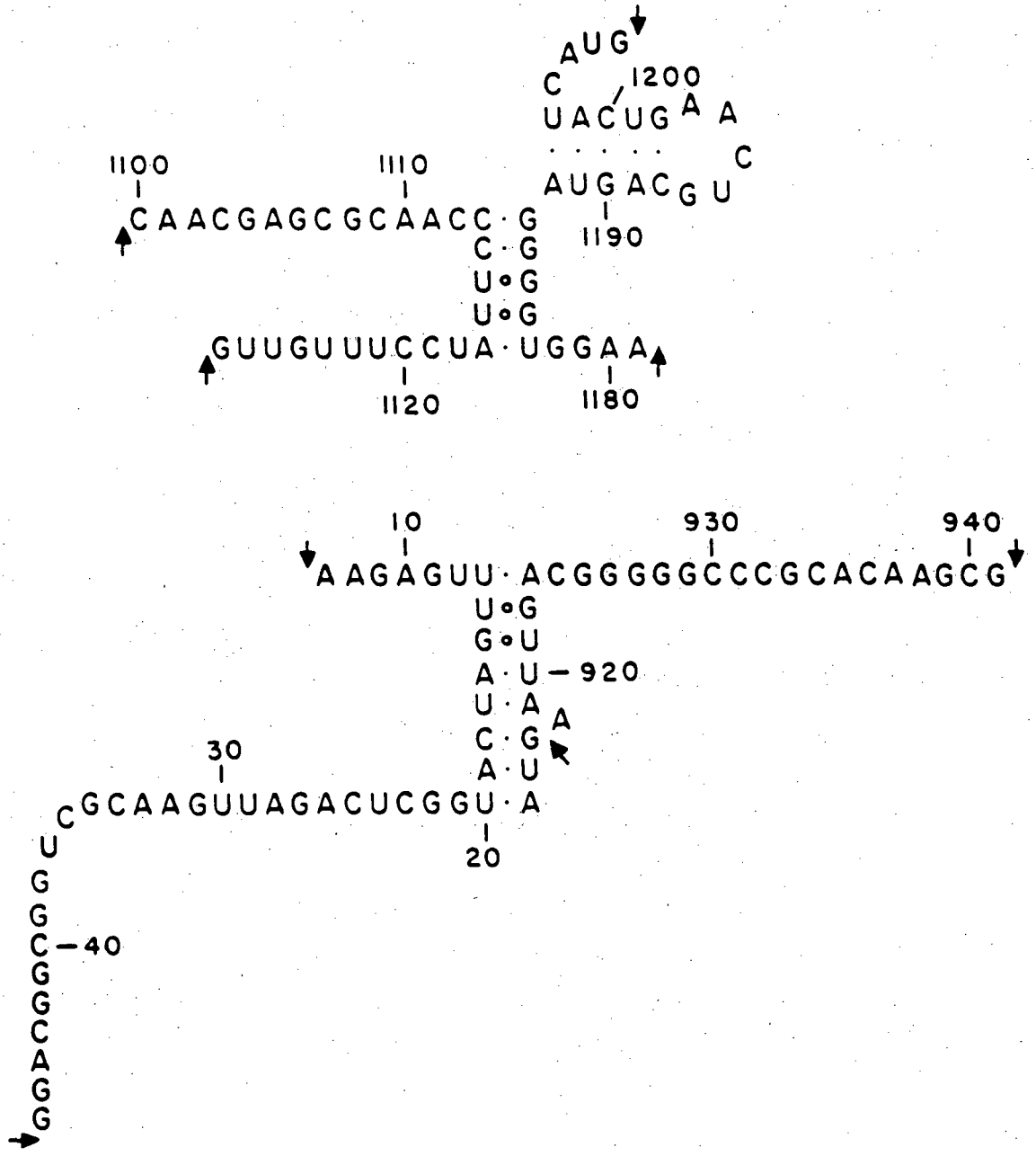
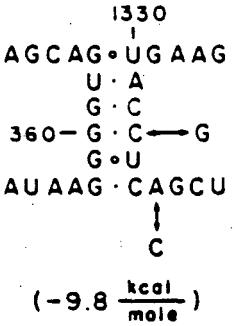

AUG
C
1200
UACUG A A
C
AUGAC GU
1190
1100
1110
CAACGAGCGCAACC·G
C·G
U∘G
U∘G
GUUGUUUCCUA·UGGAA
1120
1180
10
930
940
AAGAGUU·ACGGGGGCCCGCACAAGCG
U∘G
G∘U
A·U—920
U·A
A
C·G
A·U
30
CGCAAGUUAGACUCGGU·A
20
U
G
G
C—40
G
G
C
A
G
G

Fig. 4

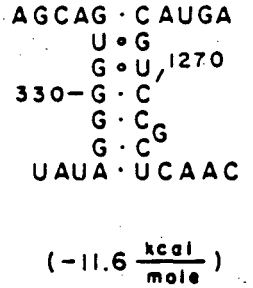
358 – 1330

Prokaryotes

E. coli

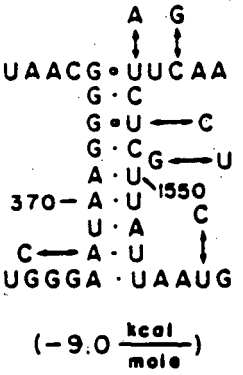


Z. mays (chloroplast)

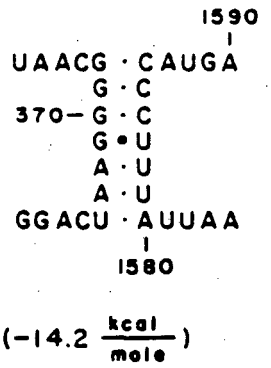


Eukaryotes

S. cerevisiae

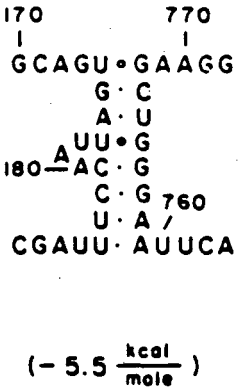


X. laevis



Mitochondria

Human



Yeast

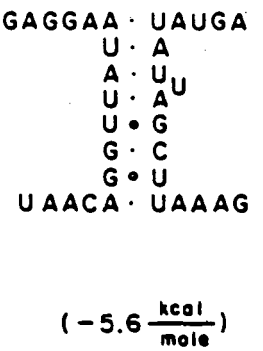
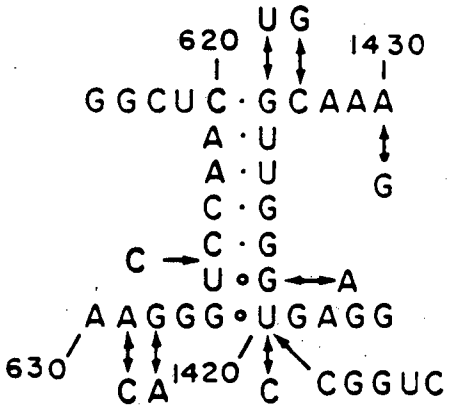


Fig. 5

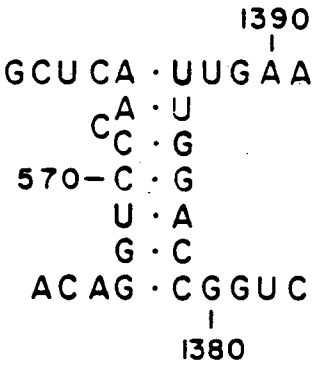
625-1420

Prokaryotes

E. coli

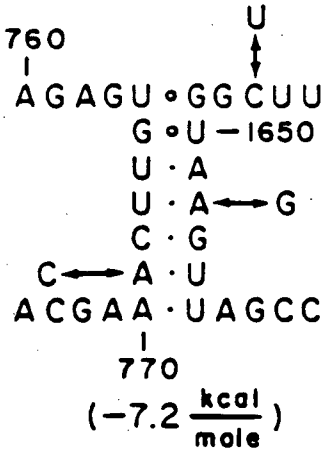


Z. mays (chloroplast)

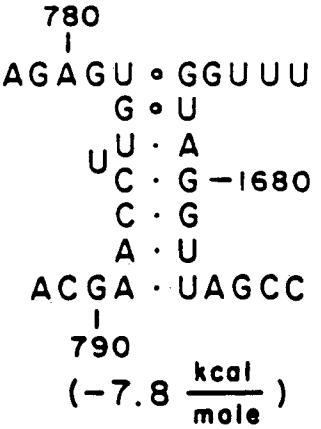


Eukaryotes

S. cerevisiae



X. laevis



Mitochondria

Human

Yeast

Fig. 6

956-1506

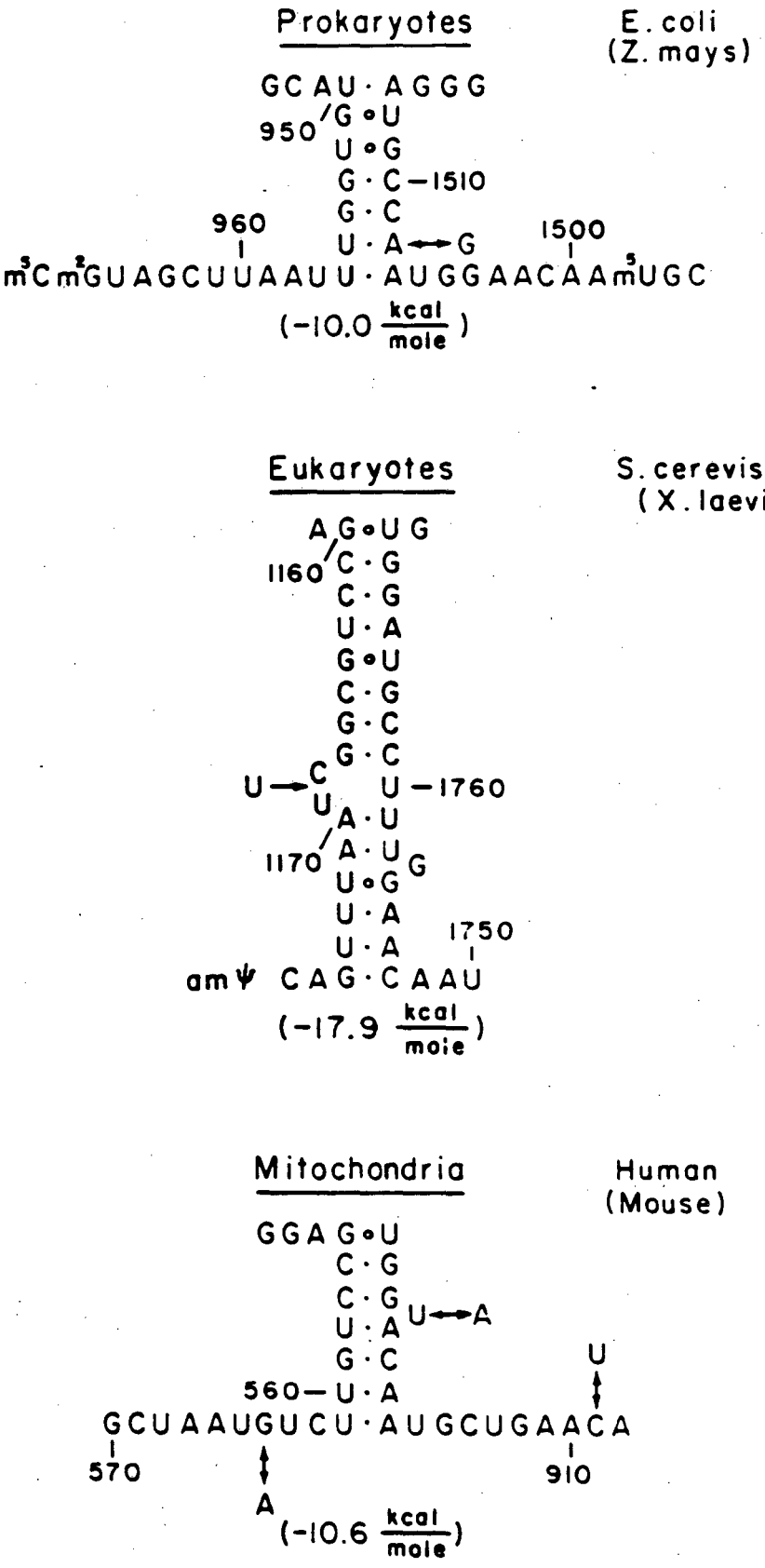


Fig. 7

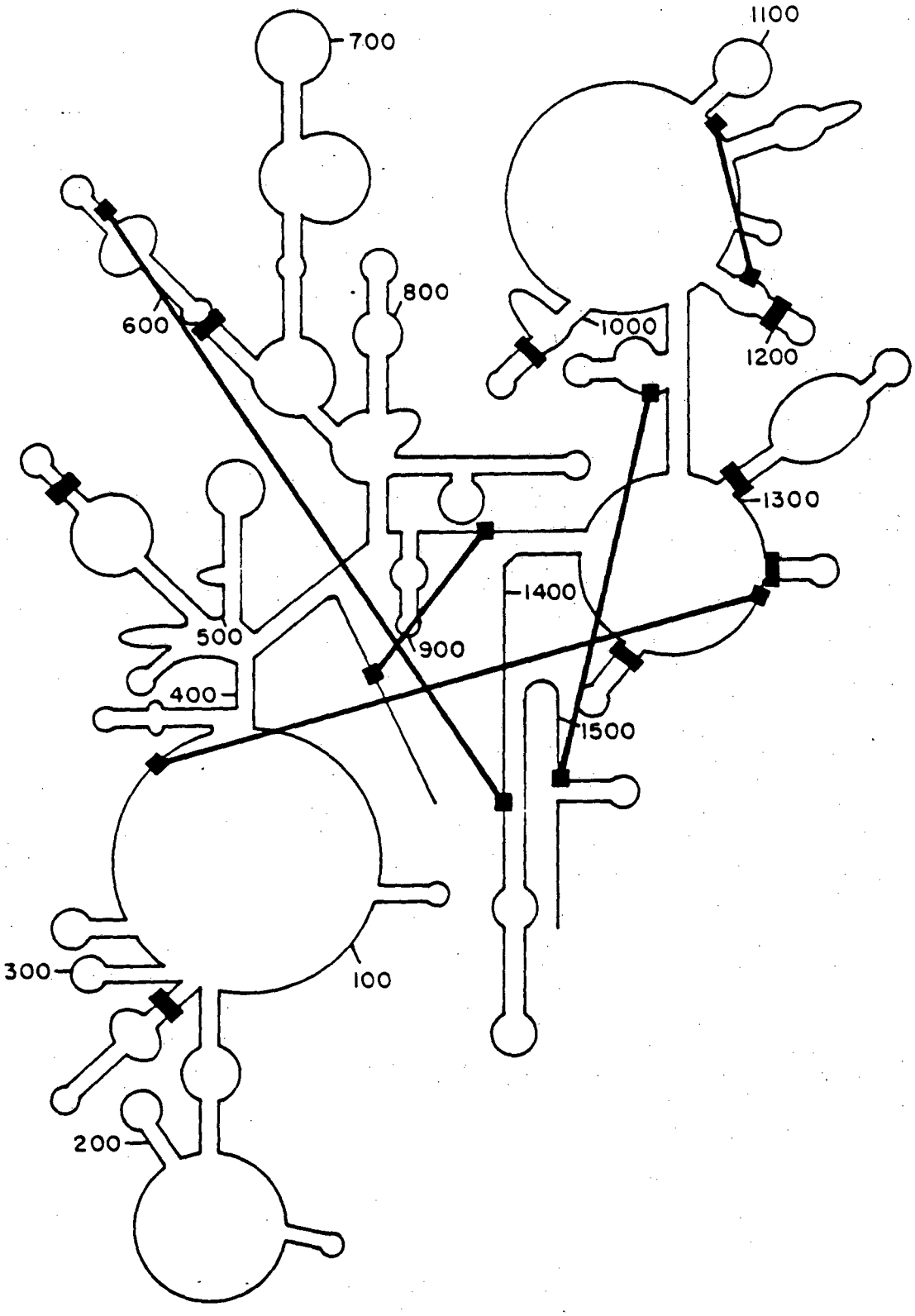

700
1100
800
600
1000
1200
1300
1400
500
900
400
1500
300
100
200

Fig. 8

This report was done with support from the Department of Energy. Any conclusions or opinions expressed in this report represent solely those of the author(s) and not necessarily those of The Regents of the University of California, the Lawrence Berkeley Laboratory or the Department of Energy.

Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.

TECHNICAL INFORMATION DEPARTMENT
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720