Biochemistry

© Copyright 1988 by the American Chemical Society Volume 27, Number 12 June 14, 1988

Perspectives in Biochemistry

The Emerging Three-Dimensional Structure and Function of 16S Ribosomal RNA

Richard Brimacombe

Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, D-1000 Berlin-Dahlem, West Germany

Received February 2, 1988

The ribosome is the organelle on which protein biosynthesis occurs in every living organism, and this central function in the cell has made it an object of intensive investigation for more than 2 decades. Research on ribosomes has many facets, and those with which this paper is concerned—namely, the study of structure–function relationships in the *Escherichia coli* ribosome—can be roughly classified into three groups, concerned with work on the ribosomal proteins, the ribosomal RNA, and the various steps in the ribosomal function, respectively. Given the complexity of the ribosome, it was inevitable that for a long time these different research interests steadily diverged relative to one another, without very much real contact taking place between the three areas. Fairly recently, however, developments in the field have reached the point where an exciting new phase of convergence has begun, centered on the study of the in situ three-dimensional structure of the ribosomal RNA. In this paper I briefly review some of these developments and describe the current state of knowledge of the structure of the 16S RNA from the smaller (30S) subunit of the *E. coli* ribosome.

The ribosome is of a very awkward size for structural studies. On the one hand, it is too small to be examined in sufficient detail by electron microscopy, but on the other hand, it is still too large and its structure too irregular for an analysis at atomic resolution by X-ray crystallography [although progress toward this goal is currently being made [e.g., Yonath and Wittmann (1988)]]. The three-dimensional structure therefore has to be put together piece by piece, with information gleaned from a variety of techniques and sources, with correspondingly varied levels of resolution. Since the protein moiety of the *E. coli* 30S subunit consists of 21 separate molecules (named S1–S21), it was possible from an early stage to carry out low-resolution studies on the topographical arrangement of the proteins, without necessarily needing to know the individual amino acid sequences. These studies will be summarized in their appropriate context below. In contrast, the mere fact that the ribosomal RNA is a single large molecule [the *E. coli* 16S RNA is 1542 nucleotides long (Brosius et al., 1978)] has the result that very little can be done with it in the way of structural or topographical investigation without reference to its primary sequence. However, for the same reason, once the sequence became available, the structural studies on the RNA tended a priori to be made at a higher level of resolution (viz. at the nucleotide level) than those with the ribosomal proteins. Thus, whereas much of the low-resolution data on the proteins have been available for quite some time, the upsurge in high-resolution data on the RNA has been rapid and recent.

Primary Sequence and Secondary Structure. The *E. coli* 16S RNA was the first large ribosomal RNA molecule to be sequenced (Brosius et al., 1978; Carbon et al., 1978), and since then sequences have been determined (for the most part via the corresponding ribosomal DNA) for well over 50 small subunit RNA molecules, covering the entire evolutionary spectrum [see Huysmans and De Wachter (1986) for a recent summary]. The sequence data soon revealed that the primary structure of ribosomal RNA has been highly conserved. To give just one example, the *E. coli* 16S RNA shows 75% homology to the corresponding RNA from maize chloroplast ribosomes (Schwartz & Kössel, 1980), and this degree of homology is ideal for the application of the "sequence comparison approach" for secondary structure determination, a method that was first used with the ribosome by Fox and Woese (1975) to derive the secondary structure of 5S RNA. In this method, a putative double helix in the secondary structure is tested by comparing the relevant *E. coli* sequence with the corresponding region in the sequence from another organism. If the proposed secondary structural element is correct, then base changes between the two sequences being compared must "compensate" (e.g., an A-U pair becoming a G-C pair) in order to maintain the integrity of the double helix.

In combination with experimental approaches to secondary structure determination, such as the isolation of base-paired fragments after partial nuclease digestion of the ribosomal RNA (Glotz & Brimacombe, 1980) or analysis of the patterns of digestion produced by single- or double-strand-specific nucleases (Stiegler et al., 1981), this phylogenetic approach quickly led to the establishment of convincing secondary

0006-2960/88/0427-4207$01.50/0 © 1988 American Chemical Society

structure models for the 16S RNA (and also the 23S RNA) from *E. coli* [reviewed by Maly and Brimacombe (1983) and Noller (1984)], with the various double-helical elements in the structure by now being supported by many thousands of compensating base-pair changes in the various sequenced species. More recently, methods for "higher order structure analysis" have been developed, which allow the secondary structure models to be tested and verified at the nucleotide level [Moazed et al., 1986; reviewed by Ehresmann et al. (1987)]. A further consequence of the sequence comparisons was the finding that a number of short stretches of sequence within the RNA molecules are almost universally conserved, whereas other regions are highly variable. Moreover, the considerable differences in length that occur between the ribosomal RNA molecules from different classes of organism are accommodated by insertions or deletions (relative to the *E. coli* sequence) in the latter variable regions of the structures [e.g., Brimacombe et al. (1983) and Huysmans and De Wachter (1986)].

Our version of the secondary structure of the *E. coli* 16S RNA [which differs only in a few details from the corresponding models of Noller (1984) and Huysmans and De Wachter (1986)] is illustrated in Figure 1, and it can be seen that some 45 separate double-helical elements are involved. Thus, the question of determining the three-dimensional structure of the RNA can to a first approximation be reduced to the problem of establishing the spatial distribution of these helical elements. Figure 1 also includes some sets of topographical data relating to this spatial distribution, and these are discussed below as they arise.

Tertiary Folding of the RNA. In principle, the same phylogenetic approach that was used to derive the secondary structure of the RNA can also be applied to the search for tertiary contacts, insofar as these involve Watson–Crick base pairs. Some pairs of compensating base changes have been found that are strongly suggestive of such contacts (Gutell et al., 1986), but at the moment it does not appear that this approach will yield very much information, largely because the tertiary contacts in the RNA are likely to consist only of single or at the most two or three base pairs. Similarly, the methods of "higher order structure analysis" already mentioned (Ehresmann et al., 1987)—although they are a powerful tool for testing an established structure, as in the case of tRNA (Peattie & Gilbert, 1980)—are not very helpful in deducing specific tertiary interactions in large RNA molecules with unknown structure.

The most widely used approach to the tertiary folding problem has been the application of intra-RNA cross-linking techniques, to search for neighborhoods between different regions of the RNA structure. There are however technical problems here, one of which is that the RNA is chemically less reactive than the proteins, with the result that attempts to introduce intra-RNA cross-links into the ribosomal subunits can lead to many unwanted side reactions. In order to circumvent this problem and simplify the analysis of the cross-linked products, a number of research groups (Expert-Bezançon et al., 1983; Thompson & Hearst, 1983; Wollenzien et al., 1985) have studied the patterns of cross-linking in the ribosomal RNA that are obtained when isolated (phenol-extracted) 16S RNA is used as the substrate for cross-linking, as opposed to intact ribosomal subunits. This presupposes that the RNA has the same tertiary stucture in the presence or absence of the ribosomal proteins, which is clearly a dangerous assumption. A second problem is that it is obviously important in this type of cross-linking approach to be able to pinpoint the sites of cross-linking accurately. Some of the groups just mentioned have relied to a large extent on electron microscopy to locate their intra-RNA cross-links (Wollenzien & Cantor, 1982; Wollenzien et al., 1985), a method that unfortunately does not offer a sufficiently high level of resolution. A first three-dimensional model of the 16S RNA was proposed in 1985, based primarily on these data (Expert-Bezançon & Wollenzien, 1985).

Our own cross-linking experiments with the ribosomal RNA have been based on the premise that the approach is only valid if intact subunits are used as the cross-linking substrate, and we have further extended this principle to the introduction of cross-links in vivo into growing *E coli* cells, by simple ultraviolet irradiation of the cell cultures (Stiege et al., 1986). Subsequently, the cross-linked 30S subunits can be isolated, the RNA cut into pieces of suitable length by partial nuclease digestion, and the cross-linked fragments isolated by two-dimensional gel electrophoresis. By use of uniformly labeled RNA, the precise sites of cross-linking within the isolated fragments can be determined by classical "fingerprinting" techniques. In any intra-RNA cross-linking study, the cross-links found belong to two classes, those that are "within" the secondary structure (and which serve to confirm the latter) and those that connect two separate secondary structural elements. The latter class of cross-link is clearly the most useful for establishing tertiary neighborhoods within the RNA, and the cross-links belonging to this class that we have so far identified are drawn superimposed on the secondary structure of the 16S RNA in Figure 1. This set of data includes the in vivo cross-links just mentioned and also results obtained from 30S subunits chemically cross-linked in vitro by using a simple bifunctional nitrogen mustard derivative as the cross-linking agent (Atmadja et al., 1986).

The intra-RNA cross-link data shown in Figure 1 begin to constrain the 16S RNA into a three-dimensional shape, but they give no indication as to how the RNA is oriented within the 30S subunit. Here a number of "reference points" have been established by the use of immune electron microscopy. The 16S RNA contains nine modified nucleotides according to Carbon et al. (1979) (included in Figure 1), and Van Knippenberg et al. (1984) have identified a further modified base at position 1517, immediately adjacent to the dimethylated adenine residues in the loop end of helix 45 (Figure 1). Antibodies have been raised against the dimethylated adenine residues, as well as against the methylated guanine residue at position 527 in helix 18, and the positions of these antibodies bound to their respective modified nucleotides have been localized on the 30S subunit surface by electron microscopy (Politz & Glitz, 1977; Trempe et al., 1982). In a similar manner, the 5′ and 3′ ends of the 16S molecule have been located, by appropriate haptenization reactions followed by antibody binding and electron microscopy (Mochalova et al., 1982; Lührmann et al., 1981). A further nucleotide that has been mapped in this way is C-1400, by virtue of the fact that this particular cytidine residue can be cross-linked (see later) by ultraviolet irradiation to a haptenized tRNA molecule bound to the ribosome (Gornicki et al., 1984). The electron microscopically located positions in the 16S RNA are indicated in Figure 1.

While on the subject of modified nucleotides, it should be mentioned that, since most of the sequencing data has been obtained via ribosomal DNA (Huysmans & De Wachter, 1986), our knowledge of the modification patterns in the ribosomal RNA is far from complete. Nevertheless, a detailed study of the modified nucleotides in the 3′-proximal region of

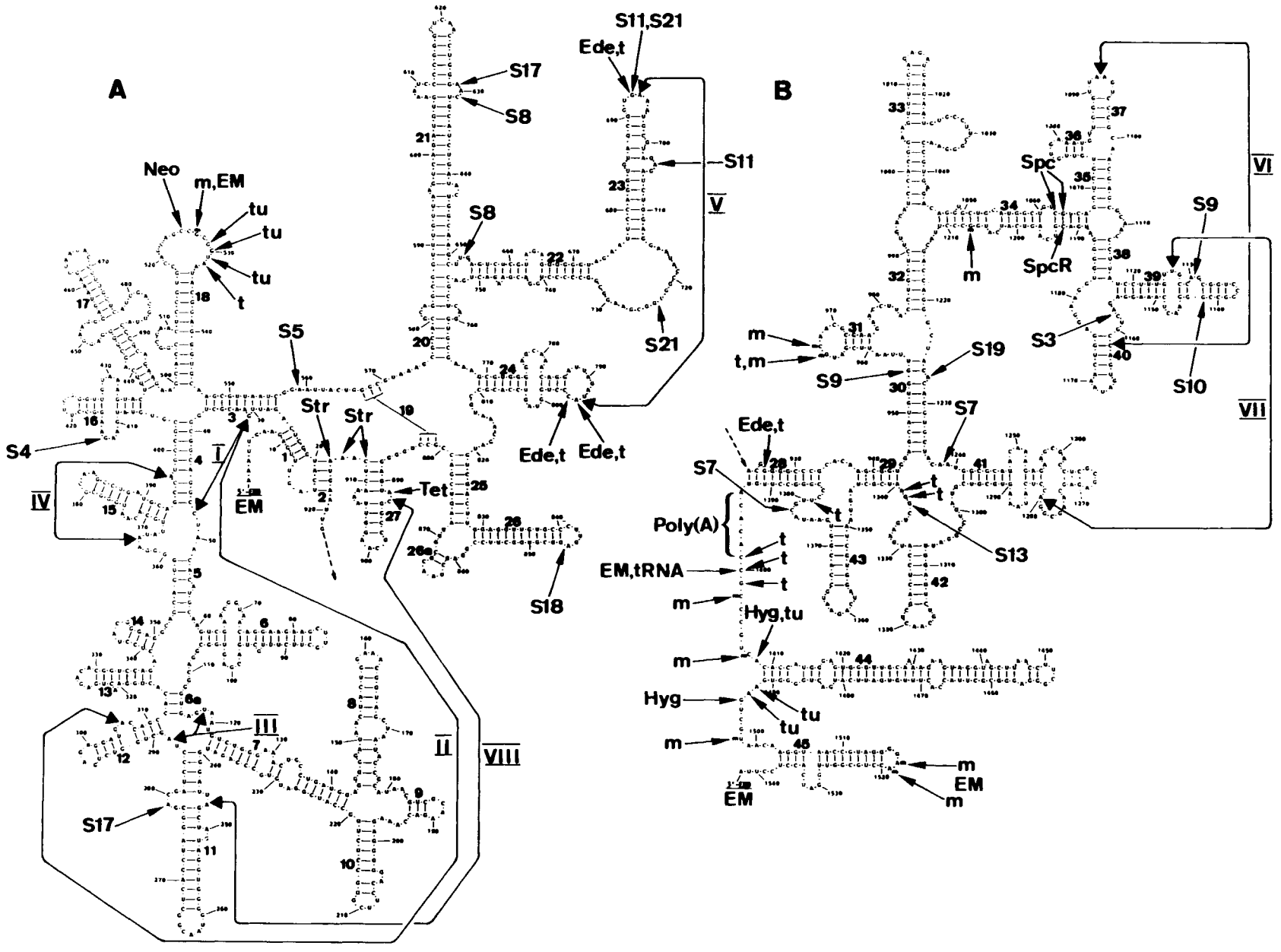

FIGURE 1: Secondary structure of *E. coli* 16S RNA showing structural and functional data. The secondary structure is divided into two halves (A and B), and the helical regions are numbered 1–45 from the 5′ end. Sites of cross-linking to ribosomal proteins (denoted by the appropriate protein number) and intra-RNA cross-links (Roman numerals) are taken from Brimacombe et al. (1988), except for cross-link VIII (W. Stiege and K. Stade, unpublished results). Modified bases (Carbon et al., 1979) are denoted by m and sites located electron microscopically by EM [cf. Gornicki et al. (1984)]. Positions implicated in tRNA binding (Moazed & Noller, 1986) are marked tu or t, according to whether the binding was in the presence or absence of poly(U), respectively. tRNA indicates the site of cross-linking of tRNA to 16S RNA (Prince et al., 1982), with poly(A) similarly indicating the cross-link site of the latter (Stiege et al., 1988). Positions implicated in antibiotic binding (Moazed & Noller, 1987) are denoted by Neo (neomycin), Ede (edeine), Tet (tetracycline), Str (streptomycin), Spc (spectinomycin), and Hyg (hygromycin). SpcR is a site of mutation causing resistance to spectinomycin (Sigmund et al., 1984).

the 16S RNA has shown a very high degree of conservation in this area of the molecule (Van Knippenberg et al., 1984). In addition to modified bases, it is worth noting that eukaryotic small subunit RNA molecules contain a relatively large number of 2′-O-methylated nucleotides and pseudouridine residues, most of which have now been located in the RNA from higher eukaryotes [e.g., Maden (1986)].

The combined results of the intra-RNA cross-linking and immune electron microscopic approaches are not by themselves sufficient to draw the 16S RNA together into the compact and specific shape that it must occupy within the 30S subunit. More importantly, these data do not give any information concerning the arrangement of the RNA relative to the ribosomal proteins. Since the topographical arrangement of the proteins themselves is very well documented, it is appropriate to digress at this point and give a brief description of the current status of this topic.

Topography of Ribosomal Proteins. A number of different approaches have been applied to the study of the spatial distribution of the ribosomal proteins [see Wittmann (1983) for review], the three principal methods being interprotein cross-linking, immune electron microscopy, and low-angle neutron scattering. Identification of pairs of proteins that can be cross-linked together gives a direct measure of those proteins that are close neighbors in space, and many such pairs have been described [e.g., Lambert et al. (1983)]. The method has its pitfalls, not least of which is the danger of misidentifying the proteins involved, as a result of the very similar mobilities of many of the proteins in most separation systems. This particular danger can be overcome, and the resolution of the method simultaneously improved, if the actual amino acids concerned in the individual cross-links are determined. So far, however, this has only been achieved in a few cases (Brockmöller & Kamp, 1986).

Antigenic sites corresponding to the majority of the *E. coli* ribosomal proteins have been mapped on the surface of the 30S subunit by immune electron microscopy (Stöffler & Stöffler-Meilicke, 1986; Oakes et al., 1986), in a manner precisely analogous to that described above for the localization of specific nucleotides in the 16S RNA. This method relies for its success on the fact that the 30S subunit has a very characteristic shape, being divided into a "head" and a "body", which comprise roughly one-third and two-thirds of the particle, respectively. The body is extended upward on one side of the subunit, forming a large protuberance often referred to as the "platform".

The cross-linking and electron microscopic approaches are both effectively concerned with points on the surfaces of the proteins and/or the subunits. In contrast, the neutron scattering approach is more concerned with the internal "architecture" of the ribosome and gives information as to the relative positions of the protein centers of mass. The method relies on a well-established property of the ribosomal subunits, namely, that functionally active ribosomes can be reconstituted from their isolated protein and RNA components (Held et al., 1974). Thus, proteins or RNA can be prepared from ribosomes out of cells grown in a fully or partially deuteriated medium, and these deuteriated components can subsequently be reconstituted into hybrid particles in which (for example) two of the proteins are deuteriated and all other components are protonated. Neutron scattering measurements allow the distance between the mass centers of the two deuteriated components to be computed. If a series of such determinations are made, the relative positions of the proteins can be triangulated, and this research has recently been completed for the *E. coli* 30S subunit, with the result that the centers of mass of all of the 21 proteins are now mapped (Capel et al., 1987). A comparison with the cross-linking and electron microscopic data shows in general a high level of agreement, although there are a few discrepancies [see Moore et al. (1986) for discussion]. As already noted in the previous section, the RNA and protein arrangements cannot be considered independently from one another, and this neutron scattering map has played an important role in the determination of the three-dimensional arrangement of the 16S ribosomal RNA.

RNA–Protein Interaction. The classical approach to the question of RNA–protein interactions in the ribosome has been to try to define "binding sites" for single proteins on the 16S RNA. The method, however, has only had limited success, because although a number of proteins are capable of binding independently to the 16S RNA, the subsequent partial nuclease digestion procedures designed to reveal a protein-protected RNA region only yielded useful data in a very few cases, most notably with protein S8 (Gregory et al., 1984). In other cases the binding sites found were either inordinately large [such as with protein S4 (Zimmermann, 1980)] or else tended to disintegrate, depending on the particular protein and RNA region under study. The approach can be extended to probe the differential protection effects on the RNA of groups of proteins, such as S6, S8, S15, and S18 in various combinations (Gregory et al., 1984) or, similarly, S7, S9, S14, and S19 (Wiener & Brimacombe, 1987).

A more recent and powerful variant of the binding site approach is the "foot-printing" technique. Here the accessibility of the 16S RNA to nucleases or modifying reagents is examined in the presence or absence of the protein concerned, and the sites on the RNA that have been "hit" (and correspondingly those that are protected by the protein) are analyzed by primer extension, using reverse transcriptase together with a synthetic oligodeoxynucleotide primer or set of primers, complementary to strategically chosen regions of the 16S sequence. This method has given detailed information on the binding sites of protein S4 (Stern et al., 1986) and S8 (Mougel et al., 1987) complexed with 16S RNA and is currently being used by Noller's group to study the binding sites on the 16S RNA of a number of proteins in complete or partially reconstituted 30S subunits [see meeting report in Wickens and Dahlberg (1987)].

Here again, cross-linking techniques offer an alternative source of data, in this case involving cross-links between the proteins and the RNA, generated in situ in 30S subunits. In principle, the RNA–protein cross-linking approach is just as simple as the corresponding interprotein or intra-RNA cross-linking approaches. After generating the cross-links with a suitable bifunctional reagent, the cross-linked subunits are partially digested with nuclease, and complexes consisting of proteins cross-linked to RNA fragments are isolated and analyzed. In practice, however, the successful application of this method proved to be very difficult, largely because the cross-linked RNA–protein complexes show a marked tendency to aggregate and do not run cleanly in standard separation systems. We have recently been able to overcome this problem by developing a three-step separation procedure in which the final (and crucial) step is a two-dimensional gel electrophoresis in the presence of high concentrations of salt. With this new system, it has been possible to pinpoint a large number of RNA–protein cross-link sites on the 16S RNA involving a majority of the 30S ribosomal proteins (Osswald et al., 1987; Greuer et al., 1987). These cross-link sites, combining the results obtained with several different bifunctional reagents,

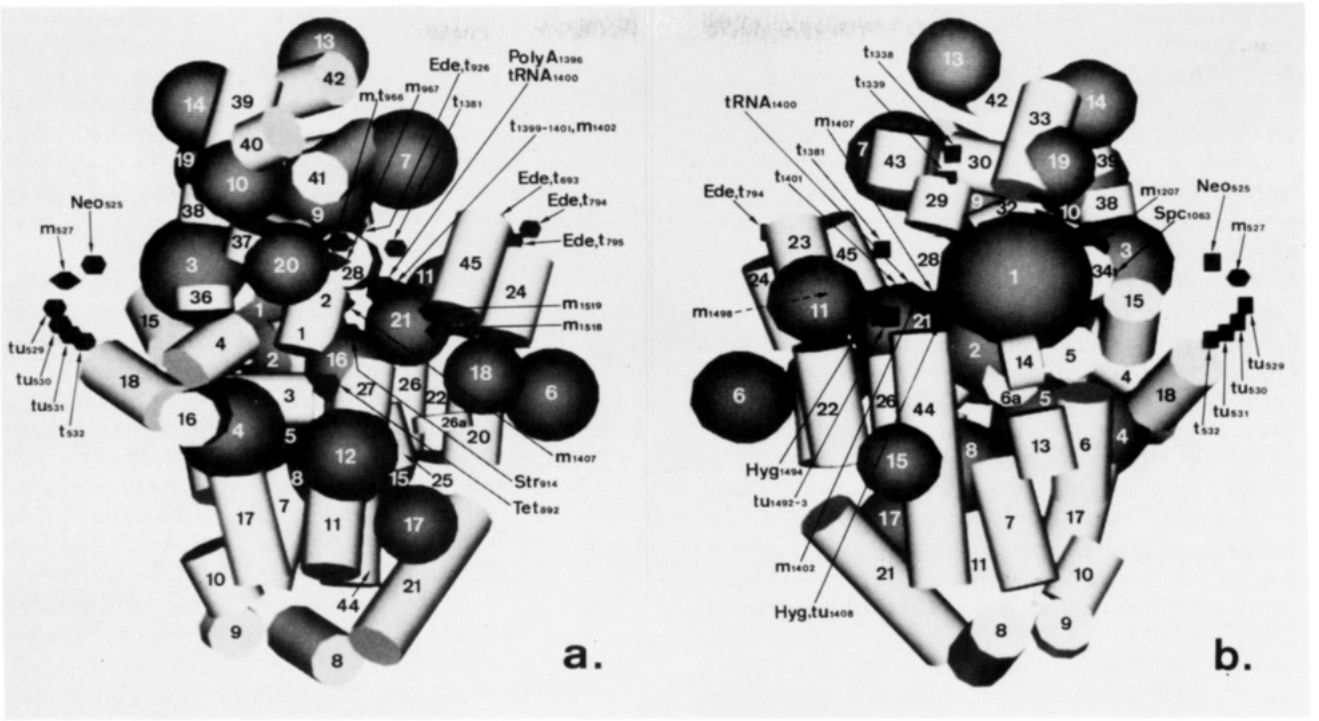

FIGURE 2: Two views of the model of the 30S ribosomal subunit. Helical regions of the RNA (cylinders with black numbers) are numbered as in Figure 1. Proteins (spheres with white numbers) are arranged according to the neutron scattering data of Capel et al. (1987) and fitted to the RNA according to Schüler and Brimacombe (1988). Small black polygons indicate functional sites, labeled as in Figure 1.

are included in Figure 1, and it is these data that have provided the "missing link" for relating the RNA arrangement to that of the proteins (Brimacombe et al., 1988).

The Three-Dimensional Model of the 16S RNA. There are essentially two reasons for building models of the RNA at this stage. One is to test whether all the available data can indeed be incorporated into a single coherent structure, and the second is then to see what can be learned from the model with regard to the three-dimensional juxtaposition of structural or functional sites in different parts of the molecule.

The model that we have derived (Brimacombe et al., 1988) is based on the secondary structure of the 16S RNA (Figure 1), which is then constrained into three dimensions by incorporation of the intra-RNA cross-links [but without taking account of those cross-links already discussed that were derived from cross-linking studies made with isolated 16S RNA [cf. Expert-Bezançon and Wollenzien (1985)]]. Next, the RNA is oriented within the 30S subunit with the help of the various immune electron microscopic locations of specific nucleotides, and finally it is maneuvered into the compact shape of the 30S subunit and fitted to the protein arrangement (Capel et al., 1987) with the help of the RNA–protein cross-links (Figure 1). More recently we have used computer graphics in combination with a least-squares fitting procedure to generate the best fit between the latest neutron scattering map of Capel et al. (1987) and the RNA–protein cross-link sites in the model (Schüler & Brimacombe, 1988). Two views of the resulting computer graphics structure are presented in Figure 2, which also shows the positions of various functional sites in the model, which will be discussed below.

The model (Figure 2) has overall dimensions of 220 Å × 140 Å × 90 Å, which agree well with electron microscopic estimates for the 30S subunit (Lake et al., 1974). Furthermore, the shape of the model is recognizably the same as that seen in electron micrographs, with the head, body, and platform regions, the latter being visible on the right in Figure 2a, together with the "cleft" separating the platform from the head. Ramakrishnan (1986) has estimated from neutron scattering experiments that the mass centers of RNA and protein in the 30S subunit are separated by approximately 25 Å, and calculation of the corresponding displacement in the model (taking into account only the double-helical regions of the RNA, i.e., the "cylinders" in Figure 2) gives a value of 20 Å (Schüler & Brimacombe, 1988), indicating that the protein–RNA distribution in the model structure must closely reflect the real situation. Many other more detailed features of the model also agree very closely (Brimacombe et al., 1988) with various sets of experimental data, such as the RNA–protein binding sites (Gregory et al., 1984; Stern et al., 1986; Wiener & Brimacombe, 1987), tertiary interactions in the RNA (Gutell et al., 1986), or sites on the RNA that are accessible to chemical modification in the 30S subunit (Moazed et al., 1986).

The degree of constraint imposed on the individual helices in the model varies considerably. Whereas some helices (such as helix 39, Figure 1) are fixed fairly rigidly by virtue of their content of topographical information, others (such as helices 13 and 14, for example) are only fixed insofar as the "base" of the helical element is constrained by neighboring structural elements, the helix itself being conformationally unrestricted. There is thus clearly considerable room for local refinements and improvements to the structure. Nonetheless, the constraints on the folding of the RNA are sufficient to enable the model to reveal a number of hitherto unobserved juxtapositions between different parts of the molecule. Of these features the two most interesting are (a) the regions of the structure where deletions or insertions in the ribosomal RNA from other classes of organism occur (relative to the *E. coli* 16S RNA) and (b) the distribution of sites on the RNA that have been implicated in ribosomal function.

Deletions and Insertions. It has already been noted above that the ribosomal RNA from different types of organism varies considerably in its overall length (Brimacombe et al., 1983; Huysmans & De Wachter, 1986). The 18S RNA from eukaryotic cytoplasmic ribosomes is some 300 nucleotides longer than the *E. coli* 16S RNA, and these extra nucleotides

are accounted for by two large "insertions" in the regions of helices 7 and 21 (Figure 1), respectively, together with two small insertions to helices 39 and 44. The two large insertions and the extension to helix 44 all combine to extend the structure in a downward direction (cf. Figure 2), giving rise to the lobes seen by electron microscopy at the base of the eukaryotic ribosomal subunit [e.g., Oakes et al. (1986)]. Electron micrographs also indicate the presence of a "beak-like protuberance" on the head of the eukaryotic subunit, and the position of this beak correlates with the position of helix 39 in the model.

In contrast to the eukaryotic cytoplasmic ribosomal RNA, chloroplast and mitochondrial ribosomes have RNA molecules that are shorter than the *E. coli* 16S RNA. In maize chloroplasts (Schwartz & Kössel, 1980), helices 6, 10, and 17 (Figure 1) are largely "deleted", and although these helices are widely separated in the secondary structure, in the three-dimensional structure they form a compact "subdomain" at the base of the model (Figure 2b, lower right). The deletions in mammalian mitochondrial ribosomal RNA are considerably more drastic (the RNA is about 600 nucleotides shorter than its *E. coli* counterpart), but here again they are concentrated into distinct domains in the three-dimensional structure, the largest set of deletions representing a virtual "amputation" of the entire lower third of the subunit (Brimacombe et al., 1988) and the remaining deletions being grouped together on the left-hand side of the head of the subunit (cf. Figure 2a).

In contrast to these variable regions of the structure, many of the most highly conserved sequence regions in the 16S RNA are concentrated in a broad band in the "throat" area of the subunit, where the head joins the body. This area is known from a number of electron microscopic studies to be the functionally vital "decoding" region of the 30S subunit [e.g., Gornicki et al. (1984)].

Functional Sites. The involvement of the 3′-terminus of the 16S RNA in initiation of protein synthesis (Shine & Dalgarno, 1974) is a concept that has been generally accepted for some years, with clear experimental support [see, e.g., Steitz and Jakes (1975)]. In contrast, functionally important sites at internal positions in the RNA [where there are a number of regions that are conserved at least as strongly as the 3′-terminal sequence [cf. Huysmans and De Wachter (1986)]] have not been so amenable to investigation. Thus, although there are many papers in the older ribosomal literature where the 16S RNA has been identified as the target in (for example) affinity-labeling experiments, progress has been blocked in the past by the lack of suitable methodology for analyzing the target sites on the RNA. This situation has now been changed, both by the availability of better techniques for cross-link site analysis and by the availability of the "footprinting" method already described above, using primer extension with reverse transcriptase to locate the relevant sites on the RNA.

Moazed and Noller (1986) have made a detailed study by the latter method of the bases in 16S RNA that are protected against chemical modification by ribosome-bound tRNA, in the presence or absence of poly(U). The sites that they found are widely distributed in the 16S RNA sequence and are indicated in Figure 1. In a similar series of experiments, the same authors (Moazed & Noller, 1987) analyzed the sites that are protected by various antibiotics bound to the 30S subunit, and these sites are also included in Figure 1. The antibiotic footprinting data showed an excellent agreement with sites of mutation in the 16S RNA causing resistance to the same antibiotics [e.g., spectinomycin resistance at position 1192 (Sigmund et al., 1984) or streptomycin resistance at position 912 (Montandon et al., 1985)], and Moazed and Noller (1987) also noted a strong correlation between several of the antibiotic sites and the sites protected by bound tRNA. The antibiotics concerned interfere with various steps of the protein biosynthetic process [summarized by Moazed and Noller (1987)], and so an overlap with the sites of tRNA binding is a satisfying result. However, with the help of the three-dimensional model (Figure 2), this correlation can be considerably extended. What the model shows is that all of the antibiotic and tRNA sites are clustered into two distinct areas of the structure and that furthermore all of the modified bases in the 16S RNA (Figure 1) are concentrated in the same two regions (Figure 2). The larger of these two clusters is centered around the cleft region, and the smaller group is on the opposite side of the subunit.

Cross-linking studies have shown that the hypermodified base in the anticodon loop of valyl-tRNA can be linked to position C-1400 in the 16S RNA when the tRNA is bound to the ribosomal P site (Prince et al., 1982), and a photoaffinity label attached to the same base in tRNA was also cross-linked to C-1400 when the tRNA was bound to the A site (Ciesiolska et al., 1985). Recently, we have been able to identify a site of cross-linking to a poly(A) message in the presence of lysyl-tRNA at positions 1394–1399 (Figure 1) in the 16S RNA (Stiege et al., 1988). These data suggest that the cluster of functional sites in the cleft region (Figure 2) does indeed represent the actual decoding site [in agreement with the electron microscopic studies (Gornicki et al., 1984)], whereas the smaller group of sites is close to the location on the 30S subunit where elongation factor Tu (which brings the charged tRNA molecules to the ribosome in the form of a ternary complex) has been mapped by immune electron microscopy (Girshovich et al., 1986). The distribution of these functional sites is discussed in more detail elsewhere (Stiege et al., 1988), together with a possible mechanism for the movement of tRNA through the subunit.

Outlook. The three-dimensional model of the 16S RNA (Figure 2), although it was put together in our laboratory, does in fact represent a compilation of data from many different methods and laboratories. As suggested at the beginning of this paper, the model now offers a focus for bringing together structural results relating to both the RNA and protein moieties of the 30S subunit, as well as providing a first framework for examining functional data at the nucleotide level in three dimensions. The picture should become progressively clearer in parallel with progressive refinements to the structure, and the model should also begin to suggest experiments, such as positions in the RNA sequence where site-directed mutagenesis techniques [e.g., Zwieb and Dahlberg (1984)] could usefully be applied, in a more coordinated manner than has hitherto been possible.

The next phase in the structural research will be to develop a similar model for the 50S ribosomal subunit. The relevant experimental programs, analogous to those described here, are already well under way, and with the experience gained from the 30S subunit it should not take more than a few years to generate a 50S model at the same level of resolution. We will then finally be in a position to begin studying the fine structure of the mechanics of protein biosynthesis in the complete functional unit, namely, the 70S ribosome.

REFERENCES

Atmadja, J., Steige, W., Zobawa, M., Greuer, B., Osswald, M., & Brimacombe, R. (1986) *Nucleic Acids Res. 14*, 659–673.

Brimacombe, R., Maly, P., & Zwieb, C. (1983) *Prog. Nucleic Acid Res. 28*, 1–48.

Brimacombe, R., Atmadja, J., Stiege, W., & Schüler, D. (1988) *J. Mol. Biol. 199*, 115–136.

Brockmöller, J., & Kamp, R. M. (1986) *Biol. Chem. Hoppe-Seyler 367*, 925–935.

Brosius, J., Palmer, M. L., Kennedy, P. J., & Noller, H. F. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4801–4805.

Capel, M. S., Engelman, D. M., Freeborn, B. R., Kjeldgaard, M., Langer, J. A., Ramakrishnan, V., Schindler, D. G., Schneider, D. K., Schoenborn, B. P., Sillers, I. Y., Yabuki, S., & Moore, P. B. (1987) *Science (Washington, D.C.) 238*, 1403–1406.

Carbon, P., Ehresmann, C., Ehresmann, B., & Ebel, J. P. (1978) *FEBS Lett. 94*, 152–156.

Carbon, P., Ehresmann, C., Ehresmann, B., & Ebel, J. P. (1979) *Eur. J. Biochem. 100*, 399–410.

Ciesiolska, J., Gornicki, P., & Ofengand, J. (1985) *Biochemistry 24*, 4931–4938.

Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J. P., & Ehresmann, B. (1987) *Nucleic Acids Res. 15*, 9109–9128.

Expert-Bezançon, A., & Wollenzien, P. (1985) *J. Mol. Biol. 184*, 53–66.

Expert-Bezançon, A., Milet, M., & Carbon, P. (1983) *Eur. J. Biochem. 136*, 267–274.

Fox, G., & Woese, C. R. (1975) *Nature (London) 256*, 505–507.

Girshovich, A. S., Bochkareva, E. S., & Vasiliev, V. D. (1986) *FEBS Lett. 197*, 192–198.

Glotz, C., & Brimacombe, R. (1980) *Nucleic Acids Res. 8*, 2377–2395.

Gornicki, P., Nurse, K., Hellmann, W., Boublik, M., & Ofengand, J. (1984) *J. Biol. Chem. 259*, 10493–10498.

Gregory, R. J., Zeller, M. L., Thurlow, D. L., Gourse, R. L., Stark, M. J. R., Dahlberg, A. E., & Zimmermann, R. A. (1984) *J. Mol. Biol. 178*, 287–302.

Greuer, B., Osswald, M., Brimacombe, R., & Stöffler, G. (1987) *Nucleic Acids Res. 15*, 3241–3255.

Gutell, R. R., Noller, H. F., & Woese, C. R. (1986) *EMBO J. 5*, 1111–1113.

Held, W. A., Ballou, B., Mizushima, S., & Nomura, M. (1974) *J. Biol. Chem. 249*, 3103–3111.

Huysmans, E., & De Wachter, R. (1986) *Nucleic Acids Res. 13*, r73–r118.

Lake, J. A., Sabatini, D. D., & Nonomura, Y. (1974) in *Ribosomes* (Nomura, M., Tissières, A., & Lengyel, P., Eds.) pp 543–557, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Lambert, J. M., Boileau, G., Cover, J. A., & Traut, R. R. (1983) *Biochemistry 22*, 3913–3920.

Lührmann, R., Stöffler-Meilicke, M., & Stöffler, G. (1981) *Mol. Gen. Genet. 182*, 369–376.

Maden, B. E. H. (1986) *J. Mol. Biol. 189*, 681–699.

Maly, P., & Brimacombe, R. (1983) *Nucleic Acids Res. 11*, 7263–7286.

Moazed, D., & Noller, H. F. (1986) *Cell (Cambridge, Mass.) 47*, 985–994.

Moazed, D., & Noller, H. F. (1987) *Nature (London) 327*, 389–394.

Moazed, D., Stern, S., & Noller, H. F. (1986) *J. Mol. Biol. 187*, 399–416.

Mochalova, L. V., Shatsky, I. N., Bogdanov, A. A., & Vasiliev, V. D. (1982) *J. Mol. Biol. 159*, 637–650.

Montandon, P. E., Nicolas, P., Schürmann, P., & Stutz, E. (1985) *Nucleic Acids Res. 13*, 4299–4310.

Moore, P. B., Capel, M., Kjeldgaard, M., & Engelman, D. M. (1986) in *Structure, Function, and Genetics of Ribosomes* (Hardesty, B., & Kramer, G., Eds.) pp 87–100, Springer-Verlag, New York and Heidelberg.

Mougel, M., Eyermann, F., Westhof, E., Romby, P., Expert-Bezançon, A., Ebel, J. P., Ehresmann, B., & Ehresmann, C. (1987) *J. Mol. Biol. 198*, 91–107.

Noller, H. F. (1984) *Annu. Rev. Biochem. 53*, 119–162.

Oakes, M., Henderson, E., Scheinman, A., Clark, M., & Lake, J. A. (1986) in *Structure, Function, and Genetics of Ribosomes* (Hardesty, B., & Kramer, G., Eds.) pp 47–67, Springer-Verlag, New York and Heidelberg.

Osswald, M., Greuer, B., Brimacombe, R., Stöffler, G., Bäumert, H., & Fasold, H. (1987) *Nucleic Acids Res. 15*, 3221–3240.

Peattie, D., & Gilbert, W. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 4679–4682.

Politz, S. M., & Glitz, D. G. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 1468–1472.

Prince, J. B., Taylor, B. H., Thurlow, D. L., Ofengand, J., & Zimmermann, R. A. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 5450–5454.

Ramakrishnan, V. (1986) *Science (Washington, D.C.) 231*, 1562–1564.

Schüler, D., & Brimacombe, R. (1988) *EMBO J.* (in press).

Schwartz, Z., & Kössel, H. (1980) *Nature (London) 283*, 739–742.

Shine, J., & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 1342–1346.

Sigmund, C. D., Ettayebi, M., & Morgan, E. A. (1984) *Nucleic Acids Res. 12*, 4653–4663.

Steitz, J. A., & Jakes, K. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 4734–4738.

Stern, S., Wilson, R. C., & Noller, H. F. (1986) *J. Mol. Biol. 192*, 101–110.

Stiege, W., Atmadja, J., Zobawa, M., & Brimacombe, R. (1986) *J. Mol. Biol. 191*, 135–138.

Stiege, W., Stade, K., Schüler, D., & Brimacombe, R. (1988) *Nucleic Acids Res.* (in press).

Stiegler, P., Carbon, P., Zuker, M., Ebel, J. P., & Ehresmann, C. (1981) *Nucleic Acids Res. 9*, 2153–2172.

Stöffler, G., & Stöffler-Meilicke, M. (1986) in *Structure, Function, and Genetics of Ribosomes* (Hardesty, B., & Kramer, G., Eds.) pp 28–46, Springer-Verlag, New York and Heidelberg.

Thompson, J. F., & Hearst, J. E. (1983) *Cell (Cambridge, Mass.) 32*, 1355–1365.

Trempe, M. R., Ohgi, K., & Glitz, D. G. (1982) *J. Biol. Chem. 257*, 9822–9829.

Van Knippenberg, P. H., Van Kimmenade, J. M. A., & Heus, H. A. (1984) *Nucleic Acids Res. 12*, 2595–2604.

Wickens, M. P., & Dahlberg, J. E. (1987) *Cell (Cambridge, Mass.) 51*, 339–342.

Wiener, L., & Brimacombe, R. (1987) *Nucleic Acids Res. 15*, 3653–3670.

Wittmann, H. G. (1983) *Annu. Rev. Biochem. 52*, 35–65.

Wollenzien, P. L., & Cantor, C. R. (1982) *J. Mol. Biol. 159*, 151–166.

Wollenzien, P. L., Murphy, R. F., Cantor, C. R., Expert-Bezançon, A., & Hayes, D. (1985) *J. Mol. Biol. 184*, 67–80.

Yonath, A., & Wittmann, H. G. (1988) *Methods Enzymol.* (in press).
Zimmermann, R. A. (1980) in *Ribosomes* (Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahan, L., & Nomura, M., Eds.) pp 135-169, University Park Press, Baltimore, MD.
Zwieb, C., & Dahlberg, A. E. (1984) *Nucleic Acids Res. 12*, 4361-4375.

Accelerated Publications

Correction of the cDNA-Derived Protein Sequence of Prostatic Spermine Binding Protein: Pivotal Role of Tandem Mass Spectrometry in Sequence Analysis†

Robert J. Anderegg,‡,§ Steven A. Carr,*,‡ I. Yih Huang,‡ Richard A. Hiipakka,∥ Chawnshang Chang,∥ and Shutsung Liao∥

Department of Physical and Structural Chemistry, L-940, and Department of Analytical Chemistry, L-950, Smith Kline and French Laboratories, Box 1539, King of Prussia, Pennsylvania 19406, and Ben May Institute and Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, Illinois 60637

Received March 17, 1988; Revised Manuscript Received April 18, 1988

ABSTRACT: Spermine binding protein (SBP) is a rat ventral prostate protein that binds various polyamines, and the level of this protein and its mRNA is regulated by androgens. Previously, the cDNA for SBP was cloned and sequenced and an amino acid sequence deduced from the cDNA. Data from partial amino acid sequencing of the purified protein were consistent with the amino acid sequence deduced from the cDNA. However, the amino terminus of the protein was blocked, and therefore, direct protein sequence information confirming the cDNA reading frame of this region could not be obtained by Edman degradation. We have now employed an integrated approach using fast atom bombardment mass spectrometry, tandem mass spectrometry, and conventional sequencing methodologies to establish the amino-terminal sequence of the protein and to identify an amino acid sequence (35 residues) present in the purified protein but missing from the amino acid sequence deduced from cDNA clones for this protein. The missing piece of cDNA corresponds to an exon found in mouse genomic clones for a protein similar to rat SBP. Therefore, the cDNA clones for rat SBP may represent splicing variants that lack the sequence information of one exon. The blocked amino terminus of the protein was identified as 5-oxopyrrolidine-2-carboxylic acid. Mass spectrometry also provided evidence regarding glycosylation of the protein. The first of two potential glycosylation sites clearly carries carbohydrate; the second site is, at most, only partially glycosylated.

Mass spectrometry (MS)[1] is now a well-accepted technique in the overall structural analysis of peptides and proteins. The recent dramatic increase in the use of MS in such studies is largely due to the advent of fast atom bombardment (FAB) (Barber et al., 1982), an ionization technique that provides the molecular weights of as little as picomole amounts of peptides present in complex mixtures without the need for extensive chromatographic purification or chemical derivatization prior to analysis [for recent reviews, see Biemann and Martin (1987) and Carr et al., 1988)]. The FABMS peptide mapping procedure can be used to rapidly corroborate protein sequences obtained by DNA or cDNA sequencing. Peptide signals that do not map into the predicted sequence may indicate the presence of cloning errors, sequencing errors, or posttranslational modifications. In favorable cases the modification or error may be localized to a specific region of the otherwise correct protein sequence. Because sequence information is generally very limited or not obtainable by FABMS alone, the strategy described above is only applicable to proteins of known primary structure.

Tandem mass spectrometry employing two consecutive mass analyzers can provide structural information on individual peptides in complex mixtures, and therefore, this technique has great potential in the sequencing of peptides derived from proteins (Biemann & Martin, 1987; Hunt et al., 1986; Johnson & Biemann, 1987; Crabb et al., 1986; Biemann & Scoble, 1987; Carr et al., 1988). Protonated molecular ions $(M + H)^+$ of peptides in a mixture can be selectively fragmented, and under favorable circumstances, the sequences can be deduced

† This work was supported by grants from the National Institutes of Health to R.J.A. (GM 32602) and S.L. (AM 09461) and from Smith Kline and French Laboratories to S.A.C., R.J.A., and I.Y.H. These results were presented in preliminary form at the 35th Annual Conference on Mass Spectrometry and Allied Topics, May 24-29, 1987, Denver, CO (Carr, et al., 1987a), and at the 78th Annual Meeting of the American Society of Biological Chemists, June 7-11, 1987, Philadelphia, PA (Carr et al., 1987b).

* To whom correspondence should be addressed.
‡ Smith Kline and French Laboratories.
§ On sabbatical leave from the University of Maine, Orono, ME.
∥ The University of Chicago.

[1] Abbreviations: FABMS, fast atom bombardment mass spectrometry; MS, mass spectrometry; FAB, fast atom bombardment; SBP, spermine binding protein; HPLC, high-performance liquid chromatography; PITC, phenyl isothiocyanate; PTH, phenylthiohydantoin; TFA, trifluoroacetic acid; Pca, 5-oxopyrrolidine-2-carboxylic acid; Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid; DTT, dithiothreitol; RP-HPLC, reversed-phase high-performance liquid chromatography.

0006-2960/88/0427-4214$01.50/0 © 1988 American Chemical Society