of velocity and concentration fluctuations would be nonsensical for instantaneous fluctuations. As is well known, the association applies only to concentration fluctuations smoothed over a velocity relaxation time (less than a picosecond for the mobile ions). If we examine the relative concentration fluctuations on the two sides of a plane of discontinuity in the equilibrium (mean) concentration, we would require identity in the tangential derivatives to be established only after one or a few velocity relaxation times. So much is implicitly assumed at the very beginning, when we write down the Smoluchowski equation.

The constancy that we have used in the derivation of boundary conditions at the surface of thin double layers goes further than a requirement of continuity across a plane. We require that the tangential components of ∇p_i are constant along the normal to a finite, albeit thin, double layer. Proof of this constancy in $\nabla_t p_i$ must follow the mathematically identical inference for the electrical analogue (constancy of the tangential part of the electrical field). What is required in these derivations is smoothness, i.e., approximate constancy, in all three components of ∇p_i over tangential displacements large compared to the thickness of the double layer. It is not excluded, indeed it is expected, that the normal component $\mathbf{n}\cdot\nabla p_i$ will vary rapidly across the double layer; only tangential variation of $\mathbf{n}\cdot\nabla p_i$ must be smooth. It is a consequence of this condition of tangential smoothness, for p_i and h_i as well as for Φ, that the tangential derivatives of all three quantities are constant with respect to displacements normal to the double layer. Achievement of tangential smoothness in ∇p_i, or in its individual parts ∇h_i and $\nabla\Phi$, will require a time interval which will be much larger than the velocity relaxation time. It will be the time for a salt ion to diffuse over distances large compared to the Debye length, i.e., it will be the Maxwell relaxation time. As we noted at several earlier points in the discussion,[1] our calculations are restricted to processes slow on that time scale.

In conclusion, any paradox arising from intuition about coupling or lack of it, between instantaneous (equilibrium) concentration fluctuations inside and outside the double layer, has little bearing on the dynamical calculations which are the subject of this work.

References and Notes

(1) M. Fixman, *J. Chem. Phys.*, in press. A discussion of background, especially the work of Dukhin, will be found here.
(2) C. Hornick and G. Weill, *Biopolymers*, **10**, 2345 (1971).
(3) F. van der Touw and M. Mandel, *Biophys. Chem.*, **2**, 218 (1974).
(4) We do not wish to be offensively critical or puristic. The experimental practice of subtracting a small steady state conductivity, and interpretation of the remainder according to equations derived for nonelectrolyte solutions, is extremely plausible in some circumstances. However, systems and frequencies for which the imaginary part of the complex dielectric constant is comparable to or larger than the real part require a dynamical study.
(5) M. Hogan, N. Dattagupta, and D. M. Crothers, *Proc. Natl. Acad. Sci. U.S.A.*, **75**, 195 (1978).
(6) K. Kikuchi and K. Yoshioka, *Biopolymers*, **15**, 583 (1976).
(7) The limit of the tangential flux in the double layer at $z = \pm L$ equals the limit of the normal flux into the end plates at $r = a$, and the second limit is 0.
(8) J. Skolnick and M. Fixman, *Macromolecules*, **11**, 867 (1978).
(9) L. D. Landau and E. M. Lifshitz, "Electrodynamics of Continuous Media", Pergamon Press, London, 1960, p 18.
(10) The electric field of a dipolar charge distribution oriented perpendicular to the cylinder varies as $\nabla\nabla \ln r$, close to the cylinder. r is the perpendicular distance to the cylinder. Far from the cylinder the field of any dipole varies as $\nabla\nabla R^{-1}$, where R is the distance to the cylinder midpoint. The field arises solely from dielectric discontinuities if polarization of the double layer is neglected.
(11) G. S. Manning, *Biophys. Chem.*, **9**, 65 (1978).
(12) F. Oosawa, "Polyelectrolytes", Marcel Dekker, New York, 1971. Oosawa's calculation incorporates an interaction energy between counterions that is evaluated in practice from a free volume model. The radius of the free volume is apparently proportional to the Debye length. We find this procedure rather obscure, but it gives the same result as the use of a screened (Debye–Hückel) interaction energy, with a short distance cutoff proportional to a. (The cutoff is necessitated by the restriction to one-dimensional charge distributions.)
(13) J. Th. G. Overbeek and P. H. Wiersema in "Electrophoresis", Vol. II, M. Bier, Ed., Academic Press, New York, 1967, p 1.
(14) E. Charney and J. B. Milstien, *Biopolymers*, **17**, 1629 (1978), and private communication from E. Charney of work in press.

Molecular Dissection of an Enzyme that Recognizes Transfer RNA[1]

Paul R. Schimmel,* Diane LaDage Meléndez, and Scott D. Putney[2]

Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

Received November 29, 1979

ABSTRACT: An analysis of one member of a class of enzymes—aminoacyl-tRNA synthetases—is being carried out. For this purpose a dissection of the polypeptide chain and of the gene for the enzyme is done, making heavy use of recombinant DNA technology. The enzyme Ala-tRNA synthetase specifically attaches alanine to alanine specific transfer RNA chains in the first step of protein biosynthesis. It is a tetramer of a 95 000 dalton polypeptide chain. Sites for the amino acid activation reaction, whereby ATP and alanine condense to form alanyladenylate, have been localized to the amino terminal half of the protein. The gene (*alaS*) for the protein has been cloned with a pBR322 vector and a restriction enzyme map has been constructed of the recombinant plasmid. With the gene in hand it is possible to manipulate the polypeptide structure through a scheme which gives localized mutagenesis at defined sites in the gene. In this way, and with an appropriate selection scheme, enzyme molecules with altered tRNA recognition specificity can be constructed. Rapid DNA sequencing permits ready identification of polypeptide chain sequences (deduced from the DNA sequences) and a quick determination of the nature of any mutations that are introduced.

Aminoacyl-tRNA synthetases catalyze the first step in protein synthesis whereby amino acids are attached to their cognate transfer RNA chains according to the reactions[3]

0024-9297/80/2213-0716$01.00/0

$$E + AA + ATP \rightleftarrows E{\cdot}AA{\sim}AMP + PP_i \quad (1)$$

$$E{\cdot}AA{\sim}AMP + tRNA \rightleftarrows AA\text{-}tRNA + AMP + E \quad (2)$$

where E is enzyme, AA, ATP, PP_i, and tRNA denote amino acid, adenosine triphosphate, pyrophosphate, and transfer RNA, and AA~AMP is the aminoacyl adenylate that is firmly bound to the enzyme. For each amino acid there is one or more distinct tRNA species and a specific aminoacyl-tRNA synthetase. The job of the synthetase is to match properly each amino acid with its transfer RNA. The tRNA's are single stranded nucleic acids generally comprised of 73 to 93 nucleotides[4,5] which, through intramolecular hydrogen bonding between base pairs, organize themselves into a specific secondary and tertiary three-dimensional structure.[4,6-8] Within each tRNA is a trinucleotide sequence, the anticodon, which is complementary to the codon of the amino acid which it represents. The attachment of amino acids to transfer RNA's is the first place in the outworkings of the genetic code where an amino acid is given correspondence in terms of a trinucleotide sequence.[3,4,9] Thus, the synthetases establish the fundamental connection between nucleotide sequences and amino acids that is used universally by all living organisms, from the simplest procaryotes to the most complex eukaryotes. Surprisingly, as a general rule the synthetases do not simply recognize the anticodon sequence and use only this as the basis for matching the amino acids with the proper trinucleotide sequence; the recognition process involves other parts of the tRNA structure and in some cases may not even involve the anticodon.[3,9,10]

These enzymes are undoubtedly an ancient component of biological evolution that probably appeared well over a billion years ago. Because the various enzymes carry out the same general reactions and differ only in their specificities, it is plausible that structural relationships and homologies exist among them, although this is yet to be proven. We would like to understand the structural organization of these enzymes and the structural relationships between them. From this we may be able to understand how the recognition of transfer RNAs is achieved and whether, in the linear sequence of amino acids, there are domains corresponding to the sites for amino acid, ATP, and transfer RNA. If this is true, then one can imagine that, among this class of enzymes, diversity in specificity is achieved by concentrating amino acid sequence variations in specific domains of the structure. This would be analogous to the situation with immunoglobulins which achieve diversity and antigen binding specificity by sequence variations in specific parts of the polypeptide chain (i.e., the variable region[11]).

To perform this kind of analysis at the molecular level is a formidable task. This is because many of the synthetases are made up of rather long polypeptide chains (as much as 1000 amino acids[3]) and also because our methods for mapping active sites in proteins are cumbersome. For example, while much progress has been made in amino acid sequence analysis over the past several years, it can still take years to complete the sequence of very long polypeptide chains. Nevertheless, the primary structure determination is a first step in any investigation that attempts to probe structure and function at the molecular level. And the problem is compounded greatly when one wishes to make comparisons between a whole class of such proteins. It is the purpose of this article to describe some of the approaches we are now taking to achieve the molecular dissection of an aminoacyl-tRNA synthetase. These approaches make heavy use of recombinant DNA technology, recent rapid DNA sequencing methods, and in vitro methods for introducing specific amino acid changes into the polypeptide sequence. These approaches rely on methods for dissection of the gene for an aminoacyl-tRNA synthetase and this, in turn, is used as a means for obtaining information on, and alteration of, the polypeptide chain produced by the gene. These methods have enormous advantages over more classical approaches, both in terms of the rapidity at which structural information is obtained and in terms of performing structural alterations and manipulations.

Characterization of Alanine tRNA Synthetase

We undertook an analysis of Ala-tRNA synthetase because this enzyme has advantages in certain genetic selection methods we planned to use (see below). However, little was known about this enzyme and this necessitated that, at the outset, we carry out a basic characterization.

The enzyme was purified from *E. coli* by using a strain we had constructed that overproduces the enzyme. This was done by inserting the gene for the enzyme on to an autonomously replicating recombinant plasmid, as described below. In strains carrying this plasmid, it is possible to arrange conditions so that cells produce as much as 8–10% of their total soluble protein as Ala-tRNA synthetase.[12] This enables us to obtain large amounts of the enzyme from relatively small quantities of cells.

The enzyme was purified by standard methods and first characterized with respect to its molecular weight and subunit composition. By utilizing gel exclusion chromatography,[13,14] together with known molecular weight markers, we estimated the molecular weight at 380 000 daltons. When the protein was analyzed in a dissociating SDS stacking gel electrophoresis system,[15] a single band with a molecular weight of 95 000 was observed. These results suggest that the protein is a tetramer of four identical polypeptide chains and thus has a quaternary structure of α_4.[16] This conclusion was confirmed by cross-linking experiments where it was shown that a stable tetramer, but not a larger species, could be generated by treating the enzyme with the bifunctional cross-linking reagent, dimethyl suberimidate.[17]

This result places Ala-tRNA synthetase as the largest known bacterial aminoacyl-tRNA synthetase. In general, synthetases have subunit structures of α, α_2, $\alpha_2\beta_2$, and α_4, with subunit polypeptide chains ranging in molecular weight from 37 000 to 120 000.[3] This diversity of subunit structure may be somewhat superficial, because numerous studies have suggested that those synthetases which contain large polypeptide chains may have sequence repeats within these chains.[18-22] That is, a large α-enzyme polypeptide may be analogous to a covalent dimeric structure, because of sequence repeats in the α-enzyme. Until sequence relationships among the various subunits of the different synthetases are worked out, the meaning of the quaternary structure heterogeneity of the enzymes will remain an open question.

In the course of these investigations, we discovered that a fragment of 48 000 daltons could be generated from the native enzyme.[16] This fragment is liberated by the presence of trace amounts of a protease. The fragment was purified and found to chromatograph on molecular sieve columns as a free particle, and thus has no strong interfragment interactions. In order to determine which part of the native 95 000-dalton subunit gives rise to the 48 000-dalton fragment, the *N*-terminal amino acid sequence of both the native subunit and the fragment was determined. The sequence at the *N*-terminus of the native protein is Ser·Lys·Ser·Thr·Ala·Glu·Ile·Arg·Gln·Ala·Phe. An analysis of the 48 000-dalton fragment gave exactly the

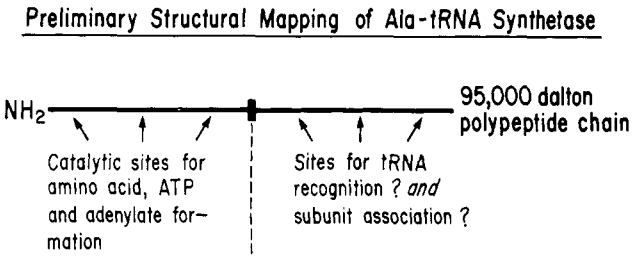

Figure 1. Schematic illustration of preliminary structural mapping of *E. coli* Ala-tRNA synthetase.

same amino acid terminal sequence.[16] Therefore, there is little question that the fragment is derived from the amino terminal end of the native subunit.

Studies were also done on the catalytic activity of the fragment. It was found that the fragment efficiently carries out the amino acid activation reaction whereby aminoacyl adenylate is formed (eq 1A), but aminoacylation of tRNA is not observed (eq 1B). Apparently the fragment lacks a portion of the structure required for the interaction with tRNA, although the exact nature of this deficiency is not clear. In any event, the results suggest that the portion of the polypeptide necessary for aminoacyl adenylate formation can be localized to the amino terminal half of the native subunit.

Figure 1 shows a schematic illustration of the information obtained thus far. Sites for aminoacyl adenylate formation are localized on the *N*-terminal half of the protein, while those for tRNA recognition and aminoacylation may be contained, at least in part, on the other half of the fragment. In addition, because the fragment undergoes no associations, we surmise that the carboxyl-half of the protein contains the information necessary for subunit interactions. But we have not yet been able to isolate the carboxyl-terminal half in order to confirm some of these conclusions.

Thus, using methods of protein chemistry we have been able to achieve a partial dissection of Ala-tRNA synthetase and to establish some simple aspects about its structural organization. But to analyze further the structural organization of this enzyme by standard methods of protein chemistry is a tedious process which is somewhat restricted in its potential scope. For these reasons, we have turned to isolation and analysis of the gene (*alaS*) itself. This is far easier to analyze and, as described below, can be manipulated in a virtually limitless number of ways as a means to gain deeper insight into the structural organization and catalytic mechanism of the protein.

Cloning and Restriction Enzyme Analysis of the Gene for Ala-tRNA Synthetase

The gene (*alaS*) for Ala-tRNA synthetase was isolated by standard recombinant DNA techniques. In brief, *E. coli* chromosomal DNA was split into defined pieces with one or more restriction enzymes.[23] These enzymes cleave at specific nucleotide sequences and, therefore, give specific DNA fragments. The enzymes which we employ for cloning procedures recognize a six-base sequence and, therefore, generate fragments of an average length of $4^6 \approx 4000$ nucleotides. These pieces were then ligated into a circular plasmid vector (pBR322[24,25]) to give a pool of recombinant plasmids that were then transformed into an appropriate recipient where selection was done for the gene of interest. For example, we used a recipient which has a defective Ala-tRNA synthetase and which, as a result, will not grow, under some conditions. By selecting (after transformation with the pool of chromosomal pieces that were inserted into pBR322) for cells that could grow in

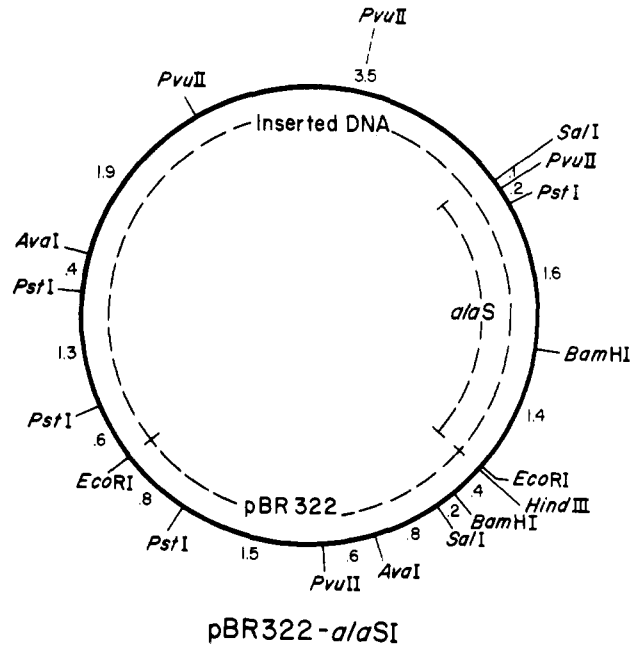

Figure 2. Schematic illustration of the recombinant plasmid pBR322-*alaSI*. The cloning vector pBR322, the piece of DNA inserted from the chromosome, and the part corresponding to *alaS* are each indicated. Restriction enzyme sites are also designated and the numbers indicate the sizes (in units of 1000 base pairs) of the various DNA segments. In the case of one of the *Pvu*II sites there is a small uncertainity (±400 base pairs) in its location and this site is therefore designated by a dashed line. The restriction enzyme sites in the pBR322 portion of the plasmid are based on those given in reference 25.

spite of their Ala-tRNA synthetase defect, we were able to identify a clone which had acquired a viable Ala-tRNA synthetase and, therefore, had picked up the recombinant plasmid that contains *alaS* (pBR322-*alaS*[12]). We used this "shotgun" transformation procedure to clone *alaS* from a pool of chromosomal DNA which had been digested with the restriction enzyme *EcoRI*. It is then a simple matter to isolate large quantities of the plasmid pBR322-*alaS* from cells which carry it.

Figure 2 gives a schematic illustration of the recombinant plasmid pBR322-*alaSI*. The part of the recombinant plasmid that is derived from the cloning vector, pBR322, is indicated. This cloning vector contains an autonomous replicon which enables the recombinant plasmid to multiply independently within the cell.[24] Inserted into this plasmid is a piece of chromosomal DNA that we have designated as "inserted DNA". This is a large piece of DNA (approximately 11 000 base pairs) while the vector itself is about 4200 base pairs.[24,25] The diagram in Figure 2 indicates where various restriction enzymes cleave the recombinant plasmid; this restriction analysis is useful for the further analysis which is described below. By chopping the inserted DNA into large pieces with restriction enzymes, it is possible to reclone portions of the inserted DNA in order to identify which part contains *alaS*. This procedure was followed to establish the location in pBR322-*alaSI* of the gene (*alaS*) for the synthetase; this is also designated in Figure 2. Subsequently, this segment (*alaS*) has been removed and cloned directly into part of the pBR322 vector, without the remaining extraneous inserted DNA. This plasmid we have designated pBR322-*alaSII*.

With cells containing pBR322-*alaSI* we have been able to obtain conditions where the autonomously replicating plasmid gives a large overdose of Ala-tRNA synthetase.

As mentioned above, it is possible to have cells produce up to 8–10% of their total soluble protein as this enzyme. This not only greatly increases the yield of enzyme from preparations, but also greatly simplifies the purification procedure because only a 10- to 20-fold purification is required to obtain homogeneous enzyme.

However, for our purposes, the main advantage of having the gene isolated on a plasmid is not simply to get greater yields of protein, but rather because, as described below, determination of the structural domains of the enzyme and active site mapping can be done more easily when we work with the gene rather than with the protein.

Approach for Active Site Mapping

There are two major advantages to approaching structure-function questions at the level of the DNA rather than at that of the protein. First, the sequence of the DNA can be determined far more rapidly than that of the protein. It is possible to do 200 to 500 nucleotides per day with the use of rapid DNA sequencing methods.[26,27] From the DNA sequence, the sequence of the protein can be deduced. In order to be sure that the protein sequence derived from the DNA sequence is accurate, it is important to have the sequence of a few peptides within the protein. For example, the sequence at the *N*-terminus indicates where in the DNA sequence the protein actually begins. If some random peptides are isolated from a proteolytic digest of the protein, and these peptides are sequenced, it should be possible to find the corresponding sequences within the DNA. If this is done for a few peptides that are scattered in the primary structure, and there is a perfect match between the DNA sequence and the peptide sequence, then it is likely that most of the entire DNA sequence is reliable. This is because an error in the DNA sequence resulting from a missed base, for example, would throw the triplet codons out of phase and result in an inability to line up peptide sequences "downstream" from the error within the DNA sequence. However, there is another built-in safeguard against missed or extra bases in the DNA sequence because they are apt to generate premature stop signals. Because 3 out of the 64 trinucleotide codons are stops, the random probability of encountering a stop is about 5%. If a stop is encountered before enough DNA has been read to give a polypeptide corresponding in length to that of the protein, then it is likely an error in the sequence (such as an extra or a missed base) has introduced a phase shift.

At present we have sequenced about 2000 nucleotides from the piece of DNA designated *alaS* in Figure 2. Altogether, this piece is comprised of about 3200 nucleotides which is slightly greater than what we estimate is necessary to encode the 95 000-dalton subunit of the synthetase. In addition to the *N*-terminal sequence which we have already determined, sections of the polypeptide chain are also being sequenced by using mass spectrometry methods in collaboration with Professor K. Biemann.[28]

The second major advantage of approaching structure-function questions at the level of the gene has to do with our analysis of active site regions and structural domains. With the gene in hand, we can freely introduce nucleotide base substitutions which, in turn, will give rise to specific amino acid changes. This is possible because of well-developed in vitro mutagenesis procedures and because mutagenesis can be concentrated at a specific local region within the gene. In this way, a systematic study can be made of amino acid substitutions along the entire sequence of the protein.

One approach is illustrated in Figure 3. The *alaS* segment is removed from the pBR322 cloning vector and subdivided into specific pieces with the aid of restriction

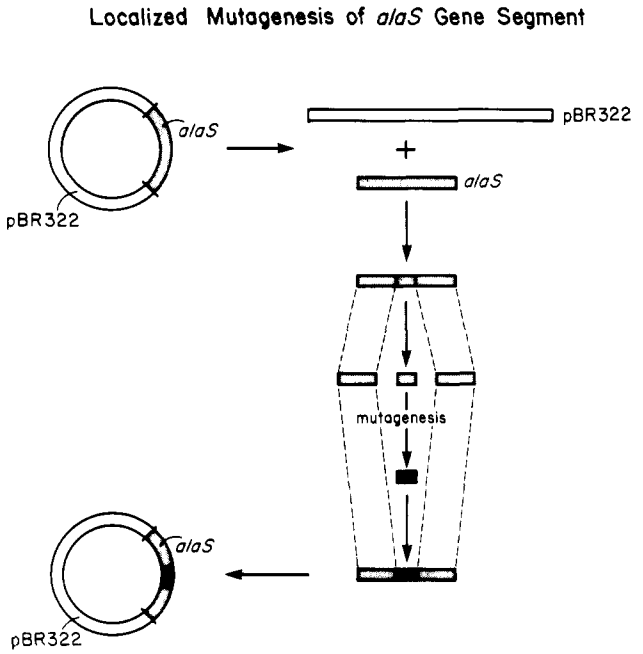

Figure 3. Scheme for localized mutagenesis of a small, defined segment of *alaS* which is fused with pBR322.

enzymes. One piece is removed and mutagenized in vitro, using one or more standard mutagens such as hydroxylamine.[29] The mutagenized piece is then reconnected with the remainder of the *alaS* gene and the reconstituted gene is in turn reinserted into the pBR322 vector. The reconstituted plasmid is then transformed into an appropriate recipient and selection is done for the phenotype of interest. Because of the availability of an extensive number of restriction enzymes,[23,30] it is possible to subdivide *alaS* DNA into an almost limitless number of pieces that can be systematically investigated (e.g., mutagenized). In this way, mutations are concentrated in one small region and, clearly, the entire gene can be examined piecemeal by this approach.

The key part of this procedure is the selection method used to obtain a particular phenotype. These can range from relatively simple to quite sophisticated. For example, assume that it is desired to locate within the sequence the positions of residues which affect the interaction of enzyme with its specific amino acid. The defect might be expressed as a highly elevated dissociation constant for the enzyme–amino acid complex. The selection could be done for cells which grow only in the presence of high concentrations of the amino acid, but not the lower, more common, physiological concentrations. This and other kinds of selections have been used to isolate mutant synthetases with altered affinities for their respective amino acids.[31,32]

In a more complex problem, which we have approached, we have attempted to introduce mutations into *alaS* which would actually change its specificity for the tRNA. Normally, this synthetase only attaches alanine to $tRNA^{Ala}$. However, based on structural work in our laboratory and in other laboratories, it has been proposed that the tRNA binding sites on synthetases have a great deal in common.[3,9,10,33,34] It is envisioned that there is a general framework of amino acids which binds to the tRNA and that amino acids placed in specific positions give rise to specificity. Thus, one might be able to achieve a specificity change by a few amino acid substitutions in the tRNA binding site of the synthetase. With this in mind, we have mutagenized the gene for *alaS* and selected for synthetase molecules which will attach alanine to $tRNA^{Ile}$.[35] Although

there are some complications associated with this system, using a special selection scheme we have been able to select cells which, upon examining the Ala-tRNA synthetase activity in cell extracts, have enzyme activity that can attach alanine to $tRNA^{Ile}$. That is, an actual tRNA recognition specificity alteration has been introduced. This not only gives support to the idea that the synthetase binding clefts can be manipulated so as to alter their actual specificity, but also illustrates the power of these kinds of recombinant DNA approaches.

If the mutagenesis is carried out as shown in Figure 3, then identification of the site of mutagenesis is relatively straightforward. The small piece from the gene that was mutagenized is reisolated and sequenced by rapid DNA sequencing methods. In this way the base substitution (or substitutions) can be quickly identified and related to amino acid changes in the protein. Moreover, even without sequencing, one can quickly learn whether mutations of a particular type always cluster in the same part of the genome. For example, if mutations that alter tRNA recognition specificity tend to cluster in a limited region of, say, 300 nucleotides, then it would be reasonable to assume that this is the location of the nucleic acid binding site within the primary structure. This entire region can then be sequenced. Subsequently, a similar analysis can be made of other aminoacyl-tRNA synthetase genes and, by this process, the question can be answered of whether sequence homologies exist in the portions of their structures that are concerned with tRNA recognition. Thus, an expansion of the procedure outlined in Figure 3 offers a ready means to tackle the question of structural relationships between the various aminoacyl-tRNA synthetases.

The inverse of the procedure used in Figure 3 can also be used to identify regions within the sequence that are crucial for a particular enzymatic activity or function. The entire gene segment can be mutagenized with subsequent selection for the phenotype of interest. The mutagenized gene that gives the desired phenotype can then be removed and cut into defined pieces. These pieces in turn can be exchanged for their counterparts, individually and in separate experiments, in the normal unmodified gene. These reconstituted genes can then be tested separately to see which piece or pieces are responsible for the mutant phenotype.

It should be mentioned that it is possible to "fine tune" the mutagenesis to preselected sites. In the procedure sketched in Figure 3, we envision broad, nonselective mutagenesis on a small segment of the cloned gene. However, it is possible to alter a precise site within the gene by using, for example, the procedure described by Shortle and Nathans.[36] In this method, a single-stranded nick is introduced by a sequence-specific restriction endonuclease. After removing a few nucleotides from the site of the nick in order to expose a small single-stranded segment of a few bases, mutagenesis is done with a base-specific, single-strand-specific mutagen such as sodium bisulfite.[37] The mutation is then fixed in place at the specific site by enzymatic closure of the single-stranded gap. This method is particularly powerful when the entire DNA sequence of the gene is known, and the location of important binding sites has already been at least tentatively located within the sequence. Then, one can use such a method to make defined, predetermined changes in order to answer questions about the roles of specific amino acids in the structure. But in the absence of fairly detailed information about the structure, the screening method of Figure 3 is a preferable route to determine rapidly the general function and significance of the various parts of the polypeptide chain.

Conclusions

For many years there has been an attempt to achieve a molecular understanding of protein structure and function. Flory and co-workers laid the foundation for much of this work with calculations of conformational energies and configurational statistics of polypeptide chains.[38] These studies laid the basis for further investigations which have attempted to utilize physical data, and conformational energies of amino acid residues situated within the polypeptide chains, as a basis for understanding the specific folding pattern of proteins.[39,40] As recent literature indicates, there has been considerable progress on predicting protein secondary structure from amino acid sequence, and even the basis for tertiary structure assembly patterns is beginning to be understood.[40]

As described above, going hand-in-hand with these advances in understanding of protein folding has been an interesting ability to determine and to manipulate the sequences of DNA and, consequently, of polypeptide chains. In the case of Ala-tRNA synthetase, we are actually manipulating the tRNA recognition specificity of this enzyme through amino acid substitutions, utilizing mutagenesis procedures on the gene itself.

It should be mentioned that, using an entirely different approach, Miller and co-workers have generated more than 300 altered *lac* repressor proteins carrying known amino acid replacements.[41-43] This protein, which specificially recognizes *lac* operator DNA, has been modified by genetic methods which take advantage of the suppression of various nonsense mutations by several of the known nonsense suppressors. Although these studies are not aimed at achieving actual changes in DNA sequence recognition, they have shown the wide variety of amino acid substitutions that can be introduced in this system, and this in turn has provided rich material for protein structure-function studies.

As our theoretical tools and experimental methods improve, we approach the point where an understanding of protein structure and the relationship of structure to function becomes sufficiently sophisticated that we can manipulate quite freely primary structures and, consequently, secondary and tertiary structures and biological activities, with an increasing ability to predict and engineer the outcome of these intentional manipulations.

References and Notes

(1) This work was supported by Grant No. GM 23562 from the National Institutes of Health.
(2) Predoctoral trainee of the National Institutes of Health.
(3) P. R. Schimmel and D. Söll, *Annu. Rev. Biochem.*, **48**, 601 (1979).
(4) A. Rich and U. L. RajBhandary, *Annu. Rev. Biochem.*, **45**, 805 (1976).
(5) M. Sprinzl, F. Gruter, and D. H. Gauss, *Nucleic Acids Res.*, **5**, r15 (1978).
(6) S. H. Kim, F. L. Suddath, G. J. Quigley, A. McPherson, J. L. Sussman, A. H.-J. Wang, N. C. Seeman, and A. Rich, *Science*, **185**, 435 (1974).
(7) J. D. Robertus, J. E. Ladner, J. T. Finch, D. Rhodes, R. S. Brown, B. F. C. Clark, and A. Klug, *Nature (London)*, **250**, 546 (1974).
(8) A. Rich and S.-H. Kim, *Sci. Am.*, **238**, No. 1, 52 (1978).
(9) P. R. Schimmel, *Mol. Cell. Biochem.*, **25**, 3 (1979).
(10) P. R. Schimmel, *Adv. Enzymol. Relat. Areas Mol. Biol.*, **49**, 187 (1979).
(11) L. M. Amzel and R. J. Poljak, *Annu. Rev. Biochem.*, **48**, 961 (1979).
(12) S. D. Putney, D. L. Meléndez, and P. R. Schimmel, in preparation.
(13) J. R. Whitaker, *Anal. Chem.*, **35**, 1950 (1963).

(14) P. Andrews, *Biochemistry*, **91**, 222 (1964).
(15) U. K. Laemmli, *Nature (London)*, **227**, 680 (1970).
(16) S. D. Putney and P. R. Schimmel, in preparation.
(17) G. Davies and G. Stark, *Proc. Natl. Acad. Sci. U.S.A.*, **66**, 651 (1970).
(18) M. R. Kula, *FEBS Lett.*, **35**, 299 (1973).
(19) G. L. E. Koch, Y. Boulanger, and B. S. Hartley, *Nature (London)*, **249**, 316 (1974).
(20) R. M. Waterson and W. H. Konigsberg, *Proc. Natl. Acad. Sci. U.S.A.*, **71**, 376 (1974).
(21) C. J. Bruton, R. Jakes, and G. L. E. Koch, *FEBS Lett.*, **45**, 26 (1974).
(22) S. Robbe-Saul, F. Fasiolo, and Y. Boulanger, *FEBS Lett.*, **84**, 57 (1977).
(23) D. Nathans and H. O. Smith, *Annu. Rev. Biochem.*, **44**, 273 (1975).
(24) F. Bolivar, R. L. Rodriques, P. J. Greene, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Crosa, and S. Falkow, *Gene*, **2**, 95 (1977).
(25) J. G. Sutcliffe, *Nucleic Acids Res.*, **5**, 2721 (1978).
(26) F. Sanger, S. Nicklen, and A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.*, **74**, 5463 (1977).
(27) A. M. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.*, **74**, 560 (1970).
(28) K. Biemann, "Biochemical Application of Mass Spectrometry", Suppl. I, G. R. Waller, Ed., Wiley, New York, in press.
(29) T. Hashimoto and M. Sekiguchi, *J. Bacteriol.*, **127**, 1561 (1976).
(30) R. J. Roberts, *Gene*, **4**, 183 (1978).
(31) F. C. Neidhardt, J. Parker, and W. G. McKeever, *Annu. Rev. Microbiol.*, **29**, 215 (1975).
(32) J. E. Brenchley and L. S. Williams, *Annu. Rev. Microbiol.*, **29**, 251 (1975).
(33) P. R. Schimmel, *Acc. Chem. Res.*, **10**, 411 (1977).
(34) A. Rich and P. R. Schimmel, *Nucleic Acids Res.*, **4**, 1649 (1977).
(35) D. L. Meléndez and P. R. Schimmel, unpublished observation.
(36) D. Shortle and D. Nathans, *Proc. Natl. Acad. Sci. U.S.A.*, **75**, 2170 (1978).
(37) H. Hayatsu, *Prog. Nucleic Acid. Res. Mol. Biol.*, **16**, 75 (1976).
(38) P. J. Flory, "Statistical Mechanics of Chain Molecules", Chapter VII, Wiley, New York, 1969.
(39) H. A. Scheraga, *Pure Appl. Chem.*, **50**, 315 (1978).
(40) For a summary of recent literature see P. Y. Chou and G. D. Fasman, *Annu. Rev. Biochem.*, **47**, 251 (1978).
(41) J. H. Miller, C. Coulondre, M. Hofer, U. Schmeissner, H. Sommer, A. Schmitz, and P. Lu, *J. Mol. Biol.*, **131**, 191 (1979).
(42) J. H. Miller and U. Schmeissner, *J. Mol. Biol.*, **131**, 232 (1977).
(43) J. H. Miller, *J. Mol. Biol.*, **131**, 249 (1977).

Configurational Statistics of Polynucleotide Chains. An Updated Virtual Bond Model to Treat Effects of Base Stacking*

Wilma K. Olson

Department of Chemistry, Rutgers University, New Brunswick, New Jersey 08903.
Received February 13, 1980

ABSTRACT: An updated virtual bond scheme has been developed to include "long-range" effects of base stacking in the treatment of the spatial configurations of the polynucleotides. As a consequence of the relative rigidity of rotations about the C–O bonds (ϕ' and ϕ) of the sugar–phosphate backbone, it is possible to represent the six chemical bonds constituting each nucleotide repeating unit in terms of two virtual bonds of comparable magnitudes (spanning C–C–O–P bond fragments of the chain). The "long-range" (three-bond) interdependence of the alternate C–C and O–P bonds in the polynucleotide backbone somewhat complicates computations of chain averages compared to our previous treatments. Despite this complexity, the "long-range" interactions introduce a novel theoretical probe to deduce the solution conformation of the phosphodiester linkages ($\omega'\omega$) from the experimentally observed conformations of the C–C rotations ($\psi'\psi$). On the basis of hard-core conformational analysis and Karplus treatment of NMR coupling constants, we have modeled the helix-to-coil transition of poly(rA) over the temperature range −12 to 60 °C. At low temperatures the molecule is a flexible helix similar in conformational detail to the A-RNA family of structures. At higher temperatures as the nucleotide residues fluctuate over 12 distinct conformational domains, chain dimensions decrease in accordance with experimental data. In contrast to the conventional concept of unstacked, randomly coiling polynucleotides, these computations describe poly(rA) at high temperatures as a highly stacked system with 5 out of 12 of the nucleotides in such arrangements.

Recent applications of polymer chain statistics to the polynucleotides have provided a useful framework for relating subtle features of chemical architecture to the macroscopic properties of this system. As a consequence of geometric constraints imposed by the very complicated skeletal structure (cf. Figure 1), it is possible to represent the nucleotide repeating units comprising six chemical bonds by imaginary virtual bonds that span structural segments of fixed conformation. The virtual bond scheme first offered to treat the unperturbed dimensions of randomly coiling polynucleotides entails two such segments, one spanning four chemical bonds and the other two.[1] A later model applied to flexible nucleic acid helices involves a single virtual bond repeating unit that connects successive phosphorus atoms of the chain.[2-4] In both schemes the mutual orientation of any pair of successive virtual bonds is independent of the orientation of all other such pairs; statistical mechanical analyses of the spatial configurations of the polynucleotide as a whole are consequently facilitated.

Until now, the design of virtual bonds in polynucleotides and other biopolymers[5-8] has stemmed from the *local* conformational preferences of all pairs of *successive* rotation angles in the chain. In the two-virtual-bond scheme for polynucleotide random coils, the parameters originally determined to be flexible are the $\omega'\omega$ and $\phi\psi$ angle pairs (cf. Figure 1).[1,9] More recent studies,[10,14] however, suggest that ϕ is a "rigid" parameter confined almost exclusively to trans or t conformations and that only the ω', ω, ψ, and ψ' angles vary in polynucleotides. Because furanose puckering (or ψ') is fixed within the longer of the two

*In honor of P. J. Flory, a dear friend and teacher.

0024-9297/80/2213-0721$01.00/0