[36] Reverse Biochemistry: Methods and Applications for Synthesizing Yeast Proteins *in Vitro*

By KEVIN STRUHL

Introduction

Reverse biochemistry is a recent and relatively general approach for determining the function of proteins encoded by cloned genes and for analyzing the relationship between protein structure and function. The key feature of reverse biochemistry involves the synthesis of a desired protein from the cloned gene by transcription and translation *in vitro*. In contrast, classical biochemistry involves the identification and purification of proteins from crude cell-free extracts derived from living organisms. As is evident from the name, the relationship of reverse and classical biochemistry closely parallels the relationship between reverse and classical genetics.

The reverse biochemical approach has a number of attractive features. First, proteins can be synthesized *in vitro* in a short time, typically a few hours.[1,2] Moreover, the resulting products are radiopure and can be labeled to very high specific activities [up to 10^9 counts per minute (cpm)/μg protein]. In contrast, protein purification by classical procedures is usually very laborious and time consuming, and it is difficult to obtain radioactively labeled proteins that are functionally active. Second, any desired mutant protein can be generated and then analyzed simply by modifying the DNA template.[3,4] This is considerably easier than the isolation and purification of mutant proteins from living organisms, especially in cases where the mutant protein lacks activity. Third, because it is easy to follow their fate, radiolabeled proteins and their mutated derivatives can be very useful as substrates to assay for factors that modify,[5-7] directly interact with,[8,9] localize, or transport the protein.[5,7,10-12] Moreover, such *in vitro*

[1] D. Stueber, I. Ibrahimi, D. Cutler, B. Dobberstein, and H. Bujard, *EMBO J.* **3**, 3143 (1984).
[2] D. A. Melton, P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green, *Nucleic Acids Res.* **12**, 7035 (1984).
[3] I. A. Hope and K. Struhl, *Cell (Cambridge, Mass.)* **46**, 885 (1986).
[4] I. A. Hope and K. Struhl, *Cell (Cambridge, Mass.)* **43**, 177 (1985).
[5] D. Baker, L. Hicke, M. Rexach, M. Schleyer, and R. Schekman, *Cell (Cambridge, Mass.)* **54**, 335 (1988).
[6] G. Hawlitschek, H. Schneider, B. Schmidt, M. Tropschug, F.-U. Hartl, and W. Neupert, *Cell (Cambridge, Mass.)* **53**, 795 (1988).
[7] J. A. Rothblatt and D. I. Meyer, *Cell (Cambridge, Mass.)* **44**, 619 (1986).
[8] C. J. Brandl, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2652 (1989).

synthesized proteins are well suited for low-resolution structural analysis such as proteolytic mapping of domains and thermodynamic stability.[13] Fourth, the ability to cosynthesize functional proteins of different sizes can greatly facilitate the determination of the subunit structure (e.g., monomer, dimer)[14] of a protein.

Despite all the advantages listed above, it is important to note the limitations of reverse biochemistry. Most importantly, the approach is suitable only for the analysis of individual proteins for which the genes have been cloned. It is clearly also inappropriate for identifying new biochemical functions or for carrying out complex biochemical processes such as DNA replication, transcription, translation, and splicing. In addition, there is always the possibility that *in vitro* synthesized proteins may differ in some structurally or functionally significant way from proteins produced *in vivo*. Finally, *in vitro* synthesis can generate only relatively small amounts of proteins (likely maximum about 1 μg) and hence is unsuitable for high-resolution structural analyses such as X-ray crystallography or NMR.

Principle of Method

The availability of a cloned gene makes it possible to synthesize the encoded protein by *in vitro* transcription and translation. In the first step of the procedure (Fig. 1A), the desired protein-coding sequences are cloned into a vector containing a promoter for SP6 or T7 RNA polymerase.[15] The protein-coding sequence can be obtained from a cDNA clone, or, for the vast majority of yeast genes that lack introns, it can be generated from cloned genomic DNA. In accord with the rules for efficient translation in eukaryotic organisms, it is essential that the correct initiation codon is the 5′-proximal AUG in the mRNA to be synthesized.[16,17] In certain situations, the initial cDNA or genomic clone might have to be modified to optimize the lengths of the 5′- or 3′-untranslated sequences adjacent to the

[9] N. F. Lue, D. I. Chasman, A. R. Buchman, and R. D. Kornberg, *Mol. Cell. Biol.* **7**, 3446 (1987).
[10] W. Hansen, P. D. Garcia, and P. Walter, *Cell (Cambridge, Mass.)* **45**, 397 (1986).
[11] A. P. G. M. vanLoon, A. W. Brandli, and G. Schatz, *Cell (Cambridge, Mass.)* **44**, 801 (1986).
[12] M. Mueckler and H. F. Lodish, *Cell (Cambridge, Mass.)* **44**, 629 (1986).
[13] I. A. Hope, S. Mahadevan, and K. Struhl, *Nature (London)* **333**, 635 (1988).
[14] I. A. Hope and K. Struhl, *EMBO J.* **6**, 2781 (1987).
[15] P. A. Krieg and D. A. Melton, this series, Vol. 155, p. 397.
[16] M. Kozak, *Nucleic Acids Res.* **15**, 8125 (1987).
[17] A. M. Cigan and T. F. Donahue, *Gene* **59**, 1 (1987).

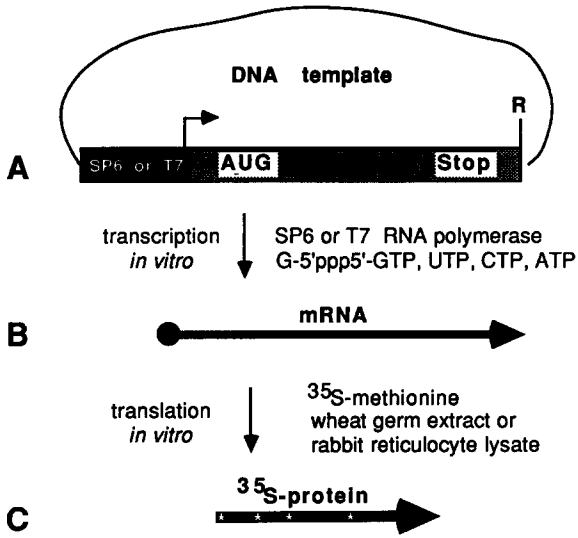


FIG. 1. Scheme for synthesizing proteins *in vitro*. (A) The template DNA contains a promoter for bacteriophage SP6 or T7 RNA polymerase (black bar; arrow indicates RNA start site) fused to a DNA fragment encoding the desired protein (gray bar with AUG and stop codons indicated) and is generally linearized by cleaving with a restriction endonuclease (R) just downstream of the protein-coding region. (B) The messenger RNA contains a 5′ cap (black circle) that is produced by transcription *in vitro*. (C) The protein labeled with ^{35}S at methionine residues (asterisks) is produced by translation *in vitro*.

protein-coding region. Given an initial DNA template that can be transcribed and translated *in vitro* to yield reasonable amounts of a desired protein, the wide array of standard recombinant DNA manipulations can be employed to generate any mutant template of interest from which any desired mutant protein can be synthesized.

Second, messenger RNA encoding the protein is produced by transcribing the DNA template with the appropriate bacteriophage RNA polymerase (Fig. 1B). This *in vitro* transcription step is performed at high concentrations of ribonucleotide triphosphates in order to obtain large amounts of full-length RNA. However, unlike standard transcription reactions, 90% of the GTP in the reaction mixture is replaced by a "capped" GTP analog such as diguanosine-GTP. In this way, 90% of the RNA synthesized contains a 5′-capped structure that is typical of eukaryotic mRNAs and is important for efficient translation.[18] Because SP6 and T7 RNA polymerases are highly specific for initiation only from their respec-

[18] P. A. Krieg and D. A. Melton, *Nucleic Acids Res.* **12**, 7057 (1984).

tive promoters,[19,20] the RNA synthesized is a single species that encodes the desired protein.

Third, the essentially pure mRNA is translated *in vitro* using wheat germ extracts or reticulocyte lysates (Fig. 1C). In most cases, translation is performed in the presence of [^{35}S]methionine so that the protein is synthesized as a radiolabeled species. Most importantly, because the translation reaction is programmed with a pure mRNA species, the only radioactively labeled species should be the protein of interest. However, it is important to remember that, although the protein synthesized *in vitro* is radiopure, it represents only a very small percentage of the protein present in the translation extract. In some cases, these unlabeled proteins in the translation extract can influence the properties of the radiolabeled protein of interest.

Once the protein is synthesized, it can be used for a wide variety of purposes. In most cases, the products of the translation reaction can be used directly without further purification. This chapter is not intended, however, to be a complete survey of the large number of current and potential applications for *in vitro* synthesized proteins. Nevertheless, the last section discusses some specific uses of reverse biochemistry such as determining whether a cloned gene encodes a specific DNA-binding protein and analyzing the structure and function of such specific DNA-binding proteins.

Methods

Design and Preparation of DNA Template

The first step is to subclone protein-coding sequences of interest into a plasmid vector that contains a promoter for bacteriophage SP6 or T7 RNA polymerase. The major consideration in designing the DNA template is that the mRNA generated by the bacteriophage RNA polymerase should permit efficient synthesis of the desired protein. Since *in vitro* translation reactions are almost always carried out in extracts from eukaryotic organisms (typically wheat germ, rabbit reticulocytes, and yeast), this means that the mRNA produced from the DNA template should fit the rules for efficient translational initiation in eukaryotes.[16,17] Obviously, synthesis of the desired protein requires that the uninterrupted protein-coding sequence be cloned in the correct orientation downstream of the bacterio-

[19] E. T. Butler and M. Chamberlin, *J. Biol. Chem.* **257**, 5772 (1982).

[20] M. Chamberlin and T. Ryan, *in* "The Enzymes," (P. Boyer, ed.), Vol. 7, p. 87. Academic Press, New York, 1972.

phage promoter. In addition, it is essential that the correct initiation codon is the 5′-proximal AUG in the mRNA to be synthesized because initiation (or reinitiation) from internal AUG codons is generally very inefficient.[21] As for most eukaryotic mRNAs, the AUG initiation codon should be relatively close (25–100 bases) to the 5′ end of the message.[16]

Because most yeast genes lack introns, protein-coding sequences are typically obtained from cloned genomic DNA. However, to ensure that the AUG initiation codon is properly situated, it is usually necessary to remove excess 5′-flanking nontranslated sequences from the initial cloned DNA. This is accomplished by cleavage at a fortuitous restriction site, sequential deletion by conventional *Bal*31 or exonuclease III–S1 nuclease procedures, or oligonucleotide-directed deletion. For some genes, synthesis of the encoded protein may be inefficient because the AUG initiation codon resides in a poor sequence context or is inaccessible owing to the secondary structure of the mRNA. In such cases, a reasonable way around the problem is to replace the normal 5′-untranslated sequences (and possibly the first few translated codons) with the equivalent region from an efficiently translated protein.

Given an appropriately designed template, plasmid DNA is prepared either by centrifugation in cesium chloride gradients or by an alternative procedure that yields relatively high-quality DNA; crude minipreparation-grade DNA may not be adequate.[15] Such DNA is then cleaved with a restriction endonuclease that cuts just downstream of termination codon [ideally 50–200 base pairs (bp)] and does not cut within the protein-coding region. Restriction sites in the polylinker of the cloning vector are often useful for this purpose, although if the site is too close (less than 50 bases) to the termination codon, the mRNA may be translated less efficiently. Such cleavage, though not essential, will usually result in higher molar amounts of the desired mRNA by minimizing the length of the transcript. It is desirable to avoid restriction enzymes that generate 3′ overhanging ends as these sometimes lead to *in vitro* transcription artifacts. Following cleavage, the DNA is purified by phenol extraction and then resuspended in Tris–EDTA (TE) buffer; at this stage the DNA should be as free of ribonucleases as possible. Typically, 10 μg of DNA is cleaved and resuspended in a volume of 50 μl; this is enough for 10 *in vitro* transcription and translation reactions.

Preparing mRNA by in Vitro Transcription

Detailed descriptions of *in vitro* transcription reactions involving bacteriophage SP6 or T7 RNA polymerase have been presented elsewhere.[15] For

[21] F. Sherman, J. W. Stewart, and A. M. Schweingruber, *Cell (Cambridge, Mass.)* **20**, 215 (1980).

the purpose of *in vitro* translation, the only significant difference is that 90% of the GTP in the reaction mixture is replaced by diguanosine-GTP so that 5′ capping of the RNA occurs concurrently with transcription. A typical transcription reaction is carried out in a volume of 25 μl and contains the components listed below. Reactions involving SP6 RNA polymerase should be set up at room temperature to avoid precipitation of the DNA template by spermidine. If desired, a small amount of labeled ribonucleotides (10 μCi) can be included to monitor the reaction.

Reaction Mix

8 μl Water (treated with diethyl pyrocarbonate)
5 μl DNA (total 1 μg)
5 μl 5× Ribonucleoside triphosphates [5 m*M* each ATP, CTP, UTP, diguanosine triphosphate (G-5′ppp5′-GTP); 0.5 m*M* GTP]
2.5 μl 10× Transcription buffer (400 m*M* Tris-HCl, pH 7.5; 60 m*M* $MgCl_2$; 100 m*M* dithiothreitol; 1 mg/ml bovine serum albumin)
2.5 μl 10 m*M* Spermidine (for SP6 RNA polymerase only; otherwise add water)
1 μl (30–60 units) Pancreatic ribonuclease inhibitor (RNasin)
1 μl (5–20 units) SP6 or T7 RNA polymerase

Transcription reactions are incubated at 40° (SP6 reactions) or 37° (T7 reactions) for 60 min. The reaction is terminated by extraction first with 25 μl buffered phenol, then twice with 2-butanol. The resulting mixture is brought to a final concentration of 2 *M* ammonium acetate, precipitated with ethanol, and washed once with ethanol. The RNA is resuspended in TE buffer containing 2 *M* ammonium acetate, reprecipitated with ethanol, washed once, and resuspended in 10 μl TE buffer. The RNA should be translated immediately or quick-frozen on dry ice and stored at −70°.

The most important consideration for a successful transcription reaction is the absence of ribonucleases throughout the procedure. A single ribonucleolytic cleavage in an RNA molecule will prevent the synthesis of the desired protein. With reasonable precautions, the transcription of full-length mRNA is rarely a problem because there are few efficient termination sequences for the bacteriophage RNA polymerases. When in doubt, the amount and quality of the RNA synthesized can be analyzed by electrophoresis in formaldehyde gels or by acid precipitation (if radiolabeled precursors are included in the reaction).

Preparing Protein by in Vitro Translation

For reasons of convenience, translation reactions are often carried out with commercially available kits that use either wheat germ extracts or rabbit reticulocyte lysates. Such commercial kits, though relatively expen-

sive, are generally recommended because the preparation of active translation extracts is not routine for most investigators and can be problematic. Procedures for making large quantities of active extracts from wheat germ[22] or rabbit reticulocytes[23] are well established, however, and they significantly reduce the cost and variability of commercial preparations.

When using commercial kits, simply add the products of the *in vitro* transcription reaction (1–10 μl of mRNA) to the appropriate reagents and follow the directions of the manufacturer; typical reactions are carried out in 30-μl volumes at room temperature for 30–60 min. In general, 1–2 μl of RNA is sufficient to produce the maximum amount of protein, although this may vary depending on the specific transcript. For each new preparation of translation extract, it is important to perform a control reaction that lacks added RNA. For most applications, the reaction mixture should contain 15 μCi of [^{35}S]methionine [1400 Ci/mmol) in order to radiolabel the protein. In some cases, however, unlabeled methionine is added so that the synthesized proteins are unlabeled or of lower specific activity. Proteins should be used as soon as possible, although they are generally stable at 0°–4° for 1 week. The relative instability of *in vitro* synthesized proteins primarily reflects the presence of proteases in the crude translation extracts. Protein preparations can be quick-frozen on dry ice and stored at −70° for longer periods of time. For this purpose, the protein should be divided into small aliquots prior to freezing; individual aliquots should be thawed only once, and then discarded after use.

To measure the amount of [^{35}S]methionine incorporated into protein, add 1 μl of the translation products to 50 μl of 0.1 *M* NaOH and incubate for 15 min at 37°. Then add 1 ml of 10% trichloroacetic acid, incubate for an additional 15 min on ice, and collect the acid-precipitated protein on glass fiber filters. By comparing the amount of incorporated [^{35}S]methionine in the sample to the control lacking added RNA, it is possible to estimate the amount of protein synthesized. Since wheat germ extracts lack any endogenous methionine, the specific activity of the protein can be directly calculated from the specific activity of the [^{35}S]methionine.[22] Reticulocyte lysates contain variable amounts of endogenous methionine, thus making the calculation more difficult.[23] In order to calculate the molar amount of protein synthesized, it is necessary to know the number of methionine residues per monomer product.

To determine the quality of the protein synthesized, 1 to 3 μl of the translation reaction should be analyzed by sodium dodecyl sulfate (SDS)–polyacrylamide gel electrophoresis; the ^{35}S-labeled proteins are easily visu-

[22] A. H. Erickson and G. Blobel, this series, Vol. 96, p. 38.
[23] R. J. Jackson and T. Hunt, this series, Vol. 96, p. 50.

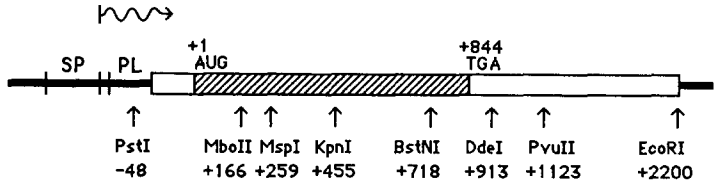
alized by fluorography and autoradiography. Ideally, the synthesis should yield a single band in 1–4 hr, the molecular weight of which corresponds to that expected for the protein of interest (Fig. 2). For some proteins (especially DNA-binding proteins), the mobility of the *in vitro* synthesized product is much slower than expected, and the apparent molecular weight can be strongly influenced by the gel conditions.[4] Further confirmation that the correct protein is being synthesized can be obtained by cleaving the template DNA at various positions within or just beyond the protein-coding sequences and examining the proteins synthesized[4] (Fig. 2).

Even in successful syntheses, minor translation products are often observed. For bands corresponding to lower molecular weight products, the most common reasons are premature termination of translation, aberrant initiation at internal AUG codons, or proteolysis. Bands representing higher molecular weight species are less common and may be due to inefficient termination at the correct stop codon or posttranslational modification. Some *in vitro* synthesis reactions yield a set of distinct products of similar molecular weight, which may arise from protein modification.

Many mRNAs can be translated efficiently either in wheat germ extracts or in rabbit reticulocyte lysates. In general, it is believed that wheat germ extracts initiate translation somewhat better than reticulocyte lysates but are more prone to premature termination (or degradation of the mRNA template). Thus, wheat germ extracts may be preferred for shorter proteins, whereas reticulocyte lysates may be the best choice for longer proteins. However, other nonpredictable factors can influence the translation efficiency for any particular protein. Wheat germ extracts contain no exogenous methionine and hence generate proteins with somewhat higher specific activity than those produced in reticulocyte lysates, which contain variable amounts of exogenous methionine. A final consideration is that the activity of proteins synthesized *in vitro* may vary owing to differences in modification or the amounts of critical cofactors or auxiliary proteins. Thus, the choice of translation extract can have profound effects on the amount and/or functional properties of protein synthesized.

For particular applications, it may be highly advantageous to use *in vitro* translation extracts from *Saccharomyces cerevisiae* itself[10,24] (see [37] and [46] in this volume). Obviously, such extracts must be employed in studies of the yeast translational machinery itself or in studies of processes that are intimately connected with translation such as some cases of protein transport. One major benefit of employing yeast extracts is that proteins are synthesized in a homologous system that is more conducive for correct posttranslational modifications and more likely to contain auxil-

[24] K. Moldave and E. Gasior, this series, Vol. 101, p. 644.



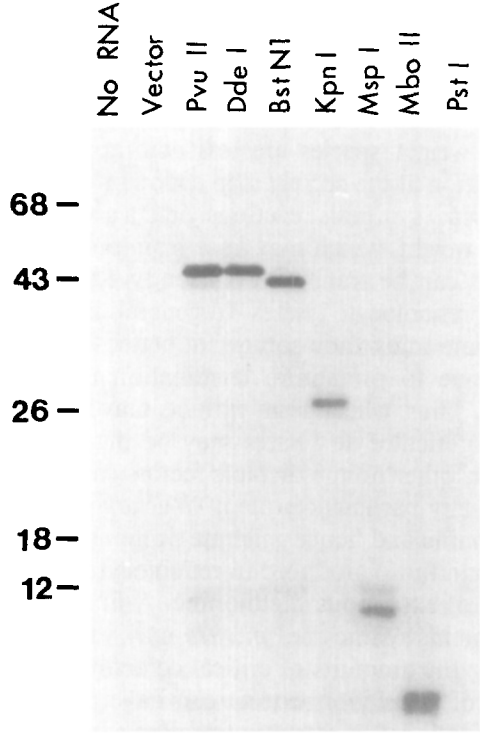


FIG. 2. Analysis by electrophoresis in denaturing gels of proteins synthesized *in vitro*. Shown above is a diagram of the template containing a bacteriophage SP6 promoter (SP) fused to a DNA fragment (open bar) encoding a protein (shaded region) with sites for restriction endonucleases as indicated. Shown below is an autoradiogram in which proteins were synthesized from template DNA cleaved with the indicated enzymes and examined by SDS–PAGE [mobilities of molecular weight markers ($\times 10^{-3}$) indicated at left]. Cleavage within the protein-coding region results in synthesis of proteins truncated at their carboxyl termini. (From Ref. 3.)

iary proteins necessary for function. In this sense, the proteins synthesized can be analyzed under conditions that are as physiologically significant as possible. Another powerful aspect is the potential to synthesize proteins in extracts from yeast strains that contain essentially any mutation of interest. Such a fusion between reverse biochemistry and classical and molecular genetics could provide a new way to examine how specific gene products affect the structure and function of a protein of interest.

The main disadvantage of *in vitro* translation extracts from yeast is that they are not commercially available, and there is considerably less experience in generating and using them. At this time, it is difficult to assess whether yeast extracts are equally suitable for efficient synthesis of a wide range of proteins. For making yeast extracts, it is probably advantageous to utilize *pept4* strains that lack the major proteases.[25] Another disadvantage of using a homologous translation system is that the extract itself will probably contain the protein of interest (unless the extract is generated from the appropriate mutant strain); this might complicate the interpretation of an experiment. Also, the presence of auxiliary proteins in the extract might make it difficult to determine if the protein of interest is directly or indirectly involved in a particular process.

Applications

Analysis of DNA-Binding Proteins

Reverse biochemistry is often used for determining whether a cloned gene encodes a specific DNA-binding protein. For this purpose, DNA-binding activity can be detected by incubating the labeled protein with appropriate DNA fragments, then separating the protein–DNA complexes from unbound protein and unbound DNA by electrophoresis in native acrylamide gels.[4] This DNA-binding assay is essentially the reverse of the standard mobility shift assay in which unlabeled proteins are examined for their ability to retard the mobility of a ^{32}P-labeled DNA fragment.[26-28] However, the "reverse assay" has the advantage that the fate of the ^{35}S-labeled protein is followed directly, thus making it possible to examine individual DNA fragments for both specific and nonspecific binding.

DNA fragments to be tested are prepared by restriction endonuclease cleavage of plasmid DNA followed by phenol extraction. Such fragments can be tested simultaneously, or they can be purified individually by

[25] E. W. Jones, *Annu. Rev. Genet.* **18**, 233 (1984).
[26] M. Garner and A. Revzin, *Nucleic Acids Res.* **9**, 3047 (1981).
[27] M. Fried and D. Crothers, *Nucleic Acids Res.* **9**, 6505 (1981).
[28] F. Strauss and A. Varshavsky, *Cell (Cambridge, Mass.)* **37**, 889 (1984).

agarose or acrylamide gel electrophoresis. Ideally, the specific binding site should reside on a fragment with a length between 50 and 500 bp and which differs from that of any other fragment generated by restriction cleavage. A typical DNA-binding reaction is carried out in a volume of 15 μl and contains the components indicated below. Of course, in order to interpret the results, it is essential to carry out parallel control reactions that contain no DNA and that contain nonspecific DNA lacking a binding site.

Reaction Mix

5 μl Water
3 μl 5× Binding buffer (100 m*M* Tris, pH 7.4; 250 m*M* KCl; 15 m*M* $MgCl_2$; 5 m*M* EDTA; 500 μg/ml gelatin)
5 μl DNA [DNA fragments each at 9 n*M*; equivalent to 0.5 μg of a 5-kilobase (kb) molecule]
1 μl 10 mg/ml Double-stranded poly(dI-dC)
1 μl ^{35}S-Labeled protein

After incubation at room temperature for 20 min, 5 μl of loading buffer (1× binding buffer containing 20% glycerol, 1 mg/ml each of xylene cyanol FF and bromphenol blue) is added, and the resulting mixture is immediately loaded on a 5% nondenaturing polyacrylamide gel. The polyacrylamide gel can either be of standard composition (30:0.8 acrylamide and bisacrylamide in 90 m*M* Tris-borate buffer, pH 8.3) or of low percentage and low ionic strength (40:0.5 acrylamide and bisacrylamide in 10 m*M* Tris-acetate buffer). Electrophoresis is performed under conditions in which the gel does not heat up above room temperature and until the bromphenol blue is near the bottom of the gel (a few hours). After electrophoresis, the gel is fixed in a solution of 45% ethanol, 10% acetic acid for 1 hr at room temperature, treated with En^{3}Hance (NEN-Du Pont, Boston, MA) for 1 hr, dried down, and autoradiographed.

As is the case with the conventional band-shift assay, the conditions for carrying out the binding reaction and electrophoresis can have an enormous impact on whether a particular protein-DNA interaction can be detected. The relative amounts of radiolabeled protein, specific DNA fragment, nonspecific DNA fragments, and bulk carrier DNA are particularly important because the presence of labeled protein means that both specific and nonspecific protein-DNA complexes are observable. Other important parameters that can be varied include the concentration of salt and divalent cations, temperature, the composition of the gel, and the ionic strength of the gel buffer.

The mobility of an individual protein depends on its charge-to-mass

ratio, a property that varies greatly among proteins and that is strongly affected by pH. Thus, the band corresponding to free protein can appear anywhere on the gel, and its location is strongly affected by the precise gel conditions. Under normal electrophoresis conditions, many radiolabeled proteins will migrate in the wrong direction owing to their positive charge at the pH of gel. This unpredictability emphasizes the necessity for proper controls to distinguish between bands corresponding to free protein and those corresponding to protein–DNA complexes (the unbound DNA is unlabeled; hence its mobility is irrelevant).

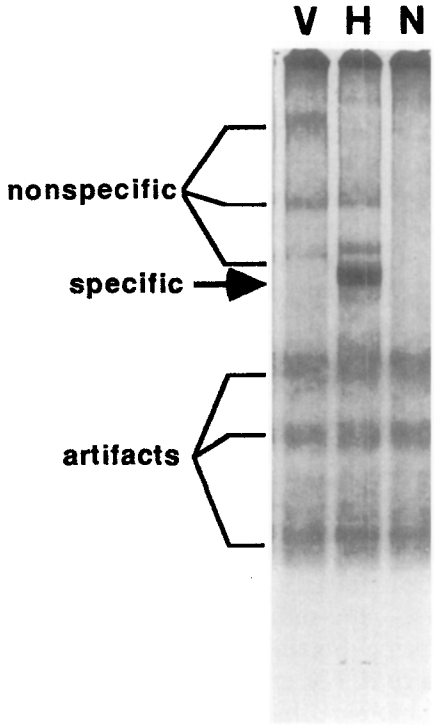


FIG. 3. Analysis of DNA-binding activity by a reverse mobility-shift assay. ^{35}S-Labeled protein was incubated in the presence of restriction fragments of plasmid DNAs that did (H) or did not (V) contain a specific binding site or in the absence of any DNA (N), and the resulting reaction products were subjected to electrophoresis in native acrylamide gels. Fragments containing specific binding sites generate dark bands (arrow), whereas fragments lacking such sites generate light bands owing to low-affinity nonspecific binding; the distinct mobilities of these bands reflect differences in the sizes of the DNA fragments. Artifact bands observed even in the absence of DNA are probably due to aberrant *in vitro* translation products.

Protein-DNA complexes are detected by the presence of new bands that do not appear in the control reaction containing protein alone (Fig. 3). Complexes involving a specific binding site should have significantly higher band intensities than those containing mutant binding sites or unrelated DNA sequences. Nonspecific, low-affinity interactions with DNA are characteristic of specific DNA-binding proteins and are observed as low-intensity bands. In considering the differences in band intensities between specific and nonspecific complexes, it is important to remember that a 400-bp DNA fragment containing a single specific binding site will have nearly 400 nonspecific binding sites. For such fragments, band intensities differing by a factor of 5 can reflect relative binding affinities that differ by a factor of 2000. The distinction between specific and nonspecific interactions can be enhanced by competition experiments in which the protein is incubated with a mixture of different sized DNA fragments.

The mobilities of protein–DNA complexes are influenced by a number of factors in addition to the electrophoretic conditions. First, complexes to larger DNA fragments migrate more slowly than complexes to smaller fragments, thus making it possible to simultaneously assay different sized DNA fragments.[4] Second, the mobility is affected by the amount of carrier DNA such as poly(dI–dC) in the reaction (C. R. Wobbe and K. Struhl, unpublished observations). In the absence of carrier DNA, the complex migrates very slowly because it contains nonspecific DNA-binding proteins from the translation extract in addition to the ^{35}S-labeled protein. As these nonspecific DNA-binding proteins are displaced by competition from the target fragment by increasing concentrations of carrier DNA, the protein–DNA complex migrates further in the gel. Third, the mobility of a protein–DNA complex is strongly influenced by the molecular weight of the protein component.[3,14] In general, complexes involving larger proteins migrate more slowly, although other factors such as shape and charge of the protein might also be involved.

Although the reverse band-shift assay just described is convenient, other methods for analyzing protein–DNA interactions can be applied to proteins synthesized *in vitro*. For example, DNA-binding activity can be assayed by the standard mobility shift assay using ^{32}P-labeled DNA,[26–28] in which case the protein does not have to be radiolabeled. Alternatively, protein–DNA complexes can be detected by immunoprecipitation.[29,30] *In vitro* synthesized proteins are also suitable for higher resolution analyses such as DNase I footprinting or methylation interference.

[29] V. Giguere, S. M. Hollenberg, M. G. Rosenfeld, and R. M. Evans, *Cell (Cambridge, Mass.)* **46**, 645 (1986).

[30] A. D. Johnson and I. Herskowitz, *Cell (Cambridge, Mass.)* **42**, 237 (1985).

Mapping Functional and Structural Domains

The region(s) of a protein necessary and sufficient for function can be mapped by generating appropriate deletions of the original DNA template, synthesizing the encoded mutant proteins, and testing them for the property of interest.[3,4] Initial mapping experiments typically involve a series of amino- or carboxyl-terminal deletions of the protein. However, proteins containing internal deletions or single amino acid substitutions can easily be tested, as can chimeric proteins. Since the amount and quality of the radiopure proteins can be easily assessed by SDS–PAGE, the failure of a mutant protein to carry out a particular function is not due to trivial reasons such as degradation. This approach has been used to localize protein regions required for DNA binding,[3,4] transcriptional activation *in vitro*,[31] ligand interactions,[29] and protein–protein associations.[8,9] However, it should be generally applicable to any functional property for which there is an appropriate assay. In general, it is unnecessary to purify the radiolabeled protein away from the vast excess of unlabeled proteins in the translation extract.

Because *in vitro* synthesized proteins are radiopure, they are well suited for mapping structural domains by partial proteolytic cleavage.[13] Unstructured regions in a protein are cleaved much more readily than independent structural domains.[32–34] In a typical experiment, a series of 10-μl reactions are performed in which a constant amount of radiolabeled protein is incubated with varying amounts of a given protease under appropriate conditions for 20 min.[13] Initially, it is useful to vary the protease concentration over a wide range (3–6 orders of magnitude) because a structurally informative amount will depend greatly on the specific protease and on the experimental conditions (a very rough average of a useful protease concentration is 10^{-4} units/μl of ^{35}S-labeled protein). The digestion products are then diluted with 20 μl of SDS sample buffer containing 500 μg/ml bovine serum albumin, boiled for 5 min, and examined by SDS–PAGE and autoradiography.

Protease cleavage sites can be mapped by cleaving a series of terminally deleted proteins.[13] For example, a band corresponding to the carboxyl-terminal region of the protein will be observed in amino-terminally deleted derivatives, whereas the band corresponding to the amino terminus will

[31] K. E. Vrana, M. E. A. Churchill, T. D. Tullius, and D. D. Brown, *Mol. Cell. Biol.* **8**, 1684 (1988).
[32] R. R. Porter, *Biochem. J.* **73**, 119 (1959).
[33] H. Jacobsen, H. Klenow, and K. Overgaard-Hansen, *Eur. J. Biochem.* **45**, 623 (1974).
[34] C. O. Pabo, R. T. Sauer, J. M. Sturtevant, and M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1608 (1979).

become progressively shorter (Fig. 4). Moreover, the presence of a given band in a set of proteolytically cleaved deletion proteins is indicative of the existence and boundaries of an independent structural domain. Further structural information can be obtained by using appropriate deletion and point mutants and a variety of proteases and by varying the cleavage conditions (e.g., temperature, ionic strength, concentration of denaturants). In interpreting the results, it is important to remember that protein fragments must contain at least one methionine residue in order to be visualized and that relative band intensities must be normalized to the number of methionines in each fragment.

Determining Subunit Structure

The availability of truncated but functional proteins can be useful for determining the subunit structure of a protein. Specifically, two derivatives of a given protein that differ in size are cosynthesized by carrying out *in vitro* transcription and translation starting from an equimolar mixture of DNA templates.[14] If, for example, the protein is a dimer, the cosynthesized mixture will generate three distinct species in a 1 : 2 : 1 molar ratio, with the intermediate species representing a heterodimer. In such an experiment, monomer proteins will generate two equimolar species, whereas tetramers will yield five species in a 1 : 4 : 6 : 4 : 1 molar ratio.

The principle of using different sized proteins to determine stoichiometry can be applied to a variety of methods for detecting the protein species

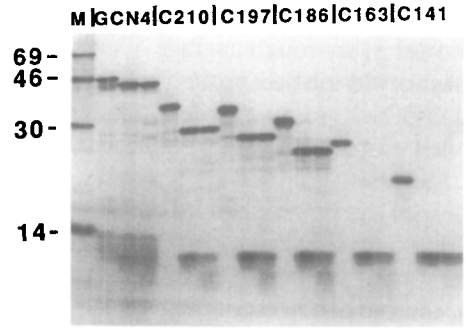


FIG. 4. Analysis of protein structure by partial proteolysis. Full-length GCN4 protein or derivatives containing the indicated number of carboxyl-terminal amino acids were incubated with 0, 10^{-4}, or 10^{-3} units of chymotrypsin for 30 min at 37° and analyzed by SDS–PAGE [molecular weight markers ($\times 10^{-3}$) indicated at left]. The arrow represents a stable proteolytic fragment that is present in all the derivatives and hence maps to the carboxyl terminus; the amino-terminal proteolytic fragments of the different derivatives vary in molecular weight.

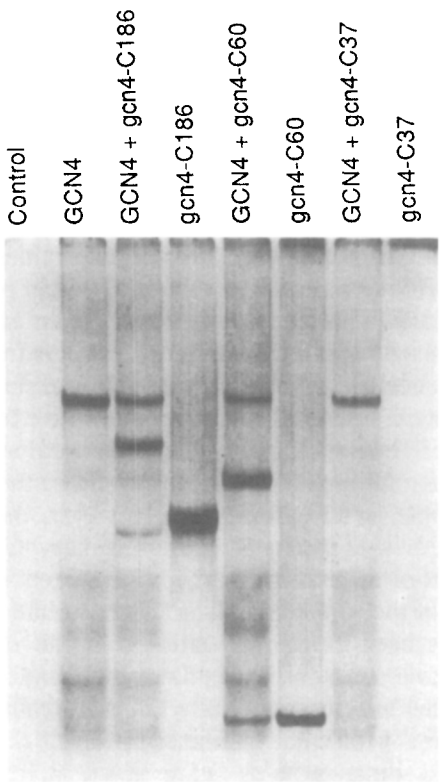


FIG. 5. Determining subunit structure by cosynthesis of differently sized proteins. Wild-type GCN4 protein and derivatives containing the indicated number of carboxyl-terminal amino acids were synthesized individually or in combination, incubated with a DNA fragment containing a specific binding site, and analyzed by electrophoresis in a native acrylamide gel (see Fig. 2). Complexes dependent on the individual proteins have different electrophoretic mobilities owing to their different molecular weights. New complexes with intermediate mobilities reflect heteromeric species; the existence of three bands in a 1:2:1 molar ratio indicates that the protein binds DNA as a dimer. (From Ref. 14.)

of interest. In the case of the reverse band-shift DNA-binding assay described above, protein–DNA complexes involving different-sized protein derivatives have different electrophoretic mobilities; thus, homo- or heteromeric complexes can be distinguished easily[14] (Fig. 5). A more general method, however, would be glutaraldehyde cross-linking followed by SDS gel electrophoresis to examine the protein species.[35]

[35] A. J. Joachimiak, R. L. Kelley, R. P. Gunsalus, and C. Yanofsky, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 668 (1983).