Reprint Series
23 March 1984, Volume 223, pp. 1299–1301

SCIENCE

Total Synthesis and Cloning of a Gene Coding for the Ribonuclease S Protein

Krishnan P. Nambiar, Joseph Stackhouse, Dora M. Stauffer, W. Poindexter Kennedy, J. K. Eldredge, and Steven A. Benner

Total Synthesis and Cloning of a Gene Coding for the Ribonuclease S Protein

Abstract. *A gene for ribonuclease S protein, has been chemically synthesized and cloned. The gene is designed to have 25 specific restriction endonuclease sites spaced at short intervals, permitting its structure to be rapidly modified. This flexibility facilitates tests of hypotheses relating the primary structure of the enzyme to its physical and catalytic behavior.*

Systematic variation of the amino acid sequence of enzymes promises to be useful for developing an understanding of structure, physical behavior, and catalysts in proteins (*1*). Such a tool is likely to be most productive when the enzymes being studied have structures known to atomic detail, catalyze reactions with rates measurable at the level of the individual reaction step, and have thermodynamic properties that can be determined at the level of microscopic reaction intermediates.

We describe here the chemical synthesis and cloning of a gene coding for the ribonuclease S protein (*2*), engineered to contain more than 25 specific restriction sites at short intervals. These restriction sites permit rapid modification of the synthetic gene by replacing sections of the gene with synthetic duplex DNA. By expression of altered genes, ribonucleases having multiple alterations in their amino acid sequences may be prepared. This altered gene has been engineered to permit rapid mutation by design, and appears to be the first synthetic gene for any enzyme (*3*).

The designed sequence of the gene for ribonuclease is shown in Fig. 1. The specific restriction sites are underlined. Multiple restriction sites and homologous sequences were removed. Sites for Eco R1 and Bam Hl (restriction endonucleases) were incorporated at the ends of the gene to facilitate introduction into cloning and expression vectors (*4*). Provisions for the expression of ribonuclease were also incorporated into the design. While active ribonuclease might be able to destroy messenger RNA (mRNA) coding for the protein, the proteolytic fragment S protein has no catalytic activity, but forms a stable non-covalent aggregate with the ribonuclease S peptide, which has full catalytic activity (*2, 5*). Thus, the ribonuclease S protein can be expressed in *Escherichia coli* as a nonactive protein. Furthermore, the S protein can be purified from other proteins of *E. coli* by affinity chromatography with the S peptide as the affinity ligand (*6*).

To construct the gene, 66 different oligonucleotides 10 to 22 residues in length were synthesized with the use of phosphoramidite reagents on a variety of solid supports (*7*), including colloidal suspensions of magnetic particles developed as supports for oligonucleotide synthesis (*8, 9*). Adapting the procedures of Beaucage and Caruthers (*8*), we prepared supports that were functionalized with 3-aminopropyltrimethoxysilane; the appropriate protected 5′-dimethoxytrityl-2′-deoxynucleoside was then linked

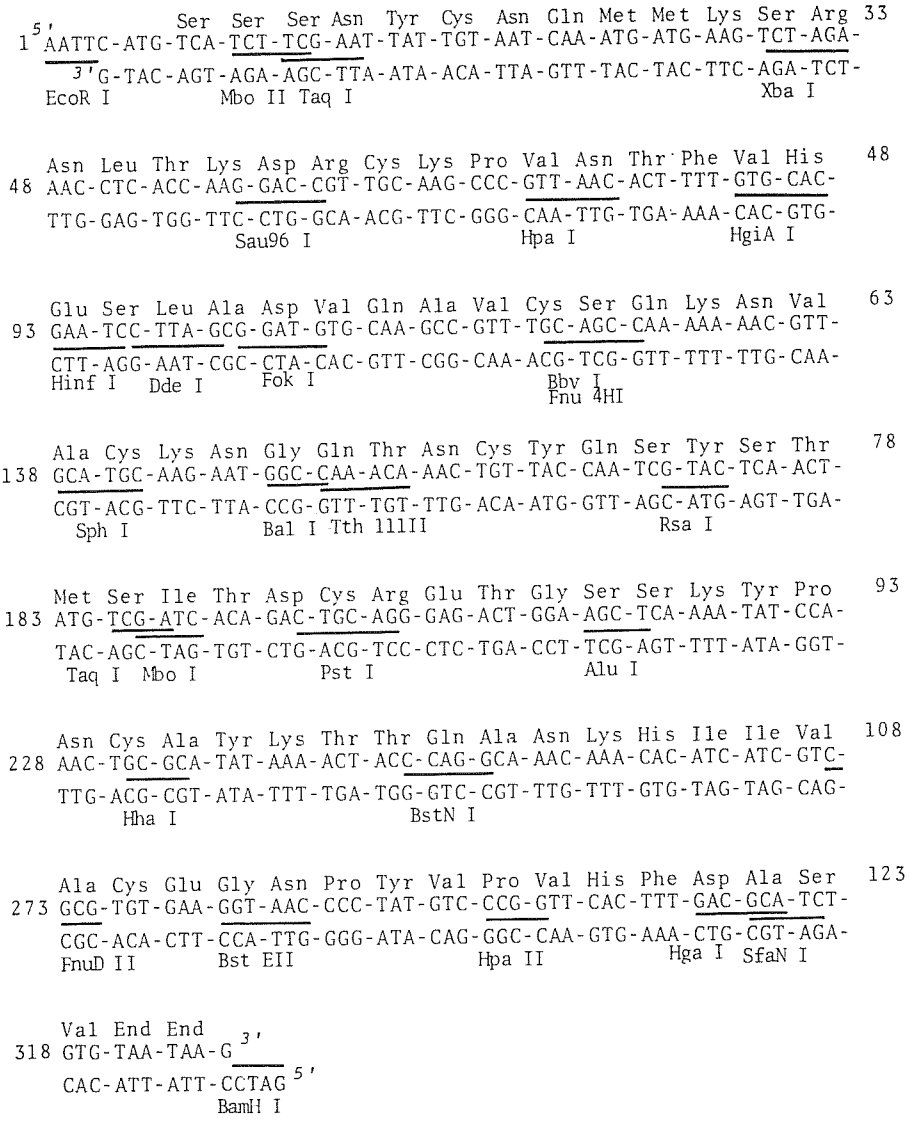

Fig. 1. Sequence of the synthetic gene coding for the ribonuclease S protein, containing about 330 base pairs. The numbering of nucleotides in the sequence is on the left; numbering of the amino acids is on the right, and corresponds to the numbering in native bovine ribonuclease. Restriction sites are underlined and labeled. Some restriction sites are not underlined, as their location adjacent to other sites makes them redundant for the purpose of cutting small segments from the gene.

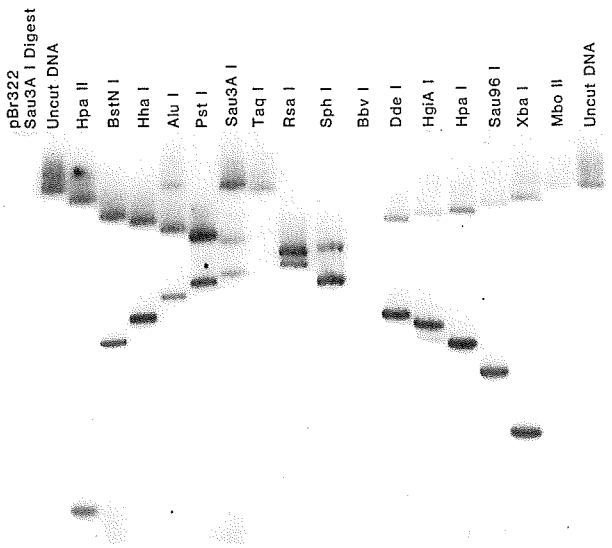

Fig. 2. Autoradiograph of a restriction digest map of the gene for ribonuclease, excised by digestion with Bam HI and Eco RI endonucleases, labeled with ^{32}P at the 5′ ends, and subjected to electrophoresis on an acrylamide (15 percent) gel.

to the support through a 3′-succinamido group. The coupling cycle consisted of detritylation with $ZnBr_2$ in a mixture of CH_3NO_2 and CH_3OH (95:5), washing with CH_3OH and CH_3NO_2 and then anhydrous CH_3CN, drying at 50°C at reduced pressure, addition of tetrazole in anhydrous CH_3CN and solid-protected 5′-dimethoxytrityl-2′deoxynucleosidyl-3′-phosphoramidite, incubation, hydrolysis [in a mixture of tetrahydrofuran, lutidine, and water (2:1:2)], oxidation with I_2 in the same solvent, capping in acetic anhydride and dimethylaminopyridine in tetrahydrofuran, and washing with tetrahydrofuran and then CH_3NO_2. Cycle times were typically 1 hour (*10*).

Synthetic oligonucleotides were released from the supports (*7*); purified by thin-layer chromatography on silica gel with a mixture of *n*-propanol, ammonia, and water (55:35:10) as eluent; 5′-phosphorylated; and sequenced (*11*) prior to subsequent ligation. Oligomers to be ligated were mixed and treated with γ-^{32}P–labeled adenosine triphosphate (ATP) (30 to 60 Ci/mmole) in the presence of polynucleotide kinase; oligomers forming the overlapping ends were treated in the same way, except that the specific activity of the ATP was tenfold higher, and added to the ligation mixture. The oligomer mixture was heated at 90°C for 5 minutes and annealed; dithiothreitol, unlabeled ATP, and T4 DNA ligase (final concentrations 20 m*M*, 0.1 to 1 m*M*, and 200 U/ml, respectively) were added. The final DNA concentration was never less than 10 μM per oligomer. The oligomers destined to become the 5′ overhanging Bam HI and Eco RI ends were added but they were not treated with the kinase in order to prevent these from self-ligating.

Oligomers were convergently ligated in eight groups, each group containing eight to ten oligomers, to yield DNA duplexes that were 40 to 50 base pairs in length, with single-stranded segments five bases in length at each of the 5′ ends (*12*). Oligomers were isolated by electrophoresis on 15 percent acrylamide gels (a mixture of acrylamide and bisacrylamide, 20:1, with 89 m*M* tris-borate buffer, *p*H 8.3, 2.5 m*M* EDTA), and sequenced (*11*). The eight DNA duplexes were joined to form the gene in two parts (*12*). Ligations of the DNA duplexes were performed under the same conditions as for the oligomers, but samples were not heated. The DNA concentration was 0.1 to 1 n*M* per duplex. The complete gene was assembled by ligating these two parts with a pBR322 plasmid that had been cut to remove the small region of DNA between Bam HI and Eco RI sites.

Transformation was effected by treatment with calcium (calcium shock) in the presence of sulfolane (*13*), and transformants were selected for ampicillin resistance. Of the clones isolated, several contained plasmids having the expected restriction digest map (Fig. 2). The structure of the cloned gene was verified by DNA sequencing (*11*).

While a number of different enzymes are being altered by means of recombinant DNA techniques, ribonuclease has certain advantages as a system for study. Six crystal structures of ribonuclease and its variants are now available (*14*), including three-dimensional structures of ribonuclease bound to reactants, products, and transition state analogs. Primary sequences for more than 40 analogous mammalian ribonucleases, mostly from herbivores, are known, indicating positions where alteration is possible and potentially interesting (*15*). There is a considerable amount of bioorganic and chemical information on ribonuclease available (*2*). Kinetic and thermodynamic methods are available making catalysis by ribonuclease subject to rigorous physical organic analysis.

In addition to being a system well suited to the study of catalysis, ribonuclease is appropriate for the study of many other important problems in biological chemistry. This enzyme represents a process that models virtually every step in protein folding. The aggregation between the S peptide and the S protein is an excellent model for chain association and has been studied by systematic alteration of S peptides (*16*). Ribonuclease is believed to have a hydrophobic nucleation site for folding (*17*). *Cis-trans* isomerization of proline is believed to determine the rate of folding (*18*). Ribonuclease is a conventional system for studying the formation of disulfide bonds (*19*). Furthermore, the protein is suited for detailed studies on the chemical nature of thermal stability of proteins (*20*)

Information from the primary amino acid sequences further guides research on this enzyme. Beintema has deduced the primary structures of ribonuclease from extinct fossil organisms that appear to be ancestral ruminants (*15*). It therefore now seems possible to prepare a series of ribonucleases that are intermediates in its recent evolutionary history. From this series, it should be apparent whether the evolution of ribonuclease is most accurately described as a gradual improvement in the protein, as a development characterized by major alterations in the protein's physical and catalytic properties, or as the random fixation of "neutral" mutations (*21*). Finally, certain ribonucleases show activity against tumors both in vitro and in vivo (*22*), and studies on the relation between the structure of the protein and its anticancer properties promises to be pharmacologically interesting.

KRISHNAN P. NAMBIAR
JOSEPH STACKHOUSE
DORA M. STAUFFER
W. POINDEXTER KENNEDY
J. K. ELDREDGE
STEVEN A. BENNER

Department of Chemistry,
Harvard University,
Cambridge, Massachusetts 02138

References and Notes

1. R. B. Wallace, M. Schold, M. J. Johnson, P. Dembek, K. Itakura, *Nucleic Acids Res.* **9**, 3647 (1981); S. Inouye, X. Soberon, T. Franceschini, K. Nakamura, K. Itakura, M. Inouye, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3438 (1982); A. D. Charles, A. E. Gautier, M. D. Edge, J. R. Knowles, *J. Biol. Chem.* **257**, 7930 (1982); D. Shortle, D. Koshland, G. M. Weinstock, D.

Botstein, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5375 (1980); I. S. Sigal, B. G. Harwood, R. Arentzen, *ibid.* **79**, 7159 (1982); G. Winter, A. R. Fersht, A. J. Wilkinson, M. Zoller, M. Smith, *Nature (London)* **299**, 756 (1982); J. E. Villafranca *et al.*, *Science* **222**, 782 (1983).
2. F. M. Richards and H. W. Wyckoff, in *The Enzymes*, P.D. Boyer, Ed. (Academic Press, New York, ed. 3, 1971), vol. 4, p. 647; P. Blackburn and S. Moore, *ibid.*, vol. 15, p. 317 (1982); R. J. MacDonald, S. J. Stary, G. H. Swift, *J. Biol. Chem.* **257**, 14582 (1982).
3. M. D. Edge *et al.*, *Nature (London)* **292**, 756 (1981); J. Smith *et al.*, *Nucleic Acids. Res.* **10**, 4467 (1982).
4. For discussions of recombinant DNA techniques, see R. Wu, Ed., *Methods Enzymol.* **68** (1979).
5. F. M. Richards and P. J. Vithayathil, *J. Biol. Chem.* **234**, 1459 (1959); T. Takahashi, M. Irie, T. Ukita, *J. Biochem. (Tokyo)* **65**, 55 (1969).
6. T. H. Gawronski and F. Wold, *Biochemistry* **11**, 442, 449 (1972).
7. K. L. Agarwal *et al.*, *Angew. Chem. Int. Ed. Engl.* **11**, 451 (1972); M. J. Gait, in *Polymer-Supported Reactions in Organic Synthesis*, R. D. Hodge and D. C. Sherrington, Eds. (Wiley, London, 1980) p. 435.
8. S. L. Beaucage and M. H. Caruthers, *Tetrahedron Lett.* **22**, 1859 (1981).
9. S. A. Benner, U.S. patent application, December 1982.
10. Many different conditions and supports were used and compared (S. A. Benner *et al.*, in preparation).
11. A. M. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 499 (1980).
12. E. L. Brown, R. Belagage, M. J. Ryan, H. G. Khorana, *ibid.* **68**, 109 (1979).
13. A modification of the method of D. Hanahan (*J. Mol. Biol.*, in press) with sulfolane substituted for dimethyl sulfoxide.
14. H. W. Wyckoff, D. Tsernoglou, A. W. Hanson, J. R. Knox, B. Lee, F. M. Richards, *J. Biol. Chem.* **245**, 345 (1970); S. Capaso, F. Giordano, C. A. Mattia, L. Mazzarella, A. Zagari, *Biopolymers* **22**, 327 (1983); A. L. Fink and G. A. Petsko, *Adv. Enzymol. Relat. Areas Mol. Biol.* **52**, 177 (1981); G. A. Petsko, personal communication.
15. J. J. Beintema and J. A. Lenstra, in *Macromolecular Sequences in Systematics and Evolutionary Biology*, M. Goodman, Ed. (Plenum, New York, 1981); personal communication.
16. K. Hofmann, J. P. Visser, F. M. Finn, *J. Am. Chem. Soc.* **92**, 2900 (1970); A. Komoriya and I. M. Chaiken, *J. Biol. Chem.* **257**, 2599 (1982); A. M. Labhardt and R. L. Baldwin, *J. Mol. Biol.* **135**, 231 (1979).
17. R. R. Matheson, Jr., and H. A. Scheraga, *Macromolecules* **11**, 819 (1978).
18. J. F. Brandts, H. R. Halvorson, M. Brennan, *Biochemistry* **14**, 4953 (1975).
19. T. E. Creighton, *J. Mol. Biol.* **113**, 329 (1977); R. R. Hantgan, G. G. Hammes, H. A. Scheraga, *Biochemistry* **13**, 3421 (1974); I. Kato and C. B. Anfinsen, *J. Biol. Chem.* **244**, 1004 (1969).
20. S. E. Zale and A. M. Klibanov, *Biotech. Bioeng.* **25**, 2221 (1983).
21. M. Kimura, *Nature (London)* **217**, 624 (1968); *ibid.* **267**, 275 (1977); *Sci. Am.* **241** (No. 5), 98 (1979); T. H. Jukes, *Science* **210**, 973 (1980).
22. C. C. Levy and T. P. Karpetsky, in *Enzymes as Drugs*, J. S. Holcenberg and J. Roberts, Eds. (Wiley, New York, 1981), pp. 103–166.
23. Supported by the Dreyfus Foundation, the Milton Fund, the Petroleum Research Fund, Alfred Bader, Harvard University, NSF grants PCM-8111659 (D.M.S.) and CHE 8213267 (J.K.E.), NIH grant 1 RO1 GM 30110-01A2 (J.S. and K.P.N.), and the Anna Fuller Fund (W.P.K.). We thank R. Meyers, J. Lingappa, and T. Maniatis for gifts of reagents, F. H. Westheimer and K. E. Bloch for equipment, and E. Brown and O. Makabe for helpful discussions. Address correspondence to S.A.B.

8 November 1983; accepted 27 January 1984