Pharmaceutical Research, Vol. 10, No. 7, 1993 Report

Identification of an *Escherichia coli* Protein Impurity in Preparations of a Recombinant Pharmaceutical

Donald O. O'Keefe,[1,3] Peter DePhillips,[2] and Mark L. Will[1]

Received October 20, 1992; accepted January 8, 1993

A host-cell protein impurity found in preparations of recombinant human acidic fibroblast growth factor (aFGF) was identified. Samples of aFGF examined by western blot analysis employing antiserum raised against an *Escherichia coli* cell lysate contained an immunoreactive protein with a molecular weight of approximately 26,000. The impurity was chromatographically isolated and the N-terminal sequence was determined. Comparing the sequence to a protein database provisionally identified the isolated impurity as the S3 ribosomal protein of *E. coli*. Monoclonal antibodies recognizing three separate epitopes of S3 confirmed the identity of the impurity in western blots of aFGF samples. The monoclonal antibodies were also used to estimate S3 levels in various preparations of aFGF.

KEY WORDS: acidic fibroblast growth factor; drug impurity; protein pharmaceutical; wound healing.

INTRODUCTION

Recombinant DNA technology has provided the pharmaceutical industry with a new source of potent and unique drugs. Like many low molecular weight drugs, recombinant protein therapeutics are frequently derived from fermentations of microorganisms. Often the recombinant protein is produced in the bacterium *Escherichia coli* and is released only when the bacteria are lysed. Cell lysis releases not only the protein of interest but also over 1800 other bacterial proteins (1). Each of these host-cell proteins is a potential impurity in the final drug product and makes the purification of the recombinant therapeutic a challenging endeavor for process engineers and generates new problems for the analytical chemist.

The many host-cell proteins may result in impurities in recombinant protein pharmaceuticals. Traditional purity analyses of low molecular weight drugs often include the identification and quantitation of impurities (2). Results are usually achieved by a combination of spectroscopic, chromatographic, and synthetic techniques. Currently, these classical approaches cannot be easily applied to protein impurity analysis. Consequently, the state of the art for protein impurity analysis consists of polyacrylamide gel electrophoresis and high-performance liquid chromatography (2,3). Unfortunately, both of these techniques have shortcomings. Neither technique allows the specific identification of protein impurities and the quantitative ability of each is poor at the low levels required for the analysis of trace impurities. To overcome these failings, bioanalysts have often employed highly sensitive and specific immunochemical methods such as western blots and ELISAs. These methods routinely detect, and often quantitate, submicrogram levels of proteins (3,4). Furthermore, if an appropriate antiserum is available, western blots can be used to identify a specific impurity. The identification of a host-cell protein impurity can be a starting point for implementing process modifications to eliminate the impurity from future lots or for developing specific assays that either quantitate the impurity or monitor the drug for lot-to-lot consistency.

Acidic fibroblast growth factor (M_r = 15,997) is a potent mitogen that also expresses chemotactic activity for endothelial cells (5). These properties have made the growth factor a potential therapy for wound healing and vascular repair and is currently under development at Merck Research Laboratories (5,6). In this paper, we present our efforts to identify and quantitate a protein impurity present in early preparations of recombinant human acidic fibroblast growth factor (aFGF). Employing a combination of N-terminal sequencing and western blot analysis, we identified the impurity as the S3 ribosomal protein of *E. coli*. Western blot analysis using anti-S3 monoclonal antibodies helped quantitate the impurity and demonstrated the production of a consistent product.

MATERIALS AND METHODS

Western Blotting. Electrophoresis was performed using either 12 or 15% polyacrylamide gels (5 × 8.5 cm) containing sodium dodecyl sulfate (SDS-PAGE) (7). The samples were reduced with β-mercaptoethanol and heated at 100°C for 5 min prior to electrophoresis. After SDS-PAGE, proteins were electrotransferred to nitrocellulose membranes (Schleicher and Schuell, Keene, NH) in a Genie blotter (Idea Scientific, Minneapolis, MN) according to the method of Towbin *et al.* but with the addition of 0.1% SDS to the transfer buffer (8). To detect host-cell protein impurities, the nitrocellulose membranes were probed as described (9) with either rabbit antiserum raised in-house to an *E. coli* cell lysate or mouse monoclonal antibodies recognizing the S3 protein of *E. coli* (10). Antibody reactive proteins were detected using either goat anti-rabbit antibody (Bio-Rad, Richmond, CA) or goat-anti mouse IgG antibody (Sigma, St. Louis, MO) conjugated to alkaline phosphatase (11,12).

Liquid Chromatography. The gradient HPLC system consisted of two Rainin Rabbit pumps, a Macintosh SE computer with Rainin Dynamax HPLC Method Manager software (Version 1.1), a Gilson Model 811 gradient mixer, and a Gilson Model 231 autosampler. The autosampler rack was chilled to 6°C by a Lauda circulating water bath. A Gilson Model 116 UV detector set at 280 nm was used for detection.

Protein Sequence Analysis. Amino terminal sequencing was performed on 115 pmol of the isolated impurity by automated Edman degradation in an Applied Biosystems 473A sequencer. The sample was subjected to 29 cycles of Edman

[1] Department of Analytical Research, Merck Research Laboratories, P.O. Box 2000, Rahway, New Jersey 07065.
[2] Department of Biochemical Process Research and Development, Merck Research Laboratories, West Point, Pennsylvania 19486.
[3] To whom correspondence should be addressed at Merck & Co., Inc., P.O. Box 2000, R80M-112, Rahway, New Jersey 07065.

0724-8741/93/0700-0975$07.00/0 © 1993 Plenum Publishing Corporation

degradation and unambiguous results were obtained for the first 16 cycles.

Other Methods. The protein concentration of the isolated impurity was determined by BCA using bovine serum albumin as the standard (13).

RESULTS

During the development of a purification process for aFGF, the purity of early preparations was evaluated by HPLC and by Coomassie blue or silver staining of polyacrylamide gels. For several aFGF preparations, protein impurities were not detected by any of these methods. Based on the silver staining intensity of minute amounts of aFGF on polyacrylamide gels, the level of any single potential impurity was estimated to be <2500 ppm (parts per million; μg impurity/g protein therapeutic). Although subnanogram levels of some proteins have been detected with silver, this method is not rigorously quantitative because silver staining of proteins can be variable (14,15). The potential for staining variability was of concern and prompted us to analyze the purity of aFGF preparations by immunochemical methods.

Immunochemical impurity analysis focused on two types of aFGF samples obtained from the final two processing steps in the initial large-scale purification of aFGF. The first type consisted of aFGF fractions eluting from a reversed-phase HPLC column, the final chromatographic step in the original purification process (Yamazaki *et al.*, manuscript in preparation). The second type of aFGF sample was from early lots of the bulk drug substance in buffered saline, a preparation of the final purified product prior to formulation. Western blot analysis of both types of samples revealed the presence of a 26,000 MW impurity (26K impurity) reactive with antiserum directed against an *E. coli* cell lysate (Fig. 1). Conversely, the 26K impurity was not detected in western blots of the bulk drug substance probed with anti-aFGF antiserum, demonstrating the impurity was derived from the *E. coli* host and immunologically unrelated to aFGF (data not shown).

Our initial approach to identify the impurity was to resolve the protein from aFGF by gel electrophoresis, transfer the impurity to a polyvinylidene fluoride (PVDF) membrane, and subject the immobilized impurity to N-terminal sequence analysis (16). This approach proved unsuccessful, the high mass ratio of aFGF to the impurity resulted in two problems. First, the excessive aFGF in these samples led to insufficient electrophoretic separation of the impurity from aFGF. Second, the amount of aFGF exceeded the binding capacity of the PVDF membrane at the site of transfer and aFGF unbound after electrotransfer bound indiscriminately to the membrane when the electroblotter was disassembled. The randomly bound aFGF interfered with the sequence analysis of the impurity. These difficulties made it necessary to isolate the impurity by chromatographic methods.

Side fractions from preparative chromatography steps were analyzed by western blot analysis with anti-*E. coli* antiserum to identify those fractions containing the 26K impurity. A number of fractions were found enriched in an *E. coli* protein with a molecular weight of approximately 26,000 (Fig. 2). One of these fractions, fraction 4, was chromatographed further by reversed-phase HPLC until the major

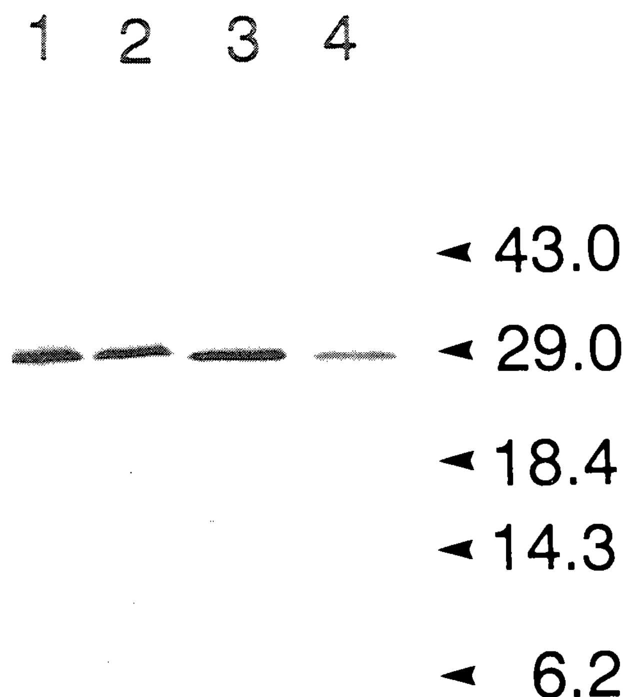

Fig. 1. Western blot analysis of aFGF preparations probed with anti-*E. coli* antiserum. Each sample contained 10 μg of aFGF. The lanes represent (1) the reversed-phase product of lot 2, (2) the reversed-phase product of lot 3, (3) the reversed-phase product of lot 4, and (4) lot 6 of the bulk drug substance. The migration of protein markers is shown on the right, with masses expressed as kilodaltons.

component was resolved from other constituents (Fig. 3). The isolated material had a molecular weight of approximately 26,000 and reacted positively with anti-*E. coli* antiserum in a western blot (Fig. 3, inset). The isolated impurity was subjected to N-terminal sequence analysis and the first 16 amino acid residues were determined and compared to the National Biomedical Research Foundation (NBRF) protein database. The amino acid sequence of the 26K impurity matched the N-terminal sequence of the S3 ribosomal protein of *E. coli* (Fig. 4) (17). The molecular weight of the S3 protein from the sequence in the database is 25,983 (233 amino acids) and agrees well with the 26,000 found for the impurity by SDS-PAGE. The congruence of the two N-terminal sequences, along with the similar molecular weights, tentatively identified the impurity as the S3 ribosomal protein.

Ideally, impurities should be identified *in situ,* but practical problems posed by the low level of the impurity described here limited this approach. Therefore, at this point the impurity's identity was only provisional. It was possible that the protein isolated from a column fraction so distant from the aFGF peak was not the impurity observed previ-

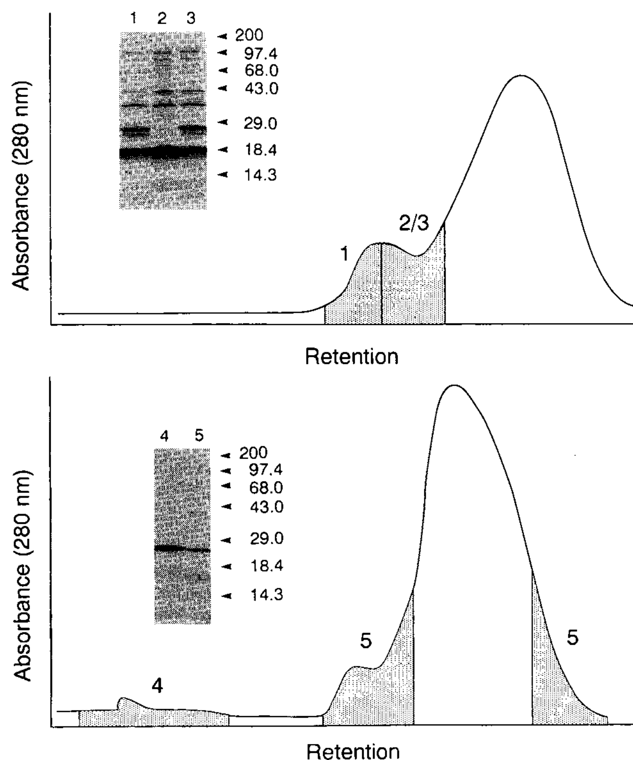

Fig. 2. Process side cuts contain the 26K impurity. Column fractions after preparative HPLC of aFGF were examined with anti-*E. coli* antiserum by western blot analysis (insets). Top: High performance ion-exchange chromatography. Bottom: High-performance affinity chromatography. The main peak in each chromatogram is aFGF. Chromatographic fractions containing an immunoreactive protein migrating at a molecular weight of approximately 26,000 are marked by the hatched areas. Fraction 4 was used to isolate the impurity. The migration of protein markers is shown at the right of each blot, with masses expressed as kilodaltons.

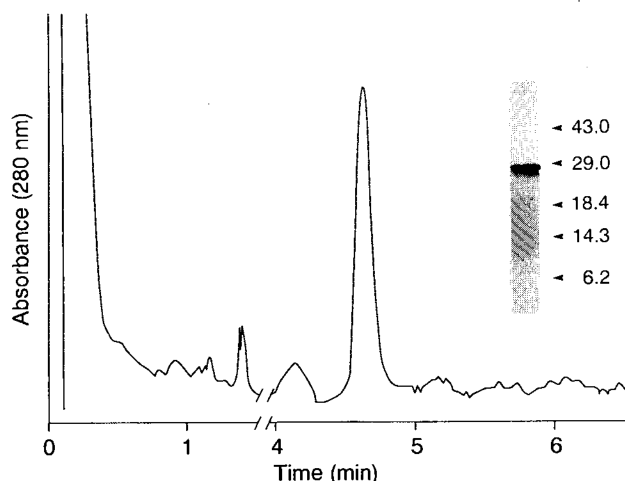

Fig. 3. Reversed-phase HPLC to isolate the *E. coli* protein impurity. Column: Poros R/H (4.6 × 100 mm; Perceptive BioSystems). Mobile phases: A, 0.1% trifluoroacetic acid in water; B, 0.1% trifluoroacetic acid in 80% acetonitrile. Gradient: 0–10 sec, 10% B; 10–50 sec, 10–35% B; 50 sec–10 min, 35–55% B. The flow rate was 4 mL/min. Absorbance was monitored at 280 nm with 0.06 AUFS. Inset: Western blot showing the isolated impurity is reactive to anti-*E. coli* antiserum. The migration of protein markers is shown at the right of the blot, with masses expressed as kilodaltons.

ously in western blots. This uncertainty compelled us to verify the impurity's identity in actual aFGF preparations. By performing individual western blots, each employing one of three monoclonal antibodies recognizing a different region of the S3 protein, the identity of the impurity in aFGF preparations was confirmed (Fig. 5 and Table I) (10).

The approximate level of S3 in a number of different aFGF preparations was determined by western blot analysis (Fig. 5). The analysis employed the S3 monoclonal antibodies, but instead could have utilized the anti-*E. coli* antiserum. We chose the former because of their higher specificity and sensitivity. Known quantities of S3 were probed to establish a standard curve for determining impurity levels in numerous aFGF preparations (Table II). Although all three monoclonal antibodies gave qualitatively similar results, antibody 373C9C3A1 was slightly more sensitive than the other two. Antibody 373C9C3A1 could detect 1 ng of S3 faintly (Fig. 5, lane 8) whereas antibodies 23E9E7 and 391B9E4E1 detected only 10 ng of S3 under the conditions employed. By this method, the reversed-phase preparations from lots 2, 3, and 4 were estimated to contain 10^3–10^4 ppm S3. Further processing reduced the level of the S3 impurity in the bulk drug substances from these lots to below the detection limit of the assay (<100 ppm) and demonstrated the production of a consistent product. Lot 6 of the bulk drug substance did retain a low level of the S3 impurity (10^2–10^3 ppm). Generally, the purity of all lots of the bulk drug substance tested during process development was >99.9%.

DISCUSSION

Intensive efforts are being put forth in the pharmaceutical industry to develop therapeutic proteins from recombinant sources. One of the inherent difficulties in these endeavors is the purification of the recombinant product from a vast number of potential protein impurities. Purity tests for these products often include SDS-PAGE and reversed-phase HPLC (15,18). SDS-PAGE has a detection limit of 0.2–100 ng, corresponding to approximately 20–10^4 ppm, depending on the staining procedure and the total quantity of protein (3,15). Similarly, reversed-phase HPLC monitored for low wavelength UV absorbance has a detection limit of 400–1000 ppm (3,19). These levels of sensitivity were not sufficient to detect the S3 impurity in the aFGF preparations described here. However, the S3 impurity was detected in western blot analyses employing antiserum raised against an *E. coli* cell lysate. For this reason, highly sensitive immunochemical techniques, such as western blots, are recommended for detecting low level host-cell protein impurities (3,18).

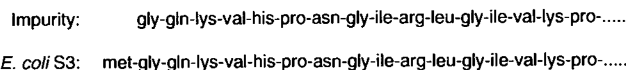

Fig. 4. N-terminal amino acid sequence of the isolated impurity. Automated Edman degradation determined the first 16 amino acids of the impurity. The first 17 residues of the S3 ribosomal protein from *E. coli* are shown for comparison. The terminal methionine in S3 is removed by the *E. coli* protein processing apparatus and thus does not appear in the impurity's sequence.

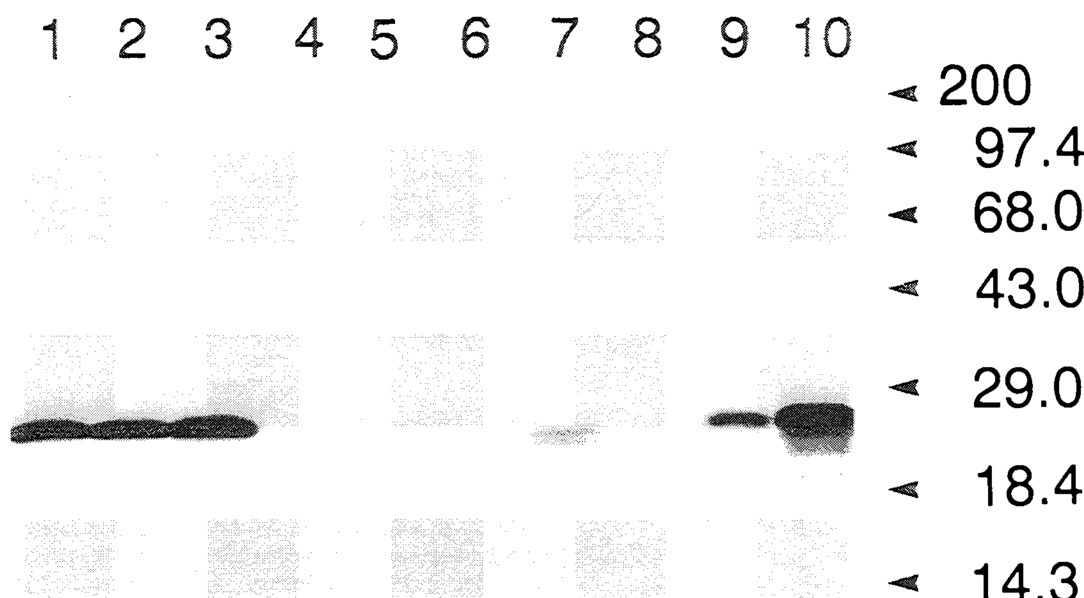

Fig. 5. Western blot analysis of aFGF preparations probed with anti-S3 monoclonal antibody 373C9C3A1. Samples containing 10 μg of aFGF were applied to the lanes as follows: (1) the reversed-phase product of lot 2; (2) the reversed-phase product of lot 3; (3) the reversed-phase product of lot 4; (4) lot 2 of the bulk drug substance; (5) lot 3 of the bulk drug substance; (6) lot 4 of the bulk drug substance; (7) lot 6 of the bulk drug substance. Known amounts of purified S3 were included as follows: (8) 1 ng, (9) 10 ng, and (10) 100 ng. The migration of protein markers is shown at the right, with masses expressed as kilodaltons.

The methodology to identify low-level protein impurities is currently ill defined. In this paper, we described one approach for identifying a host-cell protein impurity present in a recombinant therapeutic derived from *E. coli*. Our approach focused on establishing the impurity's N-terminal sequence and comparing it to the NBRF protein database. By this method, the impurity was identified as the S3 protein of *E. coli*. S3 functions in the assembly and structural maintenance of the 30S ribosomal subunit (20,21). S3 is also able to bind DNA, but no other measurable activities have been reported (22). The absence of measurable activity precluded the development of an activity assay for monitoring the process and analyzing the final product for clearance. Nevertheless, a mass assay based on western blot analysis was developed to quantitate S3 in aFGF preparations. This assay demonstrated that the level of S3 was less than 100 ppm (the limit of detection) in the bulk drug preparations of aFGF described here.

To our knowledge, the S3 protein is the first *E. coli* protein impurity identified in a recombinant pharmaceutical. This raises questions about the specific nature of *E. coli* protein impurities in general and those reported by others (4,23). Could the S3 protein or other ribosomal proteins contribute to the impurity profile of other recombinant products? Each *E. coli* cell contains 52 different species of ribosomal proteins that collectively constitute approximately 9–22% of the total cellular protein (24). Therefore, based on mass alone, *E. coli* ribosomal proteins could be an important source of host-cell impurities. The S3 protein itself does not appear to be a universal impurity in recombinant products. In another unrelated recombinant pharmaceutical we were unable to detect the S3 protein by western blot analysis. Hence, specific characteristics of the S3 protein allowed it to copurify partially with aFGF in a number of our preparations. Noteworthy here is the ability of both proteins to bind

Table I. Epitopes Recognized by Anti-S3 Monoclonal Antibodies

Antibody	Epitope[a]
373C9C3A1	1-21
23E9E7	42-133
391B9E4E1	211-232

[a] Numbers refer to specific amino acid sequences in S3.

Table II. Approximate Levels of S3 in Preparations of aFGF[a]

Sample	Monoclonal antibody		
	373C9C3A1	23E9E7	391B9E4E1
Lot 2 (RPP)[b]	10^3–10^4	10^3–10^4	10^3–10^4
Lot 3 (RPP)	10^3–10^4	10^3–10^4	10^3–10^4
Lot 4 (RPP)	10^3–10^4	10^3–10^4	10^3–10^4
Lot 2 (Bulk)[c]	$<10^2$	$<10^3$	$<10^3$
Lot 3 (Bulk)	$<10^2$	$<10^3$	$<10^3$
Lot 4 (Bulk)	$<10^2$	$<10^3$	$<10^3$
Lot 6 (Bulk)	10^2–10^3	$<10^3$	$<10^3$

[a] Levels are expressed as ppm (μg S3/g aFGF).
[b] RPP refers to the reversed-phase product of the designated lot.
[c] Bulk refers to the bulk drug substance of the designated lot.

polyanions. S3 binds DNA (22), whereas aFGF binds heparin (5).

The identification of a protein impurity can aid the process engineer and the bioanalyst in several ways. First, identification can lead to a description of the physicochemical properties of the impurity and therefore provide a basis for process modifications to eliminate it and similar impurities from future lots. For example, one property of S3 has supported the implementation of certain process modifications and we have used the S3 protein as a marker during the ongoing process optimization for aFGF to monitor the removal of low level protein impurities. Second, identifying an impurity can lead to the development of specific in-process assays that are useful for monitoring the effect of process changes, for analyzing the purity of the finished product, or for determining the production of a consistent product.

Not all host-cell protein impurities will be completely amenable to the analysis we described here because the identification of these impurities is not always possible. The NBRF database contains sequences for approximately 60% of the proteins found in *E. coli*. It is possible impurities found in other recombinant therapeutics derived from *E. coli* would not be listed in the database. Furthermore, recombinant products from either yeast or mammalian cells often have a greater potential for impurities because these sources have more host-cell proteins than *E. coli*. The database is far less complete for these cell types than it is for *E. coli*. For any of these cell sources, the absence of an impurity's sequence in the database would generally prevent the identification of the protein. Nevertheless, the N-terminal sequence of the impurity, or the isolated impurity itself, can be used to generate specific antibodies. These antibodies can then be incorporated into immunochemical assays to monitor the removal and clearance of the impurity.

ACKNOWLEDGMENTS

The authors are grateful to Dr. Larry Kahan of the University of Wisconsin—Madison for providing the monoclonal antibodies to the S3 ribosomal protein of *E. coli* and to Mr. Tim Smith for the initial analysis of aFGF preparations by SDS-PAGE.

REFERENCES

1. T. A. Phillips, V. Vaughn, P. L. Bloch, and F. C. Neidhardt. Gene-protein index of *Escherichia coli* K-12, Edition 2. In F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (eds.), Escherichia coli *and* Salmonella typhimurium—*Cellular and Molecular Biology*, American Society for Microbiology, Washington, DC, 1987, pp. 919–966.
2. F. M. Bogdansky. Considerations for the quality control of biotechnology products. *Pharm. Technol.* **11**:72–74 (1987).
3. D. O. O'Keefe and M. L. Will. Chromatographic analysis of host-cell protein impurities in pharmaceuticals derived from recombinant DNA. In S. Ahuja (ed.), *Chromatography of Pharmaceuticals*, ACS Symposium Series 512, American Chemical Society, Washington, DC, 1992, pp. 121–134.
4. V. Anicetti. Improvement and experimental validation of protein impurity immunoassays for recombinant DNA products. In C. Horváth and J. G. Nikelly (eds.), *Analytical Biotechnology: Capillary Electrophoresis and Chromatography*, ACS Symposium Series 434, American Chemical Society, Washington, DC, 1990, pp. 127–140.
5. K. A. Thomas. Fibroblast growth factors. *FASEB J.* **1**:434–440 (1987).
6. T. D. Bjornsson, M. Dryjski, J. Tluczek, R. Mennie, J. Ronan, T. N. Mellin, and K. A. Thomas. Acidic fibroblast growth factor promotes vascular repair. *Proc. Natl. Acad. Sci. USA* **88**:8651–8655 (1991).
7. U. K. Laemmli. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680–685 (1970).
8. H. Towbin, T. Staehelin, and J. Gordon. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proc. Natl. Acad. Sci USA* **76**:4350–4354 (1979).
9. R. K. Tweten and R. J. Collier. Molecular cloning and expression of gene fragments from corynebacteriophage β encoding enzymatically active peptides of diphtheria toxin. *J. Bacteriol.* **156**:680–685 (1983).
10. W. J. Syu, B. Kahan, and L. Kahan. Epitope mapping of monoclonal antibodies to *Escherichia coli* ribosomal protein S3. *J. Protein Chem.* **9**:159–167 (1990).
11. M. S. Blake, K. H. Johnston, G. J. Russell-Jones, and E. C. Gotschlich. A rapid, sensitive method for detection of alkaline phosphatase-conjugated anti-antibody on western blots. *Anal. Biochem.* **136**:175–179 (1984).
12. K.-J. Pluzek and J. Ramlau. Alkaline phosphatase labeled reagents. In O. J. Bjerrum and N. H. H. Heegaard (eds.), *CRC Handbook of Immunoblotting of Proteins*, CRC Press, Boca Raton, FL, 1988, Vol. 1, pp. 177–188.
13. P. K. Smith, R. I. Krohn, G. T. Hermanson, A. K. Mallia, F. H. Gartner, M. D. Provenzano, E. K. Fujimoto, N. M. Goeke, B. J. Olson, and D. C. Klenk. Measurement of protein using bicinchoninic acid. *Anal. Biochem.* **150**:76–85 (1985).
14. B. R. Oakley, D. R. Kirsch, and N. R. Morris. A simplified ultrasensitive silver stain for detecting proteins in polyacrylamide gels. *Anal. Biochem.* **105**:361–363 (1980).
15. R. L. Garnick, N. J. Solli, and P. A. Papa. The role of quality control in biotechnology: An analytical perspective. *Anal. Chem.* **60**:2546–2557 (1988).
16. P. Matsudaira. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* **262**:10035–10038 (1987).
17. D. Brauer and R. Röming. The primary structure of protein S3 from the small ribosomal subunit of *Escherichia coli*. *FEBS Lett.* **106**:352–357 (1979).
18. Points to Consider in the Production and Testing of New Drugs and Biologicals Produced by Recombinant DNA Technology. Office of Biologics Research and Review, Center for Drugs and Biologics, Food and Drug Administration, 1985.
19. R. G. Nielsen and E. C. Rickard. Applications of capillary zone electrophoresis to quality control. In C. Horváth and J. G. Nikelly (eds.), *Analytical Biotechnology: Capillary Electrophoresis and Chromatography*, ACS Symposium Series 434, American Chemical Society, Washington, DC, 1990, pp. 36–49.
20. V. Mandiyan, S. J. Tumminia, J. S. Wall, J. F. Hainfeld, and M. Boublik. Assembly of the *Escherichia coli* 30S ribosomal subunit reveals protein-dependent folding of the 16S rRNA domains. *Proc. Natl. Acad. Sci USA* **88**:8174–8178 (1991).
21. H. F. Noller and M. Nomura. Ribosomes. In F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (eds.), Escherichia coli *and* Salmonella typhimurium—*Cellular and Molecular Biology*, American Society for Microbiology, Washington, DC, 1987, pp. 104–125.
22. R. C. Bruckner and M. M. Cox. The histone-like H protein of *Escherichia coli* is ribosomal protein S3. *Nucleic Acids Res.* **17**:3145–3161 (1989).
23. R. P. Gooding and A. F. Bristow. Detection of host-derived contaminants in products of recombinant DNA technology in *E. coli:* A comparison of silver-staining and immunoblotting. *J. Pharm. Pharmacol.* **37**:781–786 (1985).
24. H. Bremer and P. P. Dennis. Modulation of chemical composition and other parameters of the cell by growth rate. In F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (eds.), Escherichia coli *and* Salmonella typhimurium—*Cellular and Molecular Biology*, American Society for Microbiology, Washington, DC, 1987, pp. 1527–1542.