The Human Protein Index: Relationship to Genetic Engineering

Norman G. Anderson *
Argonne National Laboratory

The central problems in genetic engineering now relate to methods for incorporating properly engineered genes into living organisms. As the relevant theoretical and technical problems relating to this work are solved, several quite different problems whose solutions will be crucial to continued scientific and commercial success will come to the fore.

The first set of problems arise from the necessity to recover proteins, often in a high state of purity, for industrial or pharmaceutical use. Methods which apply in the research laboratory often do not scale up readily. It is our view and experience that the problems of developing large-scale economically useful separations methods are usually left till last, and that few are capable of developing them for the simple reason that few researchers have either experience or facilities for such work.

The second set of problems arises from the simple fact that only a tiny fraction (possibly as little as 2%) of human proteins have been characterized. The list of proteins proposed for production by recombinant DNA techniques is therefore surprisingly small, considering that man is estimated to make between 30,000 and 50,000 different proteins (1). These are not all present in each cell type, and many of them appear in only one or a few cell types, or at a specific stage during development. Thousands of human diseases (including the over two-thousand known human genetic diseases) are thought to involve either an alteration in protein structure or composition, or the absence or excess production of a protein (2). To study these diseases, and

---

*This work is supported by the U.S. Department of Energy under contract No. W-31-109-ENG-38. This paper is co-authored by Leigh Anderson.

to test for them clinically, it will be necessary to identify the protein causally associated with each disease, to develop tests for many of them, and to produce the protein in quantity if replacement therapy is indicated. To do this it is necessary to have available analytical methods which will allow thousands of human proteins to be separated and distinguished, and to realize the full promise of the genetic engineering revolution. We therefore began the development of the protein separation and identification systems required for this work with the full realization that such analyses and separations may ultimately be the largest single components of research and development (and hence cost) in any genetic engineering endeavor.

In this paper we review some of the separations problems which will arise, and the present status of the high-resolution protein mapping required for the Human Protein Index Project, and which is directly applicable to problems in genetic engineering (3,4).

Cell Optimization

Initially organisms will be engineered to maximize chimeric proteins (i.e., the protein desired) under laboratory of pilot-plant-scale conditions using the most convenient assay. When the operation is sealed up, however, a very different set of conditions is encountered and one may discover that the organism may need to be modified accordingly. Cellular proteins, unless they are insoluble, turn over rather rapidly. The objective is to have as much chimeric protein present at the moment of harvest as possible. Arranging an insoluble product is one solution to the turnover problem. Another is to arrange a chemical or thermal switch so that the cells grow rapidly without producing the desired protein, and when a mass of cells have been produced, switch over to full product production. In this way synthesis of the chimeric protein is dissociated in time from growth of the culture to high cell concentration.

It is doubtful whether the optimization of cells for large scale production can be done in small-scale facilities. Obviously both research and production require precise on-line monitoring of both cell growth and product synthesis (5). Much remains to be done in fermenter development, especially if continuous culture systems are to be developed.

Cell Recovery

Getting from multi-thousand liter batches of effluent of fermenters to pure proteins requires a series of steps and decisions. How and at what stages should the cells be killed? Should the cells be harvested centrifugally, then lysed, or should the cells be lysed in the culture medium, and the desired protein then recovered by precipitation or adsorption?

There appear to be no centrifuges presently available to concentrate bacterial cells or granules obtained from them effectively and rapidly. The largest continuous-flow ultracentrifuges now available are the K-series titanium-rotor centrifuges. The design parameters for these were set by one of us (NGA) for influenza vaccine production (6-9), and utilized the principle of continuous-sample-flow-with-banding. For efficient recovery of cells from fermenter streams, or of protein granules from cell lysates, a quite different design is called for using new concepts for high speed unloading.

Protein Fractionation

In multi-step procedures for protein isolation, choices must be made between continuous and batch processes. Batch processes are almost invariably exploited first, and are then very gradually replaced by continuous ones (examples of this are found in steel and glass manufacture for example). While we are a very long way from the continuous production of any biologically produced product, all batch processes and plants based on them are at risk from clever process research and development. In protein isolation research, many attempts have been made to devise continuous isolation methods, for example for plasma fractionation. Three persistent problems have been the very slow rate of approach to equilibrium during fractional precipitation, difficulties in continuous recovery of precipitates, and bacterial contamination.

The most specific protein recovery methods are based on the use of affinity columns using either immobilized antibodies or specific groups including enzyme cofactors which will bind the protein to be desired. To interface such techniques to continuous streams, automatic rapid recycling affinity chromatography was developed (10,11), and has found limited use thus far in the commercial production of antigens used in clinical immunoassays. Note that affinity columns can also be used to remove specific contaminating proteins from a product.

High-Resolution Protein Mapping

As mentioned, the number of different medically important proteins now being considered as candidates for production by recombinant DNA techniques is surprisingly short relative to the total number of proteins in man. Progress is therefore critically dependent on the development of systems and methods for separating and identifying as many different human proteins as possible, and for discovering those alterations in function or amount which are causally related to disease.

High-resolution two-dimensional electrophoresis allows several thousand proteins to be separated and quantitated (1,12). The method is based on isoelectric focusing in the first dimension (a reflection of the amino acid composition of a protein) and electrophoresis in the presence of a strong ionic detergent in the second dimension (the separation in this dimension being due almost entirely to differences in molecular mass). We have developed systems which allow many analyses to be done in parallel (13-14) and have thus far run over 40,000 2-D gel analyses. Maps of the proteins of human plasma (15,16), urine (17,18), lymphocytes (19-21), red cells (22,23), muscle (24,25), hair follicles (1), and saliva (26) have been published. Internal standards for both isoelectric point (27, 28) and molecular mass (29) have been developed. This work has provided much of the basis for the proposed Human Protein Index (1,3,4).

From the point of view of genetic engineering, these are the important points about high-resolution protein mapping:

Genetic defects. High-resolution mapping makes possible a systematic approach to the problem of finding the specific proteins which are either altered structurally or changed in abundance by genetic disease. (Note

that not all mutant proteins are detected by this method since only about 1/3 of amino acid substitutions produce a charge shift.) Specific clinical tests for the newly discovered mutant molecule may then be developed, and the normal protein synthesized for possible replacement therapy.

Detection of non-genetic damage to cells or tissues. In many non-genetic diseases including infectious ones, alterations in the protein composition of tissues or body fluids may occur, and numerous examples are known. For example, the appearance of enzymes from liver cells in blood is an indication of hepatitis, while heart enzymes in the circulation generally indicate damage to heart muscle due to a myocardial infarction. High-resolution protein mapping allows many additional indicators of infection or injury to be discovered (18). These in turn may stimulate the production of additional proteins for research and diagnostic test development.

Autoimmune disease. Many of the diseases now refractory to treatment, including many degenerative ones, appear to be due to an immune reaction against normal proteins of the body. In some cases the specific protein involved has been identified (for example Hashimoto's disease involves an immune reaction to thyroglobulin). However, in many others, including arthritis, many kidney diseases, and aging, the protein antigens involved remain to be identified. The proteins separated by two-dimensional electrophoresis may be electrophoretically transferred in the third dimension to a thin support where they will react with antibodies (30). Using the antibodies from a patient with an autoimmune disease, it is possible to identify the protein on the transfer which underlies the disease. If that protein can then be produced by genetic engineering, a specific test to detect antibodies in patients can be devised, and immobilized antigen can be prepared to remove the antibodies from patients using either plasmaphoresis and retransfusion of cleared blood, or extracorporal circulation.

Cancer. The central focus of our present work is to discover proteins which are unique to different types of human cancer. Such proteins may prove useful for early cancer detection, for cancer typing, and as targets for new therapies.

Mapping genetically reengineered cells. In addition to the uses mentioned, high-resolution protein mapping is useful for the analysis of cells, whether bacteria or yeast, to see how they have actually been altered, and whether more than the desired structural genes have been introduced. It is also evident that mapping defines cells, and that high resolution maps will be central to the successful patenting of life forms.

Monitoring protein purification. Protein purification is usually monitored by determining the ratio of the desired proteins to total protein present (so called specific activity). To design effective purification methods, it is important to see and characterize impurities. Where proteins may be used in man for extended periods it is especially important to detect trace contaminant proteins and to remove them. High-resolution two-dimensional electrophoretic mapping is essential for both of these purposes.

Cloning genes for specific proteins. To work backward from a protein found in a two-dimensional pattern to the cloning of the gene for that protein and its incorporation into a suitable new host may be done by sequencing

enough of the protein to allow a DNA probe to be synthesized which may then allow isolation of the corresponding gene. Alternatively one may adopt a shotgun approach and survey a series of clones using 2-D protein mapping to find the one desired.

Data Reduction

Data acquisition, image analysis, mathematical modeling of spots on 2-D patterns, and correction of distortions between runs all require the development of very sophisticated computerized systems. These are now in routine use in our laboratory and have been described (31,32).

CONCLUSION

It is our view that present emphasis on recombinant DNA technology will, in a relatively short period of time, result in the commercial production of an ever increasing range of reagents, gene machines, and genetically engineered cells. The central problems then will be those relating to actual production and purification, to the development and validation of clinical tests, to the identification of proteins causally related to disease, and especially to proteins which may be used therapeutically. Rather few of the present commercial companies based on recombinant DNA technology will be able to make the transition to the newer areas of research and development which will be required for survival.

REFERENCES

1. Anderson, N.G. and Anderson, N.L. 1979. Molecular Anatomy. Behring Inst. Mitt. 63, 169-210.

2. Bergsma, ed. Birth Defects Compendium. 1979 2d Edition. National Foundation March of Dimes.

3. Report of Human Protein Index Task Force. December 29, 1980. (Copies available from the authors).

4. Anderson, N.G. and Anderson, N.L. 1980. Automatic Chemistry and the human protein index. J. Automatic Chemistry 2: 177-178.

5. Anderson, N.G. The development of fast analyzers. N.G. Anderson. Z. Anal. Chem. 261: 257-271.

6. Anderson, N.G., Ed. The development of zonal centrifuges. National Cancer Institute Monograph No. 21.

7. Anderson, N.G.; Walters, D.A.; Nunley, C.E.; Gibson, R.F.; Schilling, R.M.; Denny, E.C.; Cline, G.B.; Babelay, E.F.; and Peradi, T.E. 1969. K-series centrifuges. I. Development of the k-II continuous-sample-flow-with-banding centrifuge system for vaccine purification. Anal. Biochem. 32: 460-494.

8. Reimer, C.G.; Baker, R.S.; VanFrank, R.M.; Newlin, T.E.; Cline, G.B.; and Anderson, N.G. 1967. Purification of large quantities of influenza virus by density gradient centrifugation. J. Virology 1: 1207-1216.

9. Anderson, N.G. 1962. Centrifugation in gradients: history and prospects. Proc. European Symposium on Zonal Centrifugation in Density Gradients. Institut Pasteur, Paris. Spectrum 2000 4: 11-19, Paris: Editions Cite Nouvelle.

10. Anderson, N.G.; Willis, D.D.; Holladay, D.W.; Caton, J.E.; Holleman, J.W.; Eveleigh, J.W.; Attrill, J.E.; Ball, F.L.; and Anderson, N.L. 1975. Analytical techniques for cell fractions. XIX. The cyclum: an automatic system for cyclic chromatography. Anal. Biochem. 6: 159-174.

11. Anderson, N.G.; Willis, D.D.; Holladay, D.W.; Caton, J.E.; Holleman, J.W.; Eveleigh, J.W.; Attrill, J.E.; Ball, J.E.; and Anderson, N.L. 1975. Analytical techniques for cell fractions. XX. Cyclic affinity chromotography: principles and applications. Anal. Biochem. 68: 371-393.

12. O'Farrell, P.H. 1975. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. 250: 4007-4021.

13. Anderson, N.G. and Anderson, N.L. 1981. Analytical techniques for cell fractions. XXI. Two-dimensional analysis of serum and tissue proteins: Multiple isoelectric focusing. Anal. Biochem. 85: 331-340.

14. Anderson, N.D. and Anderson, N.G. 1978. Analytical techniques for cell fractions. XXII. Two-dimensional analysis of serum and tissue proteins: Multiple gradient-slab electrophoresis. Anal. Biochem. 85: 341-354.

15. Anderson, L. and Anderson, N.G. 1977. High resolution two-dimensional electrophoresis of human plasmaproteins. Proc. Nat. Acad. Sci. USA 74: 5421-5425.

16. Anderson, N.L. 1981. High resolution two-dimensional electrophoretic mapping of immunoglobulin light chains. Immunology Letters 2: 195-199.

17. Anderson, N.G.; Anderson, N.L.; Tollaksen, S.L.; Hahn, H.; Giere, F.; and Edwards, J. 1979. Analytical techniques for cell fractions. XV. Concentration and two-dimensional electrophoretic analysis of human urinary proteins. Anal. Biochem. 95: 48-61.

18. Anderson, N.G.; Anderson, N.L.; and Tollaksen, S.L. 1979. Proteins in human urine. I. Concentration and analysis by two-dimensional electrophoresis. Clinical Chemistry 25: 1199-1210.

19. Willard, K.E. and Anderson, N.L. B. Radola, ed., W. deGruyter. 1980. Alterations of two-dimensional electrophoretic maps of human peripheral blood lymphocytes induced by concanavalin A. Electrophoresis '79: 415-424.

20. Anderson, N.L.; Edwards, J.J.; Giometti, C.S.; Willard, K.E.; Tollaksen, S.L.; Nance, S.L.; Hickman, B.J.; Taylor, J.; Coulter, B.; Scandora, A.; Anderson, N.G. B. Radola, ed., W. deGruyter. 1980. High-resolution two-dimensional electrophoretic mapping of human proteins. Electrophoresis '79: 313, 328.

21. Anderson, L., Identification of mitochondrial proteins and some of their precursors in two-dimensional electrophoretic patterns of human cells. Proc. Nat. Acad. Sci. U.S., in press.

22. Edwards, J.J.; Anderson, N.G.; Nance, S.L.; and Anderson, N.L. Red cell proteins. I. Two-dimensional mapping of human erythrocyte lysate proteins. Blood 53: 1121-1132.

23. Edwards, J.J.; Hahn, H.J.; and Anderson, N.G. B. Radola, ed., W. deGruyter. 1980. Two-dimensional electrophoresis mapping of human erythrocyte proteins. Electrophoresis '79: 383-393.

24. Giometti, C.S.; Anderson, N.G.; and Anderson, N.L. 1979. Muscle protein analysis. I. Development of high resolution two-dimensional electrophoresis of skeletal muscle proteins for analysis of microbiopsy samples. Clin. Chem. 25: 1877-1884.

25. Giometti, C.S.; Barany, M.; Danon, M.J.; and Anderson, N.G. 1980. Muscle protein analysis. II. Two-dimensional electrophoresis of normal and diseased human skeletal muscle. Clin. Chem. 26: 1152-1155.

26. Giometti, C.S. and Anderson, N.G. 1980. Two-dimensional electrophoresis of human saliva. Electrophoresis '79, B. Radola, ed. W. deGruyter, pp. 396-404.

27. Anderson, N.L. and Hickman, B.J. 1979. Analytical techniques for cell fractions. XXIV. Isoelectric point standards for two-dimensional electrophoresis. Anal. Biochem. 93, 312-320.

28. Hickman, B.J.; Anderson, N.L.; Willard, K.E.; and Anderson, N.G. 1980. Internal charge standardization for two-dimensional electrophoresis. Electrophoresis '79. B. Radola, ed., W. deGruyter, 341-350.

29. Giometti, C.S.; Anderson, N.G.; Tollaksen, S.L., Edwards, J.J.; and Anderson, N.L. 1980. Analytical techniques for cell fractions. XXVII. Use of heart proteins as reference standards in two-dimensional electrophoresis. Anal. Biochem. 102: 47-58.

30. Anderson, N.L.; Nance, S.L.; Pearson, T.W.; and Anderson, N.G., in preparation. Analytical techniques for cell fractions. XXIX. Specific antiserum staining of two-dimensional electrophoretic patterns of human plasma proteins immobilized on nitrocellulose.

31. Taylor, J.; Anderson, N.L.; Coulter, B.P.; Scandora, A.E.; and Anderson, N.G. 1980. Estimation of two-dimensional electrophoretic spot intensities and positions by modeling. Electrophoresis '79, B. Radola, ed., W. deGruyter, pp. 329-339.

32. Taylor, J.; Anderson, N.L.; and Anderson, N.G. A computerized system for matching and stretching two-dimensional gel patterns represented by parameter lists. To be published in Electrophoresis '81.