Gene Synthesis, Expression, and Mutagenesis of Azurin

Thesis by

Thomas Kyu-Young Chang

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1991

(Defended December 4, 1990)

Acknowledgments

First and foremost, I would like to thank my advisors Profs. John H. Richards and Harry B. Gray. Jack's constant challenges to achieve goals, his helpful advice, and enthusiasm resulted in the work described in this thesis. Harry's unmatched insight into the blue copper proteins and his unique style of doing science were invaluable. Jack and Harry showed me that having two advisors results in getting 2^2 amount of work accomplished. Thank you very much Jack and Harry. Also, I would like to thank Prof. Jacqueline K. Barton who first pointed me toward bioinorganic research during my undergraduate days at Columbia and also served as the Chairman of my Ph.D. Committee. Thank you Jackie.

I also want to acknowledge the social and scientific interactions offered by other members of the Richards group, especially Steve Schultz, Sheila Iverson, Steve Carroll, Bill Antypas, Dave Middlemas, "Dr." (Wonhwa) Cho and Clyde Rodrigues, as well as the tennis gang for the "Fridays at 5."

I am immensely grateful to Angelo Di Bilio and Yun Sun for their help in performing the EPR and NMR experiments.

My deepest thanks go to Agnes Lew for contributing a lot of invaluable work described in this thesis and for always "trying to be funny" when we were still up working at 5 a.m. Her enthusiasm, support and encouragement are most appreciated. Thanks also for being a great friend, introducing me to things from classical music to the art of eating a buttered-toast upside-down.

Last but not least, I am indebted to my family who gave me the "homecourt" advantage: their support and understanding were there whenever I needed them.

Abstract

A synthetic gene for the blue copper protein *Pseudomonas aeruginosa* azurin has been constructed using a novel, stepwise method. This method is designed to synthesize a gene of any size reliably by building only controlled amounts of the gene in each step. Another advantage of this method is that the intermediate plasmids, which are typically present in picomolar amounts, can be amplified through plasmid preparation for editing. Using this approach, the azurin gene was synthesized in five steps.

Both the synthetic genes for azurin and *Populus nigra*, var *italica* plastocyanin have been expressed in *E. coli*. These expressions have been achieved by using a synthetic Shine-Dalgano sequence and the signal sequence for azurin, which directs the transport of the expressed proteins to the periplasmic space of *E. coli*. The membrane translocation not only facilitates the purification of azurin and plastocyanin but also seems to be required for the proper folding of these proteins. In contrast to these successful expressions, earlier efforts to express plastocyanin in the cytoplasm of *E. coli*, either directly or as a fusion protein, have been unsuccessful at yielding folded plastocyanin.

Site-saturation cassette mutagenesis was performed in azurin at Methionine 121, one of the four ligands to the copper. Variants that contain each of the other nineteen amino acids as well as the amber stop codon have been identified. Surprisingly, all the variants are stable, as judged by Western blot. Furthermore, all mutants that have been isolated at this position (Asn, Asp, Gly, His, Ile, Leu, Val) have the characteristic blue absorption near 600 nm. Despite such similarity with the wild-type azurin,

these mutants seem to have a more flexible copper center. They can lose or incorporate copper at a faster rate than the wild-type protein.

These results, along with those from EPR experiments, suggest that while Methionine 121 is important in giving stability to the copper center and in tuning the redox potential, its contribution to azurin's spectroscopic properties is small when copper is coordinated to the site. Its spectroscopic contribution apparently becomes more significant when nickel or cobalt resides in the copper site, for the electronic spectra of these derivatives differ markedly from that of azurin with copper. It is likely that the copper centers of the mutants are flexible enough to accommodate the preferred geometries of nickel and cobalt, whereas it is more rigid in the wild-type. Finally, several mutants of azurin are proposed that are designed to probe systematically specific aspects of the biological electron transfer mechanism.

Table of Contents

Acknowledgements ii

Abstract iii

Table of Contents v

List of Figures vii

List of Tables xi

Abbreviations xii

Chapter 1: Azurin and Other Blue Copper Proteins

Introduction 2

References 8

Chapter 2: Design of the Azurin Gene and Its Synthesis

Introduction 13

Materials and Methods 15

Results 31

Discussion 73

References 74

Chapter 3: Expression of Plastocyanin

Introduction 78

Materials and Methods 87

Results 95

Discussion 120

References 122

Chapter 4: Expression of Azurin

Introduction 127

Materials and Methods 129

Results 135

Discussion 155

References 157

Chapter 5: Site-directed Mutagenesis and Mutant Characterization of Azurin

Introduction 160

Materials and Methods 168

Results 172

Discussion 215

References 220

List of Figures

Chapter 1

Figure 1 Evolution of blue copper proteins 3

Figure 2 Structure of *Pseudomonas aeruginosa* azurin 6

Chapter 2

Figure 1 Modified Maxam-Gilbert sequencing strategy 26

Figure 2 Azurin DNA sequence, all *E. coli* preferred codons . . 34

Figure 3 Azurin DNA sequence, degenerate codons 37

Figure 4 Overview of azurin gene synthesis 40

Figure 5 Final azurin gene sequence 42

Figure 6 Proposed DNA sequence of *Alcaligenes denitrificans* azurin 44

Figure 7 Oligonucleotides used during azurin gene synthesis. 47

Figure 8 The five steps of the azurin gene synthesis 51

Figure 9 Restriction digest analysis of step 1 plasmid 57

Figure 10 Mismatch ligation of step 4 62

Figure 11 Results of mutations on the azurin sequence 63

Figure 12 Oligonucleotides used for mutation corrections 66

Figure 13 Mutation correction ligation scheme 68

Figure 14 Restriction digest analysis of the completed azurin gene 71

Chapter 3

Figure 1 Structure of *Populus nigra* var *italitca* plastocyanin . 79

Figure 2 Overview of the plastocyanin gene synthesis 80

Figure 3 Plastocyanin gene (pLASTO) sequence 81

Figure 4 Oligonucleotides used for pLcII ligation 96
Figure 5 pLcII ligation 97
Figure 6 Oligonucleotides used for pCFM636 ligation 99
Figure 7 pCFM636 ligation 100
Figure 8 Oligonucleotides used for pTRAP ligation 101
Figure 9 pTRAP ligation 102
Figure 10 Schematic diagram of pProAPLS 104
Figure 11 Oligonucleotides used to insert factor Xa site 105
Figure 12 Factor Xa ligation 106
Figure 13 Oligonucleotides used to insert enterokinase site 107
Figure 14 Enterokinase ligation 108
Figure 15 Sequencing gel showing factor Xa insert 109
Figure 16 Cleaved fusion protein 112
Figure 17 Oligonucleotides used to introduce the azurin signal sequence in front of the plastocyanin gene 116
Figure 18 Azurin signal sequence ligation 117
Figure 19 Oligonucleotides used to insert plastocyanin gene into pUC18 and pKK223-3 118
Figure 20 Ligation of the plastocyanin gene into pUC18 and pKK223-3 119

Chapter 4

Figure 1 Oligonucleotides for insertion of the azurin gene into pUC18 and pKK223-3 136
Figure 2 Azurin sequence in pUC18 137
Figure 3 Western blot of azurin expressed in *E. coli* 138
Figure 4 UV-Vis spectrum of recombinant azurin 140
Figure 5 HPLC data from azurin sequencing 142

Chapter 5

Figure 1 Metabolic pathway of *P. aeruginosa* 161

Figure 2 Copper center of Azurin 162

Figure 3 Cassette for site-saturation at Met 121 173

Figure 4 Three-piece ligation for site-saturation 174

Figure 5 UV-Vis spectrum of Met121Asn azurin 177

Figure 6 UV-Vis spectrum of Met121Asp azurin 178

Figure 7 UV-Vis spectrum of Met121His azurin 179

Figure 8 UV-Vis spectrum of Met121Ile azurin 180

Figure 9 UV-Vis spectrum of Met121Val azurin 181

Figure 10 CD spectrum of Met121 azurin 185

Figure 11 CD spectrum of Met121Asn azurin 186

Figure 12 CD spectrum of Met121Asp azurin 187

Figure 13 CD spectrum of Met121His azurin 188

Figure 14 CD spectrum of Met121Ile azurin 189

Figure 15 CD spectrum of Met121Val azurin 190

Figure 16 EPR spectrum of Met121Val azurin 191

Figure 17 S-band EPR spectrum of azurin 192

Figure 18 UV-Vis spectrum of Cobalt Met121 azurin 195

Figure 19 UV-Vis spectrum of Cobalt Met121Asp azurin 196

Figure 20 UV-Vis spectrum of Cobalt Met121Gly azurin 197

Figure 21 UV-Vis spectrum of Cobalt Met121Leu azurin 198

Figure 22 UV-Vis spectrum of Cobalt triple-mutant azurin . . . 199

Figure 23 UV-Vis spectrum of Cobalt Phe114Ala azurin 200

Figure 24 UV-Vis spectrum of Nickel Met121 azurin 202

Figure 25 UV-Vis spectrum of Nickel Met121Asp azurin 203

Figure 26 UV-Vis spectrum of Nickel Met121Gly azurin 204

Figure 27 UV-Vis spectrum of Nickel Met121Leu azurin 205
Figure 28 UV-Vis spectrum of Nickel triple-mutant azurin . . . 206
Figure 29 UV-Vis spectrum of Nickel Phe114Ala azurin 207
Figure 30 CD spectra of Nickel-subsituted azurins 209
Figure 31 NMR spectrum of Nickel Met121 Asp azurin 211
Figure 32 NMR spectrum of Nickel Met121Gly azurin 212
Figure 33 NMR spectrum of Nickel triple mutant 213
Figure 34 NMR spectrum of Nickel Phe114Ala azurin 214
Figure 35 Proposed Shortcut pathway through Tyr108 217
Figure 36 Electron transfer mutants coupled to Cys 112 218
Figure 37 Electron transfer mutants coupled to Met 121 219

List of Tables

Chapter 1

Table 1 Properties of blue copper proteins 4

Chapter 2

Table 1 Preference of codons in *Escherichia coli* 32

Table 2 Degenerate codon respresentation used 36

Chapter 3

Table 1 Recognition sites of site-specific proteolysis reactions . 114

Chapter 5

Table 1 Assignments of charge-transfer bands for azurin. . . . 163

Table 2 Band assignments of azurin metal derivatives 165

Table 3 Base distribution from site-saturation 175

Table 4 Summary of UV-Vis spectra of azurin mutants 180

Table 5 Spin-Hamiltonian parameters for azurin mutants . . 192

Table 6 Summary of spectra of azurin metal derivatives 208

Abbreviations

Å	angstrom
A. denitrificans	*Alcaligenes denitrificans*
ATP	adenosine 5'-triphosphate
bp	base pairs
C	cytosine
CAP	calf-intestinal alkaline phosphatase
CD	circular dichroism spectroscopy
Ci	curie
DMS	dimethyl sulfate
DMSO	dimethyl sulfoxide
DNA	deoxyribosenucleic acid
DTT	dithiothreitol
ε	extinction coefficient
E. coli	*Escherichia coli*
EDTA	ethylenediamine tetraacetic acid
EPR	electron paramagnetic resonance spectroscopy
ET	electron transfer
FPLC	fast protein liquid chromatography
G	guanine
g	grams
HCl	hydrochloric acid
HPLC	high-pressure liquid chromatography
IR	infrared absorption spectroscopy
Kb	kilobases
Kd	kilodaltons
IgG	class G immoglobulin

L	liter
M	moles/liter
mA	milliamperes
mg	milligrams
ml	milliliters
mM	millimolar
mR/h	milliroentgen per hour
mRNA	messenger RNA
mV	millivolts
MW	molecular weight
µg	micrograms
µl	microliters
µmole	micromoles
NaCl	sodium chloride
NaOAc	sodium acetate
nm	nanometer
NMR	nuclear magnetic resonance spectroscopy
O.D.	optical density
P. aeruginosa	*Pseudomonas aeruginosa*
PAGE	polyacrylamide gel electrophoresis
pmole	picomoles
PBS	phosphate-buffered saline
PMSF	phenyl methyl sulfonyl fluoride
RNA	ribonucleic acid
rpm	revolutions per minute
SDS	sodium dodecyl sulfate
T	thymine

TAE	Tris-acetate EDTA buffer
TBE	Tris-borate EDTA buffer
TEAAc	triethylammonium acetate
TPBS	PBS with tween 20
TRIS	tris(hydroxymethyl) aminomethane
TST	tris-saline tween 20 buffer
UV-Vis	ultraviolet-visible absorption spectroscopy
V	volts
v/v	volume by volume
w/v	weight by volume

Chapter 1

Azurin and Other Blue Copper Proteins

INTRODUCTION

All proteins are made up of a combination of twenty amino acids. Nature supplements this limited number of functionalities by forming a three-dimentional structure and by using various cofactors or prosthetic groups such as hemes[1] and metal ions.[2] Blue copper proteins are a family of proteins that contains type 1 copper. The most striking characteristic of a type 1 copper center is its intense-blue absorption near 600 nm arising from ligand to metal charge transfer (LMCT).[3-9] Additional telltale features include an unusually narrow hyperfine coupling constant in the EPR and a relatively high reduction potential compared to the Cu^{2+}-Cu^{+} couple.[10] Also, in contrast to a type 2 copper center, type 1 copper center is not solvent-accessible and has geometries that are in between those favored by cupric and cuprous ions.[11]

There are many elegant examples of blue copper proteins whose family tree was illustrated by Ryden[12] (Figure 1). These include laccase[13] from a fungus, stellacyanin[14,15] from Japanese lacquer trees, ascorbate reductase[16] and plastocyanin[17] from plants, ceruloplasmin[18] from humans, and azurin[19] from bacteria. In addition, new members are constantly being added to the blue copper protein family, as more proteins are isolated and characterized. Some of these proteins are summarized in Table 1.

Azurin, first discovered by Horio[27] during purification for cytochrome, has commanded scientific interest for the past 30 years. Azurin and plastocyanin are certainly the best-characterized proteins among the blue copper proteins. Several high-resolution crystal structures of *Populus nigra*, var *italica* plastocyanin structures have been solved

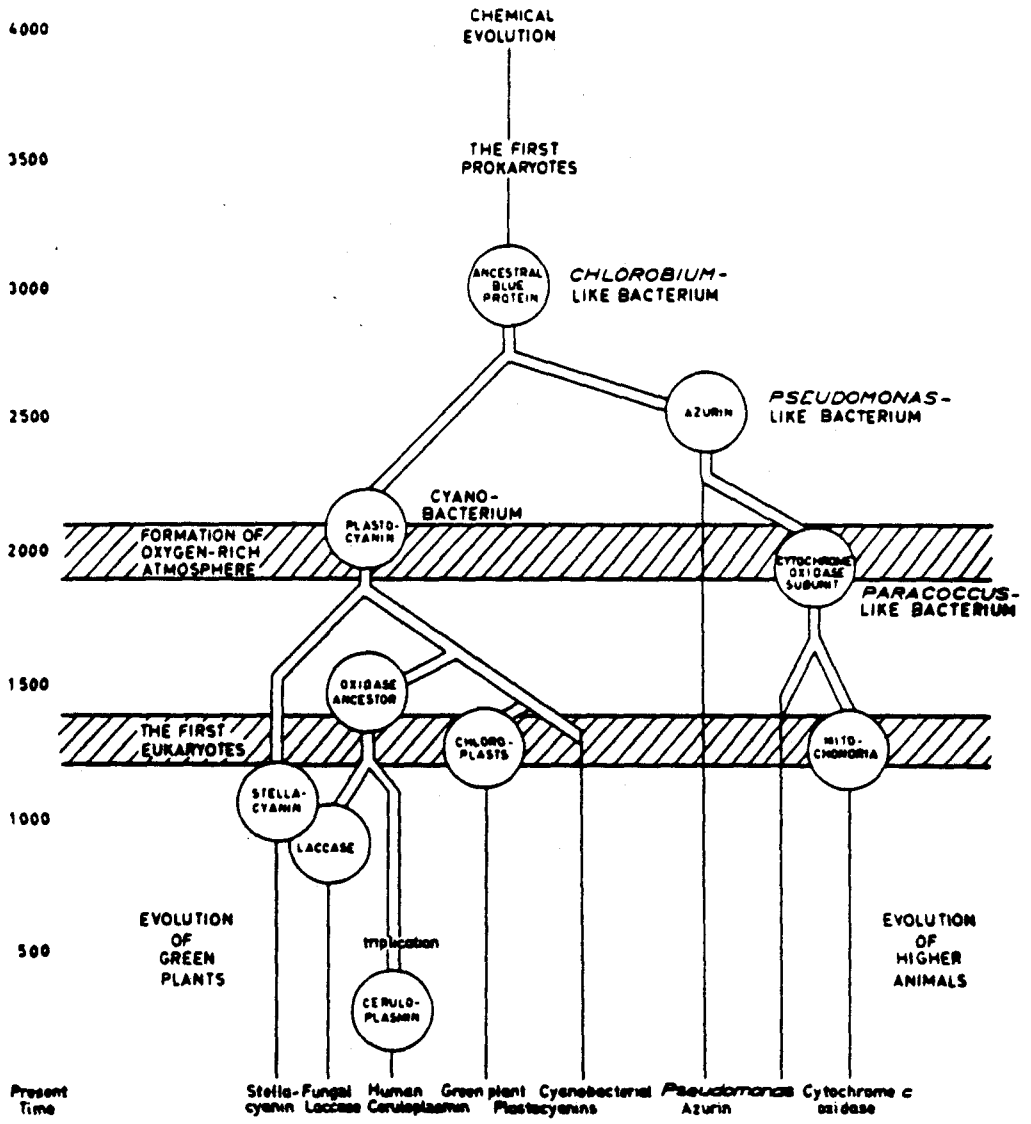


Figure 1. Evolutionary relationship among the blue copper proteins. Azurin and plastocyanin are very closely related. (From Ref. 12).

Table 1
Properties of Blue Copper Proteins

Protein	Source	MW	E°	λ_{max}
Azurin	*Pseudomonas aeruginosa*	14,600	308	625
Plastocyanin	*Phaseolus vulgaris*	10,800	360	597
Plastocyanin	spinach	10, 800	380	597
Stellacyanin	*Rhus vernicifer*	20,000	191	608
Rusticyanin	*Thiobacillo ferrodoxidan*	16,500	380	597
Amicyanin	*Parracoccus denitrificans*	11,700	180	596
Pseudoazurin	*Alacaligenes faecalis* S-6	12,000		593
Mavicyanin	squash	18,000	285	600

Sources of information are as indicated in the text.

including those of oxidized,[28] Hg^{2+}-substituted,[29] and the apo structures.[30] The structures of the reduced plastocyanin are available at six different pH values,[31] while those of azurin are known from two different souces of proteins, *Pseudomonas aeruginosa*[32] (shown in Figure 2) and *Alcaligenes denitrificans*.[33] In particular, the ligands to the copper consist of Histidine 46, Cysteine 112, Histidine 117, and Methionine 121 arranged in a roughly distorted tetrahedral geometry in the *P. aeruginosa* structure. The crystal structures of two other blue copper proteins, cupredoxin from *Alcaligenes facecalis* S-6[34] and the basic blue copper protein from cucumber,[35] show a similar tetrahedral copper center made of the same combination of amino acids. This contrasts with the copper center in the *A. denitrificans* structure, which has a longer Methionine bond (the Cu-S bond is 3.13 Å), and possibly a fifth ligand formed by the carbonyl oxygen from Glycine 45 (Cu-O bond is 3.11Å) in a trigonal bipyramidal arrangement. Comparison of the oxidized[33] and the reduced[36] structures of azurin from *A. denitrificans* confirms that the copper center undergoes only minimal structural change during reduction, thus lending support to the long-held view that the geometry of the copper center contributes to enhancing the electron transfer rate; the small change in the geometry upon reduction and the solvent-inaccessibility of the copper center lead to low inner-sphere and outer-sphere reorganization energies.[3,10,37]

To complement the crystal structure studies, these proteins have been scrutinized by various spectroscopic methods, including NMR,[38,39] UV-Vis,[10,40] EPR,[10,41,42] Resonance Raman,[43,44] differential scanning calorimetry,[45] IR,[9,46] and CD[9,47] experiments. In addition, azurin's intramolecular electron transfer kinetics[48,49] and its intermolecular interaction with cytochrome c_{551}[50] have also been extensively investigated.

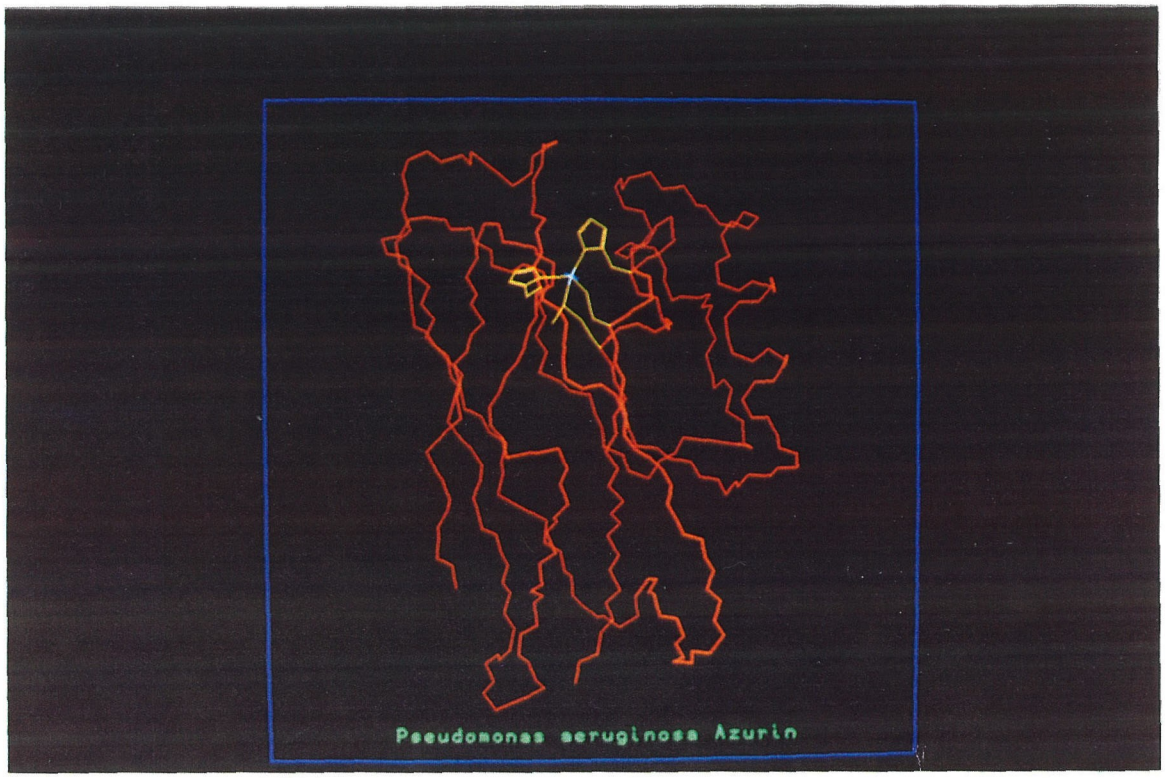


Figure 2. Structure of *P. aeruginosa* azurin. The copper center is highlighted in yellow. Note that almost the entire protein is made of β-sheets. (The picture was generated by Biograf Version 2.1, BioDesign Inc., Pasadena, California.)

Given that azurin and plastocyanin are very well-characterized proteins and attract immense interest from diverse groups of scientists, they were extremely attractive candidates for site-directed mutagenesis experiments. To carry out such experiments, the total synthesis of the genes for plastocyanin and azurin were first performed. Successful expressions of these genes allow us to investigate many aspects of these two proteins through site-directed mutagenesis. The goals which we hoped to accomplish included the following:

(1) refining the gene synthesis methodology

(2) expanding the knowledge about the expression of proteins in *E. coli*

(3) understanding structure/function relationship of proteins

(4) gaining insight into the geometry requirements of the blue copper center

(5) probing the spectroscopic dependence on the protein structure

(6) fine-tuning the Marcus equation and the parameters governing the electron transfer kinetics, and

(7) investigating protein-protein interaction.

Given these intriguing and important issues, azurin and plastocyanin deserve our attention for at least the next 30 years.

REFERENCES

1. Collman, J. P., Gagne, R. R., Reed, C. A., Robinson, W. T. and Rodley, G.A. (1974) *Proc. Natl. Acad. Sci. USA* 71, 1326-1330.
2. Williams, R. J. P. (1985) *Eur. J. Biochem.* 150, 231-248.
3. Farver, O. and Pecht, I.(1984) in *Copper Proteins and Copper Enzymes*, ed. Lontie, R., (CRC Press, Florida), Vol. 1, pp. 183-214.
4. Adman, E. T. (1985) in *Topics in Molecular and Structural Biology, Metalloproteins*, ed. Harrison, P. M. (Verlag Chemie, Weinheim), vol. 6, pp. 1-42.
5. McMillin, D. R. and Morris, M. C. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6567-6570.
6. McMillin, D. R. (1985) *J. Chem. Ed.* **62**, 997-1001.
7. Penfield, K. W., Gay, R. R., Himmelwright, R. S., Eickman, N. C., Norris, V. A., Freeman, H. C. and Solomon, E. I. (1981) *J. Am. Chem. Soc.* **103**, 4382-4388.
8. Penfield, K. W., Gerwirth, A. A. and Solomon, E. I. (1985) *J. Am. Chem. Soc.* **107**, 4519-4529.
9. Solomon, E. I., Hare, J. W., Dooley, D. M., Dawson, J. H., Stephens, P. J. and Gray, H. B. (1980) *J. Am. Chem. Soc.* **102**, 168-178.
10. Gray, H. B. and Solomon, E. I. (1981) *In Metal Ions in Biology*, ed. Spiro, T. G. (Wiley, New York), vol. 3, pp. 1-39.
11. Gray, H. B. and Malmstrom, B. G. (1983) *Comments Inorg. Chem.* **2**, 203-209.
12. Ryden, L. (1984) in *Copper Proteins and Copper Enzymes*, ed. Lontie, R., (CRC Press, Florida), Vol. 1, pp. 157-182.
13. Pecht, I. and Faraggi, M. (1971) *Nature* **233**, 116-117.

14. Bergman, C., Gandvik, E.-K., Nyman, P. O. and Strid, L. (1977) *Biochem. Biophy. Res. Comm.* **77**, 1052-1059.

15. Engeseth, H. R., Hermodson, M. A. and McMillin, D. R. (1984) *FEBS Lett.* **171**, 257-261.

16. Messerschmidt, A. and Huber, R. (1990) *Eur. J. Biochem.* **187**, 341-352.

17. Boulter, D., Haslett, B.G., Peacock, D., Ramshaw, J. A. M. and Scawen, M. D. (1977) in *International Review of Biochemistry, Plant Biochemistry II*, ed. Northcote, D. H. (Univ. Park Press, Baltimore), Vol. 13, pp. 1-40.

18. Faraggi, M. and Pecht, I. (1973) *J. Biol. Chem.* **248**, 3146.

19. Ryden, L. and Lundgren, J. (1976) *Nature (London)* **261**, 344-346.

20. Ingledew, W. J. and Cobley, J. C. (1980) *Biochim. Biophys. Acta* **590**, 14-23.

21. Sharma, K. D., Loehr, T. M., Sanders-Loehr, J., Husain, M. and Davidson, V. L. (1988) *J. Biol. Chem.* **263**, 3303-3306.

22. Murata, M., Begg, G. S., Lambron, F., Leslie, B., Simpson, R. J., Freeman, H. C. and Morgan, F. J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6434.

23. Kakutani, T., Watanabe, H., Arima, K. and Beppu, T. (1981) *J. Biochem.* **89**, 463-472.

24. Marchesini, A., Minelli, M., Merkle, H. and Kroneck, P. M. H. (1979) *Eur. J. Biochem.* **101**, 77.

25. Stigbrand, T. (1972) *FEBS Lett.* **23**, 41.

26. Trost, J. T., McManus, J. D., Freeman, J. C., Ramakrishna, B. L. and Blankenship, R. E. (1988) *Biochemistry* **27**, 7858-7863.

27. Horio, T. (1958) *J. Biochem.* **45**, 195.

28. Guss, J. M. and Freeman, H. C. (1983) *J. Mol. Biol.* **169**, 521-563.

29. Church, W. B., Guss, J. M., Potter, J. J. and Freeman, H. C. (1988) *J. Biol. Chem.* **261**, 234.

30. Garret, T. P. J., Clingeleffer, D. J., Guss, J. M., Rogers, S. J. and Freeman, H. C. (1984) *J. Biol. Chem.* **259**, 2822-2925.

31. Guss, J. M., Harrowell, P. R., Murata, M., Norris, V. A. and Freeman, H. C. (1986) *J. Mol. Biol.* **192**, 361-387.

32. Adman, E. T. and Jensen, L. H. (1981) *Isr. J. Chem.* **21**, 8-12.

33. Baker, E. N. (1988) *J. Mol. Biol.* **203**, 1071-1095.

34. Adman, E. T., Turley, S., Bramson, R., Petratos, K., Banner, D., Tsernoglou, D., Beppu, T., Watanabe, H. (1989) *J. Biol. Chem.* **264**, 87-99.

35. Guss, J. M., Merritt, E. A., Phizackerley, R. P., Hedman, B., Murata, M., Hodgson, K. O. and Freeman, H. C. (1988) *Science* **241**, 806-811.

36. Shepard, W. E. B., Anderson, B. F., Lewandaski, D. A., Norris, G. E. and Baker, E. N. (1990) *J. Am. Chem. Soc.* **112**, 7817-7819.

37. Farver, O. and Pecht, I. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6968-6972.

38. Adman, E. T., Canters, G. W., Hill, H. A. O., and Kitchen, N. A. (1982) *FEBS Lett.* **143**, 287-292.

39. Canters, G. W., Hill, H. A. O., Kitchen, N. A. and Adman, E. T. (1984) *Eur. J. Biochem.* **138**, 141-152.

40. Gewirth, A. A. and Solomon, E. I. (1988) *J. Am. Chem. Soc.* **110**, 3811-3819.

41. Groeneveld, C. M., Aasa, R., Reinhammar, B. and Canters, G. W. (1987) *J. Inorg. Biochem.* **31**, 143-154.

42. Gewirth, A. A., Cohen, S. L., Shugar, H. J. and Solomon, E. I. (1987) *Inorg. Chem.* **26**, 1133-1146.

43. Blair, D. F., Campbell, G. W., Schoonover, J. R., Chan, S. I., Gray, H. B., Malmstrom, B. G., Pecht, I., Swanson, B. I., Woodruff, W. H., Cho, W. K., English, A. M., Fry, H. A., Lum, V. and Norton, K. A. (1985) *J. Am. Chem. Soc.* **107**, 5755-5766.

44. Belhadj, M., Jean, J. M., Friesner, R. A., Schoonover, J. and Woodruff, W. H. (1990) *J. Am. Chem. Soc.* **94**, 2160-2166.

45. Engeseth, H. R. and McMillin, D. R. (1986) *Biochemistry* **25**, 2448-2455.

46. Surewicz, W. K., Szabo, A. G. and Mantsch, H. H. (1987) *Eur. J. Biochem.* **167**, 519-523.

47. Solomon, E. I., Rawlings, J., McMillin, D. R., Stephens, P. J. and Gray, H. B. (1980) *J. Am. Chem. Soc.* **98**, 8046-8048.

48. Gray, H. B. (1986) *Chem. Soc. Rev.* **15**, 17-30.

49. Margalit, R., Kostic, N. M., Che, C.-M., Blair, D. F., Chiang, H.-J., Pecht, I., Shelton, J. B., Shelton, J. R., Schroeder, W. A. and Gray, H. B. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6554-6558.

50. Silvestrini, M. C., Brunori, M., Wilson, M. T. and Darley-Usmar, V. M. (1981) *J. Inorg. Biochem.* **14**, 327-338.

Chapter 2

Design of the Azurin Gene and Its Synthesis

INTRODUCTION

Although Watson and Crick unveiled the structural nature of DNA back in 1953,[1] it was not until the discovery of restriction endonuclease,[2] which can recognize and cleave specific DNA sequences, that made molecular biology and protein engineering possible. During the past 10 years, the everexpanding knowledge about how to express proteins, coupled with the advances in such related fields as oligonucleotide synthesis and DNA manipulation, has led to enormous progress in our ability to study the structure-function relationship of proteins through site-specific modifications.[3-7]

In order to perform site-directed mutagenesis, one must possess the gene that encodes for a particular protein. To achieve this end, there are two means: to isolate the natural gene using radioactively labeled oligonucleotide probes or to construct a synthetic gene. The former method allows one to isolate not just the structural gene but also the DNA segment that codes for the control elements and any processed parts, such as the signal sequence or the pro sequence. Gene synthesis, on the other hand, gives control over the codon usage so that any rare t-RNA codon can be avoided to enhance gene expression and most importantly, one can design in a generous distribution of restriction sites that would be immensely useful in the subsequent modification of the gene.

Khorana was the first to synthesize a gene, that of alanine tRNA from yeast.[8] This landmark work has been followed by a vast volume of genes synthesized by ligating together oligonucleotides. The earliest expressed synthetic genes included those that encode for human leucocyte interferon,[9] somatostatin,[10] and insulin.[11] Some more recent examples of

successfully expressed synthetic genes in *E. coli* include the genes for a bovine growth hormone-releasing factor analog,[12] trout metallothionein,[13] sperm whale myoglobin,[14] and an HIV-1 protease.[15] Most of the successful expressions of these synthetic genes have taken place during the past four years.

As the synthesis of even long oligonucleotides (>100 bp) becomes reliable, and the techniques to assemble them together are improved, molecular gene synthesis is becoming the method of choice. This trend reflects the confidence that molecular biology has in the ability to put together oligonucleotides to obtain a gene that encodes a wide variety of enzymes or proteins. Our genes for the blue copper proteins azurin and plastocyanin are among them.

Materials and Methods

Materials

The azurin gene was constructed in *E. coli* strain LS-1,[17] an HB101 derivative, which gave good yields of plasmids. *E. coli* strains JM101Tn9 and GM48, both dam$^-$ strains, were from G. D. McFarland. The culture of *Pseudomonas aeruginosa* (ATCC 19429) used for azurin preparation was obtained from W.R. Ellis.

Restriction enzymes were purchased from Boehringer-Mannheim Biochemicals (BMB) or New England Biolabs (NEB). Polynucleotide kinase[18] was purchased from NEB, while T4 DNA ligase[19] was purchased from Bethesda Research Lab (BRL). High-concentration (19 U/μl) calf-intestinal alkaline phosphatase from BMB was of molecular biology grade. Ampicillin used for colony screening and cell growth was made by Sigma Chemical Company, purchased through the Caltech Biology stockroom. Isopropyl-β-thiogalactoside (IPTG), Tris, and other buffering reagents were purchased from BMB. Agarose for analytical gel electrophoresis was from International Biotechnologies, Inc. (IBI), while high-purity agarose for preparatory gel was bought from Seakem. For oligonucleotide purification, NACS Prepac Columns from BRL and Sephadex G-25-150 from Sigma were used. For Maxam-Gilbert sequencing, the [γ-^{32}P]-deoxyribose ATP at the activity of ~5000 Ci/mmol was purchased from DuPont New England Nuclear (NEN), while the [α-^{32}P]-deoxyribose NPT at the activity of ~3000 Ci/mmol was supplied by Amersham. For the Sanger sequencing, [α-^{35}S]-deoxyribose ATP was purchased from Amersham. U.S. Biochemical provided the Sequenase™ Sequencing kit for DNA sequence analysis. Agarose gels were photographed with Polaroid Type 667

film, while the autoradiography of sequencing gels was performed with Kodak X-Omat AR film. Water was purified through the Corning Megapure Purification system.

UV-Vis spectra were measured on either a Beckman Model DU-7 Spectrophotometer or a Hewlett-Packard Model 8452 Diode Array Spectrophotometer controlled by an IBM PC-AT computer. The gel apparati for analytical agarose gel eletrophoresis and DNA sequencing were from BRL. Sequencing gels were dried on a Bio-Rad model 483 gel dryer. Lyophilization was achieved on a Savant Speedvac or a Labconco lyophilizer. For small-scale experiments, *E. coli* cultures were grown in a New Brunswick Series 25 Incubator Shaker or in a built-in warm room maintained at 37°C. For larger scale (10 liter) preparations, a New Brunswick Magnaferm Fermentor was used. For some purification of oligonucleotides, a high pressure liquid chromatography (HPLC) system from ABI Analytical Kratos Division was used. The system consisted of ABI Spectroflow 400 Solvent Delivery System, Spectroflow 430 Gradient Former, and Spectroflow 480 Injector coupled with an LKB 2140 Rapid Spectral Detector. Data collection and analysis were performed using LKB's Wavescan software on an IBM PC-AT computer. DNA Master software (Caltech biology) was used to perform computer search of restriction sites on DNA, translation, and homology comparison.

Methods

Commonly Used Buffers and Solutions

TAE Buffer: 40 mM Tris-acetate, 1 mM EDTA

Agarose gel-loading buffer (5x recipe): 50% v/v glycerol, 0.1% w/v bromophenol blue, 1 mM EDTA

TBE Buffer: 89 mM Tris-borate, pH 8, 2 mM EDTA

Formamide loading buffer: 80% v/v formamide, 50 mM Tris-borate, pH 8, 2 mM EDTA, 0.05% w/v bromophenol blue, 0.05% w/v xylene cyanol

TE Buffer: 100 mM Tris, pH 7.5, 0.1 mM EDTA (the pH varies)

1/10 TE Buffer: 10% solution of TE Buffer

Tris-HCl: 2M Tris buffer solutions were maintained whose pH's were adjusted with HCl at several pH's

L-broth: 10 g/l bacto tryptone, 5 g/l yeast extract, 5 g/l NaCl. For plates, 15 g/l of bacto agar was added.

Ampicillin: stock solution of the sodium salt of ampicillin was made to a final concentration of 25 mg/ml, filtered through 0.22 m sterile filter from Nalgene and kept frozen at -20° C.

Oligonucleotide Purification

Oligonucleotides were synthesized using phosphoramidite chemistry[20] on an Applied Biosystem Model 380A DNA synthesizer. The usual synthesis scale was 0.2 mmole. About 1/3 of each synthesis was dissolved in 25 ml of formamide loading buffer, heated at 95°C heat block for 1 minute, and then loaded onto a denaturing 15 % polyacrylamide gel poured with TBE buffer. The gel was run at 450 volts for 15 hours. The bands were visualized by UV-shadowing: When shortwave UV light was shined through a gel slab to a fluorescent TLC plate, DNA bands present in the gel absorbed UV light and left a shadow corresponding to their positions. These bands were excised using clean razor blades. The gel pieces containing DNA were finely crushed with a spatula and then soaked in 5 ml of buffer A in a 37°C oven overnight. The gel pieces were filtered out using a spin column, and the supernatant containing DNA was passed

through NACS Prepac columns that bound DNA under low salt conditions.[21] The column was washed with buffer A, and the oligonucleotides were eluted off the column with 600 µl of buffer B. The DNA solution was desalted using a spin column packed with Sephadex G-25-150 equilibrated with water or 1/10 TE. DNA was quantitated using UV-Vis spectrophotometer, using the approximation that 1 O.D. unit corresponded to 20 µg/ml of synthesized DNA.[22] Generally, the ratio of A_{260} to A_{280} was near 1.8, indicating that the DNA was pure.

For the HPLC purification of oligonucleotides, an Aquapore column from ABI Analytical Kratos Division was used. The gradient employed two buffers. Buffer A consisted of 0.1 M TEAAc in 2% CH_3CN and 98% H_2O. Buffer B consisted of 0.1 M TEAAc in 50% CH_3CN and 50% H_2O. The gradient was increased from 100% A/0% B to 50% A/50% B over 45 minutes at the flow rate of 1.0 ml/min. The DNA samples were collected and lyophilized. They were then resuspended in 1/10 TE and quantitated.

Buffers and Solutions

NACS Buffer A: 0.2 M NaCl, 10 mM Tris-HCl, pH 7.5, 1 mM EDTA

NACS Buffer B: 1.0 M NaCl, 10 mM Tris-HCl, pH 7.5, 1 mM EDTA

TBE Buffer: 89 mM Tris-borate, pH 8, 2 mM EDTA

Formamide loading buffer: 80% v/v formamide, 50 mM Tris-borate, pH 8, 2 mM EDTA, 0.05% w/v bromophenol blue, 0.05% w/v xylene cyanol

TE Buffer: 100 mM Tris, pH 7.5, 0.1 mM EDTA (the pH varies)

1/10 TE Buffer: 10% solution of TE Buffer

Vector Preparation

Vectors were typically cleaved with two different restriction enzymes so as to avoid creating compatible overhangs, which would lead to un-

wanted self-cyclization of the plasmids during ligation. For example, pBR322[23] was cleaved with EcoRI and Ava I during step 1. The cleaved plasmid was run on a 1.2 % agarose gel using TAE buffer. When bromophenol blue dye reached the bottom of the gel, electrophoresis was terminated and the gel slab was soaked in a dilute solution of ethidium bromide. Ethidium bromide affected the relative mobility of DNA fragments and thus was not used until electrophoresis was finished. To isolate the band, International Biotechnologies Incoporated's electroeluter was originally used. This device trapped DNA into a V-shaped channel filled with 3 M sodium acetate. Later, the Elutrap device from Schleicher and Schuell was predominantly used since it gave a consistently higher yield of DNA recovery. The device was operated at 120 Volts for 5 hours for DNA of 1 kb to 5 kb. The duration was less for shorter fragments of DNA. This device trapped a DNA fragment between BT1 membranes, which were impermeable to macromolecules.[24]

Buffers and Solutions

Restriction Digest Buffer A (10x recipe): 0.33 M Tris acetate, pH 7.9, 0.1 M Mg acetate, 0.66 M potassium acetate, 5 mM DTT

Restriction Digest Buffer B (10x recipe): 0.1 M Tris-HCl, pH 8.0, 0.1 M $MgCl_2$, 1.0 M NaCl, 10 mM DTT, 10 mM β-mercaptoethanol

Restriction Digest Buffer L (10x recipe): 0.1 M Tris-HCl, pH 7.5, 0.1 M $MgCl_2$, 10 mM DTT

Restriction Digest Buffer M (10x recipe): 0.1 M Tris-HCl, pH 7.5, 0.1 M $MgCl_2$, 0.5 M NaCl, 10 mM DTT

Restriction Digest Buffer H (10x recipe): 0.1 M Tris-HCl, pH 7.5, 0.1 M $MgCl_2$, , 1.0 M NaCl, 10 mM DTT

TAE Buffer: 40 mM Tris-acetate, 1 mM EDTA

Agarose gel-loading buffer (5x recipe): 50% v/v glycerol, 0.1% w/v bromophenol blue, 1 mM EDTA

Ligation

To kinase oligonucleotides, 40 pmoles of each were incubated at 37°C in a 25 µl total volume of kinase buffer (1x) with 10 units of polynucleotide kinase for 30 minutes. Annealing was achieved by mixing the two together in 1x restriction digest buffer M. The tube containing the two complementary oligonucleotides was placed in a 95°C water bath for 2 minutes and allowed to anneal through slow cooling of the bath over an hour. This was used for the ligation. A typical ligation condition was to use 0.4 pmoles of annealed oligonucleotide duplex, 0.04 pmoles of cleaved vector, 10 units of T4 DNA ligase in 25 µl of 1x ligase buffer. But ligation attempts using additional ratios of annealed oligonucleotides to vector were also performed simultaneously.[25] Ligation was usually performed at 15°C for 12-16 hours.

Buffers and Solutions

Kinase buffer (10x recipe): 0.5 M Tris-HCl, pH 7.6, 0.1 M $MgCl_2$ and freshly added 10 mM ATP and 50 mM DTT

Ligase buffer(10x recipe): 0.5 M Tris-HCl, pH 7.4, 0.1 M $MgCl_2$ and freshly added 10 mM ATP and 50 mM DTT

Transformation

To place ligation products into *E. coli*, Hanahan's transformation method[26] was used. Typically, 2 ml of LS-1 cells were grown to saturation from a single colony. This was used to inoculate 50 ml of SOB media. When the cells reached early log phase (0.5 OD), they were pelleted using a

Beckman clinical centrifuge, spun at 2,500 rpm for 10 minutes. The collected cells were resuspended in 17 ml of TFB buffer. After sitting on ice for 10-15 minutes, the cells were again pelleted the same way. They were then suspended in 4 ml of TFB. To this solution, 140 µl of DMSO was added, and the cells were placed on ice. After 5 minutes, 140 µl of DTT (2.25 M) were added, and the cells were placed back on ice. After 10 minutes, additional 140 µl of DMSO were added and cells placed on ice for 5 more minutes. At this time, 200 µl of cells were aliquoted into several tubes. To these tubes, 3 µl of the ligation mixture were added. In addition, two controls were simultaneously performed; one tube of cells was grown with no added plasmid to see if the drug was properly selecting for transformed cells, while another tube of cells was grown with supercoiled closed circular plasmid to check to see if the transformation procedure worked. After 30 minutes on ice, the cells were heat-shocked in a 42°C bath for 90 seconds. Then 800 µl of SOC media were added to each tube, and they were incubated at 37°C for 1 hour. At this time, the cells were pelleted by spinning them briefly in a microfuge. They were resuspended in 100 µl of SOB and grown on L-plates containing 50 µg/ml ampicillin. The colonies that grew on ampicillin were picked onto a numbered grid plate and further characterized.

Media and Solutions

L-broth: 10 g/l bacto tryptone, 5 g/l yeast extract, 5 g/l NaCl. For plates, 15 g/l of bacto agar was added.

SOB media: 20 g/l tryptone, 5 g/l yeast extract, 5 g/l NaCl, 2.5 mM KCl, 10 mM $MgCl_2$ and 10 mM $MgSO_4$. This medium was made without Mg^{2+} and autoclaved. A separate solution that was 1 M in $MgCl_2$ and 1 M in

$MgSO_4$ was made and sterile filtered through a 0.22 μ filter. This was used to bring the final Mg^{2+} concentration to 20 mM in the media.

SOC media: SOB with 20 mM glucose (from 1 M stock)

DTT: 2.25 M DTT in 40 mM potassium acetate, pH 6.0

TFB (transformation buffer): 10 mM K-Mes, pH 6.2, 100 mM $RbCl_2$, 45 mM $MnCl_2 \cdot 4H_2O$, 10 mM $CaCl_2 \cdot 2H_2O$, 3 mM $Co(NH_3)_6 \cdot 3Cl$. This buffer was sterilized by filtration through a 0.22μ filter.

Plasmid preparation

Colonies from successful ligation (ones that grew on ampicillin) were picked and used to inoculate 2 ml of L-broth containing 50 μg/ml ampicillin. At the same time, they were grown also on a numbered grid plates. Saturated 2 ml culture of these cells were used to inoculate 50 ml of L-broth containing 50 μg/ml ampicillin and grown to saturation. Cells were collected using an SS-34 rotor in a DuPont Sorvall Superspeed centrifuge (either RC2B or RC5) by spinning them for 10 minutes at 6,000 rpm. The first step in the plasmid preparation procedure[27] was to add 2 ml of solution I (with freshly added lysozyme to a final concentration of 5 mg/ml), which hydrolyzed the outer cell wall. After 10 minutes at room temperature, 4 ml of solution II (freshly mixed from stock solutions) were added to lyse the cells by osmotic shock. After 5 minutes on ice, 3 ml of ice-cold solution III were added to precipitate the cellular DNA. The plasmid that remained in the supernatant was precipitated by adding 2.5x volume of ethanol. The precipitated plasmid was pelleted by centrifugation at 10,000 rpm for 30 minutes. The pellet was resuspended in 100 μl of 1/10 TE.

Buffers and Solutions

Solution I : 50 mM glucose, 10 mM EDTA, 25 mM Tris-HCl (pH 8.0) and freshly added lysozyme to a final concentration of 5 mg/ml

Solution II: 0.2 N NaOH and 1% SDS. This was made by mixing equal volumes of 0.4 N NaOH and 2% NaOH solution just prior to use.

Solution III: 3 M acetate, pH 4.8. This solution was made by mixing 5 M potassium acetate with 1.5 ml of glacial acetic acid in 28.5 ml of water.

Restriction Analysis

To confirm that a plasmid contained the restriction sites encoded in the newly added cassette, 10 µl of the plasmid from 50 ml prep were cleaved in 25 µl of appropriate buffer (according to suppliers' recommendations) with 5-10 units of restriction enzyme for 1-2 hours. The incubation temperature was 37°C for most enzymes. It was not practical to use a smaller volume for digest since the change in the salt concentration caused by evaporation would reduce the efficiency of the digest. 5 µl of agarose loading buffer were added to the digested samples, and this mixture was then run on a 60 ml 1.2 % agarose gels. The bands were visualized with ethidium bromide staining and photographed using a Polaroid MP-4 camera.

CsCl gradient plasmid preparation

When an intermediate plasmid was found to be of the expected size and contained the expected restriction sites, a larger-scale plasmid preparation was performed using CsCl gradient centrifugation.[28] Typically, 2 ml of starter cultures were grown from a single colony to saturation. This was used as the inoculant for 500 ml of L-broth containing

50 μg/ml ampicillin. When the cells reached O.D. of 0.5, 2.5 ml of 34 mg/ml stock-solution of chloramphenicol was added to a final concentration of 170 μg/ml. Chloramphenicol "amplifies" the plasmid production through suppression of protein production. The cells were grown at 37°C with chloramphenicol for another 12-16 hours. They were harvested in a GSA rotor by spinning them at 6,000 rpm for 10 minutes. Then plasmid prep was performed as described above except that 10 ml of solution I, 20 ml of solution II, and 15 ml of solution III were used to compensate for the bigger scale growth. After the addition of solution III, plasmid DNA was again ethanol-precipitated, for this yielded cleaner solutions for the ultracentrifugation run. The pellet was again spun down using a GSA rotor, 10,000 rpm for 30 minutes. This pellet was resuspended with 42 ml of CsCl stock solution. Then 250 μl of 10 mg/ml ethidium bromide were added. These samples were placed into Beckman VTi 50 tubes and balanced on an analytical balance to within 5 mg. The tubes were then heat-sealed and spun in a VTi 50 rotor in a Beckman LS series ultracentrifuge. The duration of the spin was 18-20 hours at 45,000 rpm under vacuum. After the centrifugation run, the tubes were carefully removed so as to not disturb the CsCl gradient that existed inside. The bands were then visualized by a hand-held UV lamp. Of the two bands that were visible, the lower band contained the supercoiled, closed circular plasmid, while the upper band contained the nicked, open circular plasmid and the *E. coli* chromosomal DNA. The lower band was removed with an 18 gauge needle and syringe. The needle was inserted just underneath the lower band, and the band was carefully withdrawn. This solution was extracted 3 or 4 times with water-saturated butanol until the pink color of ethidium bromide was no longer present. This solution was then dialyzed against 1/10 TE buffer, at least to

10^6 fold dilution. The yield and the purity were quantitated using a UV-Vis spectrophotometer. Typical yields were 500 to 1,000 μg of plasmid from 500 ml culture.

Buffers and Solutions

CsCl solution: 1.05 g/ml of molecular biology-grade CsCl dissolved in 30 ml of water. This solution was used to resuspend ethanol-precipitated plasmid.

Ethidium Bromide: 10 mg/ml in water

Chloramphenicol: 34 mg/ml in ethanol

Maxam-Gilbert Sequencing

Plasmids containing the partially completed azurin gene were sequenced chemically using a modified version of the method pioneered by Maxam and Gilbert;[29] instead of performing all four reactions on one strand, only the G reaction and the A reaction[30] were carried out on both strands. Because of the Watson-Crick base-pairing requirements, the G positions on the antisense strand corresponded to the C positions on the sense strand, and the A positions on the antisense strand corresponded to the T positions on the sense strand. Thus, one can obtain a complete set of information by using only these two reactions on both strands. Although this strategy eliminated the opportunity to verify the sequencing results by sequencing both strands independently, it offered the advantage of using only those reactions that are specific for a single base. The standard Maxam-Gilbert, in contrast, uses reactions that cleave more than one base (the G + A reaction and the T + C reaction), which can creat ambiguity in DNA sequence analysis. This modified Maxam-Gilbert sequencing scheme is outlined in Figure 1.[31]

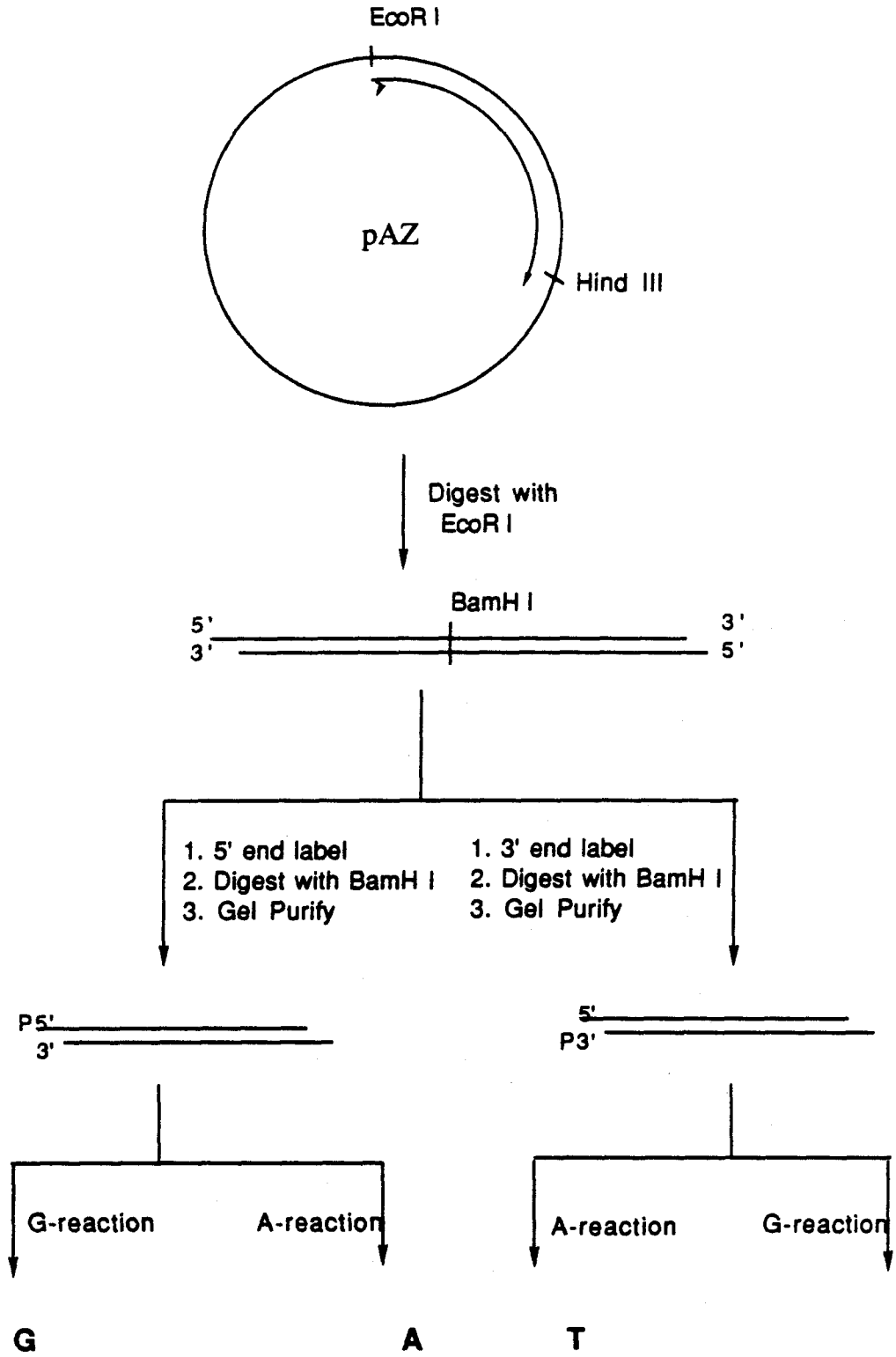


Figure 1. Modified Maxam-Gilbert sequencing strategy using only the G reaction and the A reaction.

The strategy for sequencing involved labeling a plasmid at a particular restriction site at both the 5' end, using polynucleotide kinase, and at the 3' end, using the Klenow fragment of DNA polymerase. First, two batches of 10 μg of the plasmid to be sequenced were digested with a restriction enzyme that was just upstream from the desired region to be read. The plasmid was then ethanol-precipitated and dried under vacuum.

To label at the 5' end, the precipitated plasmid was resuspended in 100 μl of CAP buffer and incubated with 2 μl of calf-intestinal alkaline phosphatase (19 U/μl) and incubated at 37°C for 30 minutes. At this time, another 2 μl of the enzyme were added, and the sample was incubated for another 30 minutes at 37°C. The solution was then extracted once each with an equal volume (100 μl) of phenol, $CHCl_3$, and ether. The plasmid was ethanol-precipitated by the addition of 10 μl of 3 M NaAc and 250 μl ethanol. DNA was pelleted using a microfuge and dried under vacuum. The pellet was resuspended in 70 μl of water, 10 μl of 10x kinase buffer, 10 μl of freshly prepared 50 mM DTT, 7 μl of polynucleotide kinase (10 U/μl), and 3 μl of [γ-^{32}P]-deoxyribose ATP. The reaction mixture was incubated at 37°C for 45 minutes. The labeled DNA was ethanol-precipitated and dried under vacuum. At this point, both 5' ends were labeled.

For the 3' end labeling, the precipitated plasmid was redissolved in 20 μl of water, 5 μl of Klenow buffer, and 5 μl of freshly prepared 50 mM DTT. To this mixture, 7 μl of [α-^{32}P]-deoxyribose NTP that is complementary to the first position of the overhang to be filled in were added along with 5 μl each of the other three dNTP's and 2 μl of the Klenow fragment of DNA polymerase (5 U/μl). This reaction was performed at room temperature for 20 minutes. Then another 2 μl of the enzyme and 1.5 μl of the last dNTP to be filled in were added, and the reaction was allowed to proceed for another

20 minutes. Then the plasmid was ethanol-precipitated as before and dried under vacuum. At this point, both 3' ends were labeled and made blunt-ended.

Since each of these labeling processes labeled the DNA at two ends, it was necessary to eliminate the other end to make only one label unique. To achieve this, a restriction enzyme that was ~1000 bp away from the first site was chosen. The reason for this was that DNA fragments of about 1000 bp were ideal; smaller pieces were difficult to handle, and largeer pieces interfered with the sequencing reactions. To cleave the labeled DNA, pellets were resuspended in 85 µl of water and digested with the chosen restriction enzyme under appropriate buffer conditions. The digest was performed for 2 hours at 37°C. The fragments were separated on a 1.2% agarose gel, stained with ethidium bromide, and eluted off the gel, using Elutrap. The DNA samples were again ethanol-precipitated using NaAc and ethanol. The total radioactivity present was measured using a Geiger-Mueller counter, and the pellet was resuspended to a final activity of about 5 mRem/hr/100 µl.

For the G reaction, 100 µl of each of the two resuspended plasmid (3' and 5') were placed into separate tubes along with 1 µl of sonicated calf thymus DNA (1 mg/ml) and 100 µl of 100 mM cacodylate buffer, pH 8. To this solution was added 1 ml of neat dimethyl sulfate (DMS) in a well-ventilated hood and allowed to react for 1-2 minutes. After this duration, 50 ml of freshly made G-stop solution was added to quench the reaction. The DNA was ethanol-precipitated, pelleted in a microfuge, and dried under vacuum.

For the A reaction, 100 µl of each of the two resuspended plasmid (3' and 5') were placed into separate tubes along with 60 µl of water and 1 µl of

sonicated calf thymus DNA (1 mg/ml). The reaction was commenced with the addition of 40 µl of 10 mM K_2PdCl_4 in 100 mM HCl, pH 2. After 30-45 minutes, the reaction was quenched by the addition of 50 µl of freshly prepared G-stop solution. The DNA was precipitated by adding 750 µl of ethanol and by chilling the solution in dry ice for 10 minutes. The DNA was again pelleted in a microfuge and dried under vacuum.

The actual strand cleavage was done by redissolving the pellet from both reactions in 50 µl of 10% piperidine. This solution was placed in a 90°C bath for 30 minutes. It was then placed in dry ice and lyophilized completely in a Speedvac. The lyophilized DNA was redissolved in an appropriate volume of formamide-loading buffer so as to achieve a final activity of 0.25 mRem/hr/µl. 2 µl of these samples were loaded on a 5% denaturing polyacrylamide gel. To read the sense strand, the following combination of labeling process and base-specific reaction was used for the four bases: for the G lane, the 5'-labeled sample with the G reaction, for the A lane, the 5'-labeled sample with the A reaction, for the T lane, the 3'-labeled sample with the A reaction, and for the C lane, the 3'-labeled sample with the G reaction. The gel was run at 50 mAmp, dried on a slab gel dryer, and autoradiographed.

Buffers and Solutions

Klenow buffer (10x recipe): 0.5 M NaCl, 60 mM Tris-HCl, pH 7.5, 6 mM $MgCl_2$

CAP buffer (10x recipe): 0.5 M Tris-HCl, pH 8.0, 1 mM EDTA. Alkaline phosphatase reaction is very sensitive to the pH of the solution, so it is important to adjust the pH of this buffer to exactly 8.0.

G-Stop buffer: 1.5 mM NaOAc, 1.0 M β-mercaptoethanol, and 40 µg/ml sonicated calf thymus DNA. This was prepared fresh each time.

Palladium solution: 100 mM K_2PdCl_4 dissolved in 100 mM HCl, pH 2.0.. This stock solution was diluted to 10 mM before use.

Sequenase sequencing

The completed azurin gene was sequenced using the Sequenase™ sequencing kit,[32] which was based on the Sanger dideoxy sequencing,[33] but which used a modified DNA polymerase. About 1 pmole of CsCl gradient purified plasmids (~3 μg) from 6 separate colonies was placed in a tube with 2-10 pmoles of 18 bp primers. This primer was about 50 bp upstream of the last ligation site, an ideal distance. The tube was placed in a boiling water bath for 4 minutes, contents quickly spun down, and frozen in dry ice. In a separate tube, a sequencing solution enough for 8 samples was made for each set of 6 plasmids by adding 16 μl of Sequenase buffer, 16 μl of 100 mM DTT, 14 μl of water, 4 μl of GTP labeling mixing, 4 μl of [α-^{35}S]-dATP, and 2 μl of Sequenase™ enzyme. To each tube containing the annealed plasmid with primer, 7.5 μl of this sequencing solution were added. After 4 minutes at room temperature, 3.5 μl of each solution were transferred into four separate wells on a Corning microtiter plate, each containing 2.5 μl of the four termination mix (G, A, T, and C). The termination reaction was performed in a 37°C incubator oven. After 5 minutes, 5 μl of formamide solution were added to each well. The entire plate was placed in a 95°C oven for 2-4 minutes. For gel electrophoresis, 4 ml from each well were used. The usual polyacrylamide gel concentration was 5% and was electrophoresed at 50 mAmp. The gel was dried on a BioRad slab gel dryer and autoradiography performed. Two promising plasmids were completely sequenced using 4 different primers.

RESULTS

General Scheme of Gene Synthesis

Our novel approach to gene synthesis involves synthesizing both strands of the gene in four separate cycles, each cycle consisting of ligation, plasmid preparation, and restriction analysis. In each step, four separate oligonucleotides are kinased and annealed together to form two sets of DNA duplexes with specific overhangs that match the overhangs of the two restriction sites into which they are to be inserted. These double-stranded oligomers are ligated into the construction vector, in our case pBR322 between the unique EcoR I and Ava I sites. Eliminating this stretch of the plasmid destroys the tetracycline resistance gene but leaves the ampicillin resistance gene intact for colony selection. The ligated products, which are present in only picomolar amounts, are transformed back into *E. coli*. This amplifies the DNA to large enough amounts so that it can be checked for the presence of a restriction site or can be sequenced. Correctly ligated plasmid can then be reopened with two restriction enzymes and the next set of oligonucleotides inserted. Each cassette contained a 10 bp-long spacer region between two adjacent restriction sites to be used, for the enzymes cannot cleave efficiently if they are directly next to each other. Such cycle forms our stepwise gene synthesis. This method was also used in our lab to synthesize the genes for plastocyanin[30] and α-lytic protease.[34]

Design of the Azurin Gene

The azurin gene was designed to optimize expression and to facilitate mutagenesis. As shown in Table 1, *E. coli* has a preference for certain codons.[35,36] For example, the preferred codon for lysine in E. coli is AAA.

Table 1

preference of Codons in *E. coli*

Ala	GCT	Leu	CTG
	GCA		CTC
	GCC		CTA
	GCG		CTT
			TTA
			TTG
Arg	CGC	Lys	AAA
	CGT		AAG
	CGA		
	CGC		
	AGA		
	AGG		
Asn	AAC	Met	ATG
	AAT		
Asp	GAC	Phe	TTC
	GAT		TTT
Cys	TGC	Pro	CCG
	TGT		CCA
			CCT
			CCC
Gln	CAG	Ser	TCT
	CAA		TCC
			AGC
			AGT
			TCA
			TCG
Glu	GAA	Thr	ACT
	GAG		ACC
			ACG
			ACA
Gly	GGT	Trp	TGG
	GGC		
	GGA		
	GGG		
His	CAC	Tyr	TAC
	CAT		TAT
Ile	ATC	Val	GTT
	ATT		GTA
	ATA		GTC
			GTG

The preferred codons are underlined. Where there is no preference, none is underlined.

This preference arises most likely from the variation in the abundance of different tRNA's. The preferred codon may differ in yeast or human. Eliminating rare codons was shown to improve the expression of tetanus toxin fragment C.[37] Consequently, the first step was to create a DNA sequence for azurin cosisting of all *E. coli*-preferred codons. Such a DNA sequence for azurin, hereafter referred to as *E. coli*-preferred sequence, is shown in Figure 2. This sequence should be most favorable theoretically for maximizing the gene expression. However, it would be difficult to introduce changes within this gene, for it provides only a limited number of restriction sites.

Our goal was to insert at least one unique restriction site every 50 bp so that only relatively short cassettes would have to be made to perform mutagenesis. Therefore, the next step of the gene design was to identify all possible restriction sites on the basis of the amino acid sequence and to introduce the most favorable ones among them into the gene. This was done by creating a degenerate codon file by reverse-translating the protein file. The degenerate codon for each amino acid is summarized in Table 2, and the degenerate codon sequence for azurin is shown in Figure 3. Computer search was done on this sequence to identify all possible sites on the gene where a restriction site can be introduced. Among these sites, the restriction sites already found on pBR322 were eliminated from consideration since unique restriction sites would be more useful. Of all the remaining candidates, the sites that required the use of less rare codons were chosen.

Once the restriction sites to be incorporated into the gene were chosen, the DNA sequence for the azurin gene was almost complete. Since the original plan was to express azurin as a fusion protein attached to

Figure 2

Proposed DNA sequence of *P. aeruginosa* azurin comprising all *E. coli*-preferred codons.

```
                                        10
Ala Glu Cys Ser Val Asp Ile Gln Gly Asn Asp Gln Met Gln Phe Asn Thr Asn Ala
GCT GAA TGC TCC GTT GAC ATC CAG GGT AAC GAC CAG ATG CAG TTC AAC ACT AAC GCT
     Bsm I
20                                      30
Ile Thr Val Asp Lys Ser Cys Lys Gln Phe Thr Val Asn Leu Ser His Pro Gly Asn
ATC ACT GTT GAC AAG AGC TGC AAG CAG TTC ACT GTT AAC CTG TCT CAC CCG GGT AAC
                                            Hpa I                 BstE II
    40                                      50
Leu Pro Lys Asn Val Met Gly His Asn Trp Val Leu Ser Thr Ala Ala Asp Met Gln
CTG CCG AAA AAC GTT ATG GGT CAC AAC TGG GTT CTG TCC ACT GCT GCT GAC ATG CAG

    60                                          70
Gly Val Val Thr Asp Gly Met Ala Ser Gly Leu Asp Lys Asp Tyr Leu Lys Pro Asp
GGT GTT GTT ACT GAC GGT ATG GCT TCC CTG GAC AAA GAC TAC CTG AAA CCG GAC GAC

            80                                      90
Asp Ser Arg Val Ile Ala His Thr Lys Leu Ile Gly Ser Gly Glu Lys Asp Ser Val
GAC TCC CGC GTT ATC GCT CAC ACT AAA CTG ATC GGT TCC GGT GAA AAA GAC TCC GTT

                100                                     110
Thr Phe Asp Val Ser Lys Leu Lys Glu Gly Glu Gln Tyr Met Phe Phe Cys Thr Phe
ACT TTC GAC GTT TCC AAA CTG AAA GAA GGT GAA CAG TAC ATG TTC TTC TGC ACT TTC

                        120
Pro Gly His Ser Ala Leu Met Lys Gly Thr Leu Thr Leu Lys Trm
CCG GGT CAC TCC GCT CTG ATG AAA GGT ACT CTG ACT CTG AAA TAG
Xma I
```

Table 2

Degenerate codon representation used in the DNA master program.

Ala	GCN	Leu	YTN
Arg	AAY	Lys	AAP
Asn	AAY	Met	ATG
Asp	GAY	Phe	TTY
Cys	TGY	Pro	CCN
Gln	CAP	Ser	RSN
Glu	GAP	Thr	ACN
Gly	GGN	Trp	TGG
His	CAY	Tyr	TAY
Ile	ATZ	Val	GTN

N = G,A,T,C	P= G, A
Z = C,A	Y= C, T
R = T, A	S = C, G

Figure 3

Proposed DNA sequence of *P. aeruginosa* azurin comprising all degenerate codons used in the search for all possible restriction sites.

```
                                        10
Ala Glu Cys Ser Val Asp Ile Gln Gly Asn Asp Gln Met Gln Phe Asn Thr Asn Ala
GCN GAP TGY RSN GTN GAY ATZ CAP GGN AAY GAY CAP ATG CAP TTY AAY ACN AAY GCN

20                                      30
Ile Thr Val Asp Lys Ser Cys Lys Gln Phe Thr Val Asn Leu Ser His Pro Gly Asn
ATZ ACN GTN GAY AAP RSN TGY AAP CAP TTY ACN GTN AAY YTN RSN CAY CCN GGN AAY

    40                                      50
Leu Pro Lys Asn Val Met Gly His Asn Trp Val Leu Ser Thr Ala Ala Asp Met Gln
YTN CCN AAP AAY GTN ATG GGN CAY AAY TGG GTN YTN RSN ACN GCN GCN GAY ATG CAP

    60                                      70
Gly Val Val Thr Asp Gly Met Ala Ser Gly Leu Asp Lys Asp Tyr Leu Lys Pro Asp
GGN GTN GTN ACN GAY GGN ATG GCN RSN GGN YTN GAY AAP GAY TAY YTN AAP CCN GAY

            80                                      90
Asp Ser Arg Val Ile Ala His Thr Lys Leu Ile Gly Ser Gly Glu Lys Asp Ser Val
GAY RSN ZGN GTN ATZ GCN CAY ACN AAP YTN ATZ GGN RSN GGN GAP AAP GAY RSN GTN

                100                                     110
Thr Phe Asp Val Ser Lys Leu Lys Glu Gly Glu Gln Tyr Met Phe Phe Cys Thr Phe
ACN TTY GAY GTN RSN AAP YTN AAP GAP GGN GAP GAP TAY ATG TTY TTY TGY ACN TTY

                    120
Pro Gly His Ser Ala Leu Met Lys Gly Thr Leu Thr Leu Lys Trm
CCN GGN CAY RSN GCN YTN ATG AAP GGN ACN YTN ACN YTN AAP TPP
```

β-lactamase with a factor Xa recognition sequence[38] at the junction, the ends were designed for easy transfer of the gene into the site between the Pvu I and Pst I sites on the ampicillin resistance gene.

The next step in the design process was to determine which fragment of the gene to add in each of the four steps. An algorithm of gene synthesis was chosen so that an about equal number of bases (~100 bp) would be introduced in each step. Because of technical limitations on the oligonucleotide synthesis back in 1985, our goal was to limit the length of each oligonucleotide to about 50 bp. The longer oligonucleotides (up to 124 bp) used in the latter half of the gene synthesis reflect the improvement in the DNA synthesis technology. A segment of the azurin gene designed to be added at each step is shown in Figure 4.

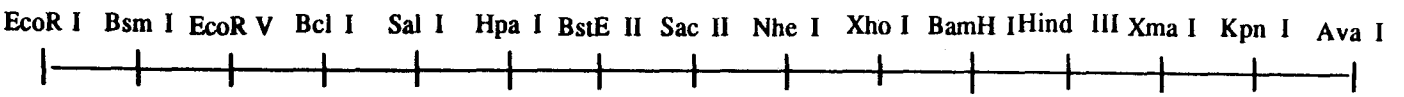
Finally, homology search was performed on the oligonucleotides to be used for self-complementarity to minimize problems during annealing and ligation. Figure 5 shows the final sequence of the azurin gene. A similar design was done on the *A. denitrificans* azurin sequence (Figure 6).[39] Since the two azurins share a high degree of amino acid sequence homology, their DNA sequences are similar. The main difference is that there is a gap of about 150 bp in the *A. denitrificans* DNA sequence, which cannot accommodate a convenient restriction site. For this reason, the gene for *P. aeruginosa* was chosen to be built.

The plasmid pBR322 was the vector in which the azurin gene was constructed. The synthesis was done between the EcoRI and the Ava I sites. Excising away the EcoR I to the Ava I segment of pBR322, which was rich in restriction sites, made many useful restriction enzymes available to be incorporated into the azurin gene as unique sites. Originally, the synthesis was to be completed in four steps, but because of a deletion that

Figure 4

Overview of azurin gene synthesis. Each of the restriction sites is unique when the gene is located between the EcoR I and the Ava I site on pBR322. The segments a-p represent the segments within the azurin gene between each restriction site. The portion of the gene added at each of the five steps is indicated.

Azurin Gene Construction



a. EcoR I-Bsm I
b. Bsm I-EcoR V
c. EcoR V-Bcl I
d. Bcl I-Sal I
e. Sal I-Hpa I
f. Hpa I-BstE II
g. BstE II-Sac II
h. Sac II-Nhe I
i. Nhe I-Xho I
j. Xho I-BamH I
k. BamH I-Hind III
l. Hind III-Xma I
m. Xma I-Kpn I
n. Kpn I-Ava I

Step 1: l-n
Step 2: a, k
Step 3: b-e
Step 4: f, j
Step 5: g-i

Figure 5

The final azurin sequence as it was designed to be constructed between the EcoR I and the Ava I site in pBR322. This design included the factor Xa recognition site as well as the restriction sites at the beginning and the end of the gene to facilitate its transfer to an expression vector.

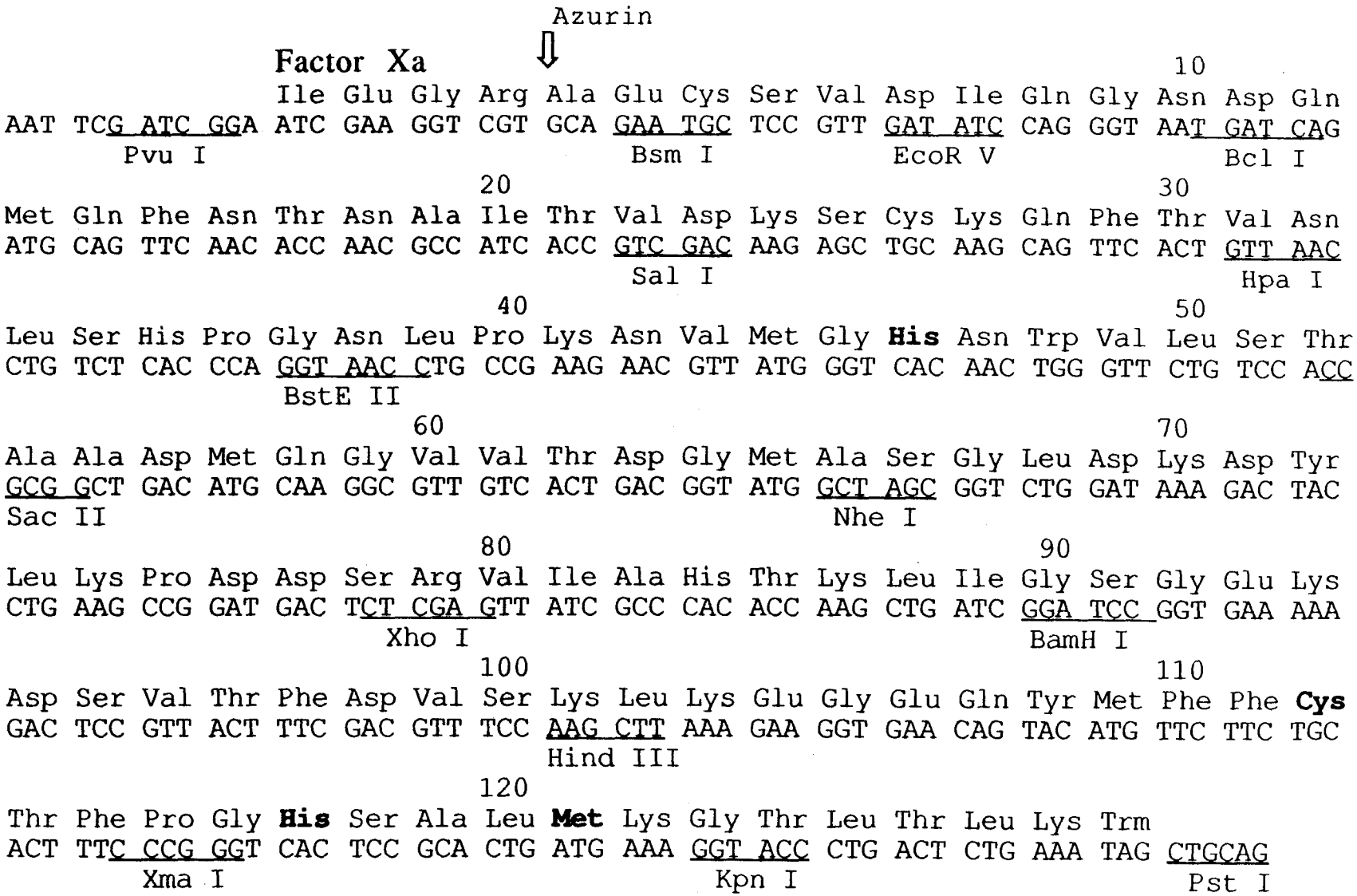

Azurin
Factor Xa
Ile Glu Gly Arg Ala Glu Cys Ser Val Asp Ile Gln Gly Asn Asp Gln
AAT TCG ATC GGA ATC GAA GGT CGT GCA GAA TGC TCC GTT GAT ATC CAG GGT AAT GAT CAG
Pvu I
Bsm I
EcoR V
Bcl I
Met Gln Phe Asn Thr Asn Ala Ile Thr Val Asp Lys Ser Cys Lys Gln Phe Thr Val Asn
ATG CAG TTC AAC ACC AAC GCC ATC ACC GTC GAC AAG AGC TGC AAG CAG TTC ACT GTT AAC
Sal I
Hpa I
Leu Ser His Pro Gly Asn Leu Pro Lys Asn Val Met Gly His Asn Trp Val Leu Ser Thr
CTG TCT CAC CCA GGT AAC CTG CCG AAG AAC GTT ATG GGT CAC AAC TGG GTT CTG TCC ACC
BstE II
Ala Ala Asp Met Gln Gly Val Val Thr Asp Gly Met Ala Ser Gly Leu Asp Lys Asp Tyr
GCG GCT GAC ATG CAA GGC GTT GTC ACT GAC GGT ATG GCT AGC GGT CTG GAT AAA GAC TAC
Sac II
Nhe I
Leu Lys Pro Asp Asp Ser Arg Val Ile Ala His Thr Lys Leu Ile Gly Ser Gly Glu Lys
CTG AAG CCG GAT GAC TCT CGA GTT ATC GCC CAC ACC AAG CTG ATC GGA TCC GGT GAA AAA
Xho I
BamH I
Asp Ser Val Thr Phe Asp Val Ser Lys Leu Lys Glu Gly Glu Gln Tyr Met Phe Phe Cys
GAC TCC GTT ACT TTC GAC GTT TCC AAG CTT AAA GAA GGT GAA CAG TAC ATG TTC TTC TGC
Hind III
Thr Phe Pro Gly His Ser Ala Leu Met Lys Gly Thr Leu Thr Leu Lys Trm
ACT TTC CCG GGT CAC TCC GCA CTG ATG AAA GGT ACC CTG ACT CTG AAA TAG CTGCAG
Xma I
Kpn I
Pst I

Figure 6

Proposed DNA sequence for *Alcaligenes denitrificans* azurin whose crystal structure is better resolved than that of *P. aeruginosa* azurin. Note the gap in the gene of 150 bp.

```
                                        10
Ala Gln Cys Glu Ala Thr Ile Glu Ser Asn Asp Ala Met Gln Tyr Asp Leu Lys Glu
GCT CAG TGC GAA GCT ACT ATC GAA TCC AAC GAC GCT ATG CAG TAC GAC CTG AAA GAA
            Bsm I
20                                      30
Met Val Val Asp Lys Ser Cys Lys Gln Phe Thr Val His Leu Lys His Val Gly Lys
ATG GTT GTC GAC AAA TCC TGC AAA CAG TTC ACT GTT CAC CTG AAA CAC GTT GGT AAA
        Sal I
   40                                       50
Met Ala Lys Ala Val Met Gly His Asn Trp Val Leu Thr Lys Glu Ala Asp Lys Glu
ATG GCT AAA GCT GTT ATG GGT CAC AAC TGG GTG TTA ACT AAA GAA GCT GAC AAA GAA
                                         Hpa I
    60                                       70
Gly Val Ala Thr Asp Gly Met Asn Ala Gly Leu Ala Gln Asp Tyr Val Lys Ala Gly
GGT GTC GCG ACT GAC GGT ATG AAC GCA GGC CTG GCT CAG GAC TAC GTT AAA GCT GGT
      Nru I                             Stu I
80                                      90
Asp Thr Arg Val Ile Ala His Thr Lys Val Ile Gly Gly Gly Glu Ser Asp Ser Val
GAC ACT CGC GTT ATC GCT CAC ACT AAA GTT ATC GGT GGT GGT GAA TCC GAC TCC GTT

                100                                     110
Thr Phe Asp Val Ser Lys Leu Thr Pro Gly Glu Ala Tyr Ala Tyr Phe Cys Ser Phe
ACT TTC GAC GTT TCC AAG CTT ACT CCG GGT GAA GCT TAC GCT TAC TTC TGC TCC TTC
                     Hind III
                     120                                        130
Pro Gly His Trp Ala Met Met Lys Gly Thr Leu Lys Leu Ser Asn Trm
CCG GGT CAC TGG GCT ATG ATG AAA GGT ACT CTG AAA CTG TCC AAC TAG
Xma I                               Kpn I
```

was later discovered upon sequencing, the synthesis was finished in five steps. In addition, mutations revealed by DNA sequencing were corrected in two separate ligations. Figure 7 shows the actual oligonucleotides used in the five-step synthesis of the azurin gene.

Gene Synthesis Step 1

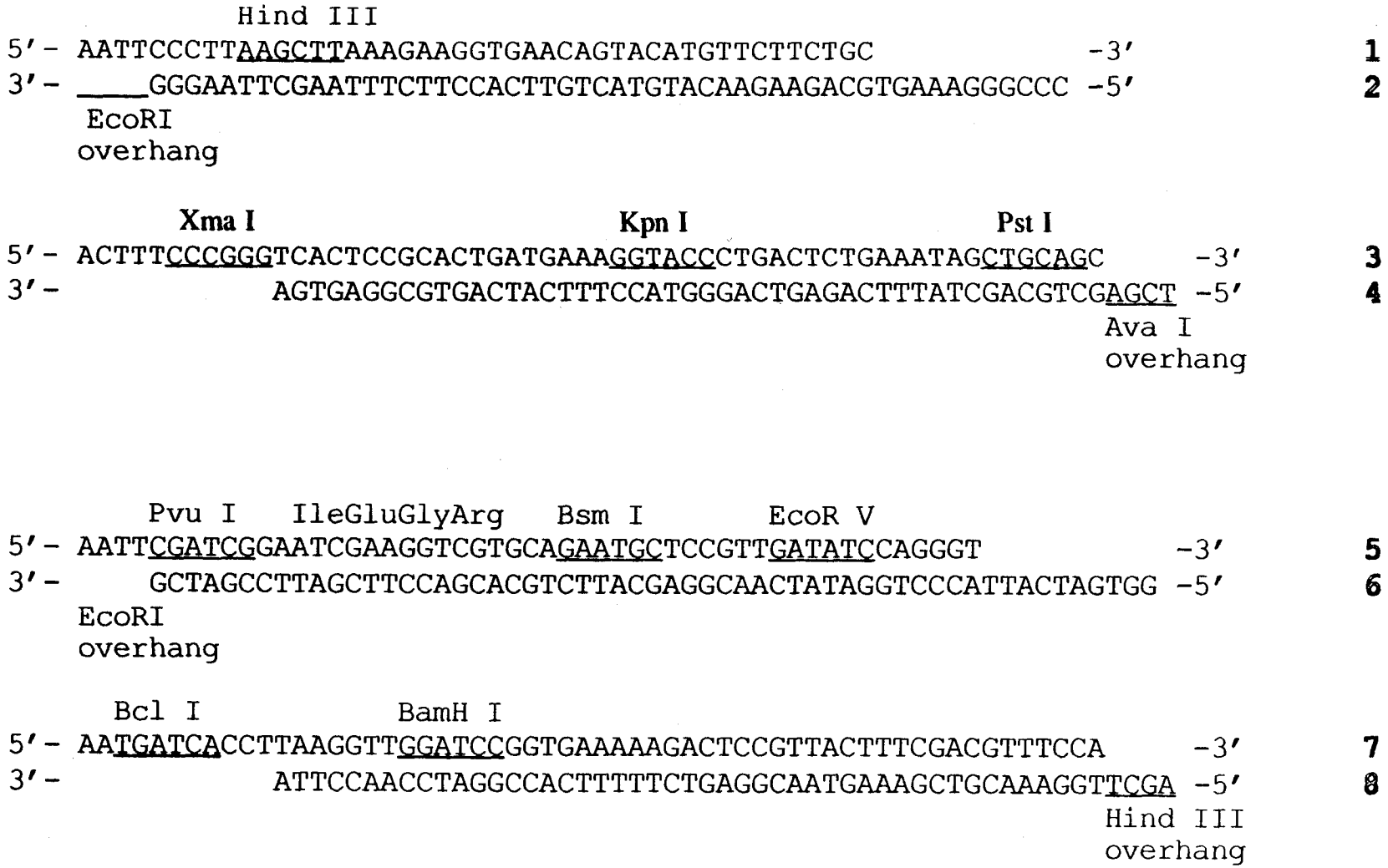
The plasmid pBR322 was cleaved with EcoR I and Ava I. The larger fragment was purified on a 1.2% agarose gel and recovered using an IBI electroeluter. Meanwhile, oligonucleotides 1-4 were synthesized, separated on a 15% polyacrylamide gel, and purified through NACS column. They were then kinased and annealed. The ligation was performed using 0.4 to 0.04 ratio of vector to oligonucleotides. *E. coli* strain LS-1 was transformed with this ligation mixture and plated on ampicillin. They were also plated on tetracycline plates to determine the amount of background parental plasmid. Because the pBR322 fragments were purified prior to ligation, no colonies grew on tetracycline. 50 ml-scale plasmid preparation was performed on six samples among the colonies that grew on ampicillin. These six plasmids were digested with Kpn I in buffer L. Kpn I was chosen since it was not present on pBR322. All six plasmids were cleaved by Kpn I and showed the expected size, indicating that step 1 was successful.

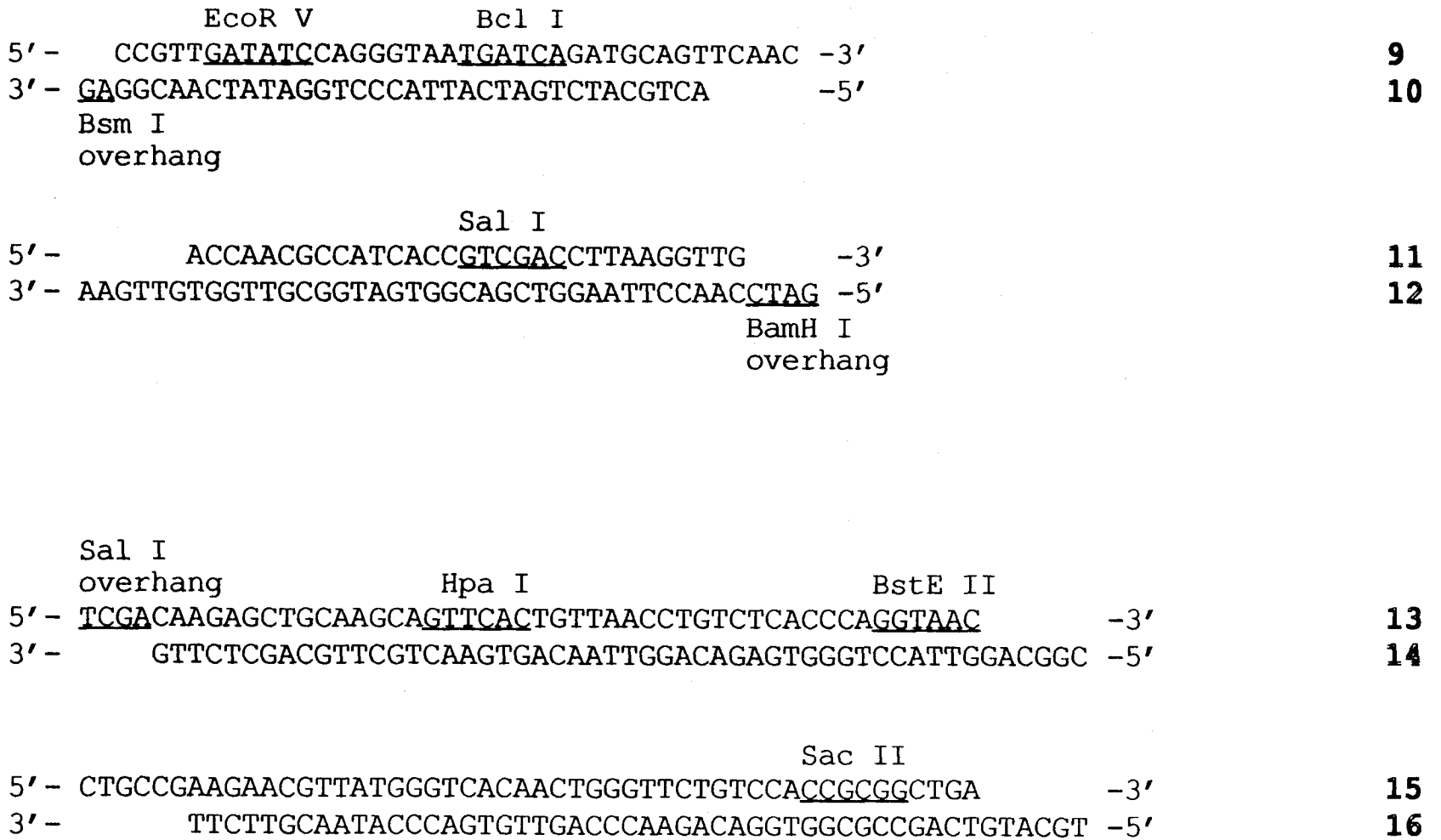
Gene Synthesis Step 2

About 5 μg of step 1 plasmid were digested with 35 units of EcoR I and 10 units of Hind III in buffer H. A greater amount of EcoR I was used since NEB claimed that pBR322 was harder to cleave with this enzyme. The DNA fragments were again separated on a 1.2% agarose gel. The larger fragment was recovered using an IBI electroeluter. Oligonucleotides 5-8,

Figure 7

The actual oligonucleotides used to synthesize the azurin gene. This was slightly altered from the original design. Each pair of oligonucleotides was designed to form a duplex cassette with compatible overhangs.


Hind III
5′- AATTCCCTTAAGCTTAAAGAAGGTGAACAGTACATGTTCTTCTGC -3′ 1
3′- ____GGGAATTCGAATTTCTTCCACTTGTCATGTACAAGAAGACGTGAAAGGGCCC -5′ 2
EcoRI
overhang
Xma I
Kpn I
Pst I
5′- ACTTTCCCGGGTCACTCCGCACTGATGAAAGGTACCCTGACTCTGAAATAGCTGCAGC -3′ 3
3′- AGTGAGGCGTGACTACTTTCCATGGGACTGAGACTTTATCGACGTCGAGCT -5′ 4
Ava I
overhang
Pvu I
IleGluGlyArg
Bsm I
EcoR V
5′- AATTCGATCGGAATCGAAGGTCGTGCAGAATGCTCCGTTGATATCCAGGGT -3′ 5
3′- GCTAGCCTTAGCTTCCAGCACGTCTTACGAGGCAACTATAGGTCCCATTACTAGTGG -5′ 6
EcoRI
overhang
Bcl I
BamH I
5′- AATGATCACCTTAAGGTTGGATCCGGTGAAAAAGACTCCGTTACTTTCGACGTTTCCA -3′ 7
3′- ATTCCAACCTAGGCCACTTTTTCTGAGGCAATGAAAGCTGCAAAGGTTCGA -5′ 8
Hind III
overhang


EcoR V
Bcl I
5′- CCGTTGATATCCAGGGTAATGATCAGATGCAGTTCAAC -3′ 9
3′- GAGGCAACTATAGGTCCCATTACTAGTCTACGTCA -5′ 10
Bsm I
overhang
Sal I
5′- ACCAACGCCATCACCGTCGACCTTAAGGTTG -3′ 11
3′- AAGTTGTGGTTGCGGTAGTGGCAGCTGGAATTCCAACCTAG -5′ 12
BamH I
overhang
Sal I
overhang
Hpa I
BstE II
5′- TCGACAAGAGCTGCAAGCAGTTCACTGTTAACCTGTCTCACCCAGGTAAC -3′ 13
3′- GTTCTCGACGTTCGTCAAGTGACAATTGGACAGAGTGGGTCCATTGGACGGC -5′ 14
Sac II
5′- CTGCCGAAGAACGTTATGGGTCACAACTGGGTTCTGTCCACCGCGGCTGA -3′ 15
3′- TTCTTGCAATACCCAGTGTTGACCCAAGACAGGTGGCGCCGACTGTACGT -5′ 16

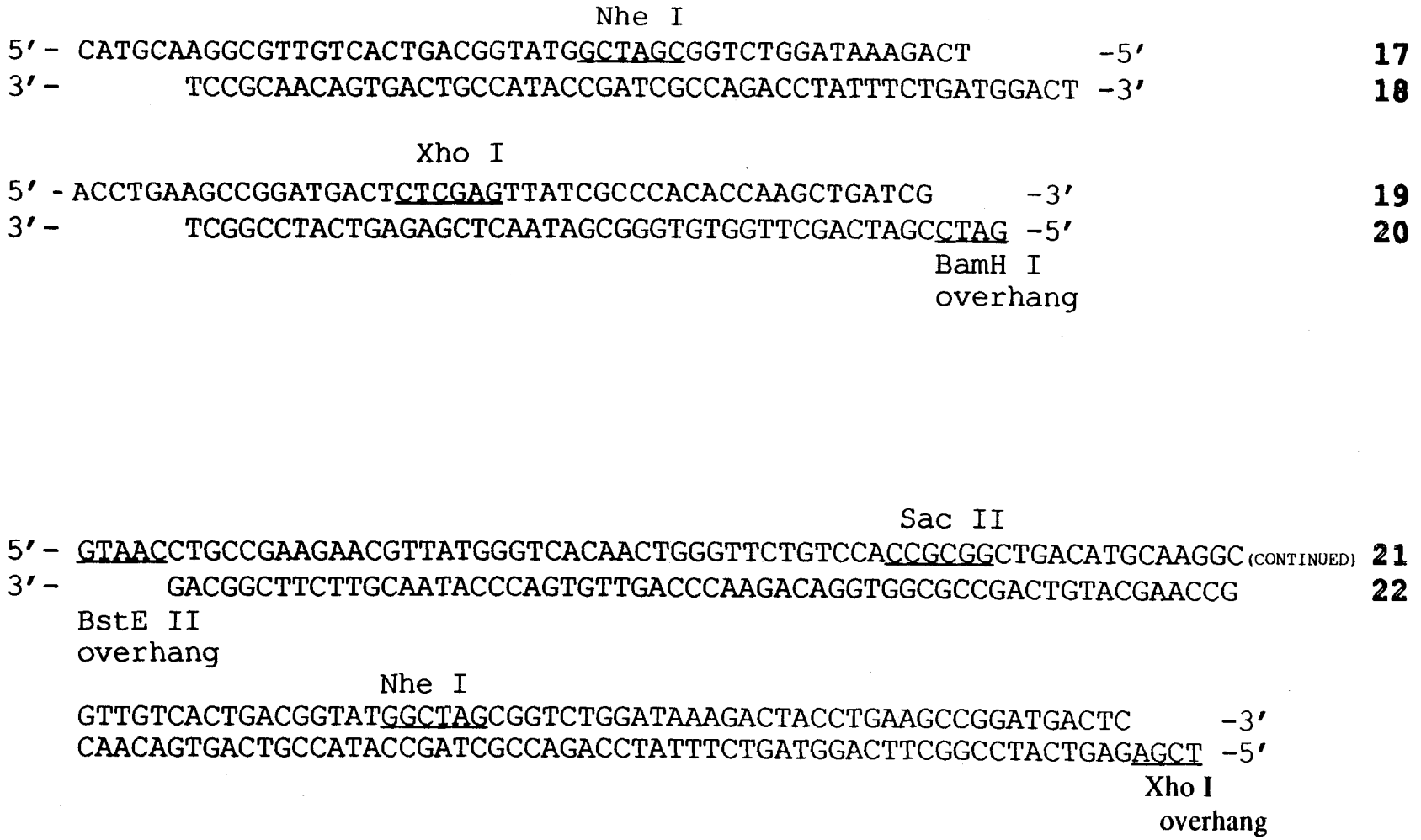
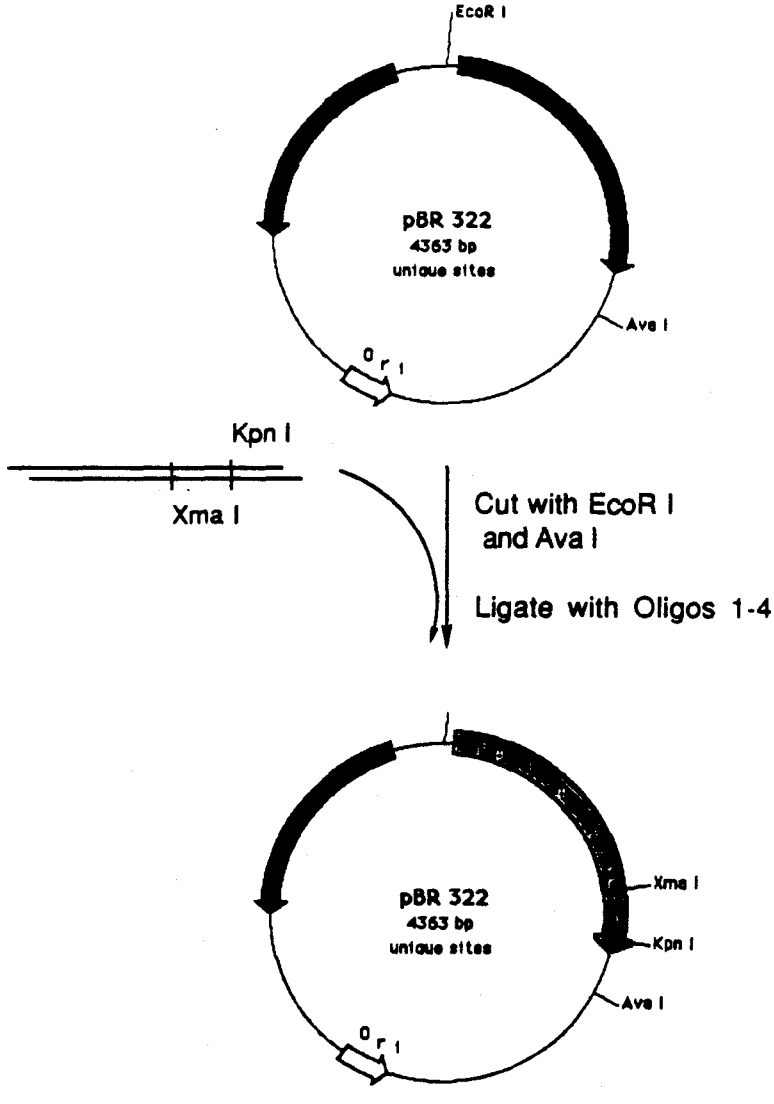
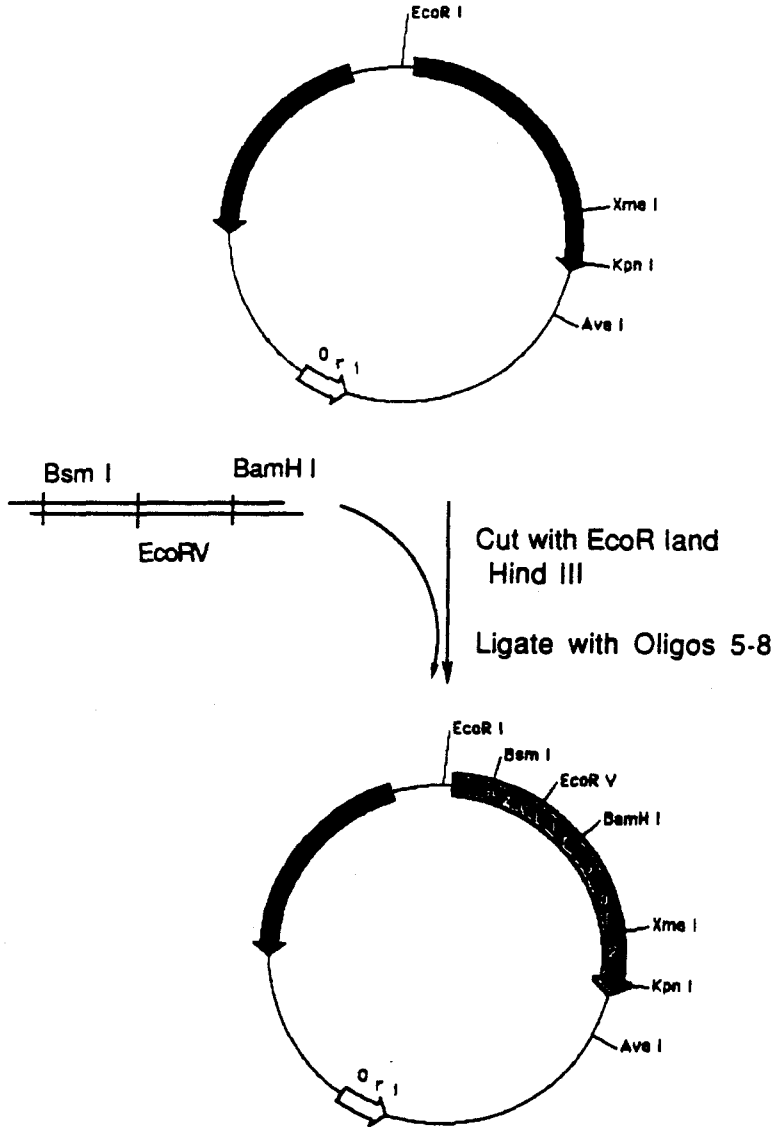
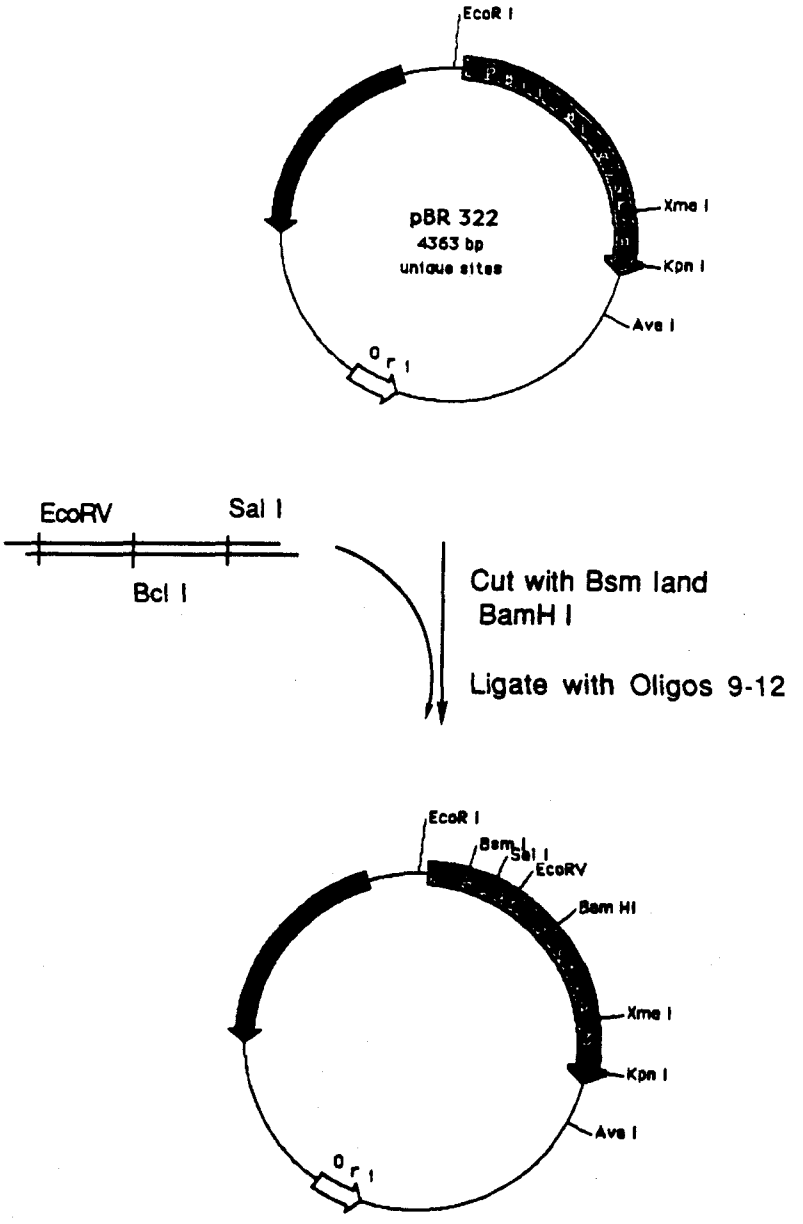

Nhe I
5′- CATGCAAGGCGTTGTCACTGACGGTATGGCTAGCGGTCTGGATAAAGACT -5′ 17
3′- TCCGCAACAGTGACTGCCATACCGATCGCCAGACCTATTTCTGATGGACT -3′ 18
Xho I
5′ -ACCTGAAGCCGGATGACTCTCGAGTTATCGCCCACACCAAGCTGATCG -3′ 19
3′- TCGGCCTACTGAGAGCTCAATAGCGGGTGTGGTTCGACTAGCCTAG -5′ 20
BamH I
overhang
Sac II
5′- GTAACCTGCCGAAGAACGTTATGGGTCACAACTGGGTTCTGTCCACCGCGGCTGACATGCAAGGC (CONTINUED) 21
3′- GACGGCTTCTTGCAATACCCAGTGTTGACCCAAGACAGGTGGCGCCGACTGTACGAACCG 22
BstE II
overhang
Nhe I
GTTGTCACTGACGGTATGGCTAGCGGTCTGGATAAAGACTACCTGAAGCCGGATGACTC -3′
CAACAGTGACTGCCATACCGATCGCCAGACCTATTTCTGATGGACTTCGGCCTACTGAGAGCT -5′
Xho I
overhang

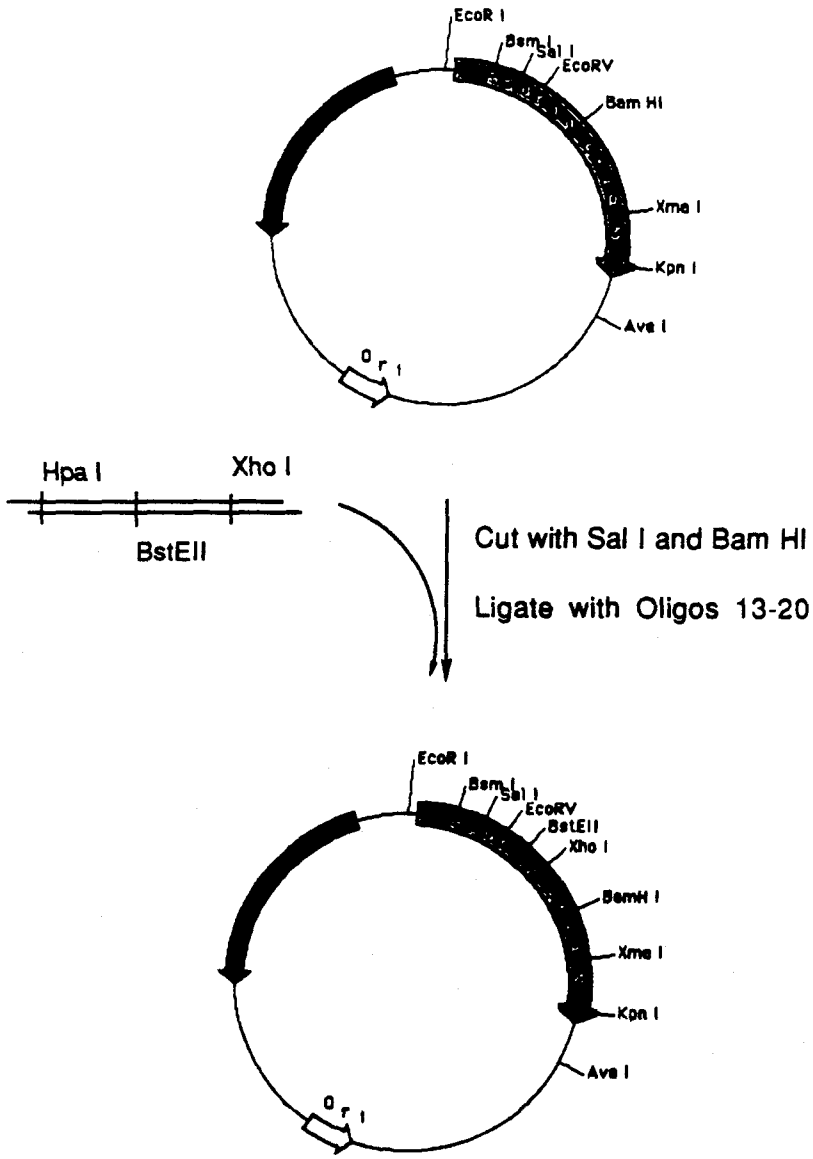
Figure 8

The five steps of azurin gene synthesis. In each of the steps, the plasmid was cleaved with two different restriction enzymes and gel-purifed. Then the annealed oligonucleotide duplexes were ligated into that site.


EcoR I
pBR 322
4363 bp
unique sites
Ava I
Kpn I
Xma I
Cut with EcoR I
and Ava I
Ligate with Oligos 1-4
pBR 322
4363 bp
unique sites
Xma I
Kpn I
Ava I


EcoR I
Xma I
Kpn I
Ave I
Ori
Bsm I
BamH I
EcoRV
Cut with EcoR I and
Hind III
Ligate with Oligos 5-8
EcoR I
Bsm I
EcoR V
BamH I
Xma I
Kpn I
Ave I
Ori


EcoR I
pBR 322
4363 bp
unique sites
Xma I
Kpn I
Ava I
Ori
EcoRV
Sal I
Bcl I
Cut with Bsm I and
BamH I
Ligate with Oligos 9-12
EcoR I
Bsm I
Sal I
EcoRV
Bam HI
Xma I
Kpn I
Ava I
Ori


EcoR I
Bam I
Sal I
EcoRV
Bam HI
Xma I
Kpn I
Ava I
Ori
Hpa I
Xho I
BstEII
Cut with Sal I and Bam HI
Ligate with Oligos 13-20
EcoR I
Bam I
Sal I
EcoRV
BstEII
Xho I
BamH I
Xma I
Kpn I
Ava I
Ori

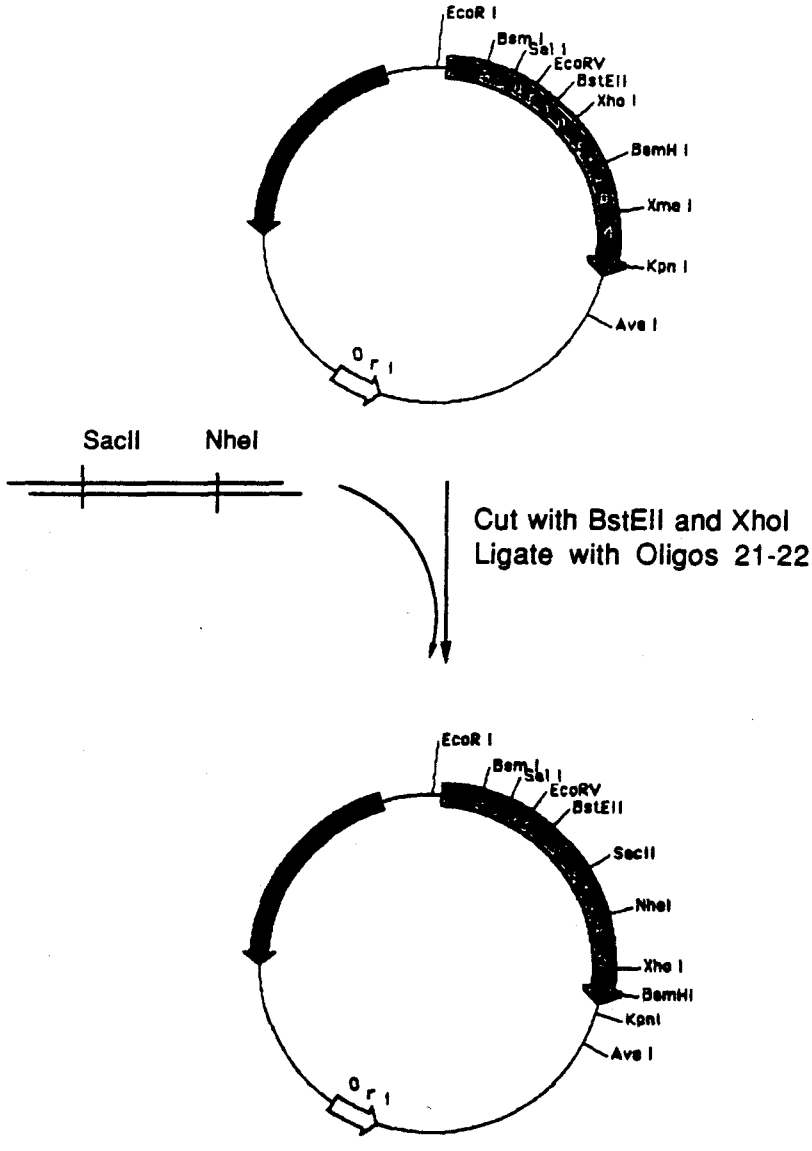
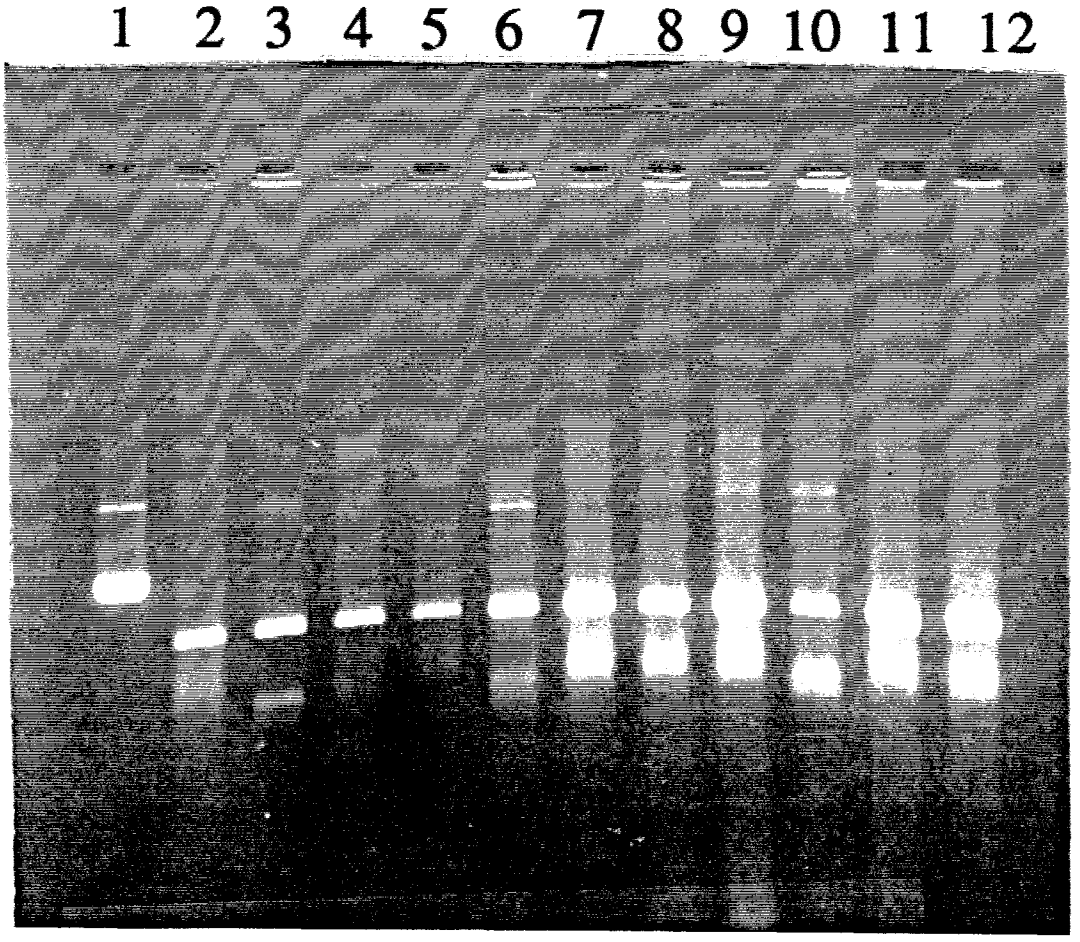

EcoR I
Bam I
Sal I
EcoRV
BstEII
Xho I
BamH I
Xma I
Kpn I
Ava I
Ori
SacII
NheI
Cut with BstEII and XhoI
Ligate with Oligos 21-22
EcoR I
Bam I
Sal I
EcoRV
BstEII
SacII
NheI
Xho I
BamHI
KpnI
Ava I
Ori

Figure 9

Restriction digest of the step 1 plasmid. Six different plasmids were digested with EcoR I (lanes 2-7) and with Kpn I, which was introduced in step 1 (lanes 8-12). Lane 1 contains undigested plasmid.

Restriction Digests of Step 1 Plasmids



which had been synthesized in advance, were gel-purified as before. They were then kinased and annealed. The ratio of 0.4 to 0.04 pmole (oligomer to vector) was again employed. The ligation mixture was transformed into LS-1 cells and plated on ampicillin plates. Plasmid preparation (50 ml) was again performed on 6 colonies. These plasmids were digested with BamH I, which was one of the 3 restriction sites added during this step. Again, all 6 plasmids were cleaved with BamH I, indicating that the cassette was ligated into the plasmid.

Gene Synthesis Step 3

Step 3 ligation was originally designed to insert the new piece between Bcl I and BamH I sites of step 2 plasmid. But for a yet unexplained reason, difficulty was encountered in cleaving the plasmid with both enzymes. Since the Bcl I recognition site contains the dam methylation site of GATC, it is sensitive to the presence of this methylase. Yet, even when the plasmid preparation was performed using a dam$^-$ strain of *E. coli* such as JM101Tn9 and GM48, the digest was never clean. The ligation product never showed the expected restriction digest patterns or sizes even after several attempts. Thus, to determine the cause of this difficulty, step 2 plasmid was purified, using CsCl-gradient and sequenced. Step 2 plasmid was sequenced using the modified Maxam-Gilbert method. The sequence was read from both the EcoR I site and the BamH I site. The sequencing revealed that the Bcl I restriction site was present. Therefore, it was determined that Bcl I was particularly sensitive to a neighboring sequence.

To circumvent this problem, it was decided that step 3 would be slightly modified and the ligation done between Bsm I and BamH I. This required a slightly longer cassette. About 10 μg of step 2 plasmid were

cleaved in buffer M with Bsm I. Then the NaCl concentration was increased to 100 mM by the addition of 5 M NaCl. BamH I was added to this mixture and incubated at 37°C. The DNA fragments were separated on 1.2% agarose gel and recovered on an Elutrap. Oligonucleotides 9-12 (modified) were purified on a 15% polyacrylamide gel and NACS column as before. The ligation was performed using the standard 0.4 to 0.04 pmole of insert-to-vector ratio. This ligation mixture was transformed into HB101 cells. Of the colonies that grew on ampicillin, plasmid preparation on 8 samples were performed on a 50 ml scale. These plasmids were digested with Sal I. Two colonies (12 and 17), which were cut by Sal I, were then digested with Bsm I/Bgl I and Sal I/Bgl I. The sizes of these plasmid were slightly bigger than the same double digests of step 2 plasmid and determined to be the correct step 3.

Gene Synthesis Step 4

Because the original design was to finish the synthesis in 4 steps, an attempt was made to complete the synthesis by inserting 4 sets of oligonucleotides totaling 198 bp. To prepare the step 3 vector, about 10 μg of this plasmid were digested in buffer H first with BamH I followed by Sal I. The DNA fragments were separated on a 1.2% agarose gel. The large piece was excised and recovered from the gel, using an elutrap. Eight oligonucleotides forming 4 double-stranded fragments were synthesized and purified from a 15% polyacrylamide gel. These were individually kinased and then annealed to their complementary strands. These were ligated into step 3 plasmid opened with BamH I and Sal I. LS-1 cells were transformed with these mixtures, and plasmid preparation performed from colonies that grew on ampicillin. The plasmids were screened by

cutting with restriction endonuclease Xho I, which was added in this step. The plasmids from several colonies could be cut by Xho I but not by some of the other enzymes introduced in this step such as Nhe I and Sac II. These plasmids were denatured and sequenced using the Sequenase™ kit. The sequencing revealed that two complementary regions ligated together and left out a substantial portion of the insert (Figure 10). This necessitated a fifth step.

Gene Synthesis Step 5

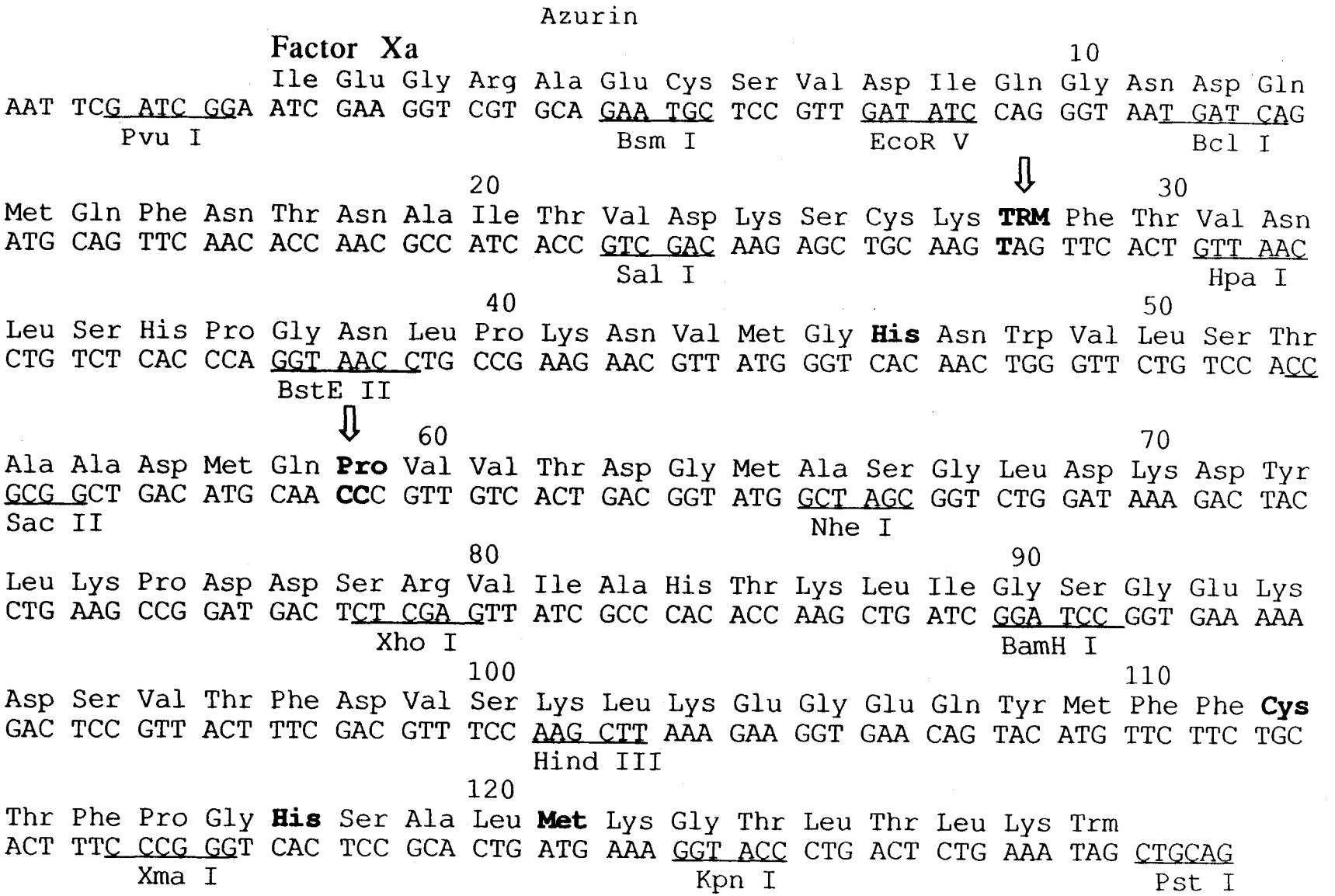
The plasmid from step 4 was cleaved with BstE II/Bgl I and also with Xho I/Bgl I for a three-piece ligation. The large fragment from the former digest and the small fragment from the latter digest were purified on a 1.2% agarose gel and an elutrap. Oligonucleotides 21 and 22 were synthesized. The ability to synthesize these long oligonucleotides, whose lengths were 124 and 123, respectively, manifested the improvement in the DNA synthesis technology. They were kinased individually and annealed together. Then 0.4 pmole of this annealed DNA was ligated with 0.04 pmole of each of the two fragments from the vector. The ligation mixture was transformed into LS-1 as before, and the colonies screened by their ability to grow in the presence of ampicillin. Plasmid preparation was performed on 6 colonies using CsCl-gradient. Two plasmids among these, colonies 1 and 23, contained the expected restriction sites and were of the correct size. They were sequenced completely using the Sequenase™ kit. Sequencing revealed the same mutations in both plasmids: CAG -> TAG (Stop) mutation near the Sal I site and a GGC -> CCC mutation near the Nhe I site (Figure 11). These needed to be corrected.

```
13      ...GGTAAC     + ACCTGAAGCCGG..19
14      ...CCATTGGACGGC         TCGGCC..20
```

Figure 10. Mismatched annealing between two remote regions of the gene resulted in a deletion of ~100 bp. The complementary bases are highlighted.

Figure 11

The results of the discovered mutations in the azurin gene's DNA sequence and its effect on the amino acid sequence are shown. The mutated bases and the resulting change in the amino acid sequence are highlighted.


Azurin
Factor Xa
Ile Glu Gly Arg Ala Glu Cys Ser Val Asp Ile Gln Gly Asn Asp Gln
AAT TCG ATC GGA ATC GAA GGT CGT GCA GAA TGC TCC GTT GAT ATC CAG GGT AAT GAT CAG
Pvu I
Bsm I
EcoR V
Bcl I
Met Gln Phe Asn Thr Asn Ala Ile Thr Val Asp Lys Ser Cys Lys TRM Phe Thr Val Asn
ATG CAG TTC AAC ACC AAC GCC ATC ACC GTC GAC AAG AGC TGC AAG TAG TTC ACT GTT AAC
Sal I
Hpa I
Leu Ser His Pro Gly Asn Leu Pro Lys Asn Val Met Gly His Asn Trp Val Leu Ser Thr
CTG TCT CAC CCA GGT AAC CTG CCG AAG AAC GTT ATG GGT CAC AAC TGG GTT CTG TCC ACC
BstE II
Ala Ala Asp Met Gln Pro Val Val Thr Asp Gly Met Ala Ser Gly Leu Asp Lys Asp Tyr
GCG GCT GAC ATG CAA CCC GTT GTC ACT GAC GGT ATG GCT AGC GGT CTG GAT AAA GAC TAC
Sac II
Nhe I
Leu Lys Pro Asp Asp Ser Arg Val Ile Ala His Thr Lys Leu Ile Gly Ser Gly Glu Lys
CTG AAG CCG GAT GAC TCT CGA GTT ATC GCC CAC ACC AAG CTG ATC GGA TCC GGT GAA AAA
Xho I
BamH I
Asp Ser Val Thr Phe Asp Val Ser Lys Leu Lys Glu Gly Glu Gln Tyr Met Phe Phe Cys
GAC TCC GTT ACT TTC GAC GTT TCC AAG CTT AAA GAA GGT GAA CAG TAC ATG TTC TTC TGC
Hind III
Thr Phe Pro Gly His Ser Ala Leu Met Lys Gly Thr Leu Thr Leu Lys Trm
ACT TTC CCG GGT CAC TCC GCA CTG ATG AAA GGT ACC CTG ACT CTG AAA TAG CTGCAG
Xma I
Kpn I
Pst I

Mutation Correction

Since the two mutations were very far apart from each other, the corrections were done in two separate ligations using the oligonucleotides shown in Figure 12. For the CCG mutation, the step 5 plasmid was cleaved with Sac II in buffer L, followed by cleavage with Xho I in 100 mM NaCl. This was dephosphorylated using calf-intestinal alkaline phosphatase prior to gel purification. The same plasmid was also cleaved with Sac II/Bgl I and Xho I/Bgl I for three-piece ligation. Oligonucleotides 23 and 24 were kinased separately and annealed together. This was inserted between the Sac II and Xho I site of step 5 plasmid, using both standard two-piece and three-piece ligations. Both ligation procedures yielded colonies with the mutation corrected.

To eliminate the TAG mutation, the plasmid with the first mistake corrected was cleaved with Sal I in buffer H, followed by Hpa I. The DNA was ethanol-precipitated, resuspended in 1x CAP buffer, and had the 5' phosphate removed using calf-intestinal alkaline phosphatase. The cleaved plasmid was then purified on a 1.2% agarose gel and purified with an elutrap. In addition to the standard ligation conditions, another ligation was done with 0.08 pmole vector and 0.12 pmole of oligomers. The higher DNA concentration was attempted since the Hpa I site was blunt-ended. Such an end usually required higher concentrations of both DNA and ligase. Plasmids purified using CsCl gradient showed that the mutation was corrected, and the azurin gene was finally completed.

Figure 12

The oligonucleotides used in the two correction steps.

Nhe I

5′- GGCTGACATGCAAGGCGTTGTCACTGACGGTATGGCTAGCGGTCTGGATAAAGACTACCTGAAGC (CONTINUED) 23

3′- CCGACTGTACGTTCCGCAACAGTGACTGCCATACCGATCGCCAGACCTATTTCTGATGGACTTCG 24

Sac II
overhang (Blunt)

CGGATGACTC -3′
GCCTACTGAGAGCT -5′
Xho I
overhang

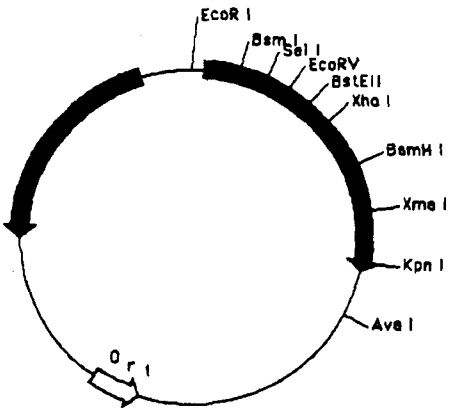
5′- TCGACAAGAGCTGCAAGCAGTTCACTGTT -3′ 25

3′- GTTCTCGACGTTCGTCAAGTGACAA -5′

Sal I Hpa I
overhang overhang (Blunt) 26

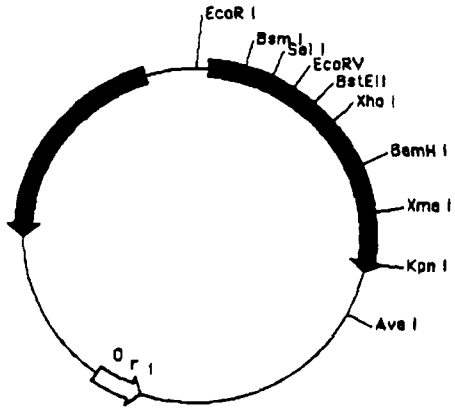
Figure 13

Ligations to correct the two mutations within the azurin gene.



1.Cut with
SacII and Xho I
2. ligate with oligonucleotides
23 + 24.

CCC (Pro) ⟹ GGC (Gly)



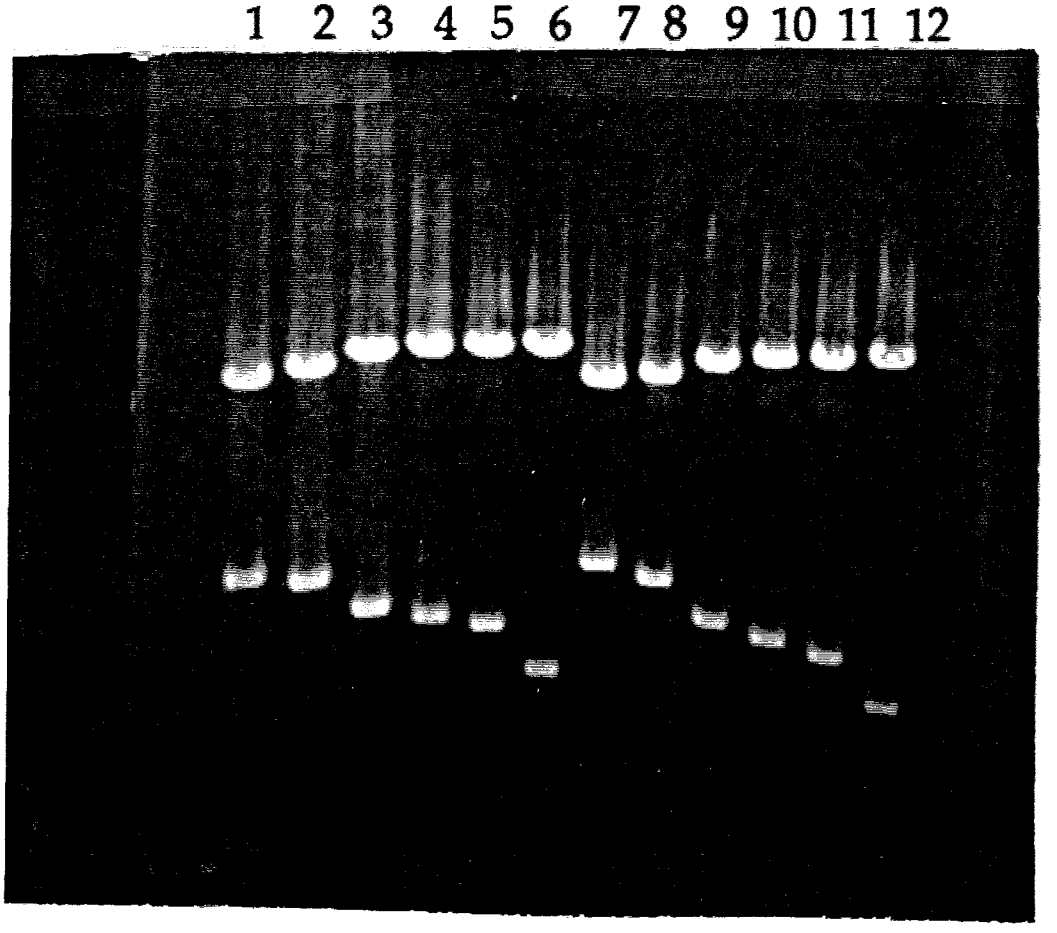
1.Cut with
SacII and Xho I
2. ligate with oligonucleotides
25 + 26.

TAG (Trm) ⟹ CAG (Gln)

Figure 14

Restriction digest of two different plasmids containing the completed azurin gene. Each lane was cleaved with successively located restriction enzymes on the azurin gene to ensure not only the presence of a restriction site but also the orientation and the order in which they appear.

Restriction Digest of Complete Azurin Gene



DISCUSSION

The azurin gene synthesis described here not only outlines a reliable way to construct a gene but also serves as a reflection on the progress of molecular biology. This method can be applied to the construction of much larger genes. Although bigger genes have been built in one-step in other labs, such one step approach often creates ligation difficulties. As seen in step 4 of our synthesis, for example, ligating in too many bases in a single step led to an unexpected and unwanted base-pairing between remote segments of the gene. Therefore, our method becomes much more favorable and reliable as the gene to be made becomes larger. This approach should be further aided by the continuing improvement in the oligonucleotide synthesis and in the diversity of restriction enzymes available. Overall, as one becomes more familiar with restriction enzymes and as the general gene manipulation techniques improve, synthesizing a gene should become a more feasible method to obtain a gene.

The mutation that introduced the amber-stop codon in the middle of the azurin gene is a particularly interesting mutation. Similar mutation was found during the plastocyanin gene synthesis. Although the tetracycline promoter was deleted from pBR322 prior to our gene synthesis, there were sequences upstream from the EcoR I site that may have served fortuitously as a promoter and a ribosomal binding site. Therefore, it is likely that azurin and plastocyanin were harmful when expressed in the cytoplasm. If there were any expression of these proteins, it could have selected for a mutant that contained a stop codon.

REFERENCES

1. Watson, J. D. and Crick, F. H. C. (1953) *Nature (London)* **171**, 737-738.
2. Nathans, D. (1979) *Science* **206**, 903-909.
3. Richards, J. H. (1986) *Nature (London)* **323**, 187.
4. Dalbadie-McFarland, G., Cohen, L. W., Riggs, A. D., Morin, A. D., Hakura, K. and Richards, J. H. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6409-6413.
5. Smith, M. (1985) *Annu. Rev. Genet.* **19**, 423-462.
6. Wells, J. A., Vasser, M. and Powers, D. B. (1985) *Gene* **34**, 315-323.
7. Knowles, J. R. (1987) *Science* **242**, 1425-1430.
8. Khorana, H. G. (1971) *J. Mol. Biol.* **72**, 209-217.
9. Edge, M. D., Greene, A. R., Heathcliffe, G. R., Meacock, P. A., Schuck, W., Scanlon, D. B., Atkinson, T. C., Newton, C. R. and Markham A. R. (1981) *Nature (London)* **292**, 756.
10. Hakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F. and Boyer, H. W. (1977) *Science* **198**, 1056-1063.
11. Goeddel, D. V., Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., Hirose, T., Kraszewski, A., Hakura, K. and Riggs, A. D. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 106-110.
12. Kirschner, R. J., Hatzenbuhler, N. T., Moseley, M. W. and Tomich, C. S. C. (1989) *J. Biotech.* **12**, 247-259.
13. Kille, P., Stephens, P., Cryer, A. and Kay, J. (1990) *Biochim. Biophys. Acta* **1048**, 178-186.
14. Springer, B. A. and Sligar, S. G. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8961-8965.

15. Louis, J. M., Wondrak, E. M., Copeland, T. D., Smith, C. A. D., Dale, M., Mora, P. T. and Oroszlan, S. (1989) *Biochim. Biophys. Res. Comm.* **159**, 87-94.

16. Beck von Bodman, S., Schuler, M. A., Jollie, D. R. and Sligar, S. G. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9443-9447.

17. Boyer, H. W. and Roulland-Dussoix, D. (1969) *J. Mol. Biol.* **41**, 459-472.

18. Richardson, C. C. (1965) *Proc. Natl. Acad. Sci. USA* **54**, 158.

19. Lehman, I. R. (1974) *Science* **186**, 790-797.

20. Beacage, S. L. and Carruthers, M. H. (1986) *Tetrahedron Lett.* **22**, 1859-1862.

21. NACS Prepac column manual, Bethesda Research Laboratories.

22. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory, New York.

23. Sutcliffe, J. G. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3737-3741.

24. Elutrap Electroelutor manual, Schleicher and Schuell, Inc.

25. International Biotechnologies Catalog, New Haven, CT.

26. Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557-580.

27. Ish-Horowicz, D. and Burke, J. F. (1981) *Nucl. Acids Res.* **9**, 2989-2998.

28. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory, New York, Vol. 1, pp. 1.42-1.45.

29. Maxam, A. M. and Gilbert, W. (1977) *Methods in Enzymology* **65**, 499-560.

30. Iverson, B. L. and Dervan, P. B. (1987) *Nucl. Acids Res.* **15**, 7823-7830.

31. Iverson, S. A. (1987) Ph.D. Thesis, California Institute of Technology, Pasadena, California.

32. Sequanase™ sequencing manual, United States Biochemical Co., Pittsburgh, PA.

33. Sanger, F., Nicklen, S. and Coulson, R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.

34. Hattori, M. M. and Sakaki, Y. (1986) *Anal. Biochem.* **152**, 232-238.

35. Perez, D. M. (1987) Ph.D. Thesis, California Institute of Technology, Pasadena, California.

36. deBoer, H. A. and Kastelein, R. A. (1986) in *Maximizing Gene Expression*, ed. Reznikoff, W. and Gold, L. (Butterworth, Boston), pp. 225-285.

37. Ikemura, T. (1981) *J. Mol. Biol.* **151**, 389-409.

38. Makoff, J., Oxer, M. D., Romanos, M. A., Fairweather, N. F. and Ballantine, S. (1989) *Nucl. Acids. Res.* **17**, 10191-10202.

Chapter 3

Expression of Plastocyanin

INTRODUCTION

Plastocyanin is an essential protein involved in the photosynthesis of green plants.[1] Its structural resemblance to azurin (Figure 1)[2] is not surprising, considering that the two proteins share a high degree of amino acid sequence homology.[3] Because of its important role in photosynthesis and its intriguing blue copper site, plastocyanin has attracted considerable interest, as evidenced by several plastocyanin genes that have been cloned.[5-11] These genes usually contain relatively long (66 amino acids for white campion) signal sequences. Among these isolated genes, only the pea plastocyanin gene,[12] which required the use of tobacco leaves as the expression host, has been expressed. A synthetic gene for *Populus nigra* var *italica* plastocyanin[4] was the first gene to be constructed in this lab using the gene synthesis method described in Chapter 2. The gene was constructed in four steps (Figures 2 and 3).

An *E. coli* bacterium contains all the biochemical machinery necessary to transcribe DNA into RNA and then to translate RNA into its corresponding polypeptide chain. This linear sequence of amino acids in turn assumes its predetermined three-dimensional structure and becomes an active protein. To optimize a gene for expression, the use of *E. coli*-preferred codon[13] was discussed in the last chapter. In addition to codon choice, other factors can be controlled to enhance transcription, translation, and the stability of an expressed protein.

A promoter is a stretch of DNA sequence to which RNA polymerase binds and initiates transcription.[14] This sequence precedes the structural gene of a protein and is not itself transcribed. The rate of transcription is

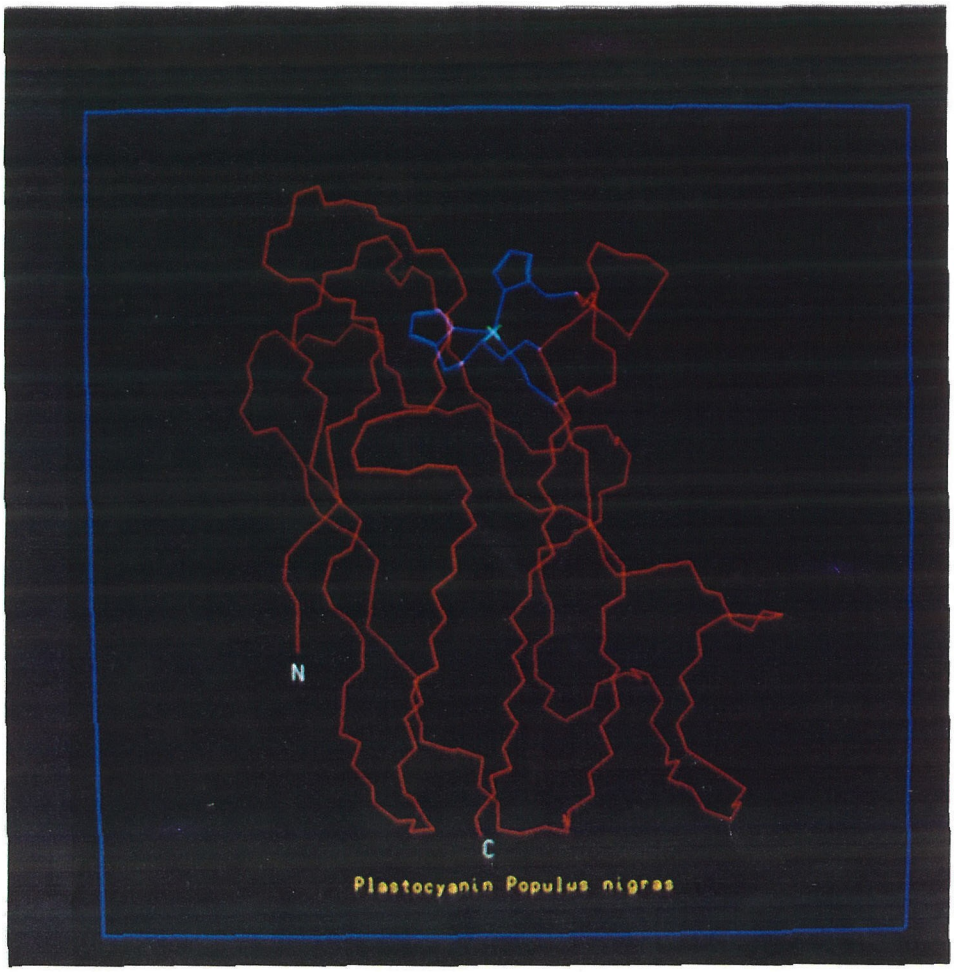
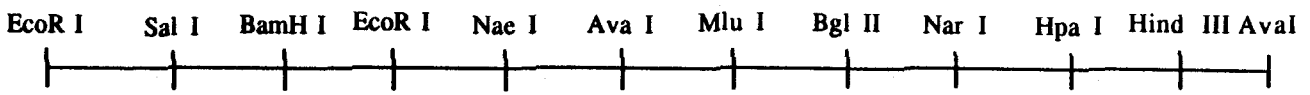


Figure 1. Structure of *Populus nigra* var *italica* plastocyanin. The copper center is highlighted. Note that the protein has all β-sheet structure.

Plastocyanin Gene Construction



a. EcoR I*-Sal I
b. Sal I-BamH I
c. BamH I-EcoR I
d. EcoR I-Nae I
e. Nae I-Ava I
f. Ava I-Mlu I
g. Mlu I-Bgl II
h. Bgl II-Nar I
i. Nar I-Hpa I
j. Hpa I-Hind III
k. Hind III-Ava I*

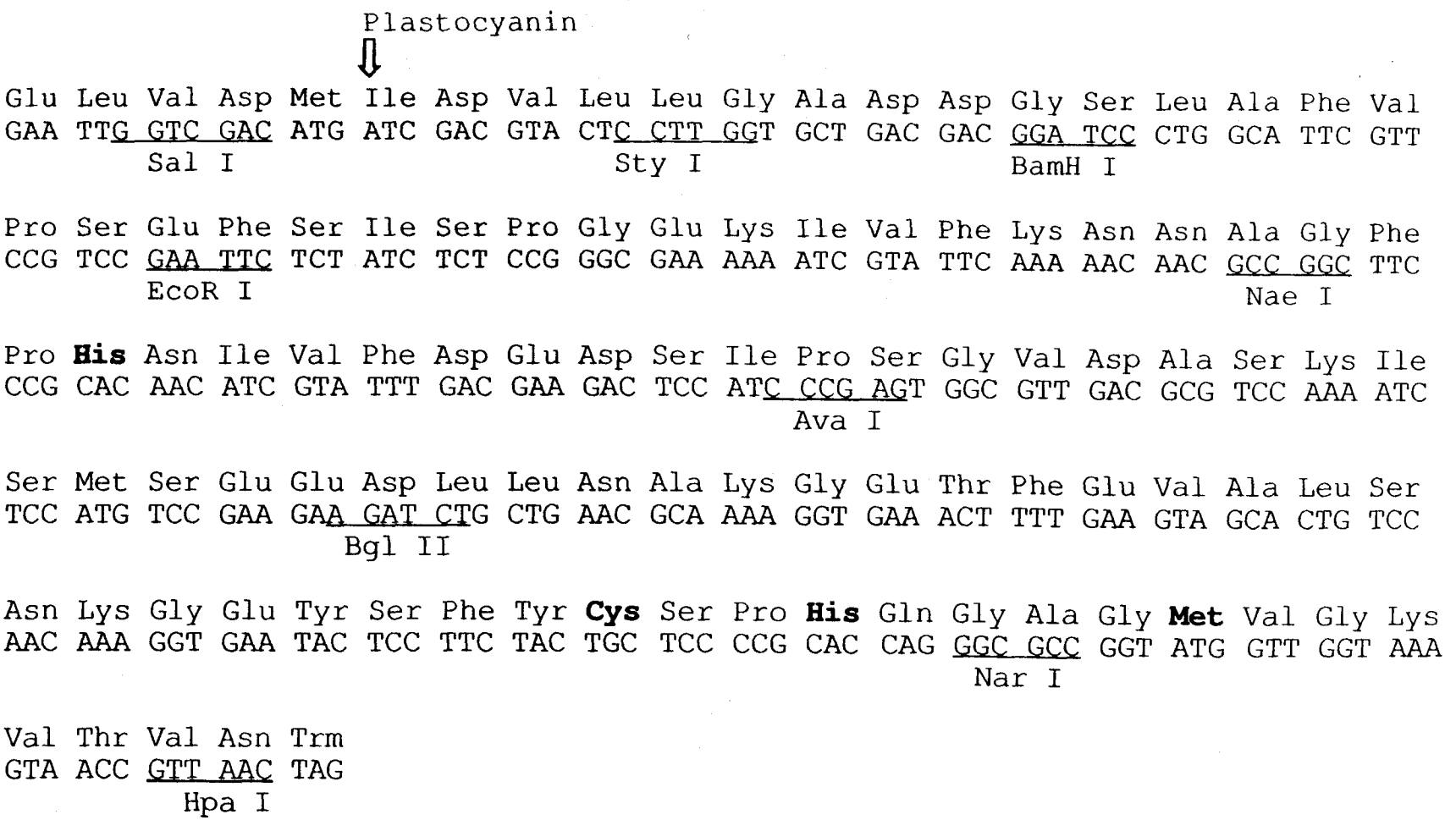
Step 1: a-c, k
Step 2: d, e
Step 3: f, g, i, j
Step 4: h

*The EcoR I and Ava I sites of pBR322 between which this gene was constructed were destroyed during ligation as to make the EcoR I and Ava I sites within the plastocyanin gene unique on the entire plasmid.

Figure 2 Overview of the plastocyanin gene synthesis. This gene was completed in 4 steps.

Figure 3

The DNA sequence of the plastocyanin gene located between the unique EcoR I and Ava I sites of pBR322. This plasmid is labeled pLASTO.


Plastocyanin
Glu Leu Val Asp Met Ile Asp Val Leu Leu Gly Ala Asp Asp Gly Ser Leu Ala Phe Val
GAA TTG GTC GAC ATG ATC GAC GTA CTC CTT GGT GCT GAC GAC GGA TCC CTG GCA TTC GTT
Sal I
Sty I
BamH I
Pro Ser Glu Phe Ser Ile Ser Pro Gly Glu Lys Ile Val Phe Lys Asn Asn Ala Gly Phe
CCG TCC GAA TTC TCT ATC TCT CCG GGC GAA AAA ATC GTA TTC AAA AAC AAC GCC GGC TTC
EcoR I
Nae I
Pro His Asn Ile Val Phe Asp Glu Asp Ser Ile Pro Ser Gly Val Asp Ala Ser Lys Ile
CCG CAC AAC ATC GTA TTT GAC GAA GAC TCC ATC CCG AGT GGC GTT GAC GCG TCC AAA ATC
Ava I
Ser Met Ser Glu Glu Asp Leu Leu Asn Ala Lys Gly Glu Thr Phe Glu Val Ala Leu Ser
TCC ATG TCC GAA GAA GAT CTG CTG AAC GCA AAA GGT GAA ACT TTT GAA GTA GCA CTG TCC
Bgl II
Asn Lys Gly Glu Tyr Ser Phe Tyr Cys Ser Pro His Gln Gly Ala Gly Met Val Gly Lys
AAC AAA GGT GAA TAC TCC TTC TAC TGC TCC CCG CAC CAG GGC GCC GGT ATG GTT GGT AAA
Nar I
Val Thr Val Asn Trm
GTA ACC GTT AAC TAG
Hpa I

influenced by the sequence of a promoter. Thus, promoter is perhaps the most important part of an expression system. Although the sequences of various *E. coli* promoters vary,[15] there are two highly conserved regions upstream of the transcriptional initiation site. These are the -35 region (TTGACA) and the -10 region (TATAAT) whose consensus sequences are shown in parentheses.[16] Mutations that lead to deviations from these consensus sequences tend to diminish the level of expression. Because of their importance in the transcriptional mechanism, several promoters have been characterized and are now widely used to construct expression vectors. These examples include the *lac*, *tac*, λP_R, λP_L, alkaline phosphatase (*Pho*A), and the T7 promoters.[17-24] Each has various advantages such as some promoters being able to initiate transcription better than others. Following the promoter, a Shine-Dalgano sequence or ribosomal binding site (RBS),[25] to which ribosome can bind, is needed to initiate translation. The consensus sequence for the ribosomal binding site is AGGAGG. Finally, promoter-RNA polymerase interaction is very much influenced by methylation of the promoter as well as the spacing between a promoter and its ribosomal binding site; 17 bp are the optimal spacing between these two regions, while the optimal spacing between the RBS and the initiation codon is 9 ± 3 bases.[26] For these reasons, transcription is believed to be dependent not only on the sequence of DNA but also on its conformation.

Despite its elegant name, a promoter is not a specialized DNA; it merely functions as a sequence of DNA to which RNA polymerase can bind stably. Thus, it is possible that the enzyme can bind to other portions of a plasmid that was not intended to be a promoter. For example, during the synthesis of the plastocyanin gene, the DNA segment preceding the EcoR I

construction site was found to contain a sequence that may have been acting as the -35 and -10 regions (ATGACA and TATCACGA) and a possible ribosomal binding site.[4] If the expression of plastocyanin was particularly deleterious to the host *E. coli* cells, the existence of such a possible promoter would explain the mutations that introduced a stop codon in both azurin and plastocyanin genes during the synthesis, as well as the low number of transforming *E. coli* cells observed during plastocyanin gene construction. These preliminary observations suggested that expressing plastocyanin was harmful to *E. coli* cells, and this pointed to the use of repressible promoters. The use of a repressible promoter would allow one to turn on the protein expression mechanism only when the optimal cell density is reached. There are several different mechanisms by which a promoter can be regulated.

One of the most common ways to repress a gene expression is through the lac repressor. An *E. coli* strain that has the lac Iq gene overproduces the lac repressor. Both the *lac* UV promoter and its derivative the *tac* promoter are among the promoters that are repressed by the lac repressor. These promoters are repressed until the addition of IPTG, at which point they are turned on. One drawback with the *lac* and the *tac* promoters is that the repression is not very tightly controlled; there is some level of protein expression even in the absence of IPTG. Two other common promoters are the λP_R and the λP_L promoters, which are repressed by a temperature-sensitive cI858 repressor. When a plasmid containing one of these promoters is harbored in an *E. coli* host possessing a cI858 gene, the repressor binds to the promoter, and there is no transcription. However, when the temperature is raised to 42°C, the repressor loses its binding ability and transcription takes place. The regulation with these promoters

is very tight; at 28°C, there is almost no transcription. The *Pho*A, on the other hand, is a promoter that is regulated by starvation. To be more specific, the *Pho*A is functional only when there is a low concentration of inorganic phosphate in the media. Finally, the T7 promoter is designed to give a high yield of protein, for when the T7 polymerase is activated, the gene inserted behind this promoter will be the predominantly expressed gene in the entire cell. But since it so avidly expresses one particular gene, the expression is often lethal to the host *E. coli* cell.

The choice of promoter, however, is not the only factor that determines how well transcription is initiated. The mRNA that is transcribed from DNA must be stable in order to lead to a high level of protein expression. The degradation of mRNA is an essential biochemical event, for an organism needs different proteins at different times. Thus, it must be able to degrade the unwanted mRNA and to synthesize the wanted ones. To this end, *E. coli* possesses polynucleotide phosphorylase and RNAseII, which have 3' -> 5' exonuclease activity.[27,28] Therefore, one way to stabilize an mRNA is to include a region at the 3' end of the structural gene, which can readily form a secondary structure, such as a transcriptional terminator. Such a stop will keep the length of the mRNA from being longer than it needs to be, which taxes the energetics of *E. coli* biosynthesis machinery and also can destabilize the mRNA. Many vectors constructed for expression purpose contain such a transcriptional stop; for example, the plasmid pKK223-3 contains the T1 and T2 termination stops.[29] In addition, there is some evidence that the formation of a secondary structure at the 5' end may contribute to mRNA stability; mutation that eliminated some likely base-pairing at the 5' end destabilizes mRNA *in vitro*.[30]

Most highly used expression vectors combine a regulated promoter, sequence to stabilize mRNA, a selectable marker, and a multiple-cloning site. But proteins must also be stable once it is expressed. Many proteins, especially partially folded proteins or proteins not native to the host, are readily degraded by *E. coli* proteases,[31] especially during shortage of amino acids[32] or when the host cells are heat-shocked.[33] For example, the *La* protease[34] is a major group of *E. coli* protease active in degrading foreign proteins whose intracellular concentration doubles at temperatures above 42°C. Consequently, using temperature-induced promoters may increase the risk of having the desired proteins degraded. Degradation problem is more severe for smaller proteins (<100 amino acids). In these cases, fusing the desired protein to a larger, readily expressed protein can stabilize the small proteins. Moreover, it has been shown that plastocyanin without Cu^{2+} is degraded rapidly in *Chlaydomonas reinhardtii,*[35] an observation that has led us to use Cu^{2+} in the growth media. Finally, some proteins require naturally present pro sequence, signal sequence, or the membrane-translocation event triggered by signal sequence for proper folding. Without them, these proteins are left unfolded and become a prime target for protease degradation.

All of these factors mentioned above were considered in our attempts to express plastocyanin. Various combinations of expression vectors with *E. coli* hosts were used. As is evident in Figure 1, plastocyanin consists of all β-sheet. This was potentially another difficulty, for most of the proteins that refold easily are made mostly of α-helices.

Materials and Methods

Materials

The expression vectors that were used to express plastocyanin were either obtained as gifts or purchased from a commercial supplier. The plasmid pBR322 was purchased from Bethesda Research Laboratories. The plasmid pLcII was obtained as a gift from K. Nagai at the Medical Research Council Centre. The plasmid PCFM636 was obtained from S. Suggs at Amgen(Thousand Oaks, California). The plasmid pTRAP was obtained as a gift from C. Craik at U.C. San Francisco. The expression vectors pRIT2T, pKK223-3, and pUC18 were purchased from Pharmacia. Two different sources of plastocyanin were used to raise anti-plastocyanin sera. The French bean plastocyanin was obtained from H. B. Gray. This was used to screen most plastocyanin expression constructs. Later, the poplar plastocyanin was obtained as a kind gift from H. Freeman(University of Sydney, Australia). This was sent to Berkeley Antibody Co. (Berkeley, California) which raised the antibody.

Restriction enzymes, polynucleotide kinase, T4 DNA ligase, and calf-intestinal alkaline phosphatase were purchased from the suppliers mentioned in Chapter 2 of this thesis. Likewise, general buffer and chemical reagents, as well as other enzymes and radioactive nucleotides were purchased from the same companies. Bovine and porcine enterokinases and their fluorogenic synthetic substrate, Gly-Asp-Asp-Asp-Asp-Lys-β-naphthylamide, were purchased from Sigma Chemical Company. Factor Xa, endopeptidase Arg-C, kallikrein, and trypsin were purchased from Boehringer-Mannheim Biochemicals. The vectastain ABC kit from Vector Laboratories (Burlingame, California) was used to develop

Western blots. DE-52 for anion-exchange chromatography was purchased from Whatman Inc. IgG-sepharose from Pharmacia was used for the affinity chromatography of the protein A-plastocyanin-fusion protein.

In addition to previously described instrumentations, protein gels were run on BRL horizontal gel electrophoresis boxes. For Western blots, proteins bands were transferred to DEAE nitrocellulose sheets (0.45 μ, purchased from Schleicher and Schuell) on a Bio-Rad Trans blotter equipped with surface electrodes. FPLC system and the columns from Pharmacia were used for protein purification. Enterokinase activity was measured on a Perkin-Elmer fluorimeter.

Methods

DNA manipulation

Oligonucleotides were synthesized on an Applied Biosystem's Model 380A DNA synthesizer and purified on a polyacrylamide gel as described in Chapter 2. The purified oligonucleotides were phosphorylated at the 5' terminus, using polynucleotide kinase. Equimolar amounts of the two oligomers were annealed together by placing them in boiling water and letting the temperature drop gradually. At the same time, the plasmid into which the oligonucleotides were to be added was cut with appropriate restriction enzymes, purified on an agarose gel, and recovered using the Elutrap. The amount and the purity of the recovered DNA were estimated on the basis of its UV absorbance. The cleaved plasmid was lyophilized and resuspended in water to a final concentration of 0.01 pmole/μl. The annealed double-stranded cassette was inserted into the opening within the plasmid, using T4 DNA ligase. A typical ligation condition used the ratio of

0.4 pmole of inserts to 0.04 pmole of vector in 25 µl of ligase buffer. For blunt end ligation, a ratio of 0.75 pmole of inserts to 0.05 pmole of vector was used.

Transformation

The Hanahan protocol described in Chapter 2 was used to transform plasmids into *E. coli*. However, when the *E. coli* cells were transformed with an expression vector that utilized heat induction, the cells were grown at 28°C for at least 2 hours after the heat shock step of transformation.

Plasmid Analysis

Initial screening for successful ligation into an expression vector was performed by restriction digest with the sites found on the plastocyanin gene. The colonies that contained these sites were then cleaved with two restriction sites so that the size of the fragment thus produced can be compared to the expected size of the fragment with and without the plastocyanin gene correctly inserted.

SDS-PAGE Electrophoresis

Protein bands were separated and visualized using SDS-PAGE electrophoresis. SDS-loading buffer was added to protein samples to a final composition of 20% v/v. This mixture was placed into a 95°C heat block for 10 minutes to eliminate aggregation and the secondary structure of the sample. This was then loaded onto a 15% SDS polyacrylamide gel poured with a 4% SDS polyacrylamide gel stack on top. The gel was electrophoresed at 10 mA until the bromophenol blue dye entered the resolving phase, at which point the current was increased to 20-30 mA. When the dye reached the bottom of the gel, the current was stopped and the

stack portion of the gel was cut with a razor blade and discarded. The resolving phase of the gel was either stained with a Coomasie dye or transferred to a nitrocellulose sheet for Western blotting.

For staining with Coomasie blue, the gel was immersed in the fixing solution as to precipitate the protein onto the gel matrix. The gel was then immersed in the Coomasie staining solution and placed in a 65°C bath for 10 minutes. Alternatively, the gel was microwaved for 1 minute. The staining solution was discarded, and the bands on the gel were visualized by destaining in the destaining buffer with frequent solution changes over several hours. The gel was preserved by being immersed in a 5% glycerol solution and wrapped in plastic wrap.

Buffers and Solution

Lower Buffer (4x recipe): 1.5 M Tris, 0.4% SDS, final pH 8.8

Upper Buffer (4x recipe): 0.5 M Tris, 0.4% SDS, final pH 6.8

SDS-PAGE running buffer: 25 mM Tris, 192 mM glycine, final pH ~8.5

SDS-loading buffer: 10 % v/v glycerol, 3% w/v sodium dodecyl sulfate, 62.5 mM Tris-HCl, pH 6.8, 1 mM EDTA, 0.05% w/v bromophenol blue, and freshly added b-mercaptoethanol to 5 % v/v

Protein fixing solution: 50% v/v methanol, 12%v/v glacial acetic acid

Coomasie blue staining solution: 0.1 % w/v coomasie brilliant blue G dye in destaining solution

Destaining Solution: 25% ethanol, 8% glacial acetic acid in water

Western blot[37,38]

For Western blotting, the cells harboring an expression vector with the plastocyanin gene correctly inserted were grown to appropriate O.D. (different for each vector) in 2 ml cultures and then induced by either IPTG,

temperature change, or other means. A 1.5 ml aliquot of this culture was spun in a microfuge for 10 seconds. The supernatant was discarded and the remaining cell pellet was resuspended in 100 µl of SDS loading buffer. The tube containing the mixture was immersed in a boiling water bath for 5 minutes to achieve cell lysis. The sample was then vortexed vigorously to reduce its viscosity. 20 µl of this mixture were loaded onto a 15% SDS-polyacrylamide gel with a 4 % SDS-polyacrylamide stack phase. The protein bands were separated as described above for a polyacrylamide gel. But instead of fixing the gel and staining it with the staining solution, the gel was immersed in the blotting buffer for 30 minutes to desalt it. Then the gel was placed directly adjacent to a nitrocellulose sheet in a sandwich. Care was taken to avoid any air bubbles, which can leave spots on the transferred sheet. The sandwich was placed into a BioRad trans-blot cell equipped with a pair of surface electrodes. This was run at 100 mA for 1-2 hours. At the conclusion of the transferring process, the nitrocellulose sheet was stained using 1000 fold dilution of the French bean plastocyanin antibody unless stated otherwise. The ABC vectastain kit utilizing goat antirabbit antibody and biotinylated, horseradish peroxidase conjugated system.

Buffers

blotting buffer: 25 mM Tris, 192 mM glycine, 20% v/v methanol

1 M Sodium Phosphate (pH 7.5): 11.04 g monobasic and 59.64 g anhydrous dibasic sodium phosphate in 500 ml of water

Phosphate-buffered saline: 10 mM sodium phosphate, pH 7.5 and 0.9% w/v NaCl

TPBS: phosphate-buffered saline with 0.05% v/v Tween 20

Substrate solution for Western: 30 mg of 4-chloro-1-naphthol dissolved in 10 ml of methanol, 25 µl 30% hydrogen peroxide, and 50 ml TPBS

Purification of protein A-plastocyanin fusion protein

E. coli strain N4830 harboring the plasmid pProAPLS (plastocyanin gene inserted into pRIT2T plasmid) was grown in 1 L of L-broth with 50 µg/ml ampicillin at 30°C. When the cells reached the O.D. of 0.5, the λP_L promoter was induced by increasing the temperature to 42°C. After 2 hours at this temperature, the cells were harvested in a GSA rotor and resuspended in 30 ml of fractionation buffer. The suspended cells were sonicated on ice for 30-second intervals until they were lysed. Upon lysis, there was a change in the color of the cell suspension from brown to creamy white, and the sound of the sonicator shifted to a lower frequency. Usually 3 or 4 cycles were needed to lyse the cells from a 1 L preparation. The lysate was not as viscous as it was during a plasmid preparation, for sonication prevented the cellular DNA from aggregating. The cell debris was removed by spinning the mixture in a SW-28 rotor at 24,000 rpm for 2 hours at 4°C in an ultracentrifuge.

A 20 cm column was packed with 20 ml of IgG-sepharose. Since the binding capacity of IgG-sepharose was 2 mg of protein A per ml of column material, the estimated maximum capacity of this column was 40 mg of protein. Newly poured column was washed with 5x bed volume (100 ml) of TST buffer to remove ethanol in which IgG-sepharose is packed. The column was equilibrated by washing successively with 2-3x bed volume each of elution buffer, TST, elution buffer, and TST. When the lysate was loaded, most of the colored (light brown) material did not bind and passed through the column. The loaded proteins were washed with 10x bed

volume of TST (200 ml) and then with 2x bed volume (50 ml) of washing buffer. The sample was eluted off the column with the elution buffer and collected on a fraction collector. The peaks were identified by monitoring the absorbance at 280 nm. The fractions containing the samples were dialyzed into Tris/Tween 20 buffer and then concentrated, using Amicon ultrafilteration device. The yield was approximately ~9 mg of protein A-plastocyanin fusion protein from 1 L growth.

Buffers and Solutions

Tris-saline Tween 20 buffer (TST): 50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 0.05% v/v Tween 20

Tris-Tween 20 buffer: same as TST except without NaCl

Fractionation buffer: 0.02 M Tris-HCl, pH 8.0, 2 M NaCl, 1 mM PMSF

Elution buffer: 0.5 M CH_3COOH, pH 3.4 adjusted with NH_4Ac

Wash buffer: 5 mM NH_4Ac, pH 5.0

L-broth: 10 g/l bacto trypton, 5 g/l yeast extract, 5 g/l NaCl

Formic Acid Cleavage

Concentrated formic acid was added to a solution containing the fusion protein to a final concentration of 70%. This mixture was incubated at 30°C. After 2 days, the mixture was dialyzed once against Tris-Tween 20 buffer and once against water. The protein sample was then lyophilized, resuspended in 1 ml of Tris-Tween 20 buffer, and loaded onto an SDS-PAGE gel for Western blotting.

Enterokinase Purification

Porcine enterokinase was suspended in 10 mM MES buffer, pH 6, 50 mM NaCl. This suspension was centrifuged in an ultracentrifuge for 45

minutes at 24,000 rpm. This was then dialyzed into the same buffer to remove small-molecular weight impurities and the salts from the enzyme preparation. The enzyme was then passed through a DE-52 anion exchange column, equilibrated with a buffer containing 50 mM Tris, pH 8 and 50 mM NaCl. Enterokinase was eluted, using 50 mM to 500 mM NaCl gradient. The peaks were collected and redialyzed into the 50 mM NaCl buffer. This time, the sample was purified through FPLC with Mono-Q anion exchange column, using the same gradient. The peaks were again collected, dialyzed into the low-salt buffer, and then purifed through FPLC with Superose-12 gel filteration column. The collected fractions were assayed for enterokinase activity on a fluorimeter, using a synthetic substrate Gly-$(Asp)_4$-Lys-β-naphthyl amide.

RESULTS

Plasmid pLcII

pLcII is an expression vector that utilizes the λ P_L promoter and has the ampicillin resistance gene as a selectable marker. This was obtained from Nagai, who used it successfully to express hemoglobin in *E. coli*.[39] Others have used it to express myoglobin[40] and a mouse monoclonal antibody with catalytic activity.[41] The pLcII plasmid was cleaved with Sal I and Hind III and gel-purified. The plastocyanin gene was ligated into this site using a Sal I-BamH I linker that contained the Ile-Glu-Gly-Arg recognition sequence of factor Xa (oligonucleotides 27 and 28, Figure 4).[42,43] The ligation scheme is shown in Figure 5. This plasmid should express plastocyanin as a fusion protein attached to the C-terminal side of an LcII gene product. The correct ligation product was identified using the restriction digest analysis with Bgl II, since only the plasmids that retained a copy of the plastocyanin gene were expected to contain this site. This was then transformed into JA221,cI857 cells, which repressed the promoter at 30°C. The cells were grown to log phase (O.D. of ~0.5) in 2 ml of L-broth containing various concentrations of $CuSO_4$ and 50 μg/ml ampicillin. At this point, the promoter was induced by increasing the temperature to 42°C. Western blot on the whole cell extract containing this construct gave no band.

Plasmid pCFM636

This plasmid shares the same λP_L promoter as the plasmid pLcII. This plasmid also has an engineered-in ribosomal binding site and the kanamycin-resistance gene. This plasmid was constructed at and obtained from Amgen. To prepare the plasmid for ligation, it was cleaved

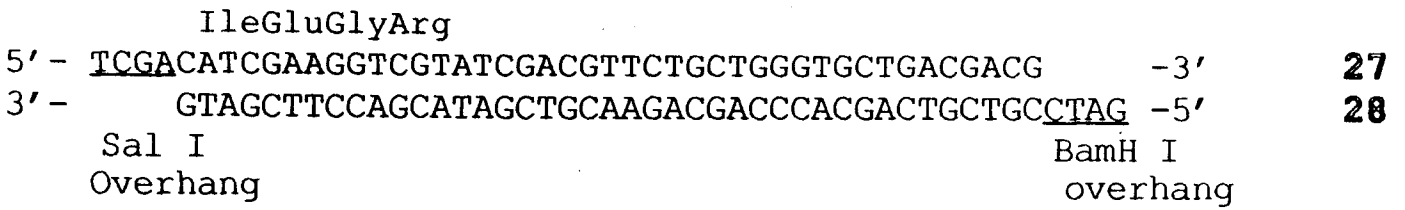


Figure 4. Oligonucleotides used to insert the plastocyanin gene into the expression vector pLcII. The Factor Xa recognition site was built in.

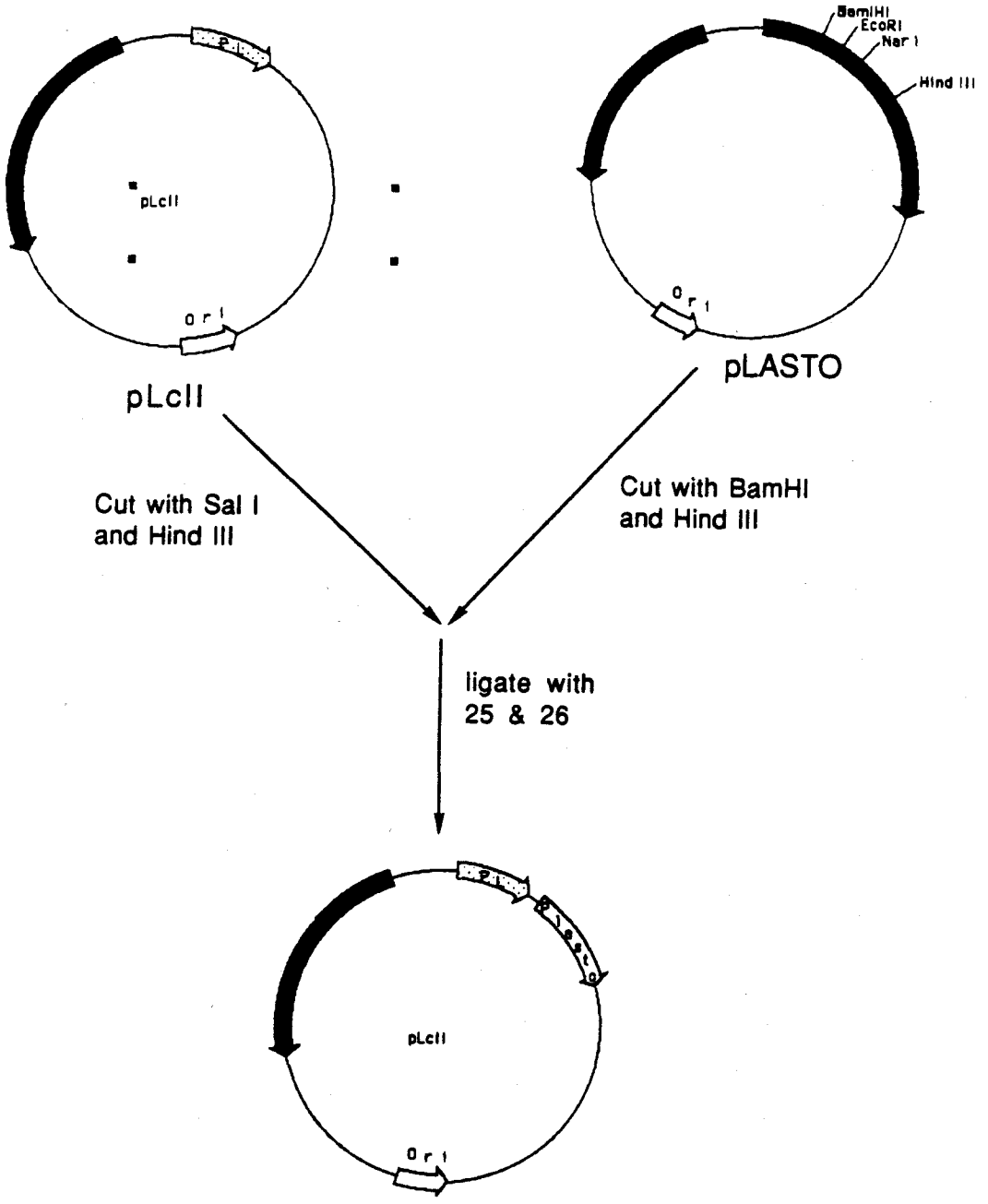


Figure 5. Ligation of plastocyanin gene into pLcII.

with Hind III and Cla I and then purified on a 1.2% agarose gel. The plasmid pLASTO was cleaved with BamH I and Hind III, and the small piece was gel-isolated. This piece was ligated into the cleaved pCFM636 using the annealed oligonucleotides 29 and 30 (Figures 6 and 7). The correctly constructed vector was identified on the basis of restriction digest with Bgl II. This enzyme cuts the parental plasmid twice, but the size of the small fragment should increase if the plastocyanin gene were correctly inserted. A correctly assembled plasmid was transformed into the *E. coli* strain JA221. The cells were grown in 2 ml of L-broth containing a range of $CuSO_4$ concentration. When the cell cultures reached the O.D. of 0.5, the growth temperature was shifted up to 42°C, and the cells were grown for another 2 hours. Western blot gave no 10 Kd band expected of plastocyanin.

Plasmid pTRAP

pTRAP stands for trypsinogen-alkaline phosphatase promoter.[44] This vector utilizes the alkaline phosphatase (*Pho*A) promoter and expresses a protein fused to the alkaline phosphatase signal sequence. To clone plastocyanin into this plasmid, pLASTO was cleaved with EcoR I and Nde I and gel-purified. pTRAP was cleaved with BamH I and Nde I and also gel-purified. The DNA fragment containing the plastocyanin gene was ligated into this site on pTRAP along with an oligonucleotide cassette that contained the factor Xa recognition sequence (Figure 8 and 9). Repeated ligation attempts failed to yield a correctly ligated plasmid.

Plasmid pRIT2T

pRIT2T is an expression vector[45-47] designed to express proteins fused to the IgG binding domain of *Staphylococal aureus* protein A.[48-50]

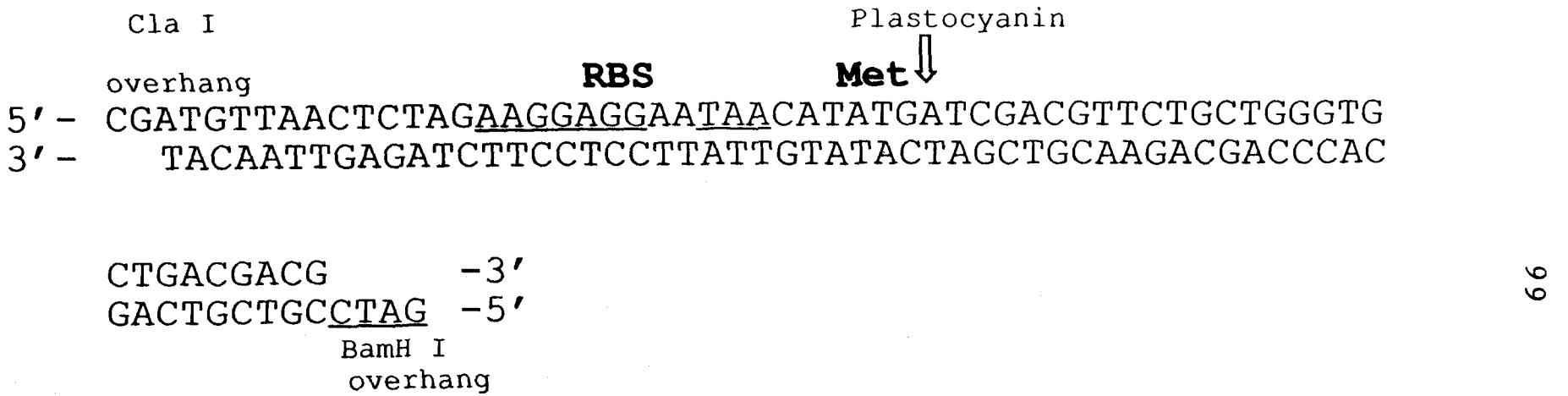


Figure 6. Oligonucleotides used to insert the plastocyanin gene into the expression vector pCFM636. A synthetic ribosomal binding site was designed in. Oligonucleotides #'s 29 and 30.

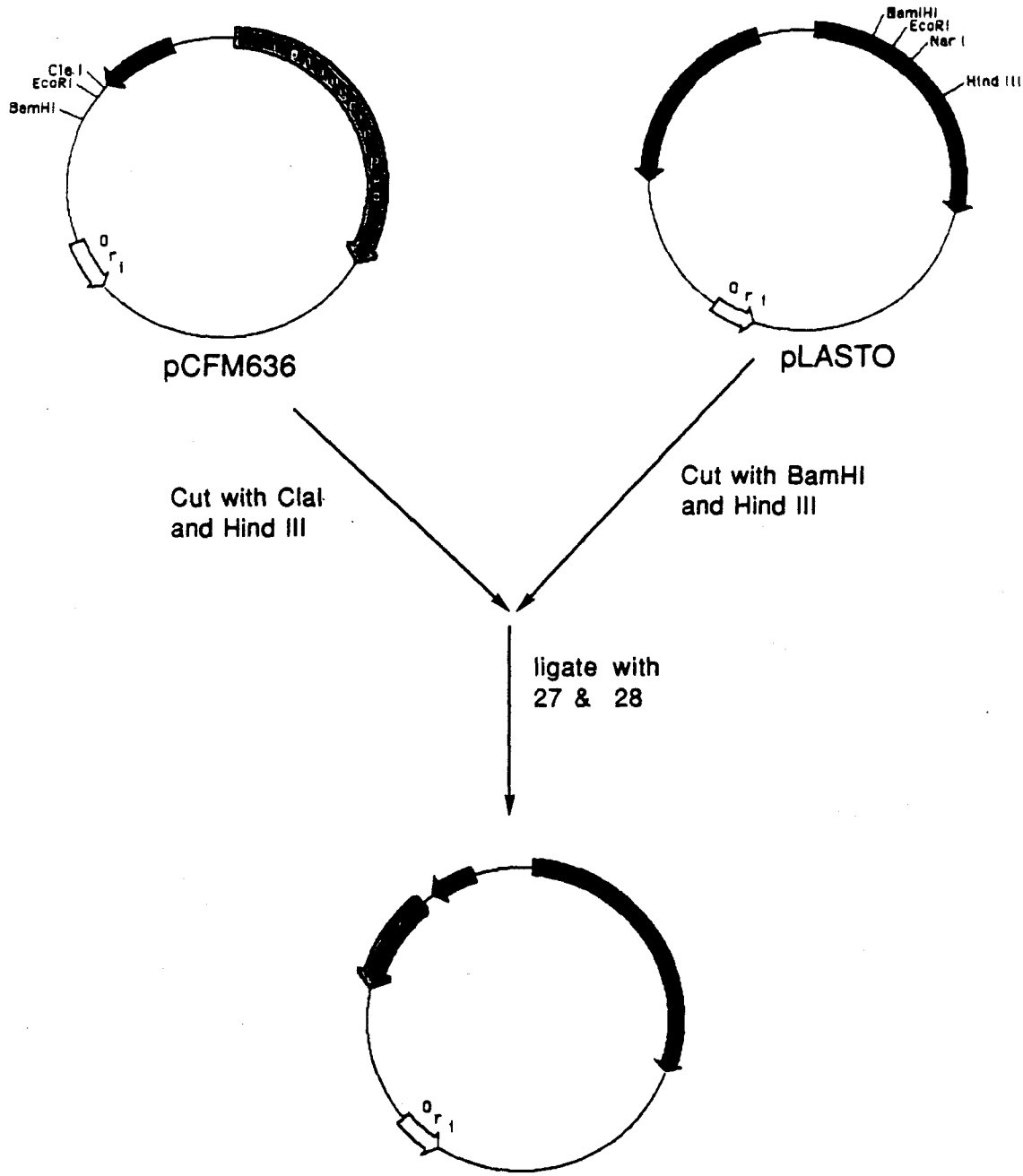


Figure 7. Ligation of plastocyanin gene into pCFM636.

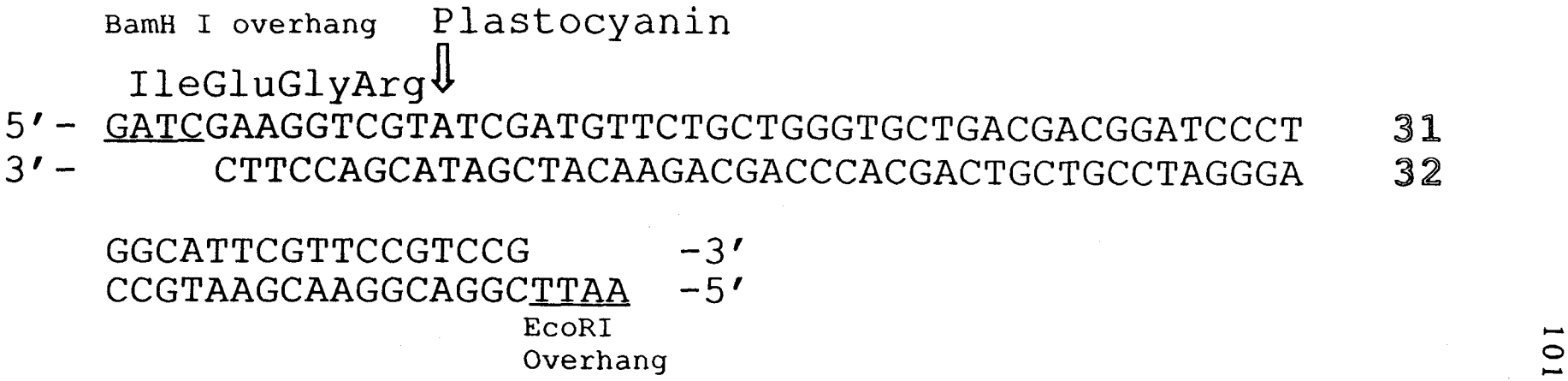


Figure 8. Oligonucleotides used to insert the plastocyanin gene into the expression vector pTRAP. This construct would have fused the alkaline phosphatase signal sequence to plastocyanin with the factor Xa recognition site in between.

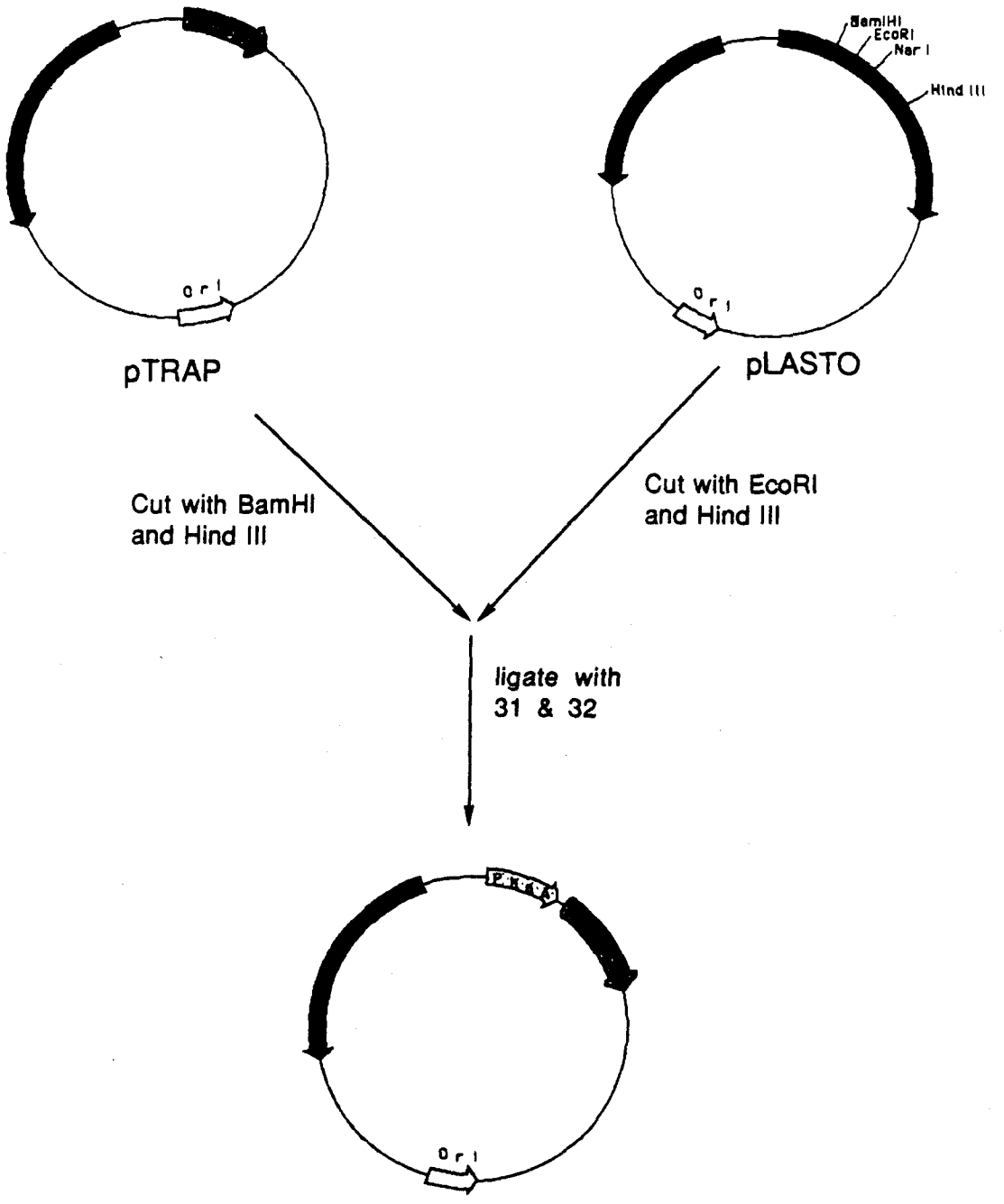


Figure 9. Ligation of the plastocyanin into pTRAP plasmid. This plasmid is under the *PhoA* promoter, and expresses proteins as a fusion protein attached to the alkaline phosphatase signal sequence.

This facilitates the purification because one can use IgG-sepharose affinity chromatography to purify protein A. The plasmid is under the control of the λ P_R promoter and has an ampicillin resistance marker. The ligation of plastocyanin gene into this vector was performed by J. Lin of Clonetech, Inc. (Palo Alto, California), and the successful Western blot is described elsewhere.[4] This original construct (pProAPLS) contained the acid labile sequence of Asp-Pro[51] near the junction between protein A and plastocyanin. To cleave the fusion protein with a more specific protease, the recognition sequences for both the factor Xa (Ile-Glu-Gly-Arg)[39] and the porcine enterokinase (Asp-Asp-Asp-Asp-Lys)[52-54] were inserted at the junction between the two proteins. All three approaches are summarized in Figure 10.

To insert the factor Xa cleavage site, the plasmid pProAPLS was cleaved with Sma I and Bam HI and purified on a 1.2% agarose gel. The cassette containing the factor Xa cleavage site (Figure 11) was ligated into this opening (Figure 12). Because this ligation involved a blunt-end ligation, several ligation attempts were needed before succeeding. The resulting plasmids were initially screened with restriction-digest analysis and then sequenced using the modified Maxam-Gilbert method. For the enterokinase cleavage site, the plasmid pProAPLS was cleaved at the two BamH I sites. Since this produced two identical overhangs, the ends were dephosphorylated to avoid self-ligation. The cassette containing the enterokinase recognition site was ligated into this opening(Figures 13 and 14). The initial screening was done again with restriction-digest analysis. The plasmids were then sequenced to ascertain whether the cassette was inserted in the correct orientation. Figure 15 shows the DNA sequencing gel that shows the factor Xa recognition site.

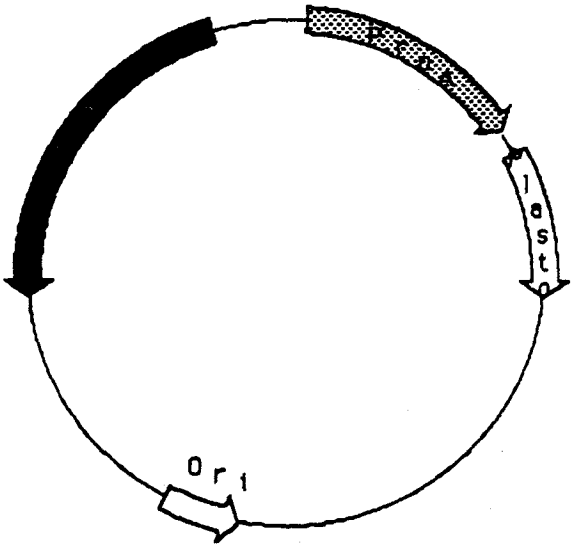


Figure 10. Schematic diagram of pProAPLS.

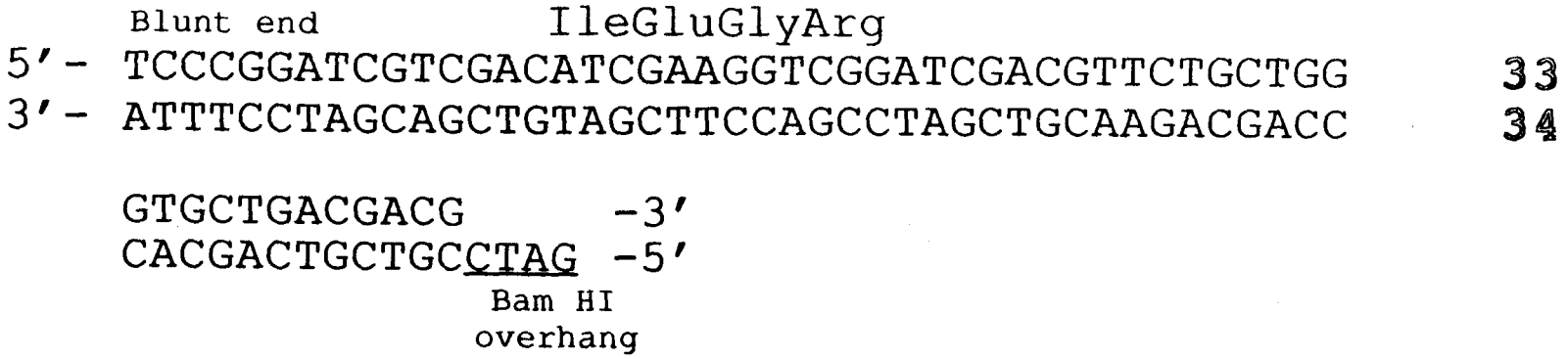
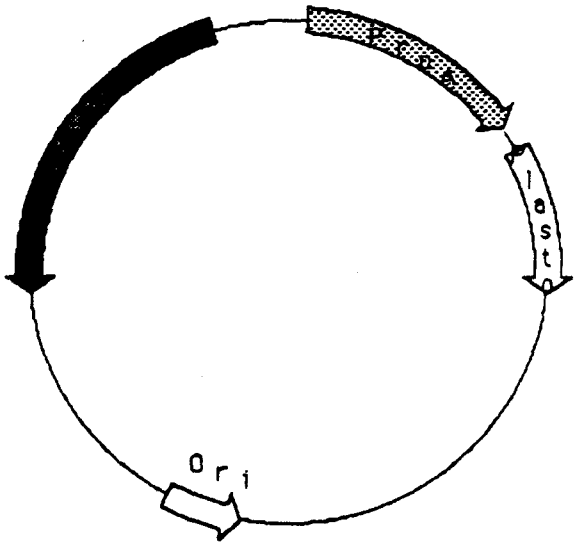


Figure 11. Oligonucleotides to introduce the factor Xa recognition site in front of the plastocyanin gene in the expression vector pProAPLS.



Cut with Sma I
and Bam HI

Ligate with oligonucleotides
33 + 34

Protein A	Xa	Plastocyanin

Figure 12. Factor Xa Ligation. The plasmid pProAPL produces a fusion protein between protein A and plastocyanin. The recognition sequence of the factor Xa has been inserted between the two proteins to facilitate cleavage.

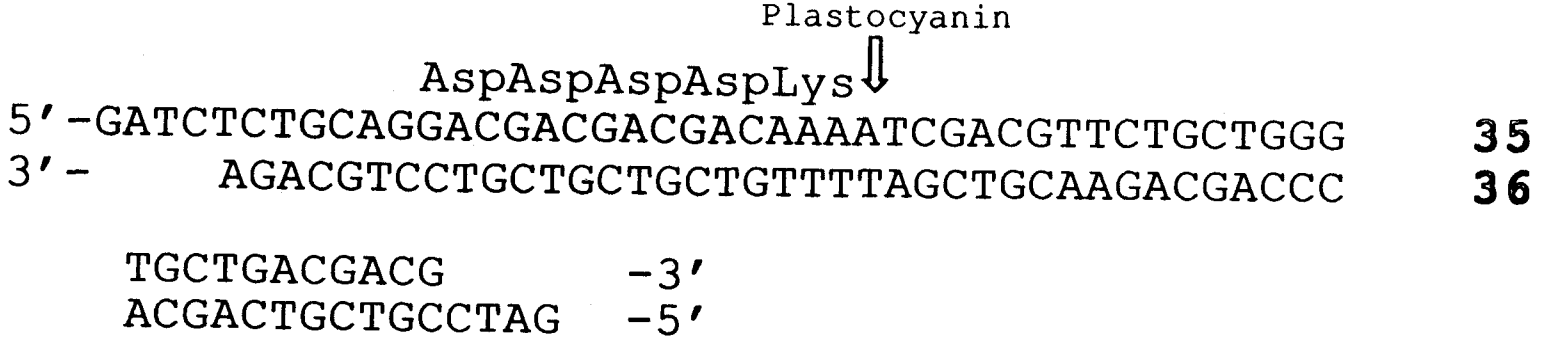


Figure 13. Oligonucleotides to introduce the enterokinase recognition site in front of the plastocyanin gene in the expression vector pProAPLS. The ligation used two BamH I sites.

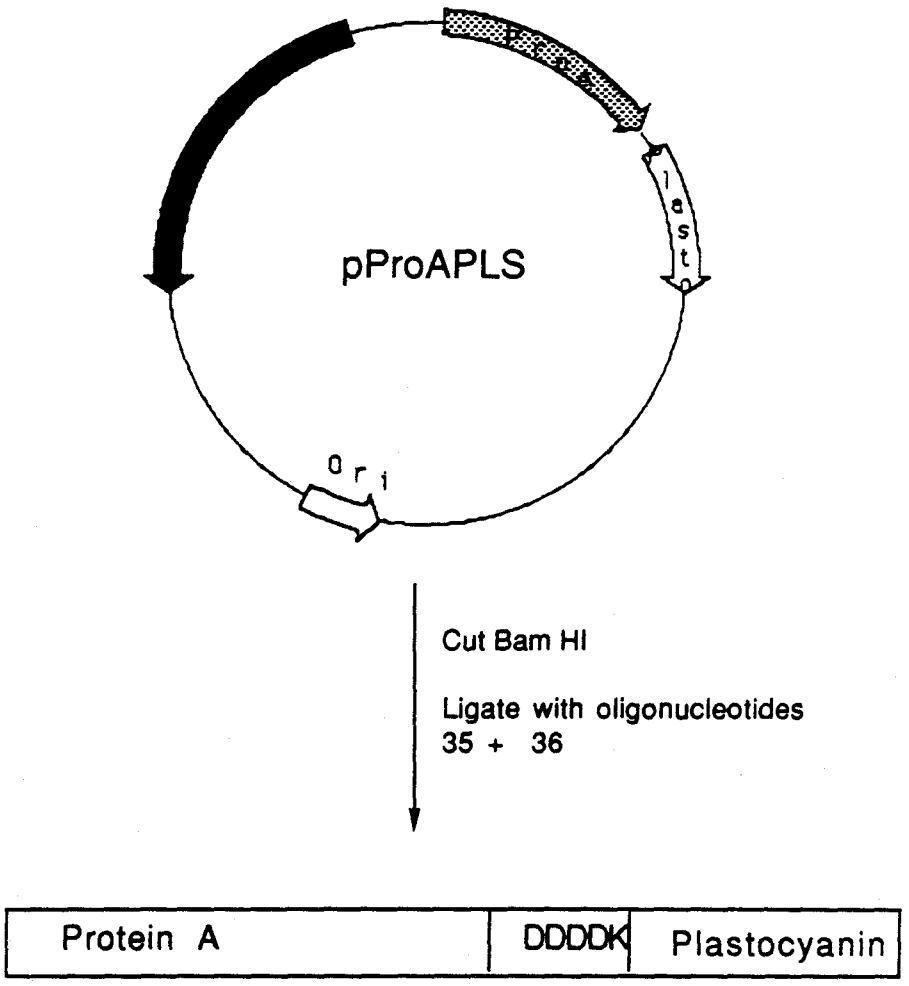
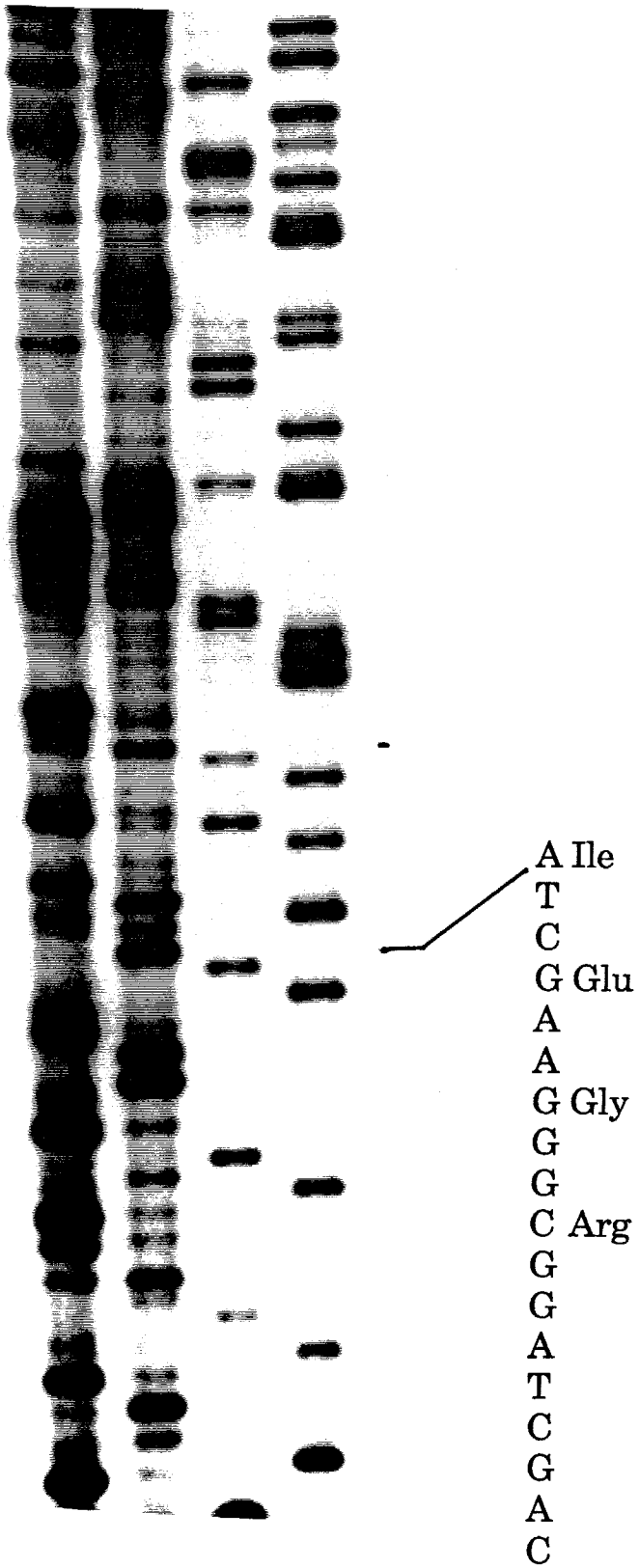


Figure 14. Ligation of the Enterokinase recognition site (DDDDK) into pProAPLS. Since this ligation is performed between two BamH I sites, the ends were dephosphorylated prior to the ligation attempt.

Figure 15

Sequencing gel of the plasmid from the factor Xa ligation. The recognition site of Ile-Glu-Gly-Arg was successfully inserted.


A Ile
T
C
G Glu
A
A
G Gly
G
G
C Arg
G
G
A
T
C
G
A
C

Cleavage of plastocyanin from the fusion protein

The initial construct of plastocyanin in pRIT2T contained the acid labile Asp-Pro bond near the junction between the two proteins. The formic acid treatment was the least desirable one among the three approaches, since treating plastocyanin with very low pH condition was sure to denature it. Furthermore, the Asp-Pro sequence left an overhang of 5 amino acids at the amino terminal of plastocyanin (Figure 10). This method did yield a faint band that could have been plastocyanin around 10 Kd, but the amount of protein was never high nor the results consistent from trial to trial. Thus, other modes of cleavage were considered, some of which are summarized in Table 1.

The use of site-specific protease was more promising. Cleavage with both factor Xa and porcine enterokinase cleaved the 41 Kd fusion protein into a 30 Kd and an 11 Kd fragments, apparently representing protein A and plastocyanin (Figure 16). Unfortunately, the identities of the bands could not be ascertained from Western blots, since both proteins stain upon treatment with antiplastocyanin. This cross-reactivity caused by protein A's affinity for IgG is one of the disadvantages of using this protein as a fusion protein. Several other proteases that may cleave at the factor Xa site were also tried. They were trypsin, endopeptidase Arg-C, thrombin, and kallikrein. Their recognition sites are summarized in Table 1. In most cases, they proved to be a too promiscuous protease to isolate an intact plastocyanin. In fact, one of the difficulties in using enterokinase was that there was considerable trypsin contamination that had to be purified through 3 different columns as detailed in the Methods section. But in the end, using the fusion-protein approach turned out to be laborious,

Figure 16

Protein A-plastocyanin cleaved with (1) formic acid, (3) trypsin, (4 and 5) enterokinase, and (6) factor Xa. Protein A with the MW of 30 Kd is shown in lane 2.

Wester Blot of Digested
Plastocyanin-Protein A Fusion

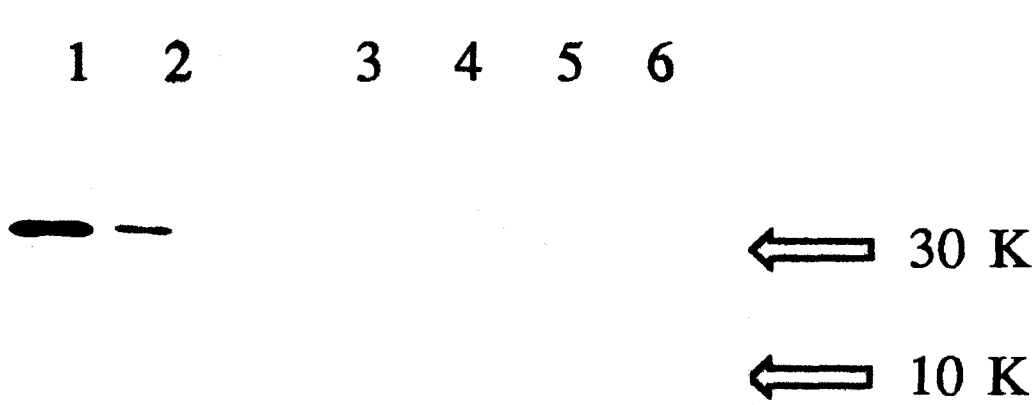


Table 1

Some commonly used site-specific proteolytic reactions

Reagent	Cleavage Site
Cyanogen bromide	Met ⇩
Formic Acid	Asp-Pro ⇩
Hydroxylamine	Asn-Gly ⇩
Enterokinase	Asp-Asp-Asp-Asp-Lys ⇩
Factor Xa	Ile-Glu-Gly-Arg ⇩
Collagenase	Pro-Val-Gly-Pro ⇩
Thrombin	Gly-Pro-Arg ⇩
Trypsin	Arg ⇩ or Lys ⇩
Endopeptidase Arg-C	Arg ⇩
H64A Subtilisin	Gly-Ala-His-Arg ⇩

since both the protease and the fusion protein had to be purified. The cleaved plastocyanin was never in great quantity, and it refused to fold.

Use of Azurin Signal Sequence to express plastocyanin

Since azurin and plastocyanin share a high degree of amino acid sequence homology and structural resemblance, it was decided that the use of azurin signal sequence might transport plastocyanin across the cell membrane into the periplasmid space of *E. coli*. To achieve this, the gene for the azurin signal sequence[55,56] was placed in front of the plastocyanin gene in pBR322. It was then cleaved, isolated, and ligated into both pUC18 and pKK223-3.

The azurin signal sequence ligation was done in two steps. First, an oligonucleotide cassette that encodes for most of the signal sequence was ligated in front of the plastocyanin gene in pBR322 (Figure 17 and 18). This was then transferred to the expression vectors pUC18 and pKK223-3, using appropriate linkers (Figures 19-20). Both constructs gave bands on Western blot stained with antipoplar plastocyanin antibody, with the pUC18 construct resulting in a much stronger band. The molecular weight marker from BRL initially led us to believe that the band that we were staining had the apparent MW of 6 Kd. But the comparison of the expressed plastocyanin to the sample isolated from the poplar tree confirmed that the MW was indeed 11 Kd as expected. Soon after this expression, a similar construct in pUC18 made by other members of this lab also showed a band on the Western. The two constructs yielded almost similar amounts of plastocyanin. Plastocyanin from the second construct was purified to homogeneity and sequenced. The sequence analysis and the UV-Vis absorbance confirmed that plastocyanin was being expressed.

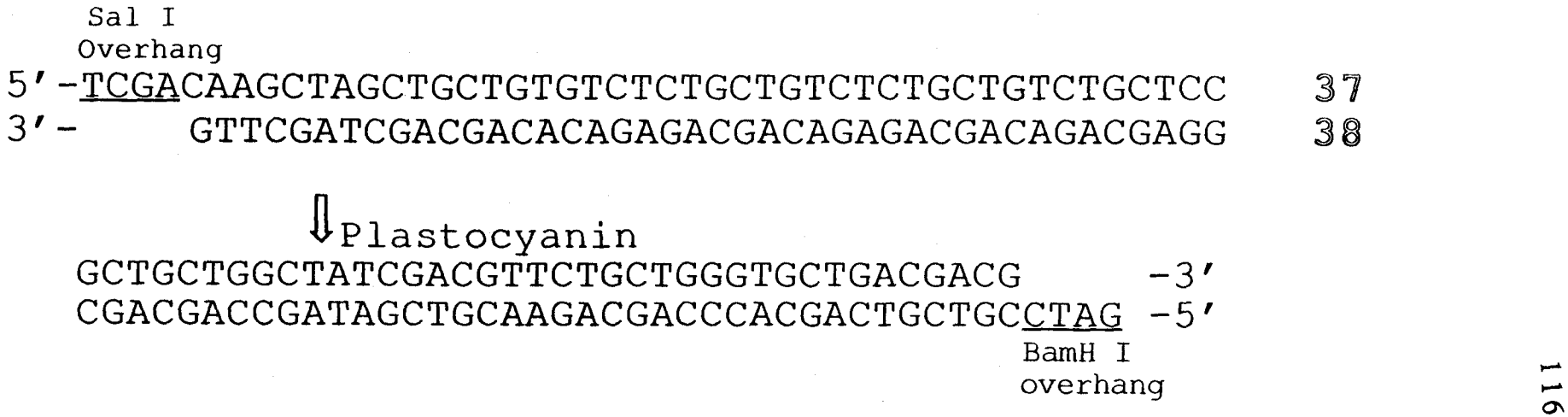


Figure 17. Oligonucleotides to introduce the azurin signal sequence in front of the plastocyanin gene in pLASTO.

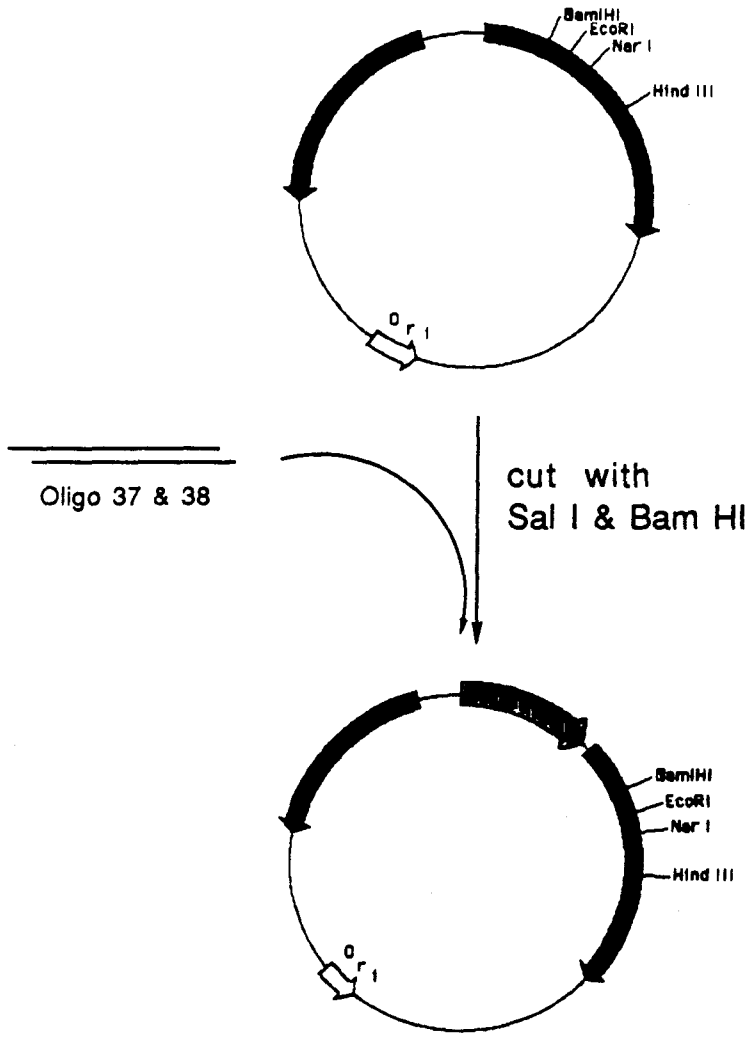


Figure 18. Azurin signal sequence ligation

Sal I
Overhang

```
                              Met
5'-AATTCAAGGAGGAATAACATAATGCTGCGTAAG        -3'     39
3'-     GTTCCTCCTTATTGTATTACGACGCATTCGATC   -5'     40
```

Figure 19. Linker to insert plastocyanin gene into pUC18 or pKK223-3 expression vvectors. The initiation methionine is included in this linker.

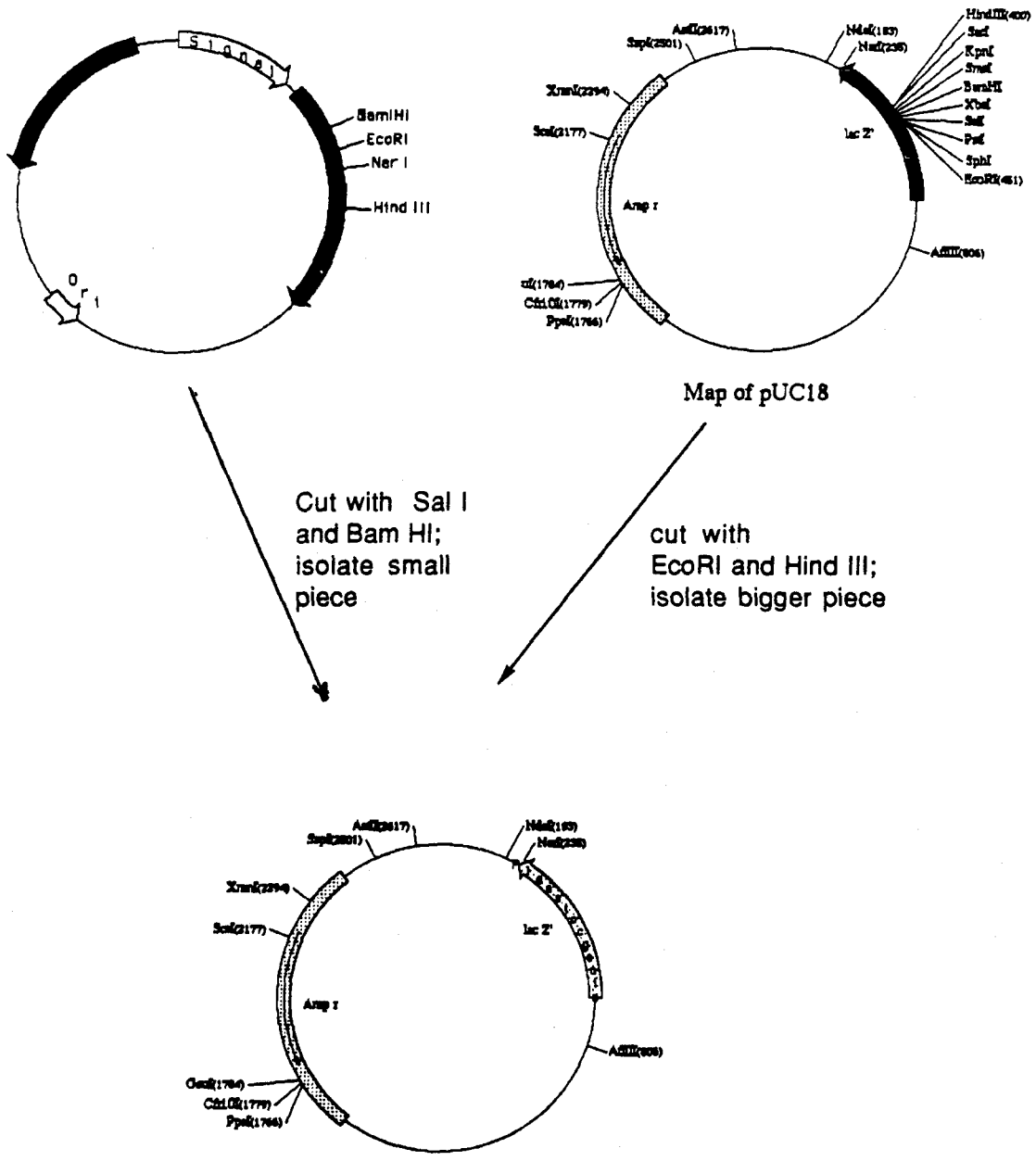


Figure 20. Ligation of the plastocyanin gene into pUC18 and pKK223-3 plasmids. The two ligations were performed the same way.

DISCUSSION

It can be concluded from our numerous unsuccessful attempts to express plastocyanin in the cytoplasm of *E. coli* that plastocyanin is unstable. Apparently, *E. coli* proteases can recognize plastocyanin and can degrade it much faster than one can purify it. When plastocyanin is expressed as a fusion protein with protein A, the latter protein most likely protects plastocyanin from the proteases. On the other hand, the difficulty that we have experienced in cleaving the fusion protein may also be attributed to protein A's protecting of plastocyanin. It is possible that the cleavage site is not accessible to proteases. Another disadvantage of using protein A as the fusion protein is that because of its affinity for IgG, its bands show on Westerns stained with any antibody. Therefore, one cannot distinguish between a real, positive staining and a protein A-antibody interaction. The expression of plastocyanin was concluded on the basis of the MW of the fusion protein's being 41 Kd (30 Kd for protein A and 11 Kd for plastocyanin) and the DNA sequence of the plasmid. The initial use of anti-French bean plastocyanin was not very sensitive. Although the antibody raised against the French bean plastocyanin does cross-react with the poplar plastocyanin, its interaction is much weaker than that observed for the antibody raised against the poplar protein. Western blots performed with the anti-French bean plastocyanin was not as sensitive as the Western blots performed with the antipoplar plastocyanin antibody.

The use of azurin signal sequence, on the other hand, was a fortuitous event. It is interesting that a bacterial-signal sequence can be used to transport a eucaryotic protein across a bacterial-cell membrane. Nevertheless, the level of expression of plastocyanin is far less than that of azurin using the same system. This may be due to several reasons.

Plastocyanin may still be degraded by bacterial proteases in the periplasm much faster than azurin does; these proteases may be recognizing some structural feature that is recognizable to them but not obvious to us. It may also be that the membrane-translocation process is more efficient for azurin than plastocyanin, so that less plastocyanin is being put into the periplasm. Another possibility is that the signal sequence may not cleave plastocyanin as well as azurin. The preferred cleavage site of a common *E. coli* signal peptidase is between Ala-X-Ala. The first sequence of plastocyanin is isoleucine, while it is alanine in azurin. But when the protein is purified, there is no doubt that *E. coli* is expressing poplar plastocyanin that is identical in every way to the native protein. Thus, this was both the first plastocyanin gene synthesized and the first plastocyanin gene successfully expressed in *E. coli*.

REFERENCES

1. Boulter, D., Haslett, B.G., Peacock, D., Ramshaw, J. A. M. and Scawen, M. D. (1977) in *International Review of Biochemistry, Plant Biochemistry II*, ed. Northcote, D. H. (Univ. Park Press, Baltimore), Vol. 13, pp. 1-40.
2. Guss, J. M. and Freeman, H. C. (1983) *J. Mol. Biol.* **169**, 521-563.
3. Ryden, L. (1984) in *Copper Proteins and Copper Enzymes*, ed. Lontie, R., (CRC Press, Florida), Vol. 1, pp. 157-182.
4. Iverson, S. A. (1988) Ph.D. Thesis, California Institute of Technology, Pasadena, California.
5. Rother, C., Jansen, T., Tyagi, A., Tittgen, J. and Herrmann, R. G. (1986) *Curr. Genet.* **11**, 171-176.
6. Nielsen, P. S. and Gausing, K. (1987) *FEBS Lett.* **225**, 159-162.
7. Smeekens, S. and Weisbeek, P. (1988) *Gene* **65**, 59-69.
8. Detlefsen, D. J., Dichersky, E. and Pecoraro, V. L. (1989) *Nucleic Acids Res.* **17**. 1414.
9. Last, D. I. and Gray, J. C. (1989) *Plant Mol. Biol.* **12**, 655-666.
10. van der Plas, J., Bovy, A., Kruyt, F., De Vrieze, G., Dassen, E., Klein, B. and Weisbeek, P. (1989) *Mol. Microbiol.* **3**, 275-284.
11. Last, D. I. and Gray, J. C. (1990) *Plant Mol. Biol.* **14**, 229-238.
12. deBoer, H. A. and Kastelein, R. A. (1986) in *Maximizing Gene Expression*, eds. Reznikoff, W. and Gold, L. (Butterworth, Boston), pp. 225-285.
13. Reznikoff, W. S. and McClure, W. R. (1986) in *Maximizing Gene Expression*, eds. Reznikoff, W. and Gold, L. (Butterworth, Boston), pp. 1-33.

14. Hawley, D. K. and McClure, W. R. (1983) *Nucleic Acids Res.* **11**, 2237-2255.

15. Raib, O. and Schwartz, M. (1984) *Annu. Rev. Genet.* **18**, 178-206.

16. Beckwith, J. (1981) *Cell* **23**, 307-308.

17. deBoer, H. A., Comstock, L. J. and Vasser, J. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 21-25.

18. Amann, E., Brosius, J. and Ptashne, M. (1983) *Gene* **25**, 167-178.

19. Remaut, E., Stranssen, P. and Fiers, W. (1981) *Gene* **15**, 81-93.

20. Oka, T., Sakamoto, S., Miyoshi, K., Fuwa, T., Yoda, K., Yamasaki, M., Tamura, G. and Miyake, T. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7212-7216.

21. Studier, F. W., Rosenberg, A. H., Dunn, J. J. and Dubendorff, J. W. (1990) *Methods in Enzymol.* **185**, 60-89.

22. Denhart, D. T. and Colasanti, J. (1988) in *Vectors: A Survey of Molecular Cloning Vectors and Their Uses*, eds. Rodriguez, R. L. and Denhardt, D. T. (Butterworth, Boston), pp. 179-203.

23 Brosius, J. (1988) in *Vectors: A Survey of Molecular Cloning Vectors and Their Uses*, eds. Rodriguez, R. L. and Denhardt, D. T. (Butterworth, Boston), pp. 205-225.

24. deBoer, H. A. and Hui, A. S. (1990) *Methods in Enzymol.* **185**, 103-114.

25. Stormo, G. D. (1986) in *Maximizing Gene Expression*, eds. Reznikoff, W. and Gold, L. (Butterworth, Boston), pp. 195-224.

26. Simon, M. and Apirion, D. (1972) *Nature (London)* **239**, 79-81.

27. Donovan, W. P. and Kushner, S. R. (1981) *Proc. Natl. Acad. Sci. USA* **83**, 120-124.

28. Brosius, J. and Holy, A. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6429-6933.

29. Shatzman, A. R. and Rosenberg, M. (1987) *Methods in Enzymol.* **152**, 166.

30. Goldberg, A. L. and Goff, S. A. (1986) in *Maximizing Gene Expression*, eds. Reznikoff, W. and Gold, L. (Butterworth, Boston), pp. 207-314.

31. Mandelstam, J. (1960) *Bacteriol. Rev.* **24**, 289-304.

32. Mount, D. W. (1980) *Annu. Rev. Genet.* **14**, 279-294.

33. Chung, C. H. and Goldberg, A. L. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 4931-4935.

34. Merchant, S. and Bogorad, L. (1986) *J. Biol. Chem.* **261**, 15850-15853.

35. Smeekens, S., de Groot, M., van Binsbergen, J. and Weisbeek, P. (1985) *Nature (London)* **317**, 456-458.

36. Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557-580.

37. Vectastain ABC Kit manual, Vector Laboratories, Burlingame, California.

38. Renart, J. and Sandoval, I. V. (1984) *Methods in Enzymol.* **104**, 455-460.

39. Nagai, K. and Thøgerson, H. C. (1984) *Nature (London)* **309**, 810-812.

40. Varadarajan, R., Szabo, A. and Boxer, S. G. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 5681-5684.

41. Baldwin, E. and Schultz, P. G. (1989) *Science*, **245**, 1104-1107.

42. Titani, K., Hermodson, M. A., Fujikawa, K., Ericson, L. H., Walsh, K. A., Neurath, H. and Davie, E. W. (1972) *Biochemistry* **11**, 4899-4903.

43. Fujikawa, K., Legaz, M. E. and Davie, E. W. (1972) *Biochemistry* **11**, 4899-4903.

44. Graf, L., Craik, C. S., Patty, A., Roczniak, S., Fleterrick, R. J. and Rutter, W. J. (1987) *Biochemistry* **26**, 2616-2623.

45. Nilsson, B. and Abrahmsen, L. (1990) *Methods in Enzymol.* **185**, 144-161.

46. Nilsson, B., Abrahmsen, L. and Uhlen, M. (1985) *EMBO J.* **4**, 1075.

47. Abrahmsen, L., Moks, T., Nilsson, B. and Uhlen, M. (1986) *Nucleic Acids Res.* **14**, 7487.

48. Uhlen, M., Guss, B., Nilsson, B., Gaatenbeck, L., Philipson, L. and Lindberg, M. (1984) *J. Biol. Chem.* **259**, 1895.

49. Moks, T., Abrahmsen, L., Nilsson, B., Hellman, U., Sjoquist, J. and Uhlen, M. (1986) *Eur. J. Biochem.* **14**, 7487.

50. Guss, B., Uhlen, M., Nilsson, B., Lindberg, M., Sjoquist, J. and Sjodahl, J. (1984) *Eur. J. Biochem.* **138**, 413.

51. Nilsson, B., Holmgren, E., Josephson, S., Gatenbeck, S., Philipson, L. and Uhlen, M. (1985) *Nucleic Acids Res.* **13**, 1151-1162.

52. Anderson, L. E., Walsh, K. A. and Neurath, H. (1977) *Biochemistry* **16**, 3354-3360.

53. Baratti, J., Maroux, S., Louvard, D. and Desnuelle, P. (1973) *Biochim. Biophys. Acta* **315**, 147-161.

54. Maroux, S., Baratti, J. and Desnuelle, P. (1971) *J. Biol. Chem.* **246**, 5031-5039.

55. Canters, G. W. (1987) *FEBS Lett.* **212**, 168-172.

56. Arvidsson, R. H. A., Nordling, M. and Lundberg, L. G. (1989) *Eur. J. Biochem.* **179**, 195-200.

57. Wolfe, P. B., Wickner, W. and Goodman, J. M. (1983) *J. Biol. Chem.* **258**, 12073.

58. Antonowicz, I., Hesford, F. J., Green, J. R., Grogg, P. and Hadorn, B. (1980) *Clinica Chimica Acta* **101**, 69-76.

Chapter 4

Expression of Azurin

Introduction

Many proteins are synthesized in one compartment of a living cell and secreted into another compartment. These proteins are synthesized as preproteins containing a short, amino-terminal extension called the signal or the leader sequence.[1] The signal sequence is usually cleaved by a membrane-bound signal peptidase upon membrane translocation. The known bacterial-signal sequences[2] show considerable variations in their primary sequences but share structural similarities;[3-5] they are typically 15 to 30 amino acids in length and composed of a positively charged amino terminal and a hydrophobic core. E. T. Kaiser has shown that the propensity to form an α-helical structure is important in their function.[6] In addition, the carboxyl terminal is important for proteolytic processing.[7] Finally, proper interaction between a signal sequence and the mature protein is needed for proper secretion; J. R. Knowles' fusion between the β-lactamase signal sequence gene and the triose phosphate isomerase gene failed to generate secreted protein.[8]

There are numerous advantages to expressing a protein using a secretory scheme. First, it removes an expressed protein from the cytoplasm, which contains many proteases that will consequently lower the amount of protein recovered. The periplasm, on the other hand, has fewer proteases. Second, proteins may not be able to fold correctly in the cytoplasm. Some proteins need the energy from the membrane-translocation process to fold. In addition, the disulfide bond may not form inside the cytoplasm, which is a very much reducing environment, hindering proteins from folding correctly even further. Third, the removal of the signal sequence by a membrane-bound signal peptidase would

conveniently yield the correct amino acid residue at the amino terminal. Fourth, it facilitates purification because there are fewer contaminating proteins present in the periplasm.[9,10]

When the natural DNA sequence of azurin was published,[11,12] the presence of the signal sequence indicated that azurin is a naturally secreted protein. Expressing azurin with its signal sequence was an attractive approach, since at the time of its gene completion, plastocyanin could be expressed only as a fusion protein.[13] Thus, a gene that codes for the 19 amino acids of the signal sequence was introduced in front of the azurin gene. In addition to the signal sequence, a ribosomal binding site had to be created as well.

The exact mechanism by which the translation takes place has not yet been definitively established. However, there are some general rules that seem to make the process more efficient.[14] For example, at least four bases from the consensus sequence of AGGAGG are needed to function as the ribosomal binding site. Among several possible initiation codons, ATG is the most preferred. The spacing between regions is also important. The optimal spacing is 9 bp between the RBS and the initiation codon, and A's and T's are favored within this spacer region. These rules were generally followed in synthesizing the azurin-expression plasmids.

Materials and Methods

Materials

E. coli strain TG-1[15] was purchased from Amersham as a lyophilized pellet. The culture of *Pseudomonas aeruginosa* (ATCC 19429) used for azurin preparation was obtained from W. R. Ellis, Jr. The expression vectors pUC18 and pKK223-3 were purchased from Pharmacia.[16]

Most restriction enzymes were purchased from Boehringer-Mannheim or New England Biolabs. An *E. coli*-expressed polynucleotide kinase was purchased from New England Biolabs, while the T4 DNA ligase was purchased from Bethesda Research Labs. Molecular biology-grade calf-intestinal alkaline phosphatase was purchased from Boehringer-Mannheim. The Vectastain ABC kit from Vector Laboratories was used to develop Western blots. The biotinylated goat-antirabbit serum from the kit was sometimes substituted with the same product from Bethesda Research labs. Sodium salt of ampicillin was purchased from Sigma, and IPTG was purchased from Boehringer-Mannheim. For conventional chromatography, Econo columns (Bio-Rad) were used. CM-52 and DE-52, were purchased from Whatman, Sephadex G-75 from Sigma, and CM-sepharose from Pharmacia. All FPLC columns were supplied by Pharmacia.

UV-Vis spectra were taken on Hewlett-Packard Model 8542 Diode Array Spectrophometer interphased with an IBM PC-AT computer for data collection. Peptide sequencing was performed on an Applied Biosystem's automated sequencer. Large-scale growths (10 L) were performed on a New Brunswick Magnagerm fermentor.

Methods

DNA manipulation

Oligonucleotides were purified on a 15% polyacrylamide gel and purified through NACS columns as described in Chapter 2. Plasmids were cleaved open with two restriction enzymes, dephosphorylated, and purified on an agarose gel as before. The ligations generally used 1:10 ratio of vector to insert and were performed at 15°C overnight.

Transformation

Hanahan's transformation protocol was followed to transform *E. coli* with plasmids as described in Chapter 2.[17]

Azurin Purification from *P. aeruginosa*[18,19]

A 6-liter culture of *P. aeruginosa* was grown anaerobically (no agitation), and this was used to inoculate 350 L of *Pseudomonas* media in a Stainless and Steel Products fermentor. The cells were again grown anaerobically for 2 days at 37°C and harvested on a Sharples continuous-flow centrifuge. The yield of azurin would be reduced if *P. aeruginosa* were grown aerobically.

Of the obtained cell paste, 300 g were resuspended in 2.6 L of acetone/dry ice bath and stirred using an overhead mechanical stirrer. After 1 hour, the cells were filtered on a Whatman #1 filter paper and then resuspended in 1.5 L of acetone. This was stirred for another 1.5 hours and filtered as before. The cells were washed with acetone and placed in a vacuum dessicator and dried for an hour to remove residual acetone. 66 g of acetone powder of *P. aeruginosa* were obtained.

Half of the acetone powder (33 g) was redissolved in 330 ml of 0.1 M NH_3Ac buffer (pH 6.5) preheated to 45°C. The cell suspension was blended in a blender and stirred for 25 minutes. 1.5 mg of DNAse I (from Sigma, at the activity of 300-600 U/mg) were added to the mixture and stirred overnight. This mixture was centrifuged in a GSA rotor at 10,000 rpm for an hour to remove the cell debris. It was spun again in an SS-34 rotor at 20,000 rpm for an additional hour. The supernatant, which was golden brown in color, was dialyzed into 50 mM NH_3Ac (pH4) at 5°C. The white precipitate that formed during the dialysis was removed by centrifugation. The solution now had a blue/brown color.

The supernatant was loaded onto a CM-52 column equilibrated with the same pH 4 buffer. The brown solution separated into a green band (cytochrome cd), two orange bands (cytochrome c_{551} and c_{151}), and a blue band (azurin). The column was washed with 50 mM NH_3Ac (pH 4) buffer, and azurin was eluted with 50 mM NH_3Ac (pH 4.45) buffer. The collected azurin fractions were dialyzed into 10 mM Tris-HCl (pH 8.7) buffer and purified through a DE-52 column, using a 0.01M to 0.05 M NaCl gradient. The blue band was collected and purified through a Sephadex G-75 column, using 10 mM Tris-HCl (pH 7.0). The yield was 33 mg of azurin.

Buffers and Media

Pseudomonas media (for 1 L): 5 g sodium citrate, 5 g $NaNO_3$, 1 g KH_2PO_4, 0.5 g $MgSO_4 \cdot 7H_2O$, 6.5 g yeast extract, 10 mg $FECl_3 \cdot 6H_2O$, 10 mg $CuSO_4 \cdot 5H_2O$, 2 g bacto tryptone

NH_3Ac buffer: ammonium acetate was dissolved to the appropriate concentration, and the pH was adjusted with acetic acid

Azurin Antibody

One mg of azurin was suspended in mineral oil to denature the protein. This was used to inoculate a New Zealand white rabbit. After 1 week, a booster shot of 330 μg was injected into the rabbit, and its serum was collected after waiting another week. This process was repeated two more times, and the serum from the third bleeding was used for subsequent Westerns.

SDS-PAGE and Western

Protein samples containing azurin were run on a 15% SDS polyacrylamide gel with a 4% stack as detailed in Chapter 3. The protein bands were transferred to a DEAE nitrocellulose sheet. The bands were visualized using 1:2000 dilution of the azurin antibody and the Vectastain ABC kit, which uses horseradish peroxide as the active enzyme.

Growth condition and periplasmic extrusion

TG-1 cells harboring pUC18AZ were grown in 2 ml of XB broth containing 50 μg/ml ampicillin. This was used to inoculate 200 ml of XB starter culture. At the same time, the fermentor was prepared with 8 L of XB with 50 μg/ml ampicillin and 1 mM $CuSO_4$. This was inoculated with the starter culture. When the O.D. reached 1.0, IPTG was added to a final concentration of 0.2 mM and induced for 3-5 hours. The cells were harvested in a GSA rotor (6000 rpm for 10 minutes); the cell pellet was resuspended in 1/10 volume (800 ml) of the sucrose solution. After 10 minutes at room temperature, the cells were again pelleted and resuspended in ice-cold 0.1 mM $MgSO_4$. 80 ml of 0.5 M NH_3Ac buffer (pH 4.1) were added along with $CuSO_4$ to a final concentration of 1 mM. The

solution turned green when copper was added. Some of the contaminating proteins precipitated at this pH. The solution was placed on ice for 10 minutes, and the precipitates were removed by centrifuging the solution at 10,000 rpm for 20 minutes. The supernatant containing azurin was filtered once through a 0.22 μ filter before being loaded onto a column.

Buffers and solutions

Sucrose solution: 20% w/v sucrose, 30 mM Tris (pH 8), 1 mM EDTA

IPTG: stock solution of 0.1 M in water

ampicillin: Sodium salt of ampicillin dissolved in water to 25 mg/ml and sterile-filtered

NH_3Ac buffer (various pH): 0.5 M ammonium acetate, pH adjusted with the addition of glacial acetic acid

XB: 25 g/l bacto tryptone, 7.5 g/l yeast extract, 30 mM $MgSO_4$. In later preparation, adding glucose to 10 mg/l helped to increase the final O.D. of the cell growth

Chromatography

The pH of the azurin containing supernatant was adjusted to 4.1 and loaded onto a 5 cm x 10 cm CM-sepharose column equilibrated with 50 mM NH_3Ac buffer, pH 4.1. The blue azurin became visible as it separated from a yellow, contaminating protein. The column was washed with the same buffer, and azurin was eluted with 50 mM NH_3Ac buffer, pH 5.0. The pH of the azurin fractions was readjusted to 4.1, and this solution was loaded onto a Mono S column on FPLC. Azurin was eluted with a linear pH gradient from 4.1 to 9.0 over 30 minutes at the flow rate of 1 ml/min. The fractions containing azurin were again pooled and passed through a Superose 12 gel filtration column using the flow rate of 0.5 ml/min. The yield of azurin

varied slightly from run to run. The maximum yield to date is about 50 mg from an 8 L culture.

RESULTS

Ligation into pUC18 and pKK223-3

The azurin gene was inserted into both pUC18 and pKK223-3. pUC18 has the *lac* UV promoter[20], while pKK223-3 has the *tac* promoter.[21] These two plasmids share the same multiple-cloning site so that these ligations can be done the same way. The azurin gene was cleaved from the pAZ plasmid with Bsm I and Pst I. This was purifed on a 1.5% gel because of its small size. This fragment was then ligated into the opening between the EcoR I and the Pst I sites within the multiple-cloning site of pUC18 and pKK223-3. The oligonucleotide cassette contained compatible overhangs for these sites and the gene that coded for the azurin signal sequence (Figure 1). The ligation scheme is shown in Figure 2. The resulting plasmids were screened by restriction-digest analysis and then sequenced completely. The plasmids that contained the azurin gene were expressed in TG-1 and ran on a Western blot. Both constructs gave a 14-Kd band on Western blots, with the pUC18 construct yielding a much stronger band. A Western blot for the pUC18 constructs is shown in Figure 3. These results showed that azurin was finally expressed.

Recombinant azurin purification and sequence conformation

Despite the appearance of the 14 Kd band on Western blots, the initial yields of azurin were very low. Originally, the TG-1 cells harboring the azurin plasmid were grown in rich media, such as L-broth, containing micromolar amounts of $CuSO_4$. Only less than 1 mg of blue azurin, however, could be isolated from a 10 L culture. There was more apo azurin present, which could not be reconstituted with subsequent additions of

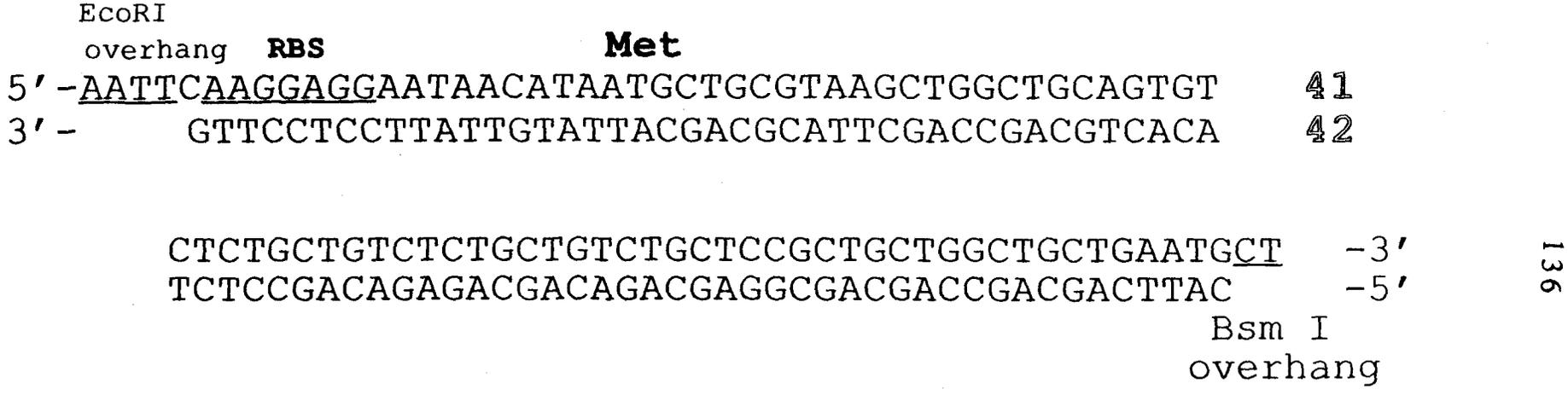


Figure 1. Oligonucleotides which contains a synthetic ribosomal binding site, spacer, and a synthetic gene for the azurin signal sequence. This was used to transfer the azurin gene into pUC18 and pKK223-3.

```
   RBS                  Met Leu Arg Lys Leu Ala Ala Val Ser Leu Leu Ser Leu Leu Ser
CAAGGAGG AATAACATA ATG CTG CGT AAG CTG GCT GCA GTG TCT CTG CTG TCT CTG CTG TCT
                                              Pst I
                          ⇩
Ala Pro Leu Leu Ala Ala Glu Cys Ser Val Asp Ile Gln Gly Asn Asp Gln Met Gln Phe
GCT CCG CTG CTG GCT GCA GAA TGC TCC GTT GAT ATC CAG GGT AAT GAT CAG ATG CAG TTC
                        Bsm I          EcoR V              Bcl I

Asn Thr Asn Ala Ile Thr Val Asp Lys Ser Cys Lys Gln Phe Thr Val Asn Leu Ser His
AAC ACC AAC GCC ATC ACC GTC GAC AAG AGC TGC AAG CAG TTC ACT GTT AAC CTG TCT CAC
                        Sal I                                Hpa I

Pro Gly Asn Leu Pro Lys Asn Val Met Gly His Asn Trp Val Leu Ser Thr Ala Ala Asp
CCA GGT AAC CTG CCG AAG AAC GTT ATG GGT CAC AAC TGG GTT CTG TCC ACC GCG GCT GAC
    BstE II                                                  Sac II

Met Gln Gly Val Val Thr Asp Gly Met Ala Ser Gly Leu Asp Lys Asp Tyr Leu Lys Pro
ATG CAA GGC GTT GTC ACT GAC GGT ATG GCT AGC GGT CTG GAT AAA GAC TAC CTG AAG CCG
                                    Nhe I

Asp Asp Ser Arg Val Ile Ala His Thr Lys Leu Ile Gly Ser Gly Glu Lys Asp Ser Val
GAT GAC TCT CGA GTT ATC GCC CAC ACC AAG CTG ATC GGA TCC GGT GAA AAA GAC TCC GTT
          Xho I                                  BamH I

Thr Phe Asp Val Ser Lys Leu Lys Glu Gly Glu Gln Tyr Met Phe Phe Cys Thr Phe Pro
ACT TTC GAC GTT TCC AAG CTT AAA GAA GGT GAA CAG TAC ATG TTC TTC TGC ACT TTC CCG
                      Hind III                                        Xma I

Gly His Ser Ala Leu Met Lys Gly Thr Leu Thr Leu Lys Trm
GGT CAC TCC GCA CTG ATG AAA GGT ACC CTG ACT CTG AAA TAG
                            Kpn I
```

Figure 2. Azurin sequence in pUC18.

Figure 3

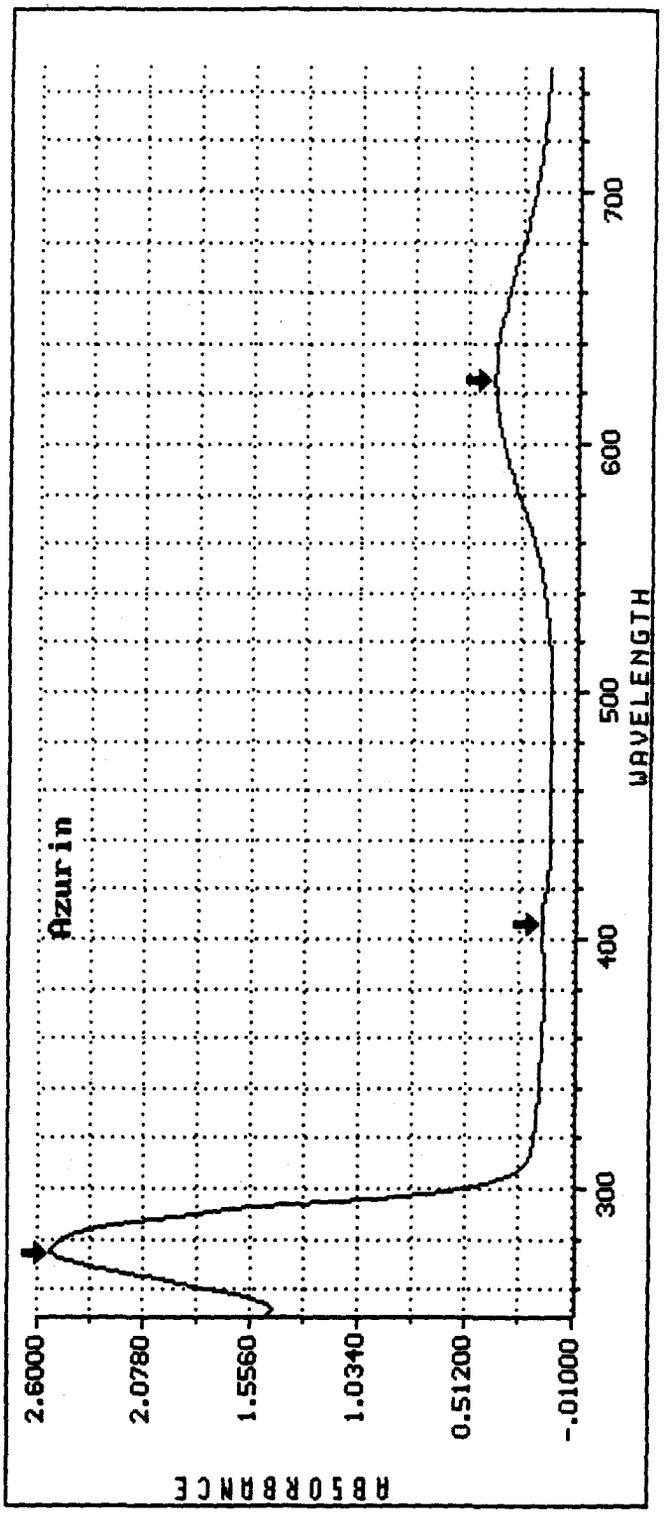
Western blot of whole cell extracts of TG-1 cells with pUC18Azu plasmids. Lane 1 is azurin purified from *P. aeruginosa* while lanes 2-5 are azurin expressed from the synthetic gene.

Western Blot of Azurin

1 2 3 4 5 6

Figure 4

UV-Vis spectrum of azurin expressed in *E. coli*.

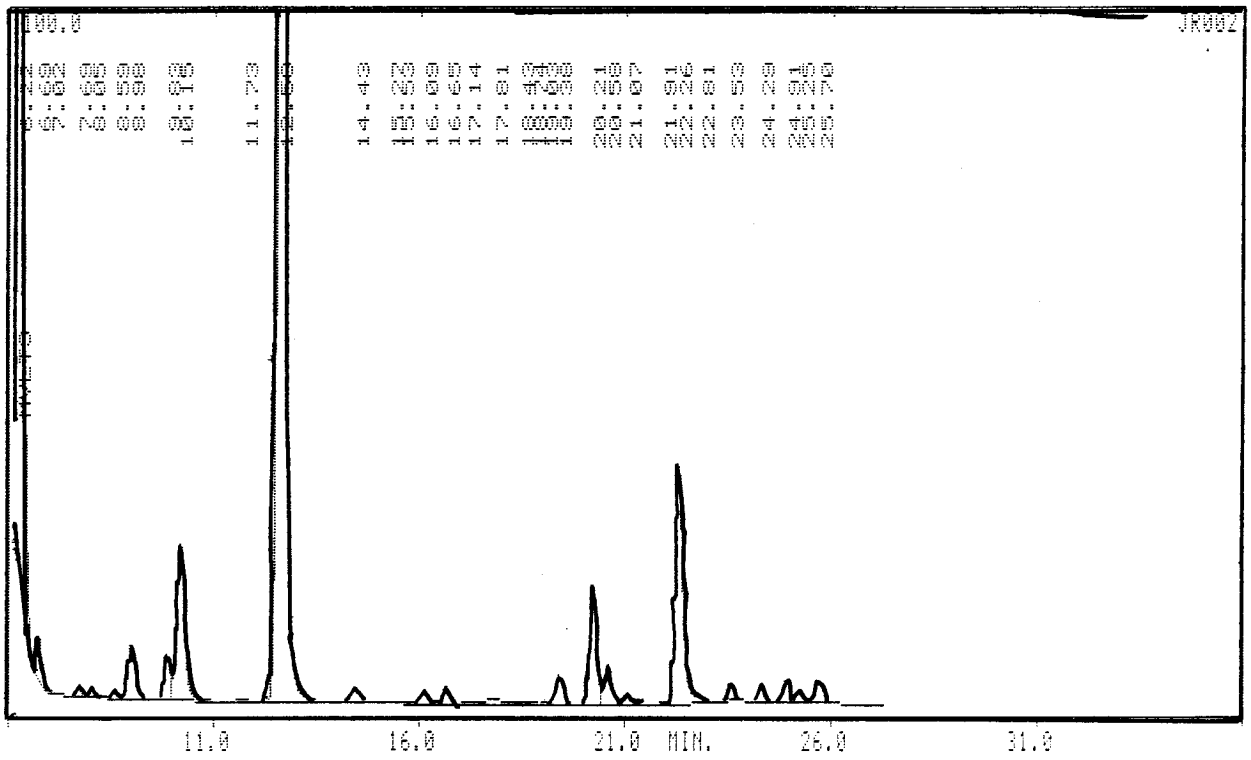


Annotated Wavelengths:

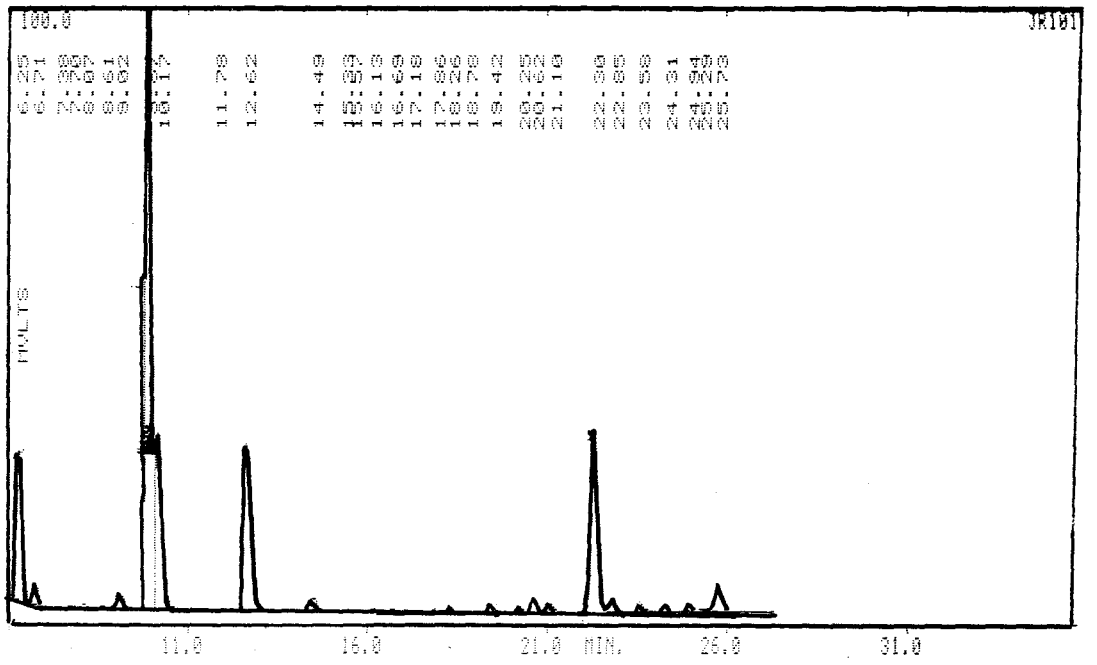
1 : Wavelength = 274	Result =	2.538254
2 : Wavelength = 406	Result =	0.144653
3 : Wavelength = 626	Result =	0.382324

Figure 5

Data from sequencing azurin expressed in *E. coli*. The first 11 residues were correctly sequenced. The only ambiguity was at Cys 3. This residue normally forms a disulfide bond, and this may have led to the ambiguous result.

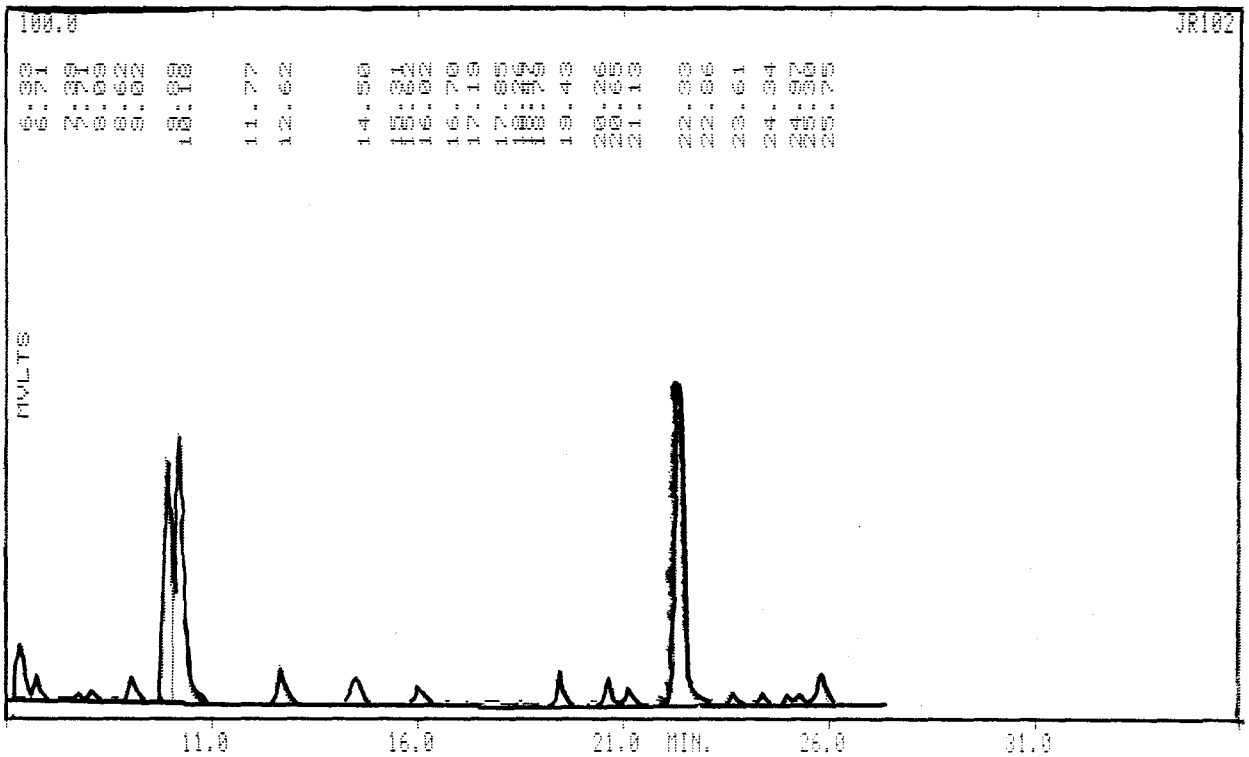


Residue 1

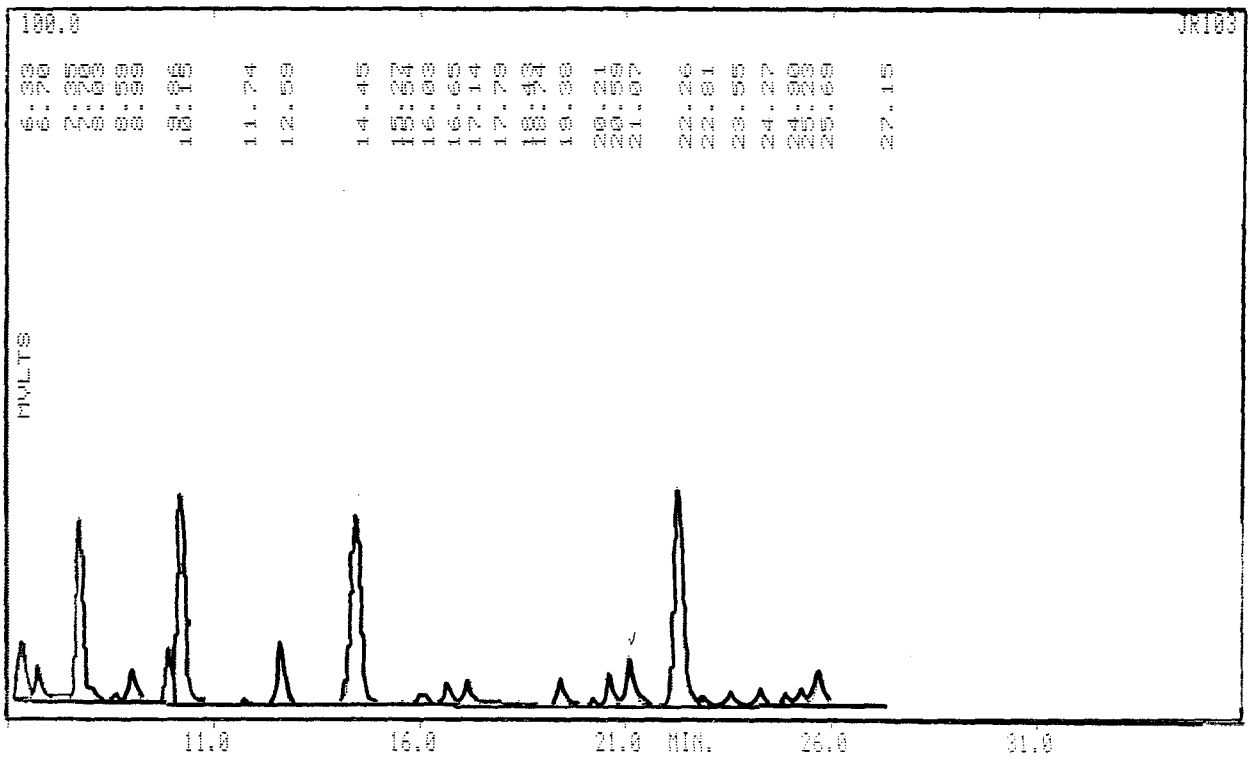


FEB. 0 9 1990

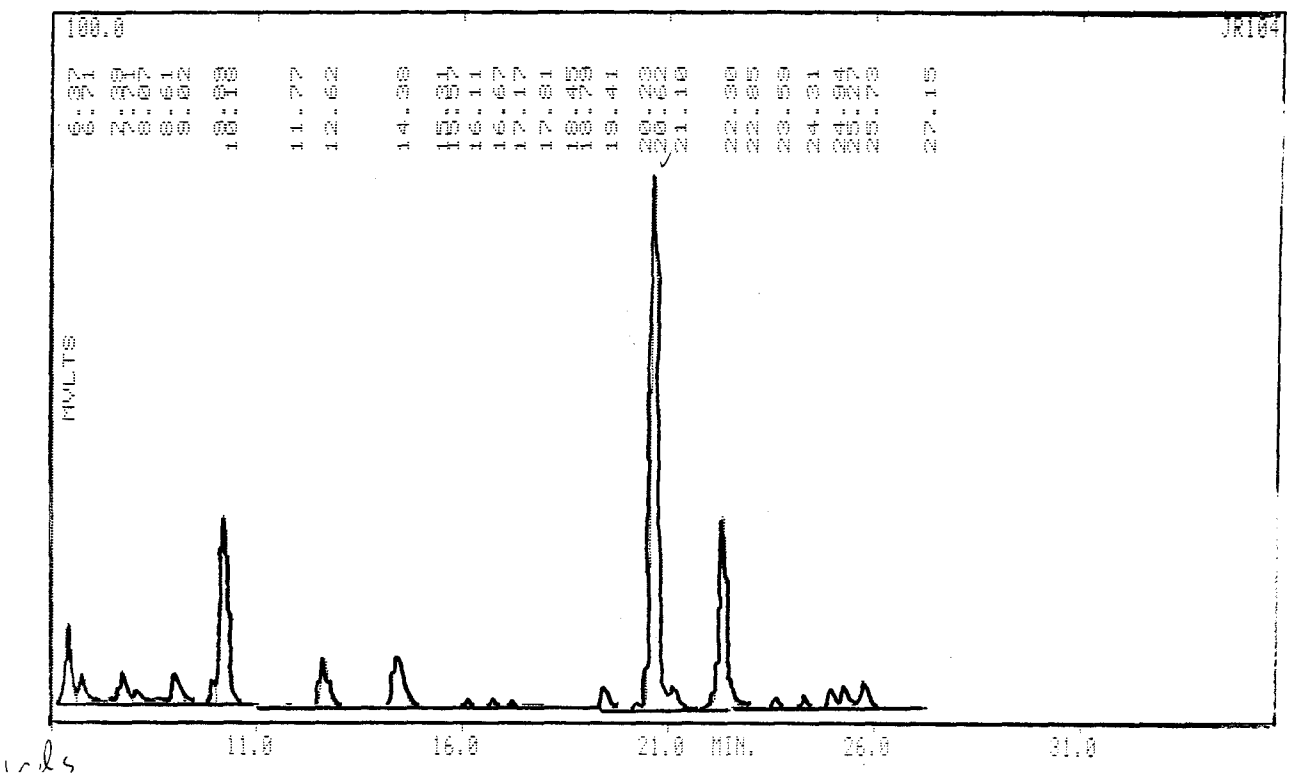
Residue 2



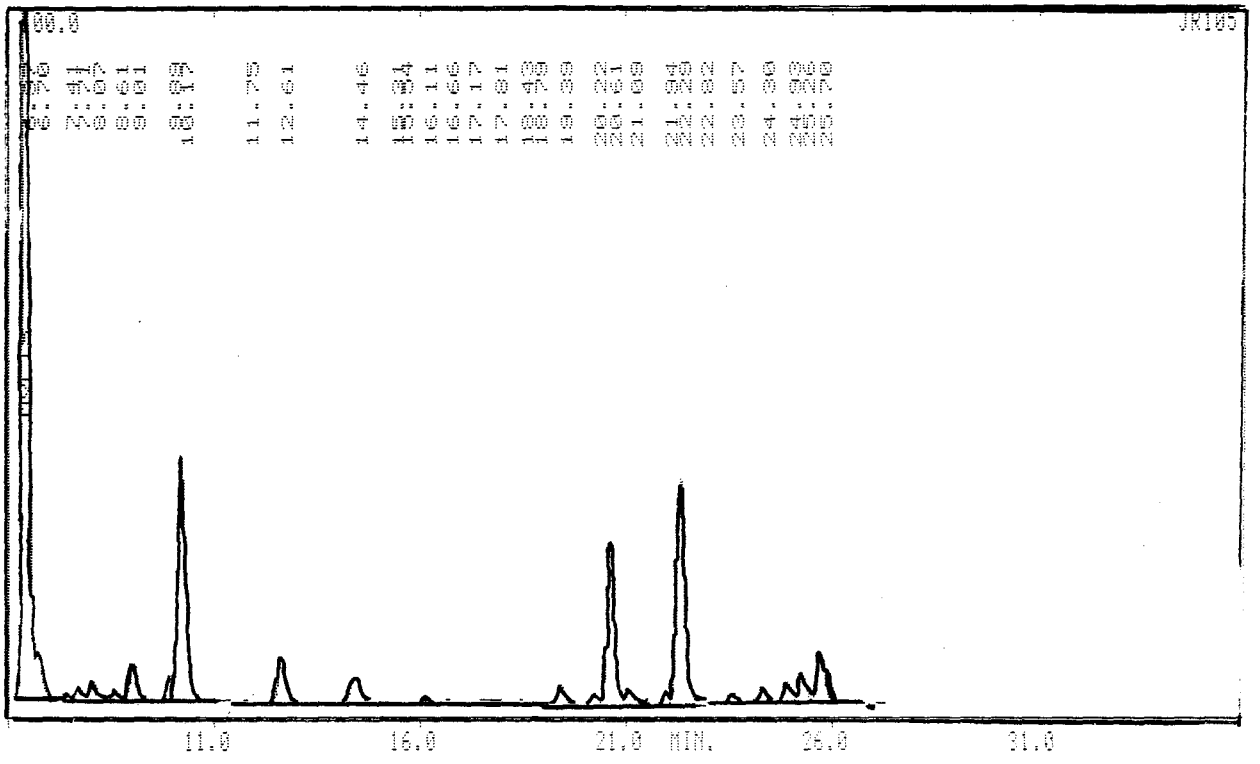
Residue 3



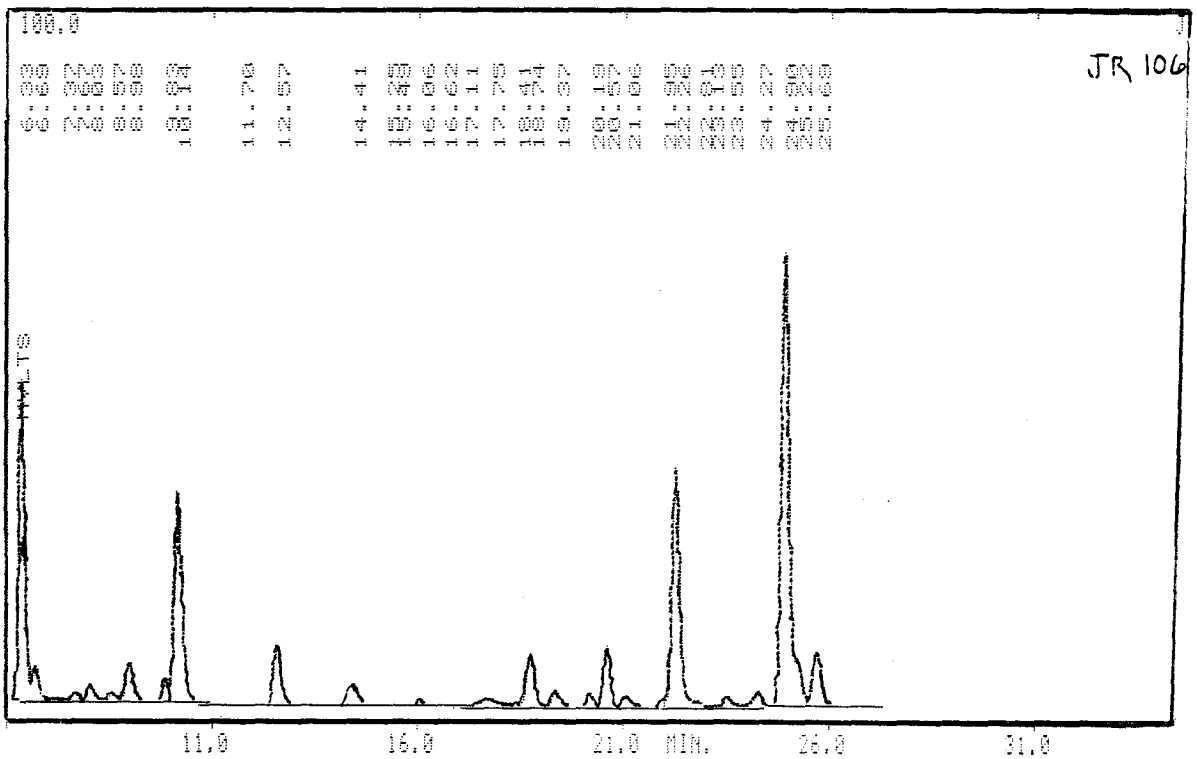
Residue 4



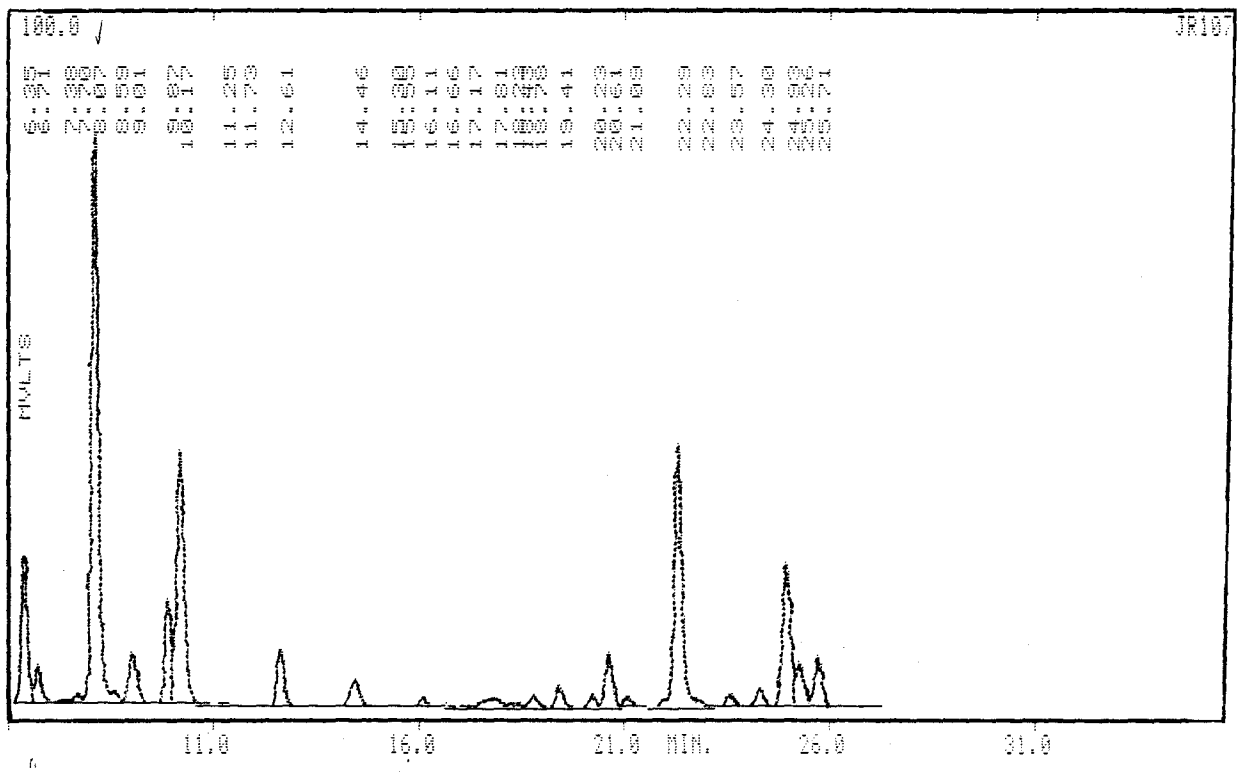
Residue 5



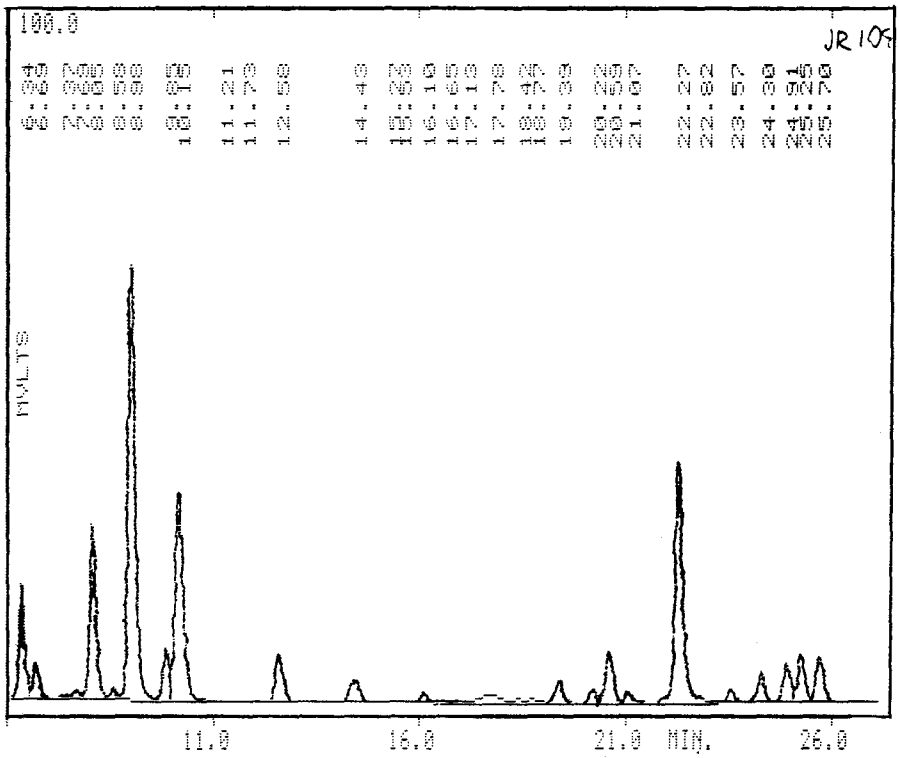
Residue 6



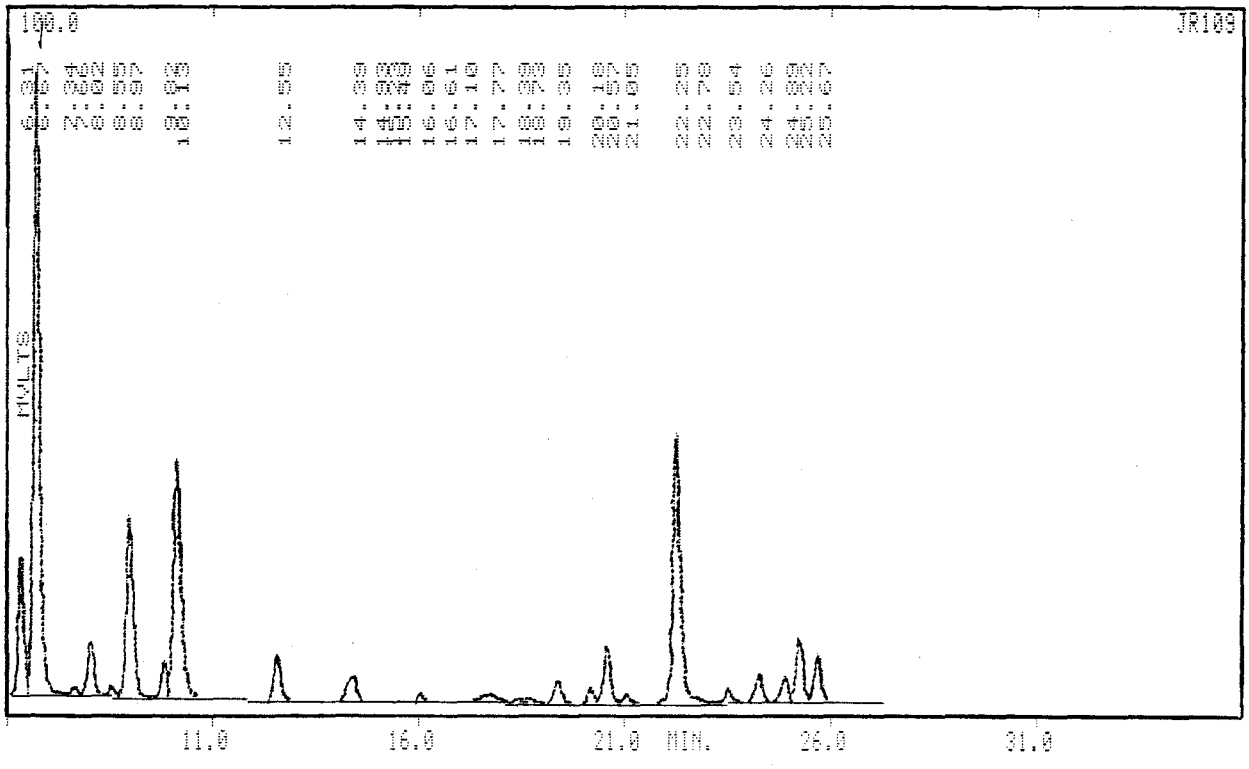
Residue 7



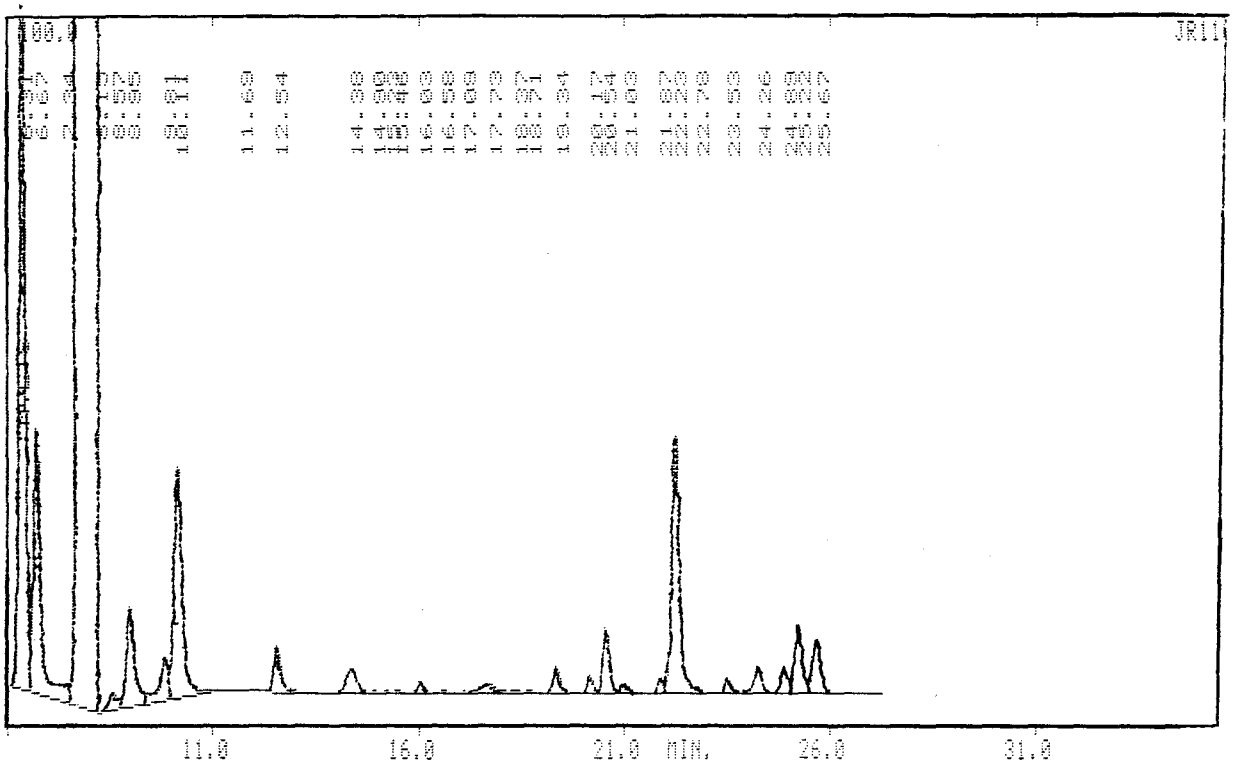
Residue 8



Residue 9



Residue 10



R.11

Residue 11

$CuSO_4$. However, when the cells were grown with 1 mM $CuSO_4$, 20 mg of azurin were recovered. The UV-Vis spectrum of this sample was identical to that of native azurin (Figure 4). Sequencing of this sample confirmed that the identities of the first 11 residues were correct and that it was processed correctly, yielding alanine at position 1. Sequencing data are shown in Figure 6. The only ambiguity in the peptide sequencing was at Cys 3; apparently, the disulfide bond was not completely reduced so that its identity could not be determined with absolute certainty. However, since the DNA sequence is correct, the amino acid at position 3 must be a cysteine. This proved that the expressed blue protein was azurin.

DISCUSSION

The expression of azurin represents the first successful expression of a synthetic gene in this lab. The correct processing of a *P. aeruginosa* signal sequence by an *E. coli* signal peptidase implies that the signal sequence mechanism between these two procaryotes are similar. Such result is expected, since the membrane-translocation mechanism is thought to be universal even between a procaryote and a eucaryote.[22,23] Plastocyanin was also expressed, using the same azurin-signal sequence soon after azurin was expressed, albeit at a lower yield. This may be due to minor structural differences between the two proteins, as discussed in Chapter 3. With plastocyanin, it would be interesting to model the differences between the azurin-signal sequence's interactions with azurin and plastocyanin. It may then be possible to engineer the azurin sequence so as to function more efficiently with plastocyanin.

Initially, we believed that growing the cells in micromolar amounts of copper would be sufficient to fully metalate all expressed azurin. The high concentration of $CuSO_4$ required in the growth indicates the possible presence of a metal pump in *E. coli*, which regulates the intracellular copper concentration. Thus, the pumping mechanism had to be overwhelmed with millimolar amounts of copper before azurin in the periplasm could incorporate copper into its binding site.

The yield of azurin was higher from the pUC18 construct than from the pKK223-3 construct. This may be due to the higher copy number that pUC18 possesses. The yield of azurin could be improved by replacing the *lac* promoter with the *tac* promoter, while using the same origin of replication to retain the high copy number of plasmids. In addition, the

yield of azurin may be further increased by improving the growth and the purification conditions. Finally, more detailed changes in the gene sequence can be introduced to favor the formation of secondary structures at the 3' and the 5' ends of the transcribed mRNA to stabilize it.[24] Currently, modifications of the plasmid are under way to improve the yields of azurin and plastocyanin.[25,26]

REFERENCES

1. Randall, L. L. and Hardy, S. J. S. (1984) *Microbiol. Rev.* **48**, 290-329..
2. Dierstein, R. and Wickner, W. (1986) *EMBO J.* **5**, 427-431.
3. Watson, M. E. E. (1984) *Nucleic Acids Res.* **12**, 5145-5164.
4. Bankaitis, V. A. and Bassford, P. J. (1984) *J. Biol. Chem.* **259**, 12193-12200.
5. Michaelis, S. and Beckwith, J. A. (1982) *Rev. Microbiol.* **36**, 435-465.
6. Kendall, D. A., Bock, S. C. and Kaiser, E. T. (1986) *Nature (London)* **321**, 706-708.
7. Koshland, D. and Botstein, D. (1980) *Cell* **20**, 749-760.
8. Kadonaga, J. T., Gautier, A. E., Strauss, D. R., Charles, A. D., Edge, M. D. and Knowles, J. R. (1984) *J. Biol. Chem.* **25**, 2149-2154.
9. Stader, J. A. and Silhavy, T. J. (1990) *Methods in Enzymol.* **185**, 166-187.
10. Nagarajan, V. (1990) *Methods in Enzymol.* **185**, 214-223.
11. Canters, G. W. (1987) *FEBS Lett.* **212**, 168-172.
12. Arvidsson, R. H. A., Nordling, M. and Lundberg, L. G. (1989) *Eur. J. Biochem.* **179**, 195-200.
13. Iverson, S. A. (1988) Ph.D. Thesis, California Institute of Technology.
14. Stormo, D. (1986) in *Maximizing Gene Expression*, eds. Reznikoff, W. and Gold, L. (Butterworths, Boston), pp. 195-224.
15. Gibson, T. J. (1984) Ph.D. Thesis, Cambridge University.
16. Brosius, J. (1987) in *Vectors: A Survey of Molecular Cloning Vectors and Their Uses*, ed. Rodriguez, R. L. and Denhardt, D. T. (Butterworths, Boston), pp. 205-226.
17. Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557-580.

18. Wharton, D. (1978) *Methods in Enzymol.* **53**, 646-661.

19. St. Clair, C. S. (1989) Ph.D. Thesis, California Institute of Technology.

20. Messing, J. (1984) *Methods in Enzymol.* **101**, 20-78.

21. Amann, E., Brosius, J. and Ptashne, M. (1983) *Gene* **25**, 167-178.

22. Schatz, G. (1986) *Nature (London)* **321**, 108-109.

23. Briggs, M. S. and Gierasch, L. M. (1986) *Adv. Prot. Chem.* **70**, 109-180.

24. Simon, M. and Apirion, D. (1972) *Nature (London)* **239**, 79-81.

25. Germanas, J. P., work in progress.

26. Kiser, C. N., work in progress.

Chapter 5

Site-directed Mutagenesis and Mutant Characterization of Azurin

INTRODUCTION

Azurin is an electron-transport protein involved in the anaerobic metabolic pathway of *P. aeruginosa* (Figure 1).[1] Its unique blue copper center has intrigued spectroscopists and inorganic chemists for a long time.[2] Using various spectroscopic investigations of azurin and its model structures, the composition and the geometry of the copper center had long been identified before they were confirmed by crystal structures.[3-5]

The copper center, similar to that of plastocyanin, consists of His 46, Cys 112, His 117, and Met 121 (Figure 2).[6] The crystal structure from the *A. denitrificans* azurin, however, shows a very long S-Cu bond for the Met 121, raising questions about its significance.[7,8] That structure also shows a possible bond between the carbonyl of Gly 45 and the copper. For these reasons, the copper centers in azurins are now considered to have a distorted trigonal planar geometry with two weaker axial interactions from Met 121 and Gly 45.

The electronic spectrum of azurin has been analyzed down to minute details. Its electronic spectrum has been Gaussian-resolved,[9] and the assignment of the bands is summarized in Table 1. The LMCT from Cys 112 is the most dominant contributor to azurin's electronic signature. These results are based on UV-Vis, CD, and model-building studies. In addition, its proton NMR[10] and Resonance Raman[11-13] peaks have also been assigned in detail. These studies allow one to delineate the metal-ligand interaction exactly. Other useful techniques include the EPR[14] to probe the nature of the metal environment, while the reduction potential has been routinely measured by either spectroelectrochemistry or cyclic

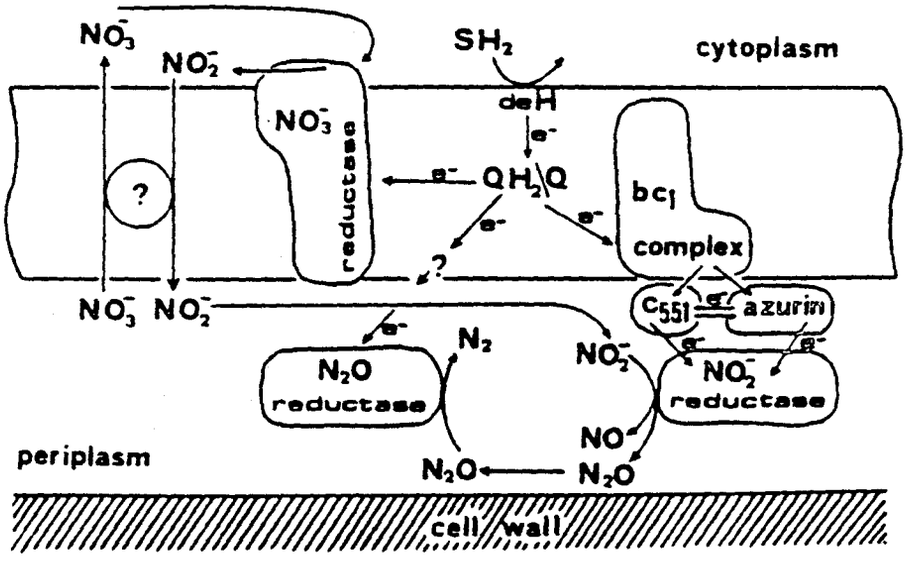


Figure 1. Postulated organization of nitrate, nitrite, and nitrous oxide reduction system in the anaerobic metabolic pathway of *P. aeruginosa*. Azurin is one of two possible reductants of nitrite reductase.

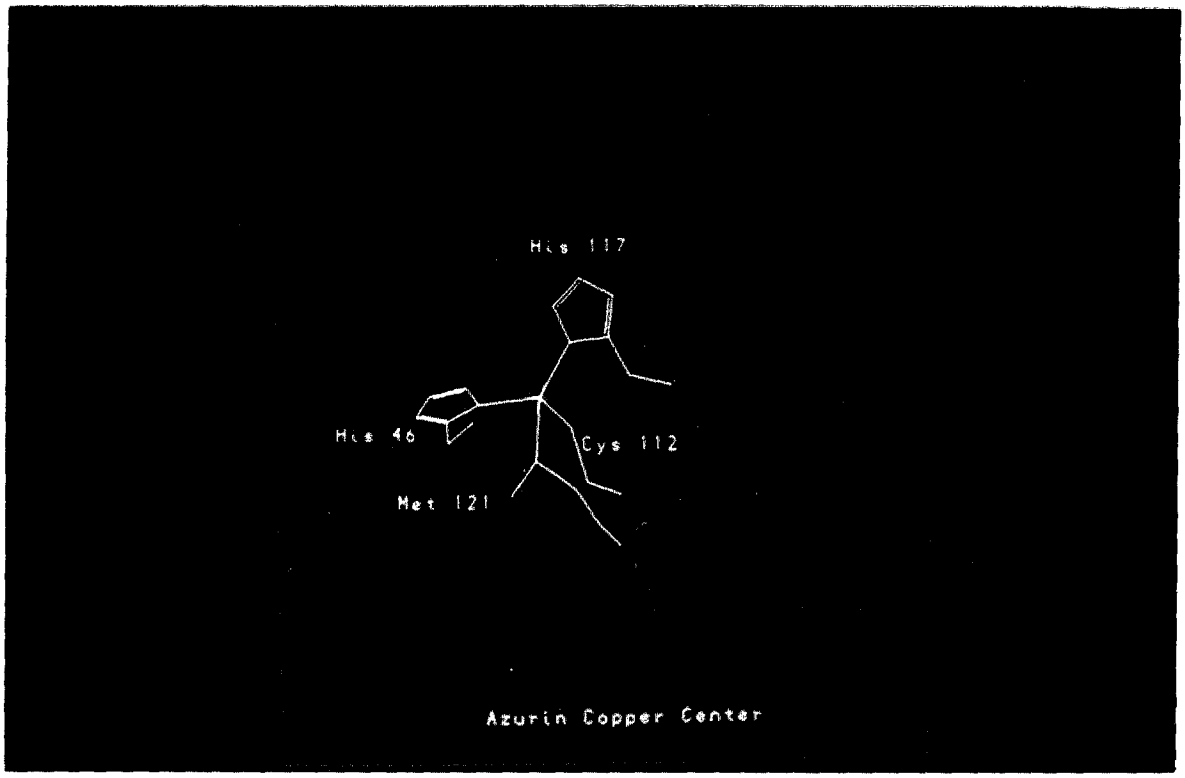


Figure 2. The copper center of azurin. The ligands consist of His 46, Cys 112, His 117, and Met 121.

Table 1

UV-Vis			CD			
υ(cm-1)	nm	e	υ(cm-1)	nm	e	Assignment
12830	779	683	12500	800	-5.9	πS -> d_{x2-y2}
15850	631	3798	16100	621	6.5	σS ->d_{x2-y2}
17650	567	504	19000	526	1.2	σS ->d_{x2-y2}
20790	481	198	21400	467	-1.8	πN->d_{x2-y2}

Assignments of charge transfer bands for azurin.

voltammetry.[15,16]

Metal substitution experiments give another spectroscopic handle in elucidating the nature of the copper center. The electronic spectra of Ni^{2+}, Mn^{2+}, and Co^{2+} derivatives of azurin have been well characterized and are summarized in Table 2.[17] In addition, McMillin has studied the relative stabilities[18] of Ni^{2+}, Zn^{2+}, Cu^{2+}, Hg^{2+}, and apo derivatives, ^{113}Cd NMR of the Cd^{2+} derivative[19] and proton NMR of the Ni^{2+} derivative,[20] and the kinetics of copper uptake by apo azurin.[21] The differences between these derivatives can be used to gain information on the nature of the copper center and the behavior of the copper ligand variants.

Another well-characterized behavior of azurin is its electron-transfer kinetics. Azurin was one of the first proteins modified by selective attachment of a Ru^{2+} label on a surface-accessible histidine.[22,23] Using flash photolysis, azurin's intramolecular electron-transfer rate through the interior of the protein has been measured. Compared to the results from azurin, an unusually low electron-transfer rate was measured for plastocyanin.[24] The wide range in the ET rates observed for these different proteins may be due to the differences in the electronic coupling in the electron-transfer pathway. Another interesting variable, the effect of aromatic residues on the electron-transfer rate, has been investigated by Hoffman.[25] These results, along with those computer modeling[26] of the purported electron pathway, are beginning to paint a clear picture of how electrons are transferred.[27-29]

The behavior of electron transfer rate is predicted by the Marcus equation:[30]

$$k_{ET} = v \exp(-\beta (d - 3)) \exp(-\Delta G^*/RT)$$

Table 2

Metal	ν(cm-1)	nm	Assignment
Ni(II)	1.78	561	πS (Cys) -> Ni
	2.00	500	πS (Cys) -> Ni
	2.28	439	σS (Cys) -> Ni
	2.55	392	πS (Met) -> Ni
	2.82	354	σS (Met) -> Ni
Co(II)	2.45	408	πS (Cys) -> Co
	2.68	373	πS (Cys) -> Co
	3.00	333	σS (Cys) -> Co

Assignments of charge transfer bands for metal-substituted azurin.

where

$$\Delta G^* = (\Delta G^\circ + \lambda)^2/4\lambda$$

Here exp(-β (d - 3)) is an electronic term that is dependent on the distance and the electronic coupling (β), which itself is influenced by the media. The λ is the nuclear reorganization energy that takes into account both the inner-sphere and the outer-sphere reorganization energies, while d is the distance between the donor and the acceptor molecules.

The Marcus equation can be used as a guide in selecting the mutants that are relevant to the investigation of electron transfer. The beauty of site-directed mutagenesis is that each variable in this equation can be varied independently, offering a systematic way to study the electron-transfer problem.

From its inception in the early eighties, site-directed mutagenesis has been a powerful means to study the structure-function relationship in proteins. Several techniques are widely used, including oligonucleotide-directed,[31,32] cassette,[33,34] and random mutagenesis.[35] Using cassette mutagenesis, for example, it is possible to change multiple sites at the same time or to introduce all other 19 amino acids at a particular position.[36] The latter method is called site-saturation, and it has been used successfully to saturate several sites in β-lactamase.[36-39] These mutagenesis techniques have proved invaluable in elucidating the enzyme mechanism,[40,41] the role of a particular residue[42], changing the catalytic activity of an enzyme,[43-45] and stabilization of expressed proteins.[46] The versatility of this technique is growing with the introduction of new techniques such as the PCR.

The azurin gene was designed specifically for site-directed mutagenesis. The abundance of restriction sites facilitates the creation of variants. Once generated, the nature of the mutant can be unsolved using the wealth of techniques available discussed above.

Some work on the mutagenesis of azurin has been reported recently. The H35K and E91Q mutants, for example, have unaltered spectroscopic properties, while in the F114A mutant, the optical band is downshifted by 7 nm and the reduction potential by 20-24 mV.[47] Another mutation at a conserved residue within the hydrophobic path, M44K causes only a minimal change in the spectroscopic properties but significantly affects the electron self-exchange rate.[48]

In this chapter, various experiments that have been performed in order to gain a better understanding of azurin's copper center are reported. Combining site-directed mutagenesis with various spectroscopic techniques allows us to investigate azurin and its properties in an unprecedented way.

Materials and Methods

Materials

All expressions were performed in *E. coli* strain TG-1. All mutants were made by the author except for the Met 121 mutants Ala and Leu, F114A, and the Triple mutant. These were obtained from B. G. Malmstrom for collaborative work. M121D was made by both groups.

Restriction enzymes, polynucleotide kinase, T4 DNA ligase, and calf-intestinal alkaline phosphatase were purchased from the same supplier as mentioned in Chapter 2.

Western blot was performed on a Bio-Rad transblotter. The cells were grown in a New Brunswick Magnaferm fermentor, and the purification was done using the same conventional and FPLC chromatographic procedures detailed in Chapter 4. UV-Vis spectra were taken on a Hewlett-Packard Model 4830 spectrophotometer. The CD spectra were taken on a JASCO Model 600 spectrophotometer, while NMR spectra were taken on a Bruker AM-500 NMR. The X-band EPR spectra were taken by A. DiBilio. The S-band EPR spectra were taken at the national ESR center located at the Medical College of Wisconsin by W. Antholine.

Methods

DNA manipulation

For the site-saturation cassette mutagenesis, oligonucleotides containing mixed bases at position 121 codon were synthesized and purified using Applied Biosystem's OPC cartridge. This method takes advantage of the dimethyltrityl group present at the end of the synthesis. The trityl group was removed after purification by treating it with acetic acid.

The plasmids were cleaved with two restriction enzymes and purified on a 1.5 % agarose gel. The oligonucleotides were individually kinased and annealed together in a boiling water bath. The ligation was performed using 0.4 pmole of annealed cassette and 0.04 pmole of cleaved plasmid in 25 µl of ligase buffer.

Transformation

The ligated plasmid was transformed into TG-1 cells using the Hanahan protocol as described in Chapter 2.

Colony screening

The plasmids from site-saturation ligation were all sequenced using the Sequenase sequencing kit. The plasmids from minipreps were used for sequencing. Cells were grown in 2 ml of L-broth containing 50 µg/ml ampicillin. 1.5 ml of this was spun down in a microfuge, and the standard alkaline lysis was performed using 100 µl of solution 1, 200 µl of solution 2, and 150 µl of ice-cold solution 3. The mixture was extracted with phenol once, chloroform once, and ethanol precipitated by the addition of 3 M NaAc (1/10 volume) and ethanol (2.5x volume). The DNA pellet was washed once with 70% ethanol and dried under vacuum. The recovered plasmid had high enough purity for sequencing. The plasmids were sequenced using the Sequenase kit as described in Chapter 2.

Protein Preparation

Large-scale (8L) protein preparations were performed as before (Chapter 4). Up to 2 mM of $CuSO_4$ were used, although this did not seem to make a difference in the yield.

Metal substitution

To replace Cu^{2+} with another metal, ~10 mg of azurin in 1 ml of NH_3Ac buffer (pH 4.1) were reduced by adding ascorbic acid (1 M) in 10 mM increments until the blue color disappeared. The amount of ascorbic acid required to reduce the protein differed completely with each mutant. The copper was then removed by dialyzing the protein sample against 3 changes of thiourea buffer. The samples were then dialyzed into Tris-HCl buffer, pH 9. The NH_3Ac buffer was originally used, but Cu^{2+} proved to be too reactive in this buffer. Azurin was then removed from the dialysis tubings, and the metal (Co^{2+}, Ni^{2+}, or Cu^{2+}) was added to a final concentration of 1 mM. After allowing the metal to be incorporated overnight, the excess metal was dialyzed away in a Tris-HCl or a phosphate buffer, pH 7.0.

NMR

NMR spectra of nickel-substituted azurins have been taken. Since a proton spectrum of a protein is rather complex, the NMR spectra of the nickel derivatives were taken with a fast pulse rate (3 or 4 pulses/sec). The normal relaxation time of the peaks is in the order of seconds, allowing all signals to be saturated except for those from the residues that are directely associated with nickel. Thus, this technique enables one to focus onto only those residues that are immediately associated with the metal center where nickel is inserted.

Computer Modeling

All mutants were modeled on BIOGRAF, versions 1.5, 2.0 or 2.1 (BioDesign Inc., Pasadena, California) running on an Evans and

Sutherland PS 390 graphic terminal controlled by a DEC MicroVax 3000 work station. This program was set up with the Dreiding force field for energy minimization.

RESULTS

Site-saturation mutagenesis

In order to generate all other 19 amino acid substitutions in one step, a site-saturation cassette was synthesized that contained all four bases at the first two bases, and G and C at the third base of the codon at position 121 (Figure 3). Then pAZ18 was cleaved in two ways: with Hind III/EcoR I and with Kpn I/EcoR I. The small piece from the first digest and the large piece from the second digest were gel-purified. The cassette was ligated with these fragments using a three-piece ligation(Figure 4). 200 colonies originating from this ligation were sequenced before all 19 amino acid variants were identified. The actual number of colonies sequenced was more than expected, since according to Poisson distribution, one needs to sequence only four times the total number of possibilities (4 x 32 = 128) to have a 95% probability of finding them all. The codon and the base distributions are summarized in Table 3. The high occurrence of ATG (Met) is obvious from this table. This was most likely caused by incomplete separation of the doubly digested plasmid from the singly digested plasmid, since they differed by only ~400 bp out of 3000 bp. Successive approximation was done to estimate that the 200 colonies contained 121 mutants and 79 parental plasmids. Here, G and C are found more frequently than A and T, perhaps because of their greater number of hydrogen bonds.

UV-Vis

The UV-Vis spectra of the Met 121 mutants Asn, Asp, His, Ile, and Val are shown in Figures 5-9 and summarized in Table 4. The peaks are shifted, but not drastically.

Figure 3. Cassette with mixed bases used to site-saturate Methionine 121 in azurin. The codon at position 121 is highlighted.

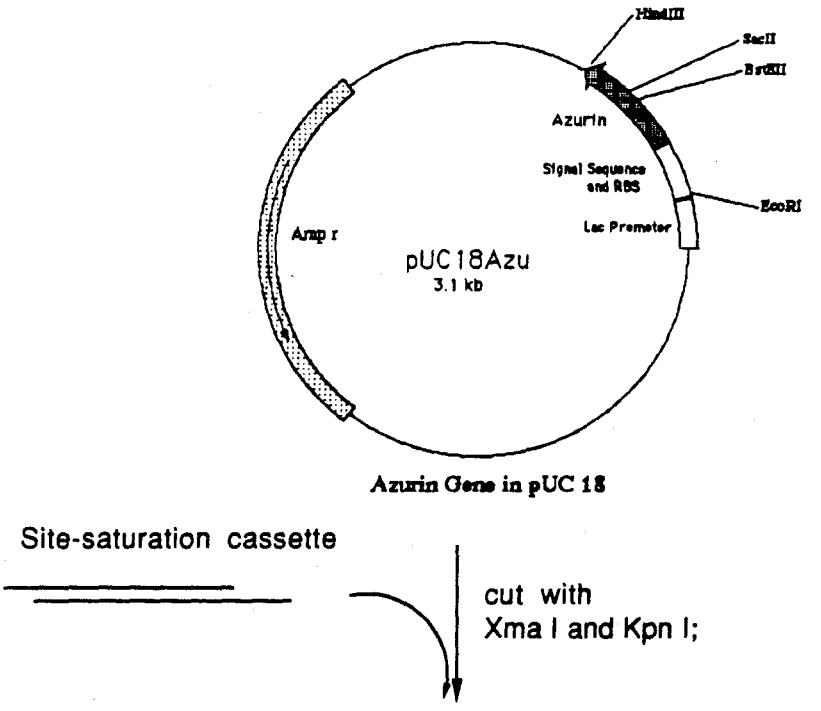


Figure 4. Ligation of site-saturation cassette into azurin expression vector.

Table 3
Base Distribution from Site-saturation

First Position	Second Position T(27)	C(50)	A(21)	G(23)
T (23)	C(F) 5 G(L) 2	C(S) 3 G(S) 3	C(Y) 3 G(Stop) 2	C(C) 2 G(W)
C (35)	C(L) 2 G(L) 3	C (P) 11 G(P) 12	C(H) 1 G(Q) 4	C(R) 1 G(R) 1
A (24)	G(I) 2 G(M) 3/56	C(T) 5 G(T) 4	C(N) 2 G(K) 3	C(S) 3 G(R) 2
G (39)	C(V) 3 G(V) 7	C(A) 6 G(A) 6	C(D) 4 G(E) 2	C(G) 8 G(G) 3

Numbers shown are the number of times that a particular base was found at each of the three positions of a DNA codon. The amino acids are represented using the standard one-letter code. Using the probability of finding a particular base at a particular site, successive approximations were made to estimate the number of contaminating wild-type plasmid ATG (approximately 53 out of 56 ATG's were contaminants).

Figure 5-9

UV-Vis spectrum of Methionine 121 variants (Asn, Asp, His, Ile, Val) of azurin.

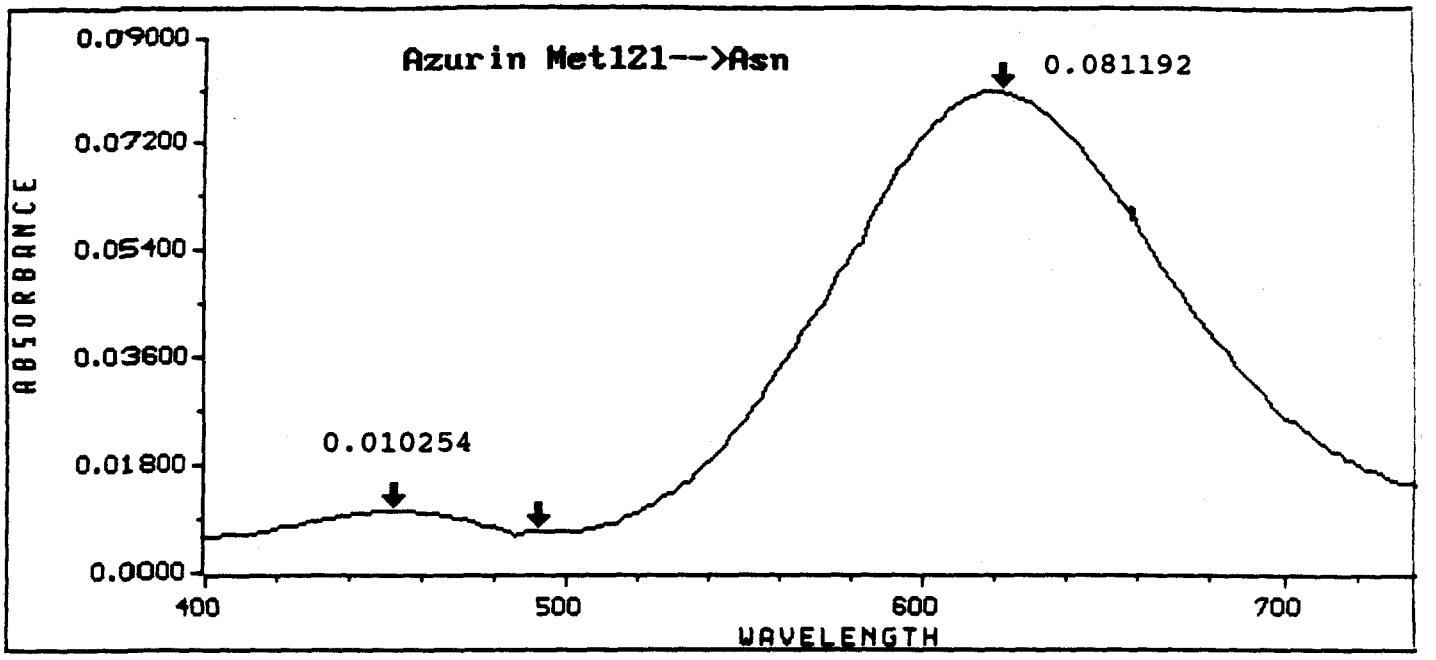
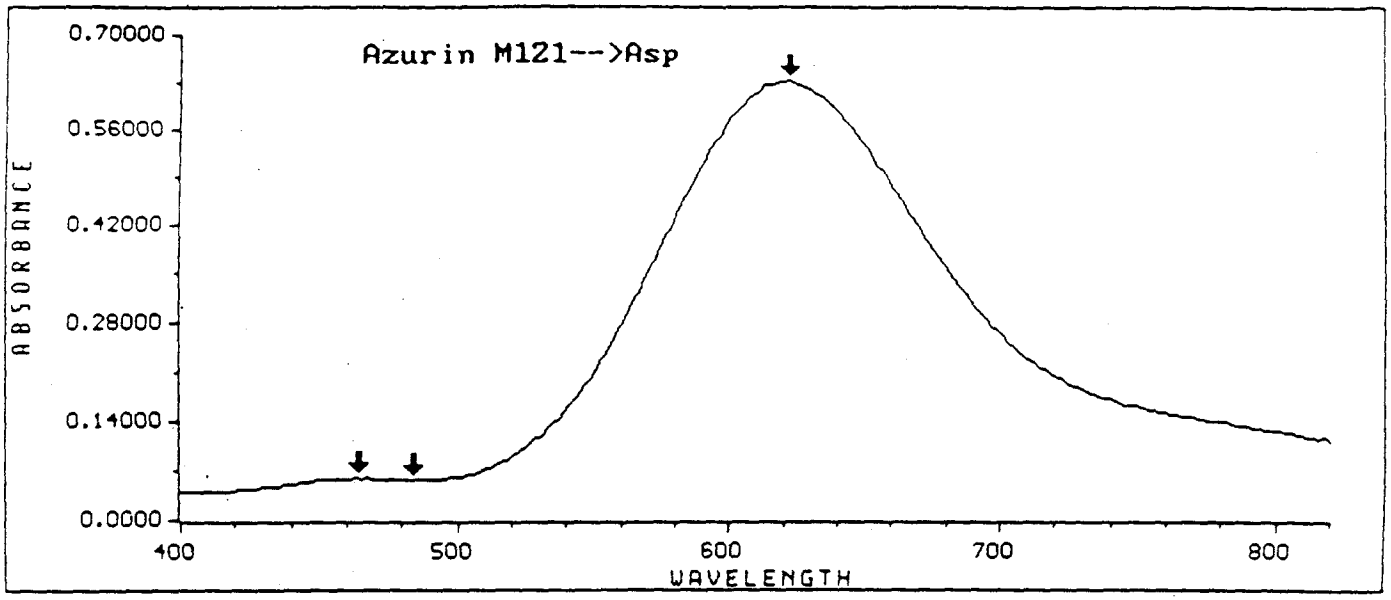


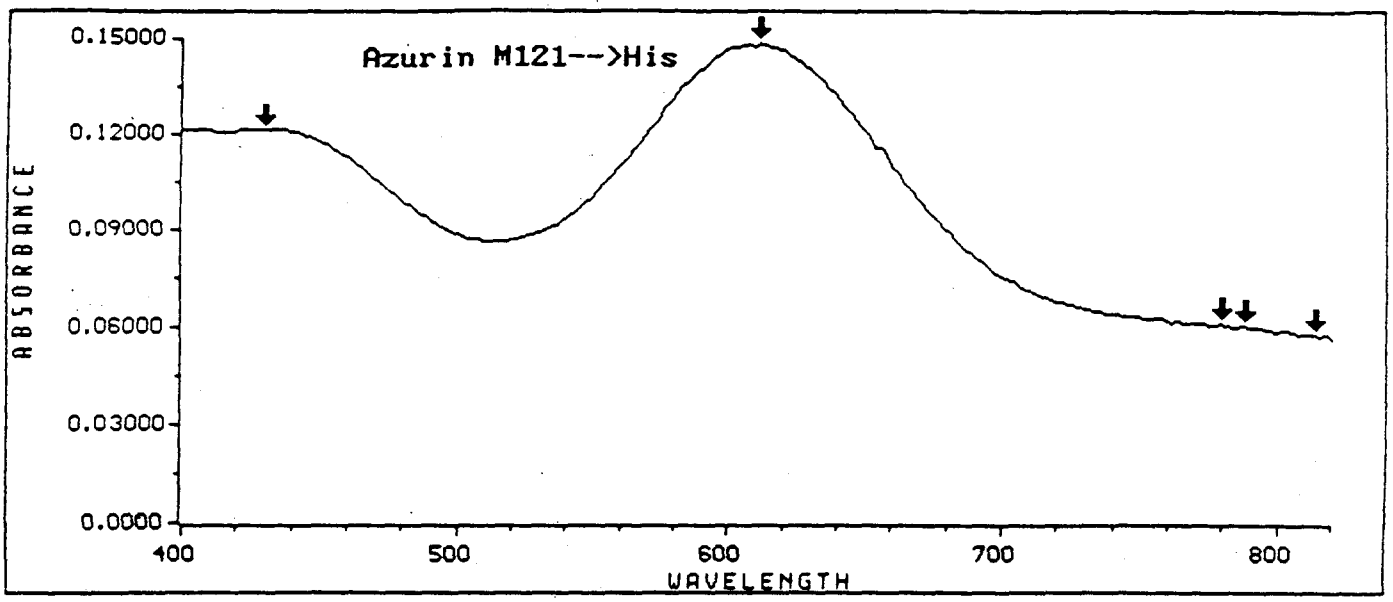
Figure 5. UV-Vis spectrum of Met121Asn azurin.



Annotated Wavelengths:

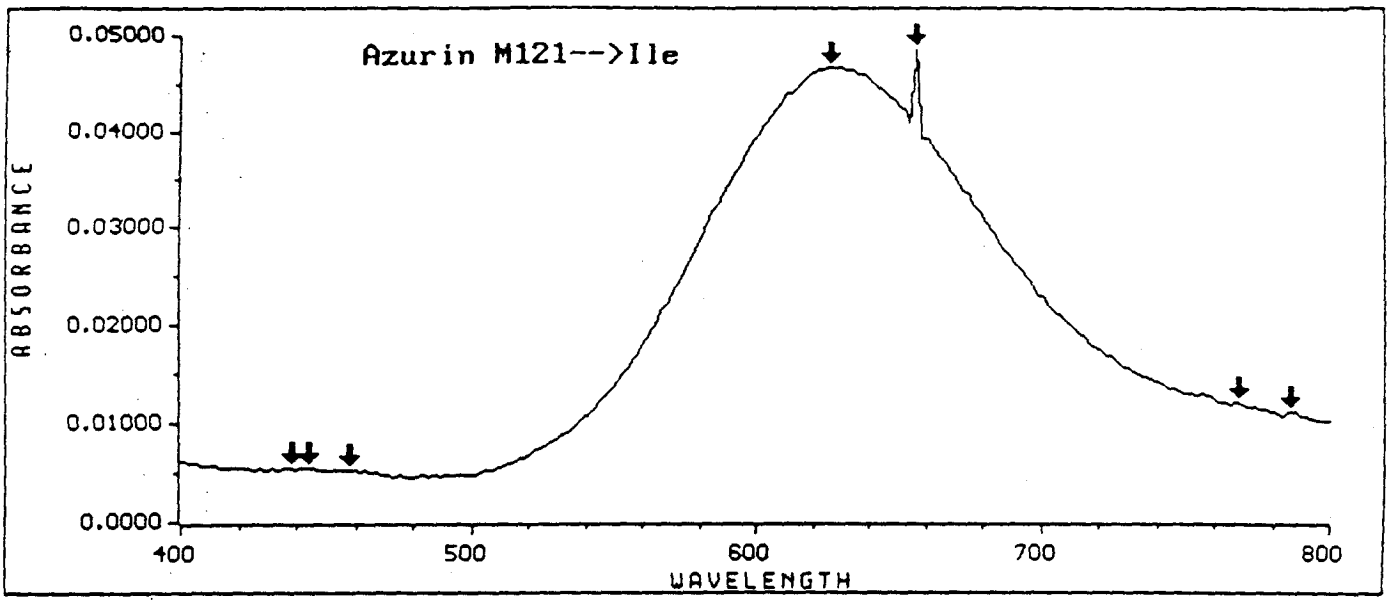
1 :	Wavelength = 464	Result =	0.058395
2 :	Wavelength = 484	Result =	0.057037
3 :	Wavelength = 622	Result =	0.631699

Figure 6. UV-Vis spectrum of Met121Asp azurin.




```
Annotated Wavelengths:
  1 : Wavelength = 430    Result =    0.121521
  2 : Wavelength = 612    Result =    0.148651
  3 : Wavelength = 780    Result =    0.061081
```

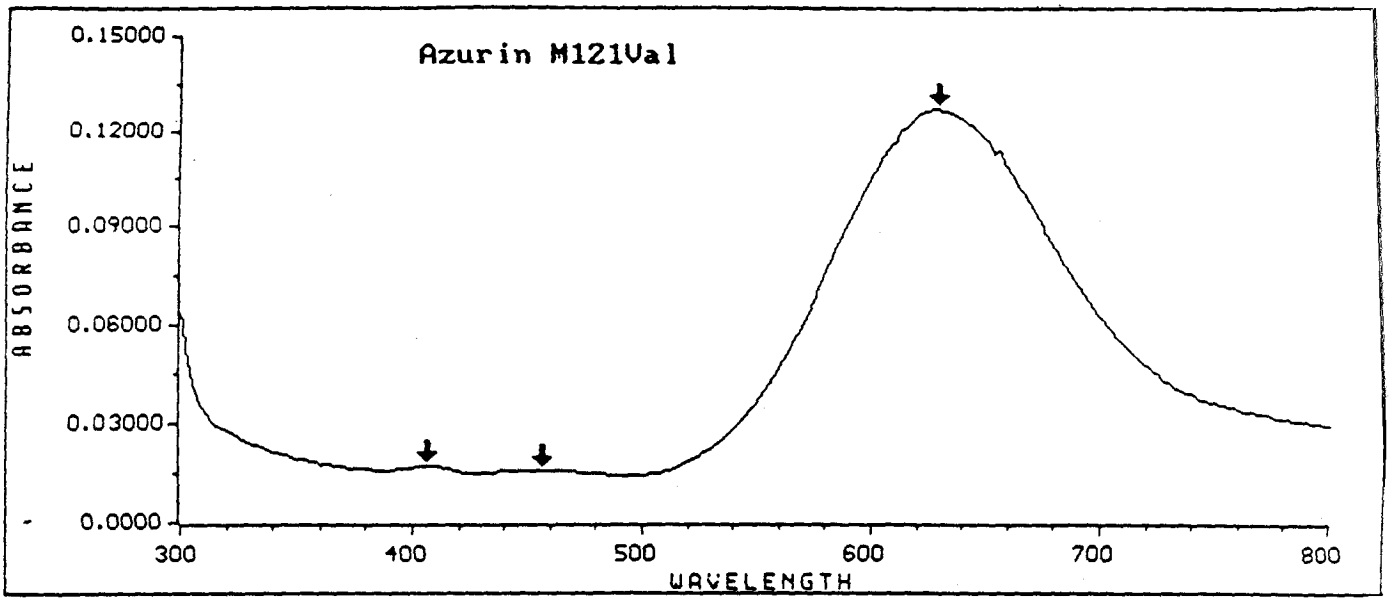
Figure 7. UV-Vis spectrum of Met121His azurin.



Annotated Wavelengths:

1 :	Wavelength = 438	Result =	0.005417
2 :	Wavelength = 444	Result =	0.005417
3 :	Wavelength = 458	Result =	0.005341
4 :	Wavelength = 626	Result =	0.046677

Figure 8. UV-Vis spectrum of Met121Ile azurin.



Annotated Wavelengths:
1 : Wavelength = 406 Result = 0.017319
2 : Wavelength = 456 Result = 0.015869
3 : Wavelength = 630 Result = 0.127152

Figure 9. UV-Vis spectrum of Met121Val azurin.

Table 4

UV-Vis absorbance maxima of the mutants

Variant	Major Peak	Minor Peak
M121(Wild type)	626	445
M121V	630	459
M121I	626	459
M121D	622	445
M121H	612	449

CD

The CD spectra (Figures 10-15) of the same mutants in the visible range show similar results; that the peaks are shifted but not by much.

EPR

The copper ligand variants show different g and a values for the X-band EPR spectrum, as summarized on Table 5. The EPR spectrum of Met121Val is shown in Figure 16. The S-band EPR spectra show high-resolution interaction at the copper site, especially when a pure isotope of copper is used. The S-band spectrum of the wild-type azurin is shown in figure 17.

UV-Vis and CD of metal-substituted azurins

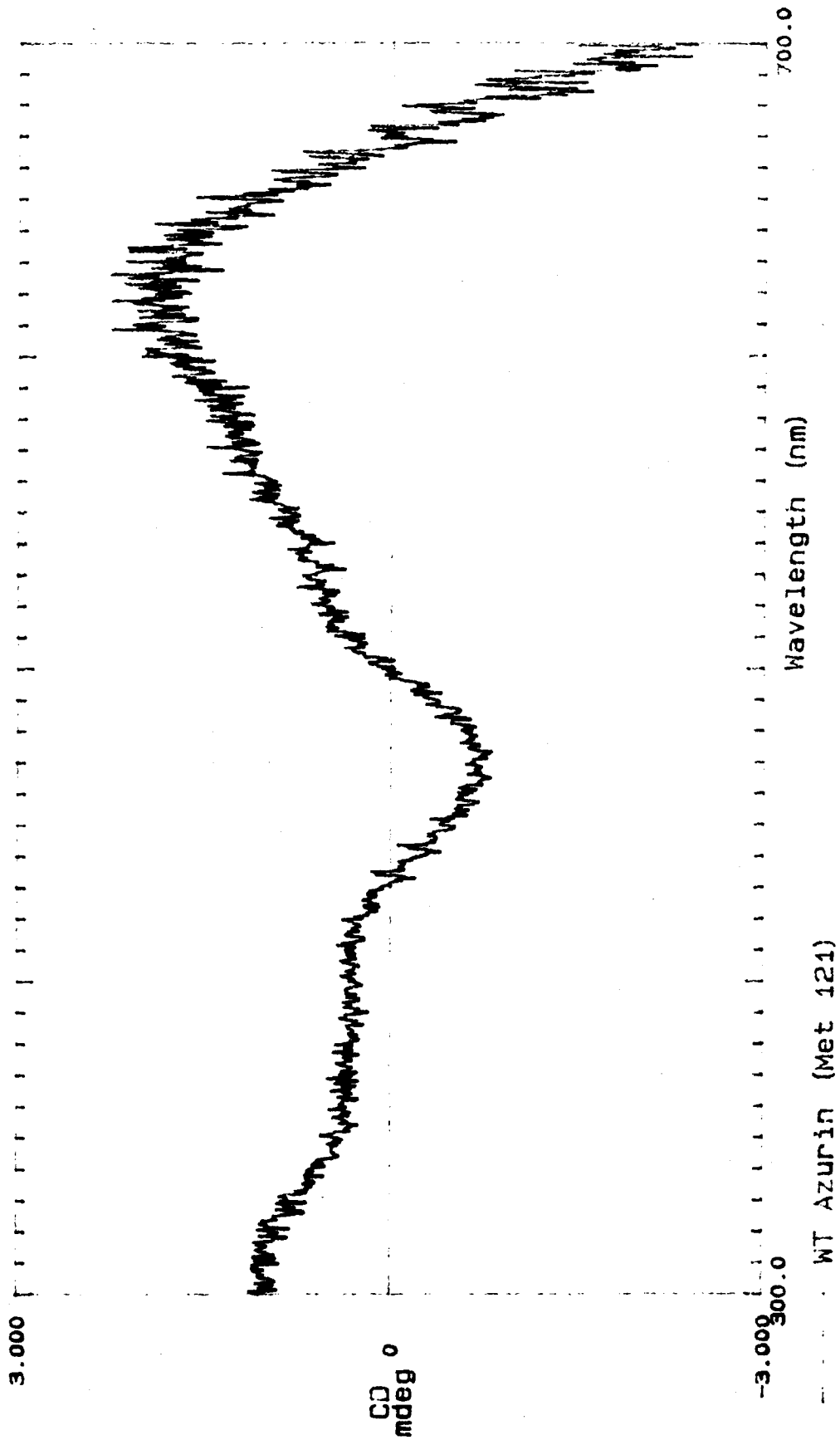
Ni and Co were substituted into M121D, M121G, M121L, F114A, and the triple mutants. Both the UV-Vis and CD spectra show more appreciable shifts compared to the copper azurins. The peaks are shown in Figures 18-28 and are summarized in Table 6.

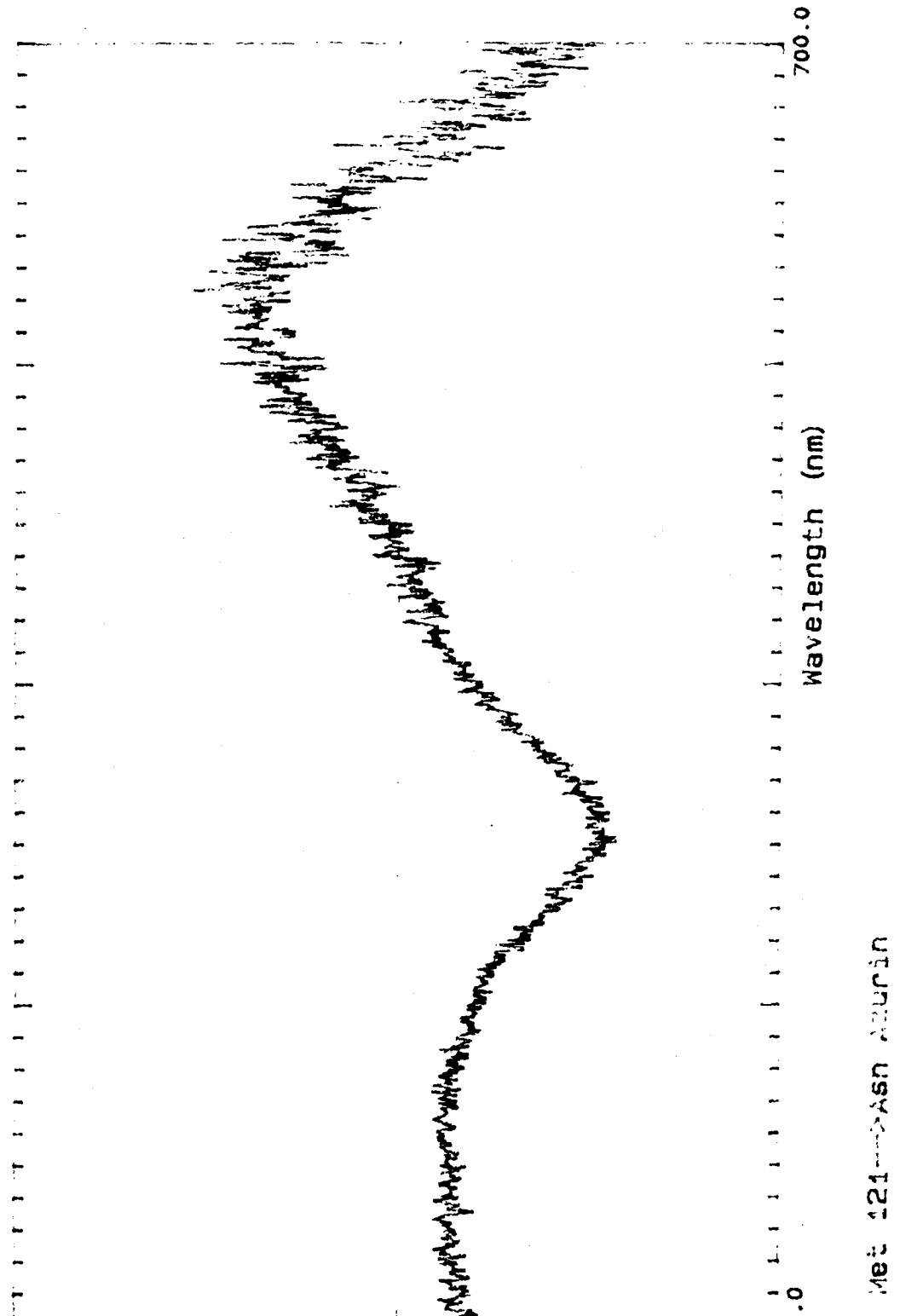
NMR of Nickel- substituted azurins

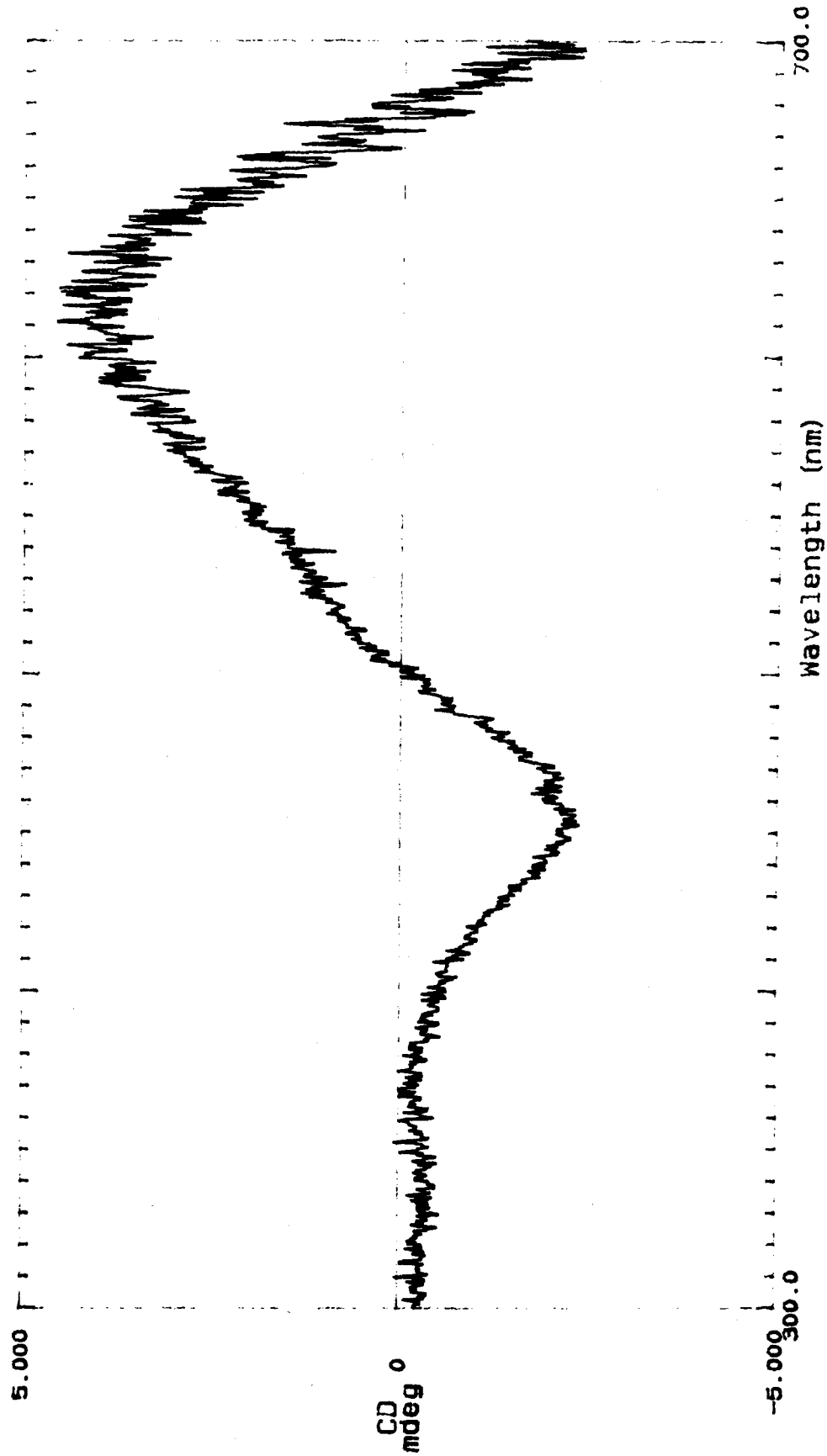
The spectrum of the Ni^{2+} substituted wild-type azurin resembles the published spectrum. Unfortunately, only the protons from the methyl group of methionine have been assigned. In the mutants lacking this methionine, the peak disappears, while for the Phe114A mutant, the peaks are retained. The copper center may be flexible in the mutants, creating a big difference between the NMR spectra of azurins with and without M 121.

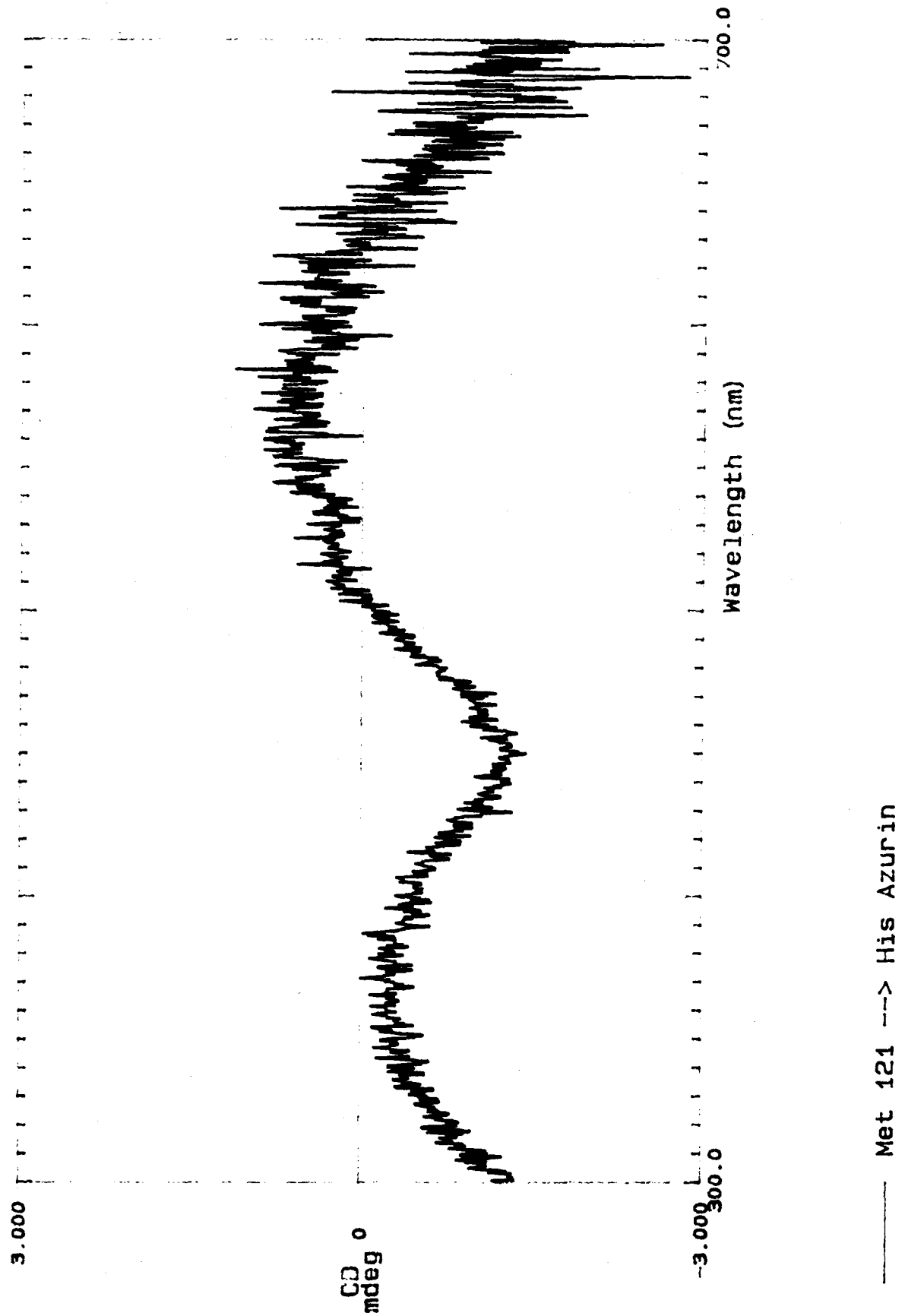
Figure 10-15

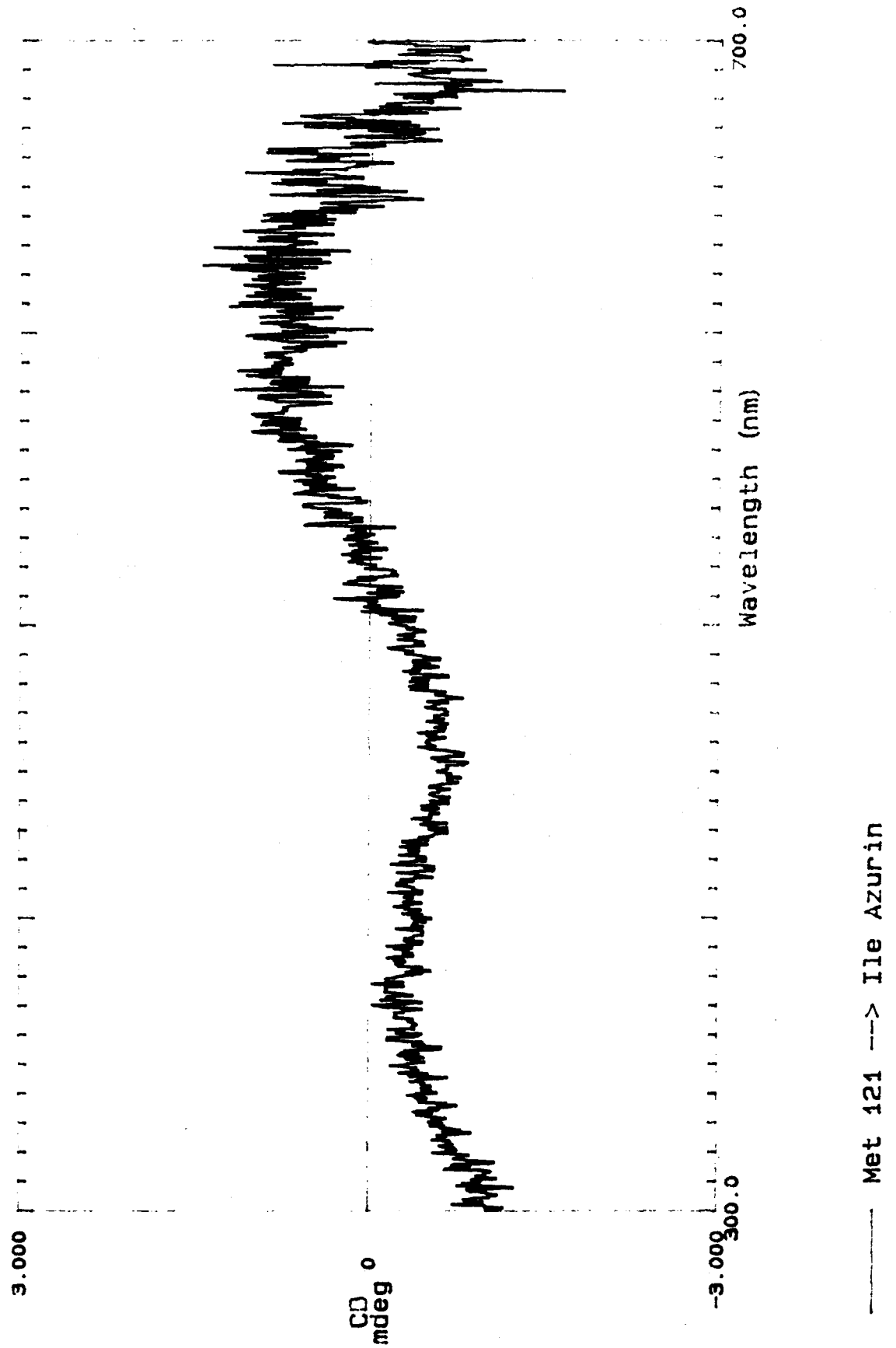
CD spectra of Methionine 121 variants (Met, Asn, Asp, His, Ile, Val) of azurin.

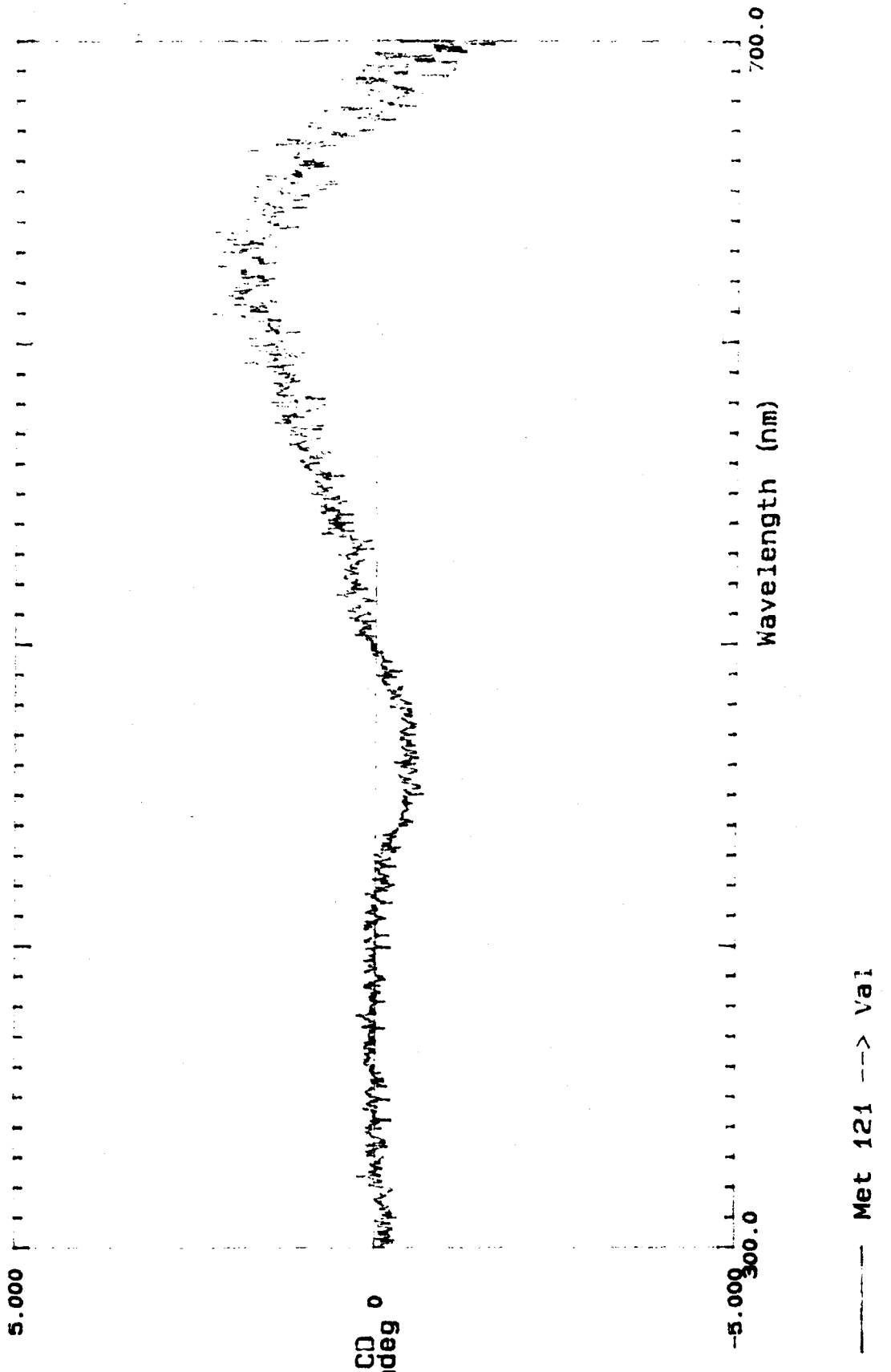

3.000
CD
mdeg 0
-3.000
300.0
700.0
Wavelength (nm)
WT Azurin (Met 121)


5.000
CD
mdeg 0
-5.000
300.0
700.0
Wavelength (nm)


5.000
CD
mdeg 0
-5.000
300.0
700.0
Wavelength (nm)
Met 121 --> Asp


3.000
CD
mdeg 0
-3.000
300.0
700.0
Wavelength (nm)
Met 121 --> His Azurin


3.000
CD
mdeg 0
-3.000
300.0
700.0
Wavelength (nm)
Met 121 --> Ile Azurin


5.000
CD
mdeg 0
-5.000
300.0
700.0
Wavelength (nm)
——— Met 121 --> Val

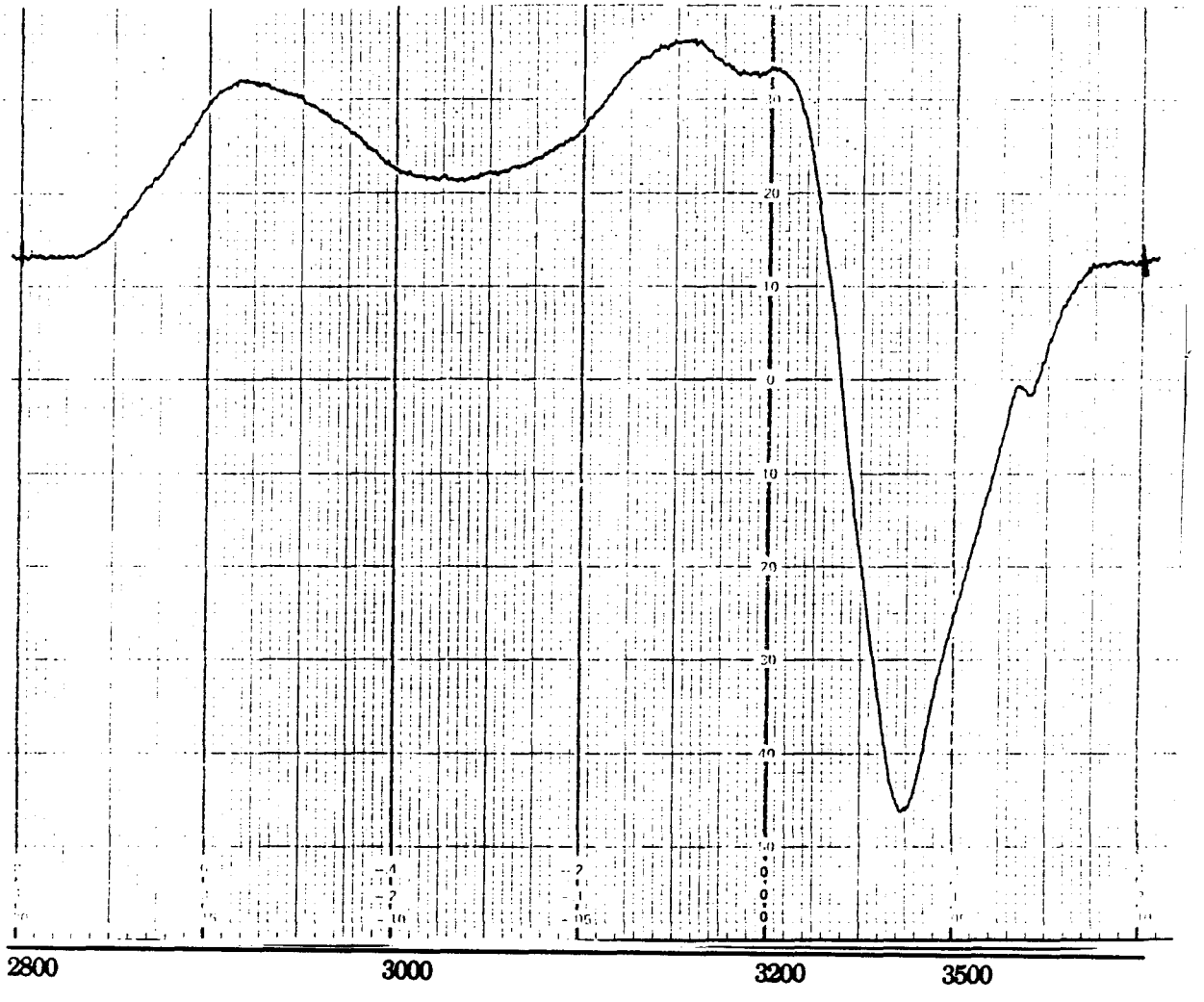


Figure 16. X-band EPR of Met121Val Azurin. υ = 9.3 GHz, T = 77°K.

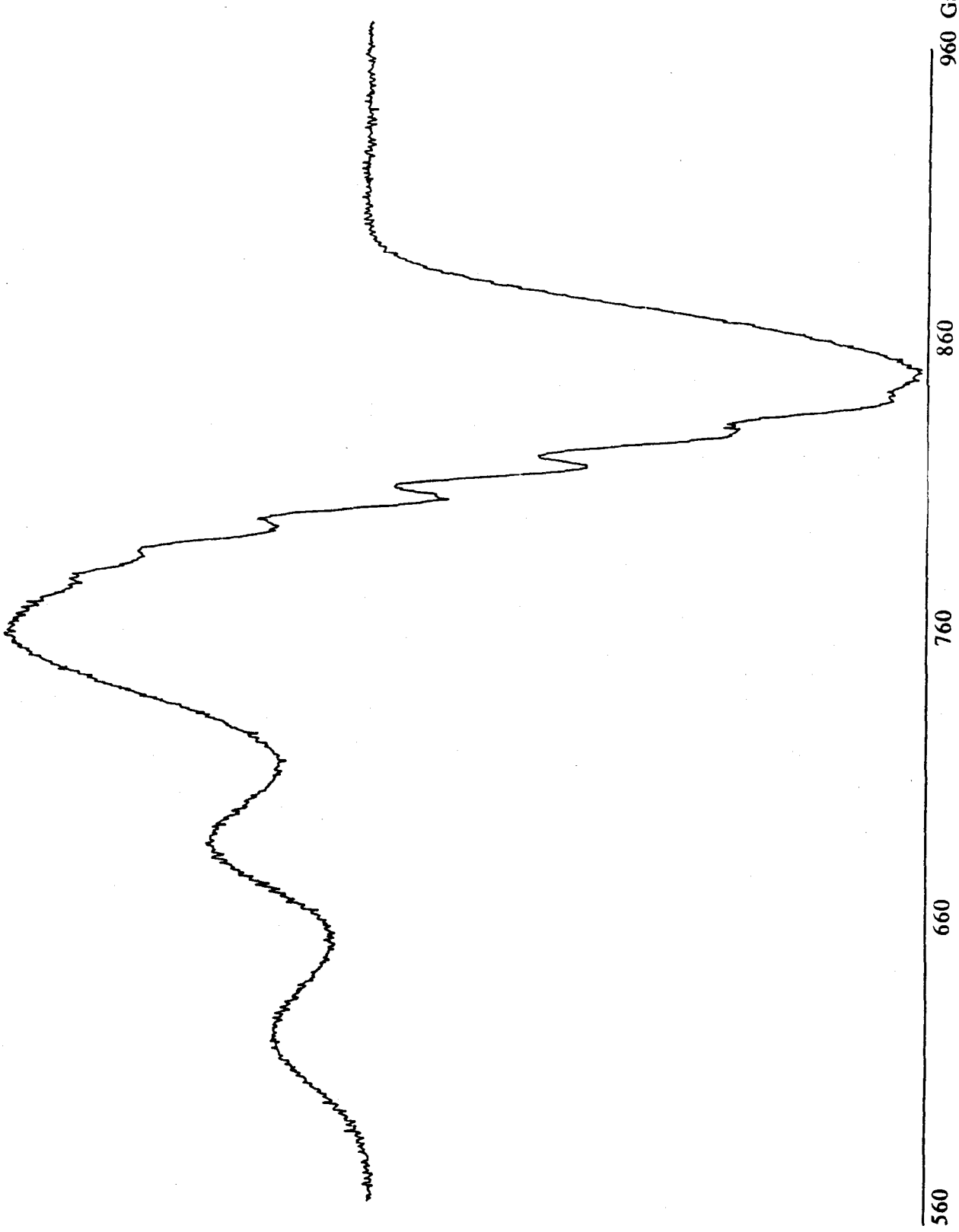


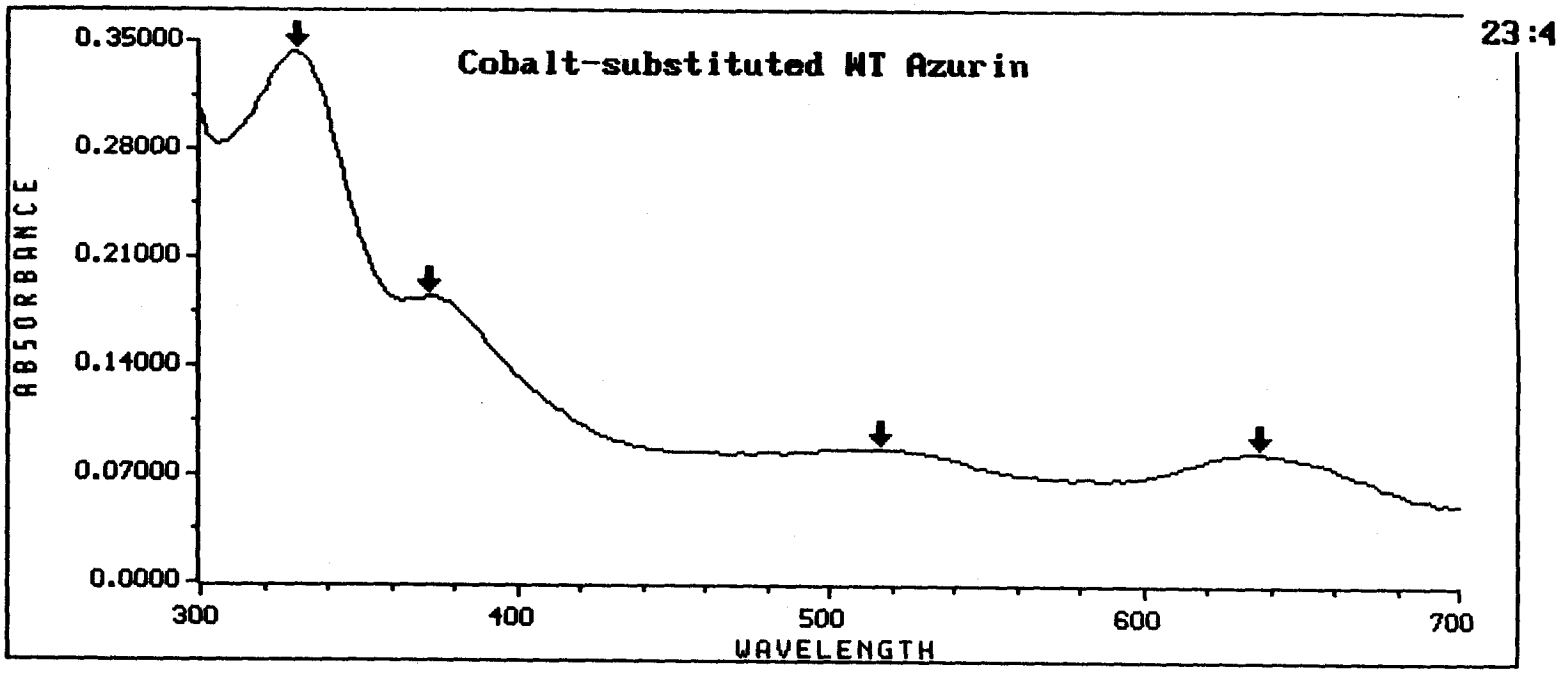
Figure 17. S-band spectrum of azurin with ^{65}Cu which drastically improves the resolution. u = 2.28 GHz, T = 16K

Table 5
Spin-Hamiltonian parameters for azurin mutants

Azurin	$g_{\|\|}$	$A_{\|\|}(10^{-4}$ cm-1)	g_{+}
Wild Type	2.27	53	2.04
M121N	2.249	36	2.056
M121I	2.246	35	2.060
M121V	2.243	34	2.064

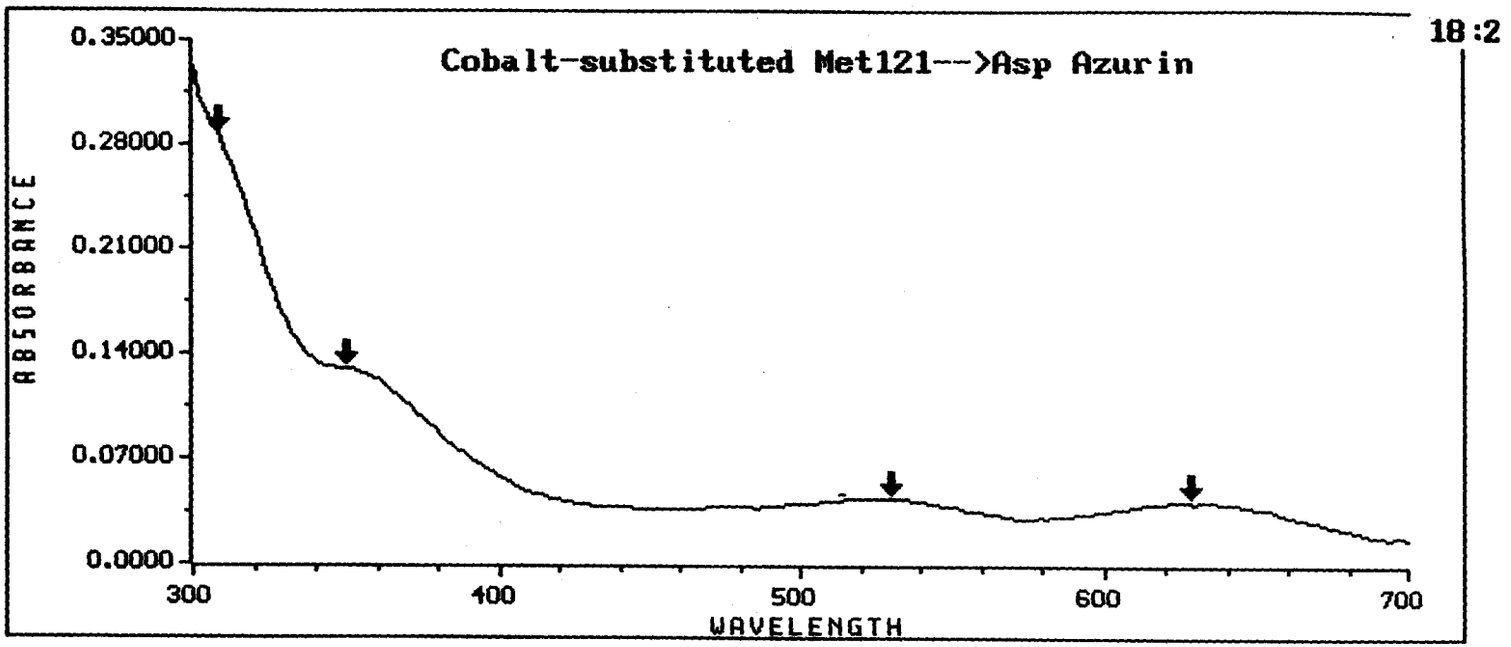
Figures 18-23

UV-Vis spectra of Cobalt-substituted azurins : M121, M121D, M121G, M121L, triple mutant, F114A



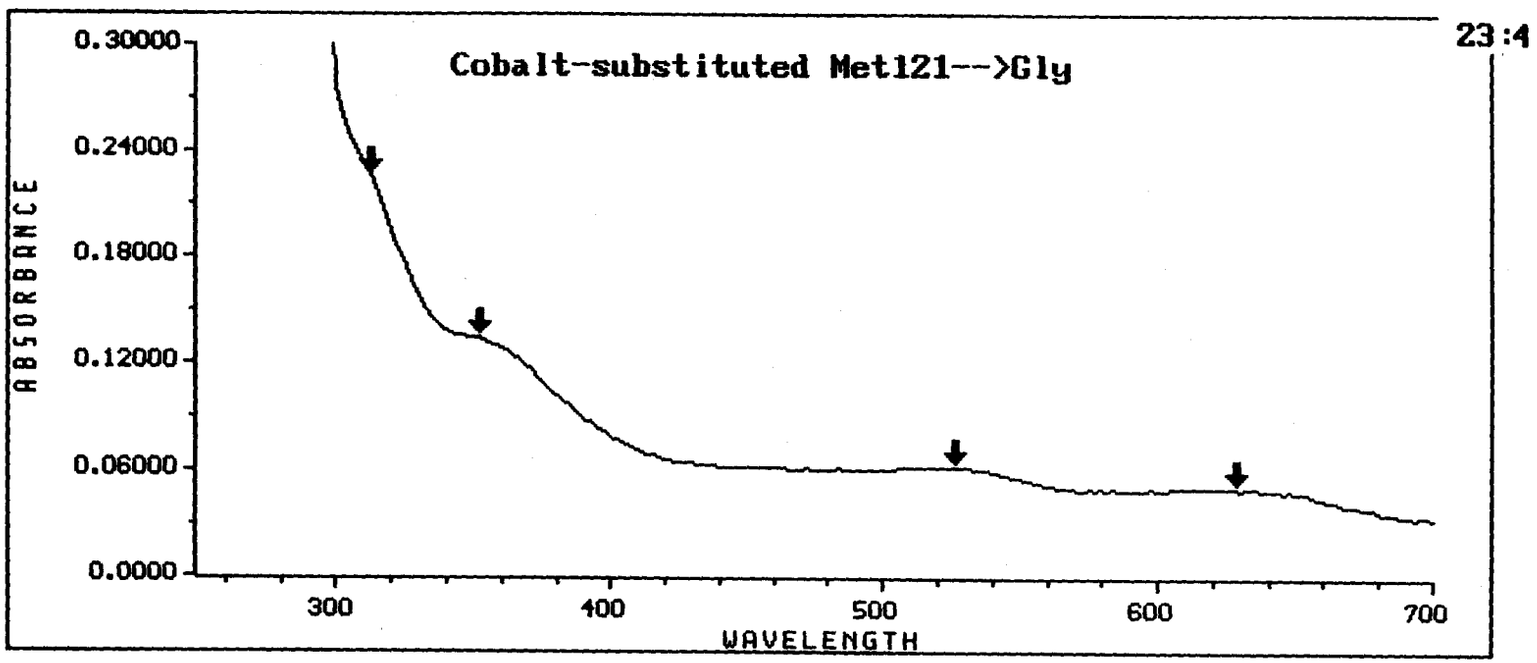
Annotated Wavelengths:

1 :	Wavelength = 330	Result =	0.343033
2 :	Wavelength = 372	Result =	0.184311
3 :	Wavelength = 516	Result =	0.087357
4 :	Wavelength = 636	Result =	0.084808



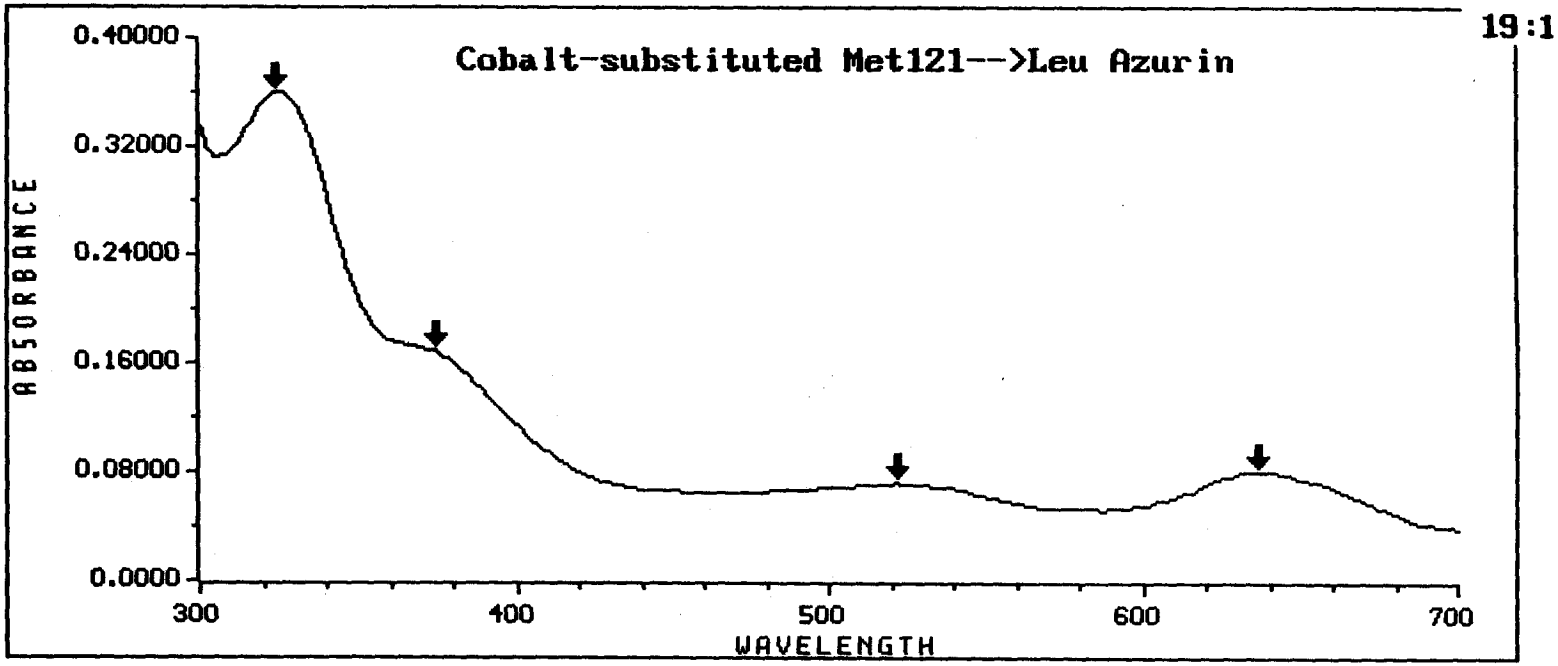
Annotated Wavelengths:

1 :	Wavelength = 628	Result =	0.041016
2 :	Wavelength = 530	Result =	0.043182
3 :	Wavelength = 350	Result =	0.130005
4 :	Wavelength = 308	Result =	0.286011



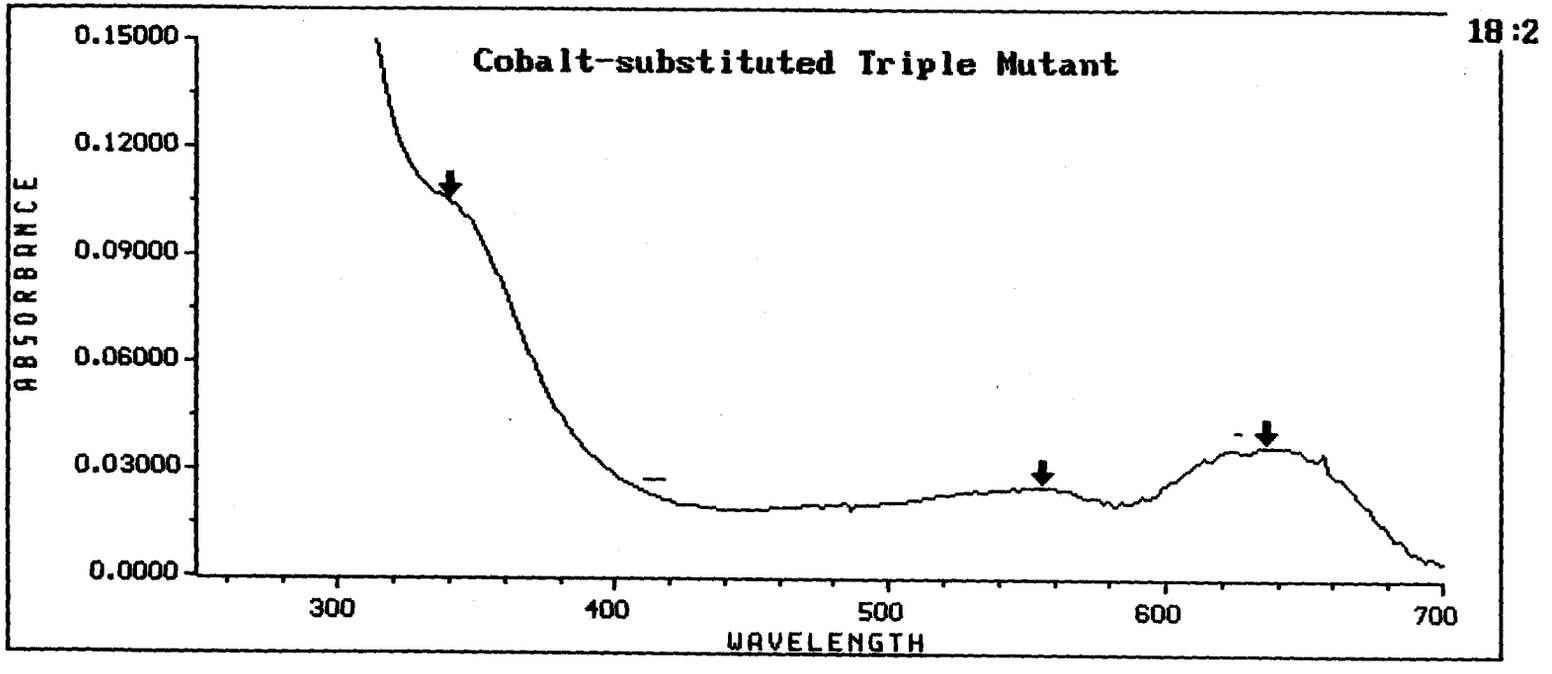
Annotated Wavelengths:

1 :	Wavelength = 312	Result =	0.225525
2 :	Wavelength = 352	Result =	0.133575
3 :	Wavelength = 526	Result =	0.061234
4 :	Wavelength = 628	Result =	0.049622



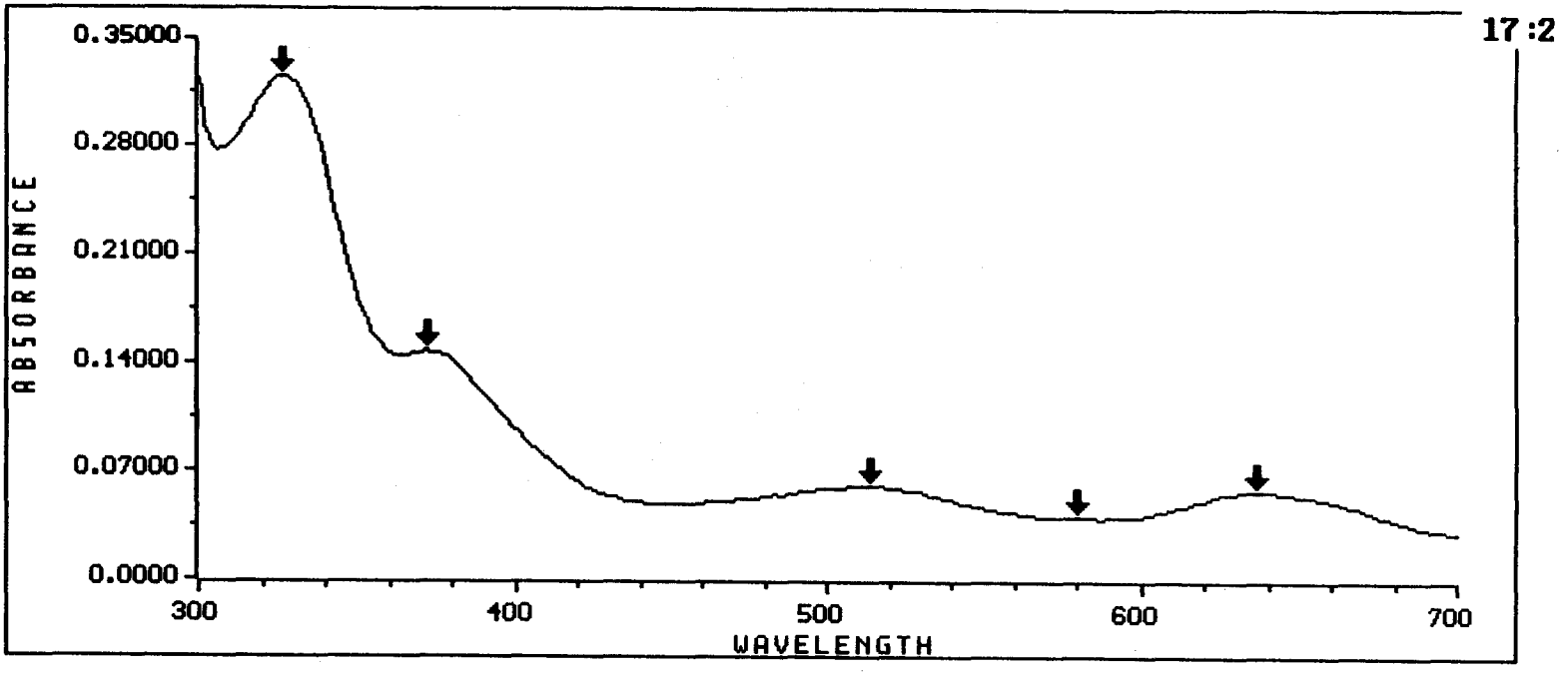
Annotated Wavelengths:

1 :	Wavelength = 324	Result =	0.360809
2 :	Wavelength = 374	Result =	0.168594
3 :	Wavelength = 522	Result =	0.071609
4 :	Wavelength = 636	Result =	0.080673



Annotated Wavelengths:

1 :	Wavelength = 340	Result =	0.105301
2 :	Wavelength = 556	Result =	0.024872
3 :	Wavelength = 636	Result =	0.036575



Annotated Wavelengths:

1 :	Wavelength = 326	Result =	0.325821
2 :	Wavelength = 372	Result =	0.147720
3 :	Wavelength = 514	Result =	0.059555
4 :	Wavelength = 580	Result =	0.039795
5 :	Wavelength = 636	Result =	0.056564

Figure 23. UV-Vis spectrum of Co F114Ala azurin.

Figures 24-28

UV-Vis spectra of Nickel-substituted azurins: M121, M121D, M121G, M121L, triple mutant, and F114A

Figure 24. UV-Vis spectrum of Ni Met121azurin.

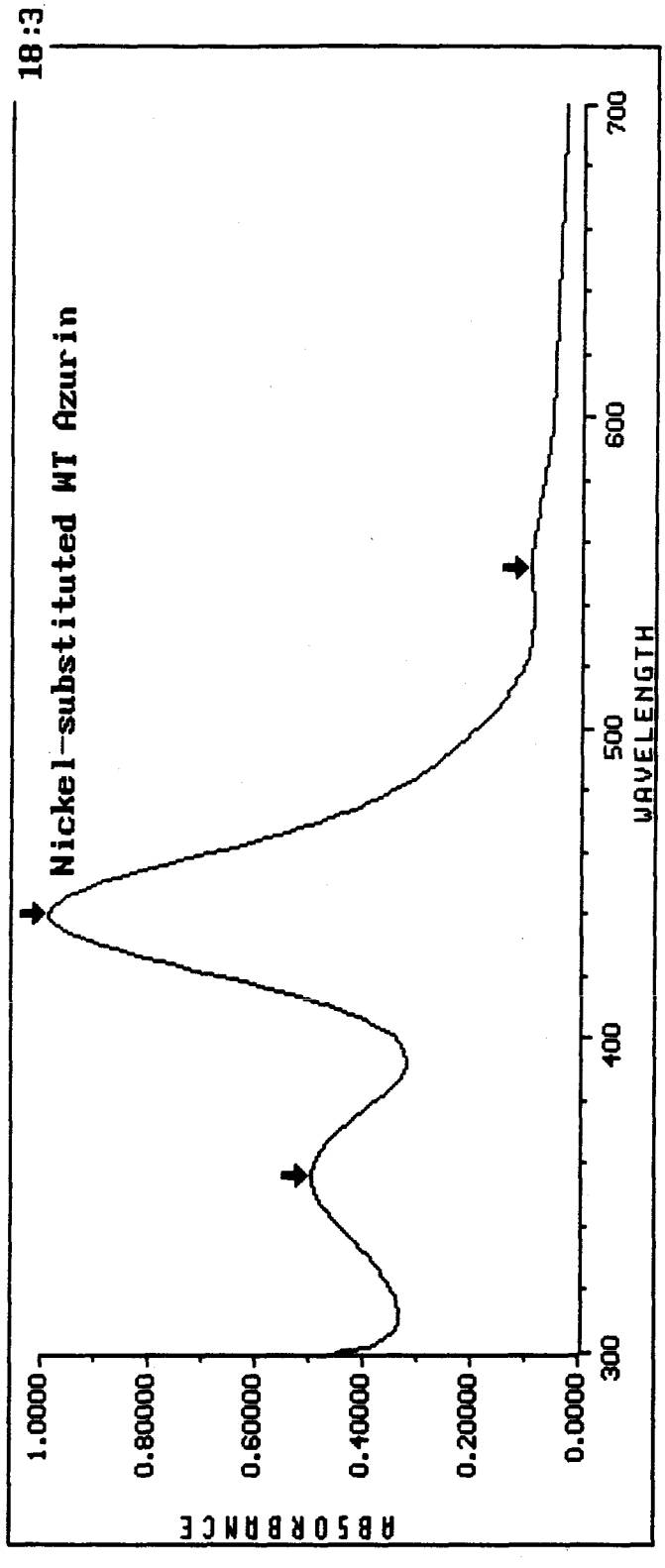


Figure 25. UV-Vis spectrum of Ni Met121Asp azurin.

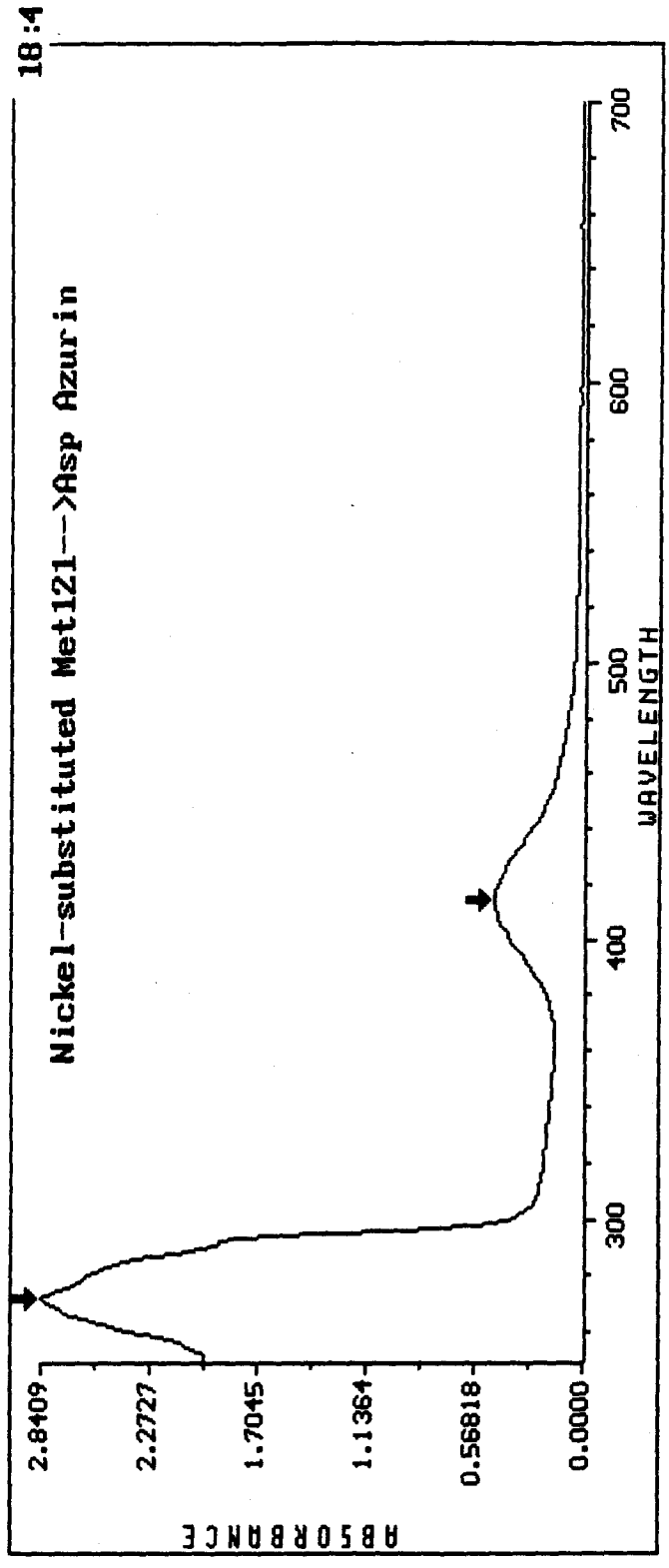


Figure 26. UV-Vis spectrum of Ni Met121Gly azurin.

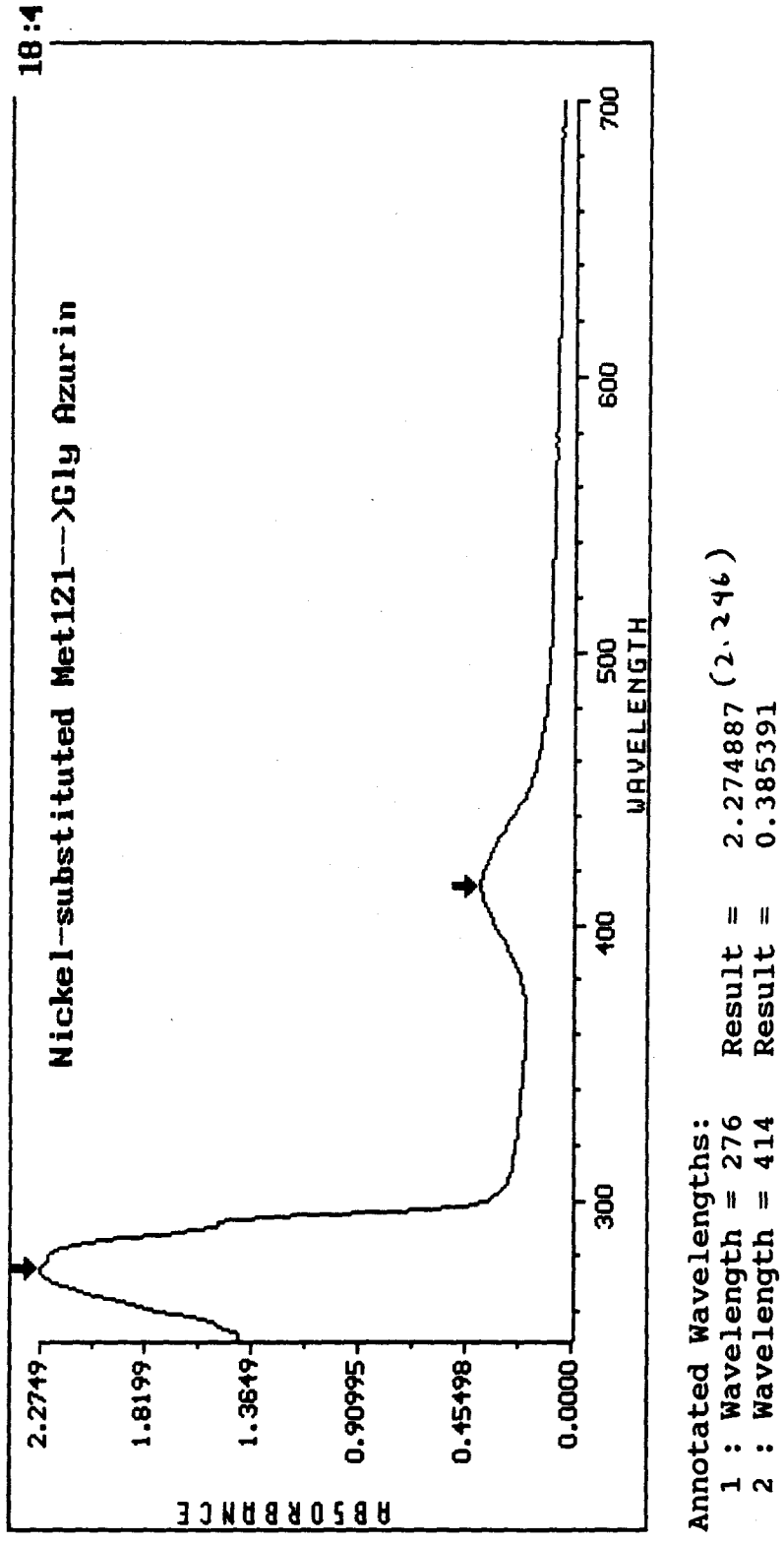


Figure 27. UV-Vis spectrum of Ni Met121Leu azurin.

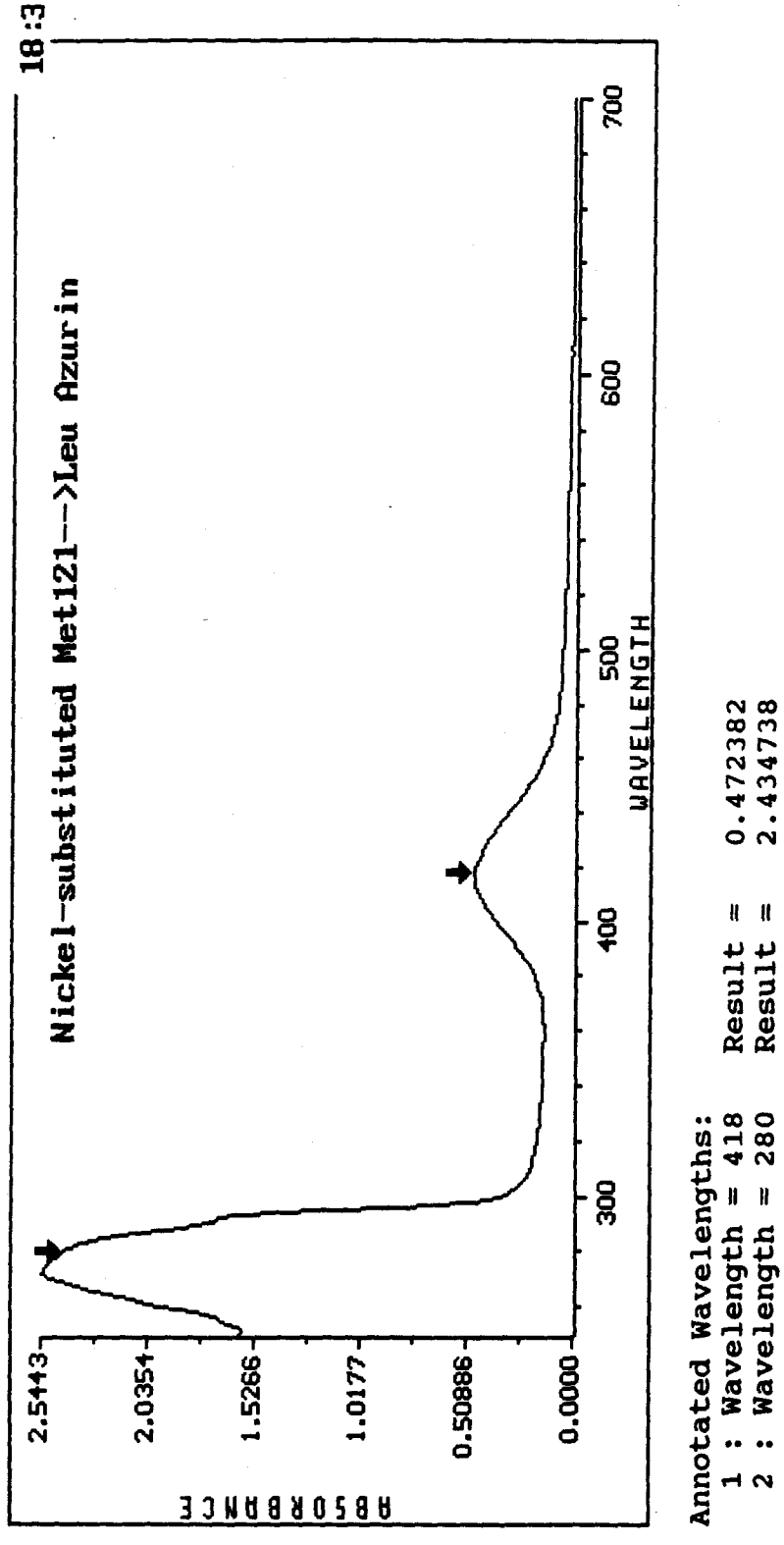


Figure 28. UV-Vis spectrum of Ni triple mutant azurin

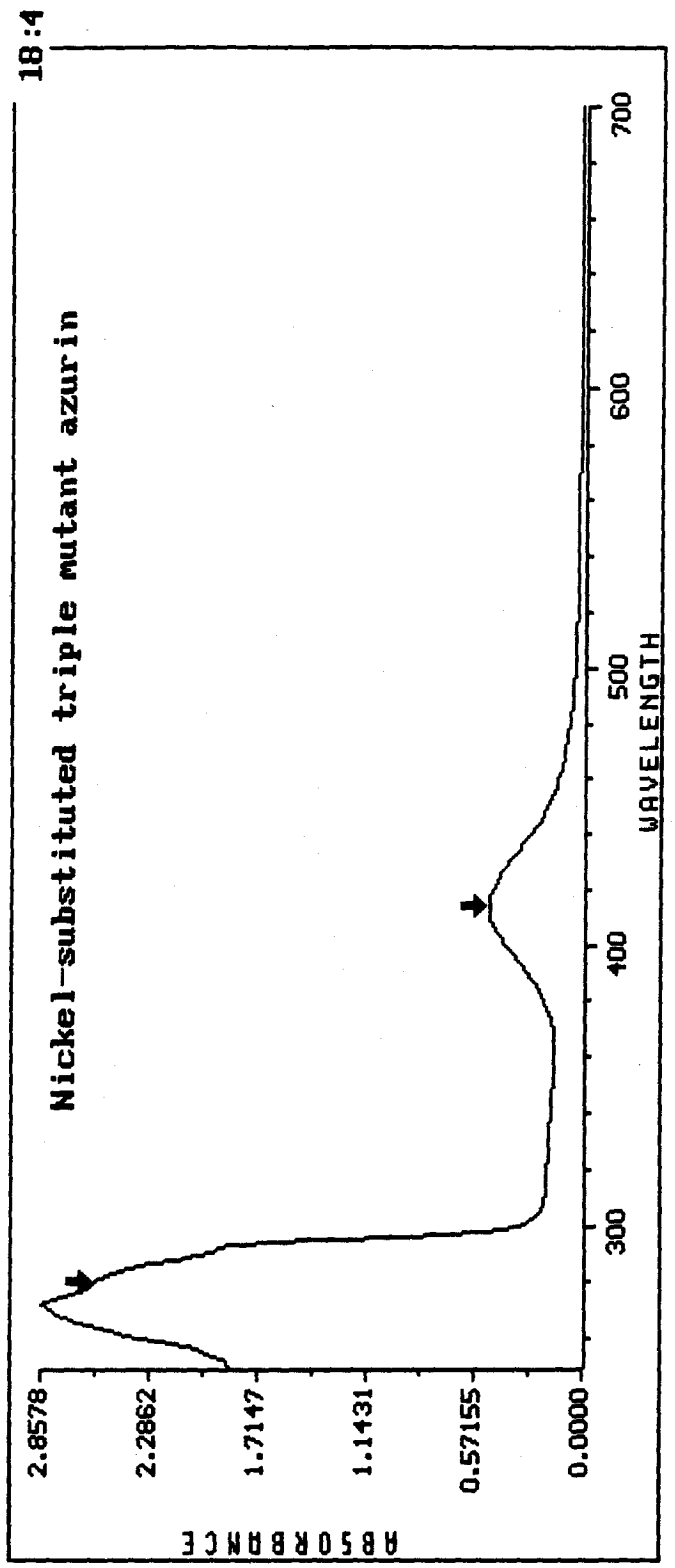


Figure 29. UV-Vis spectrum of Ni Phe114Ala azurin.

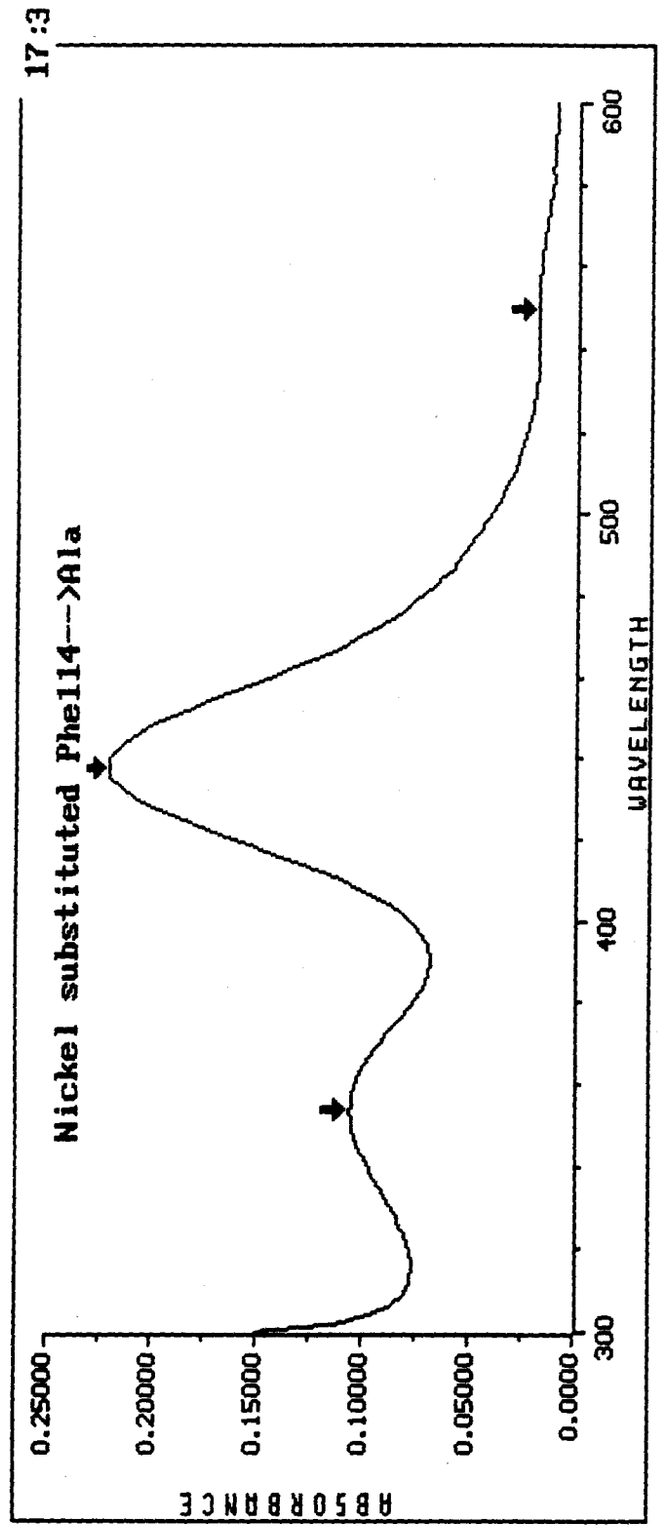


Table 6

Peaks in the Visible range of Ni- and Co-substituted azurin mutants

	Cobalt			Nickel		
Mutant	peak 1	peak 2	peak 3	peak1	peak 2	peak 3
Wild type	330	372	636	356	440	552
M121D	350	530	628		414	
M121G	352	526	628		414	
M121L	374	522	636		418	
Triple	340	556	636		414	
F114A	372	514	636	354	438	550

Figure 30

CD spectra of Nickel-substituted azurins

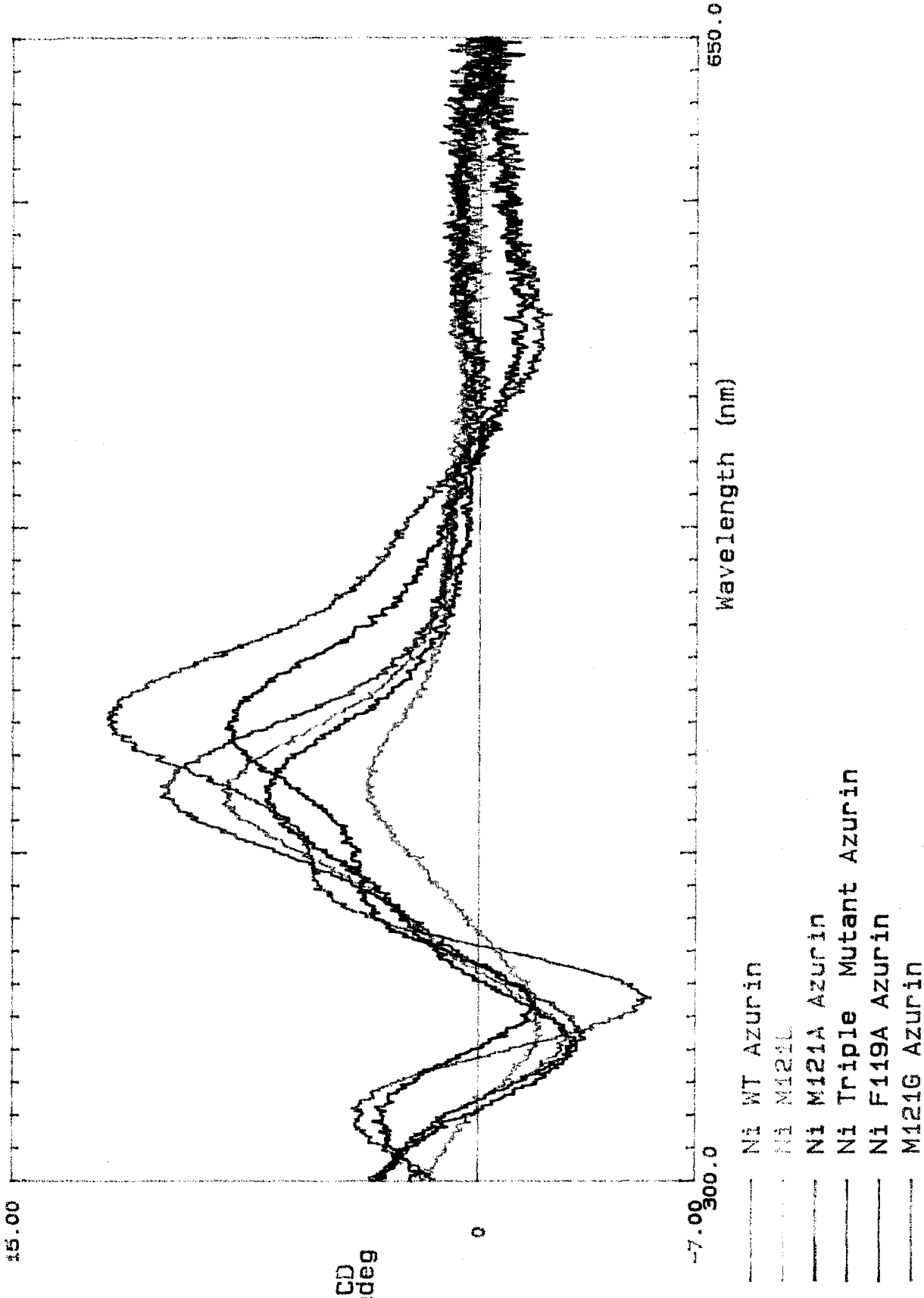

15.00
0
-7.00
CD
mdeg
300.0
650.0
Wavelength (nm)
Ni WT Azurin
Ni M121L
Ni M121A Azurin
Ni Triple Mutant Azurin
Ni F119A Azurin
M121G Azurin

Figure 31. NMR of Nickel M121Asp azurin

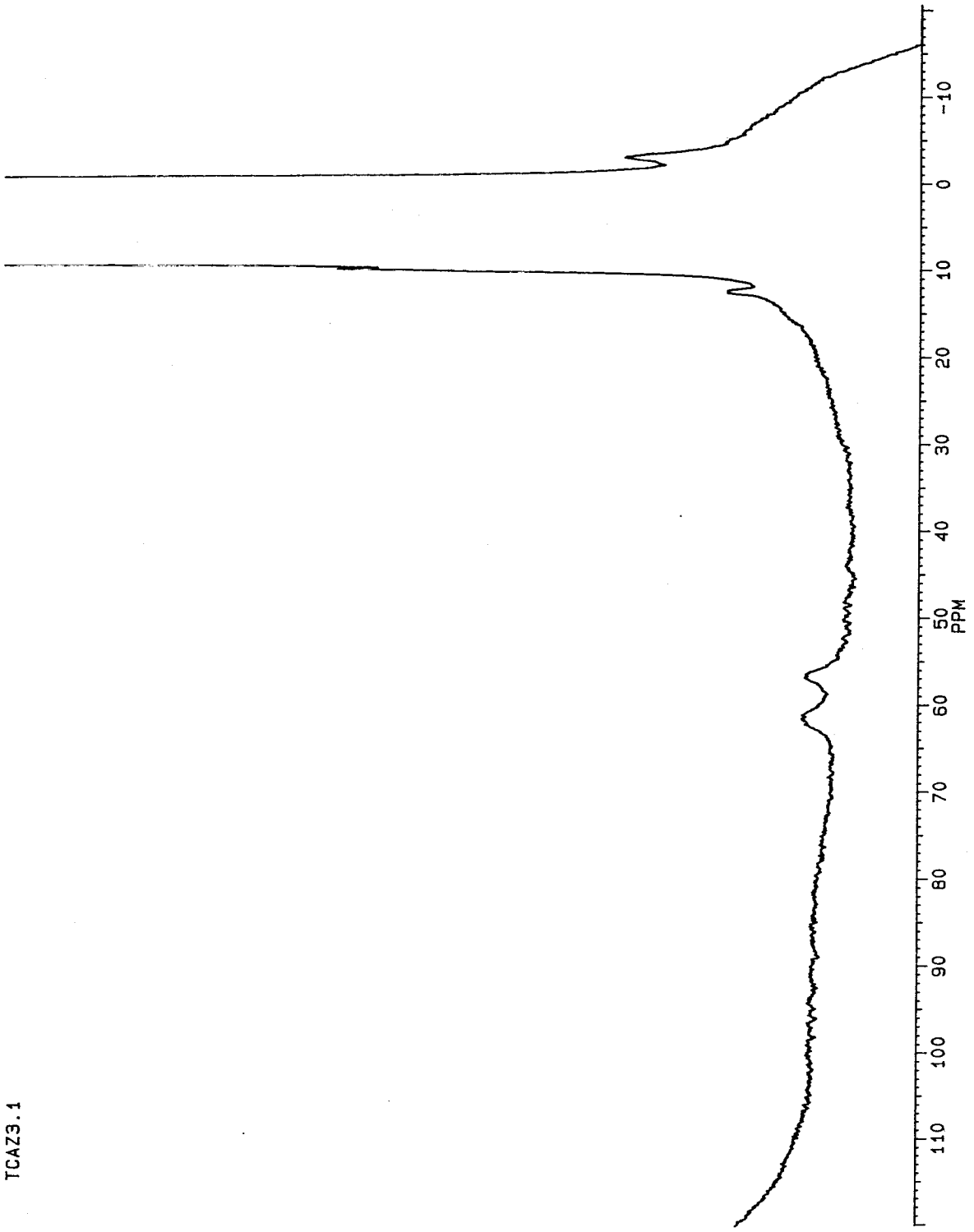


Figure 32. NMR of Nickel M121Gly azurin

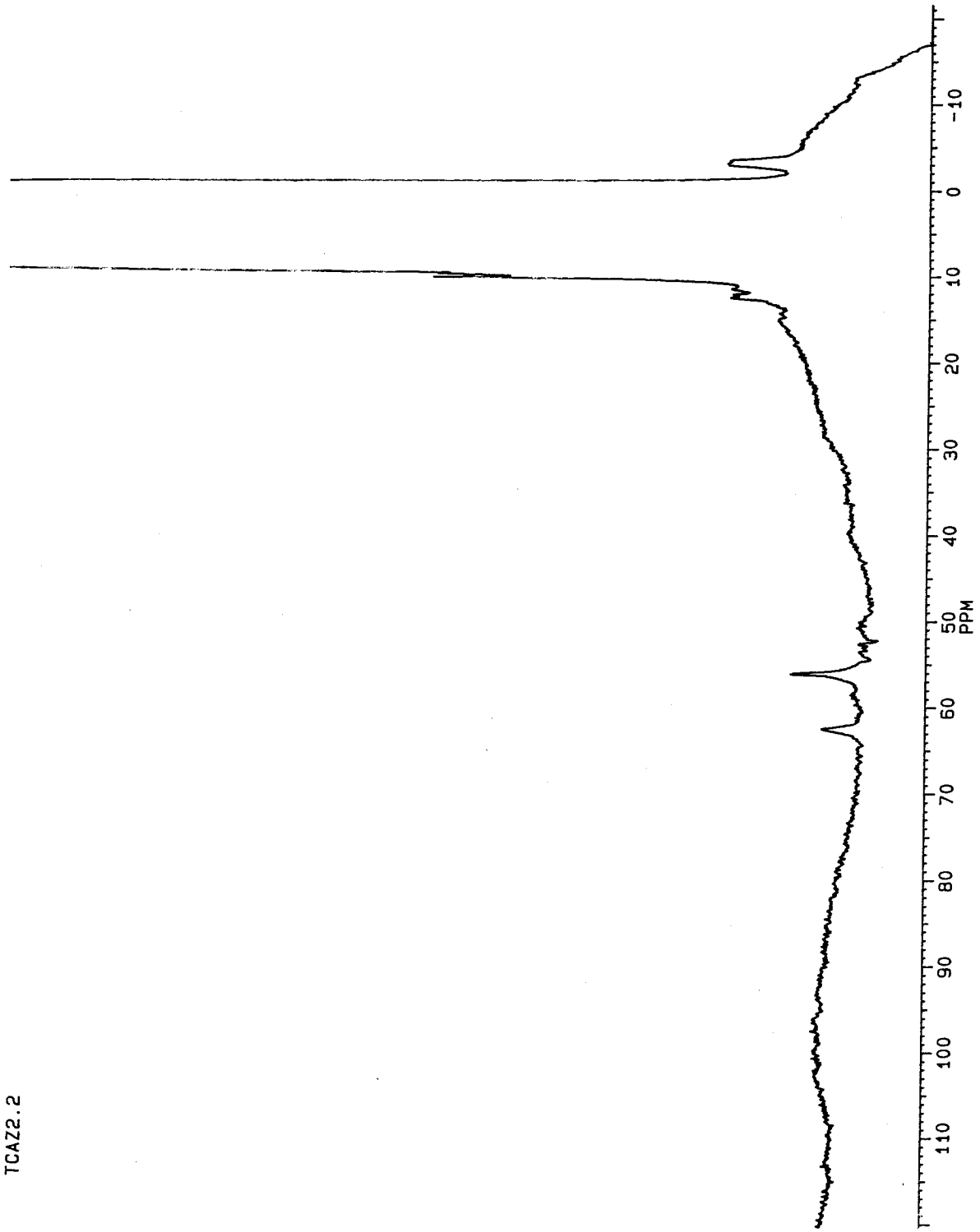


Figure 33. NMR of Nickel triple mutant of azurin

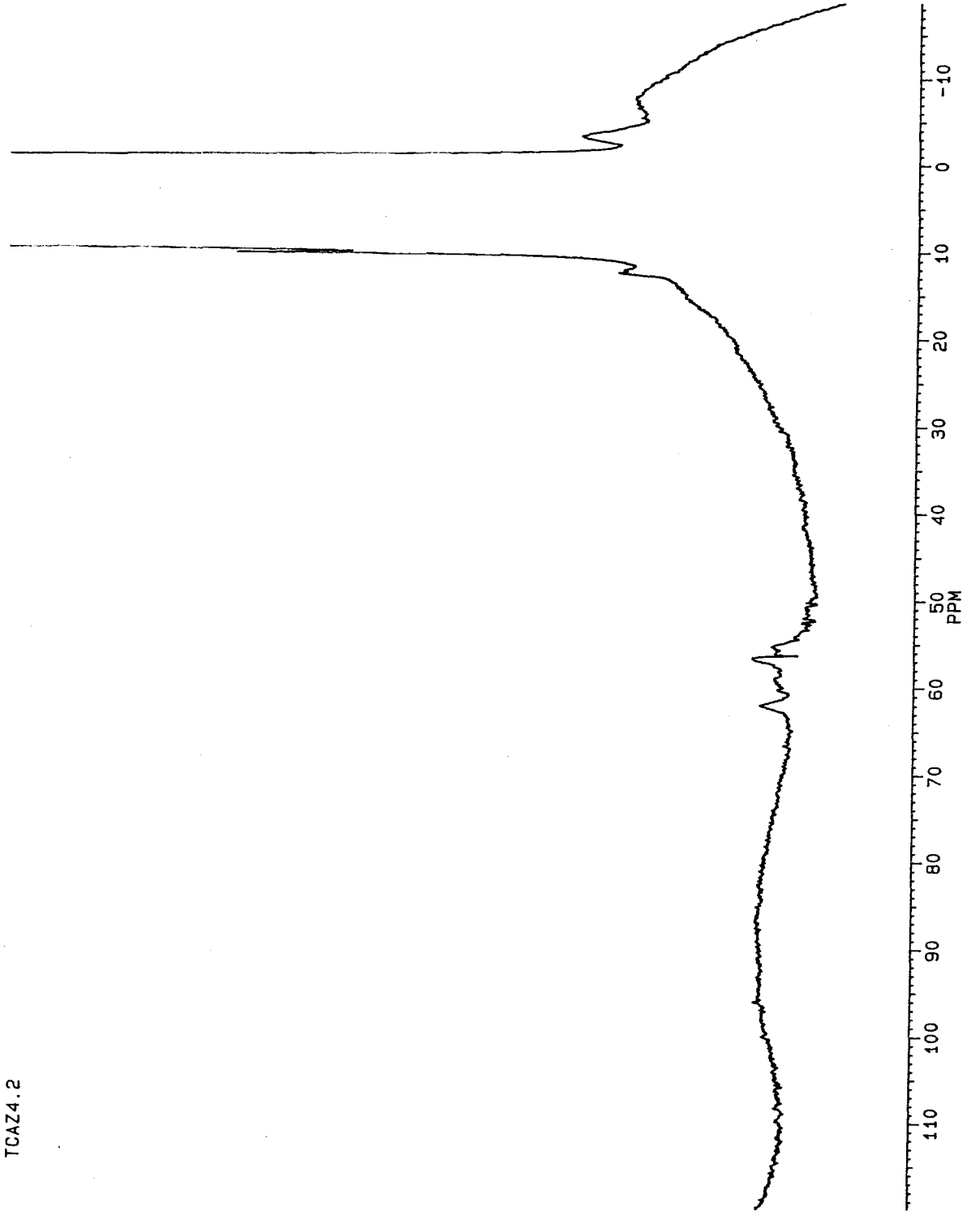
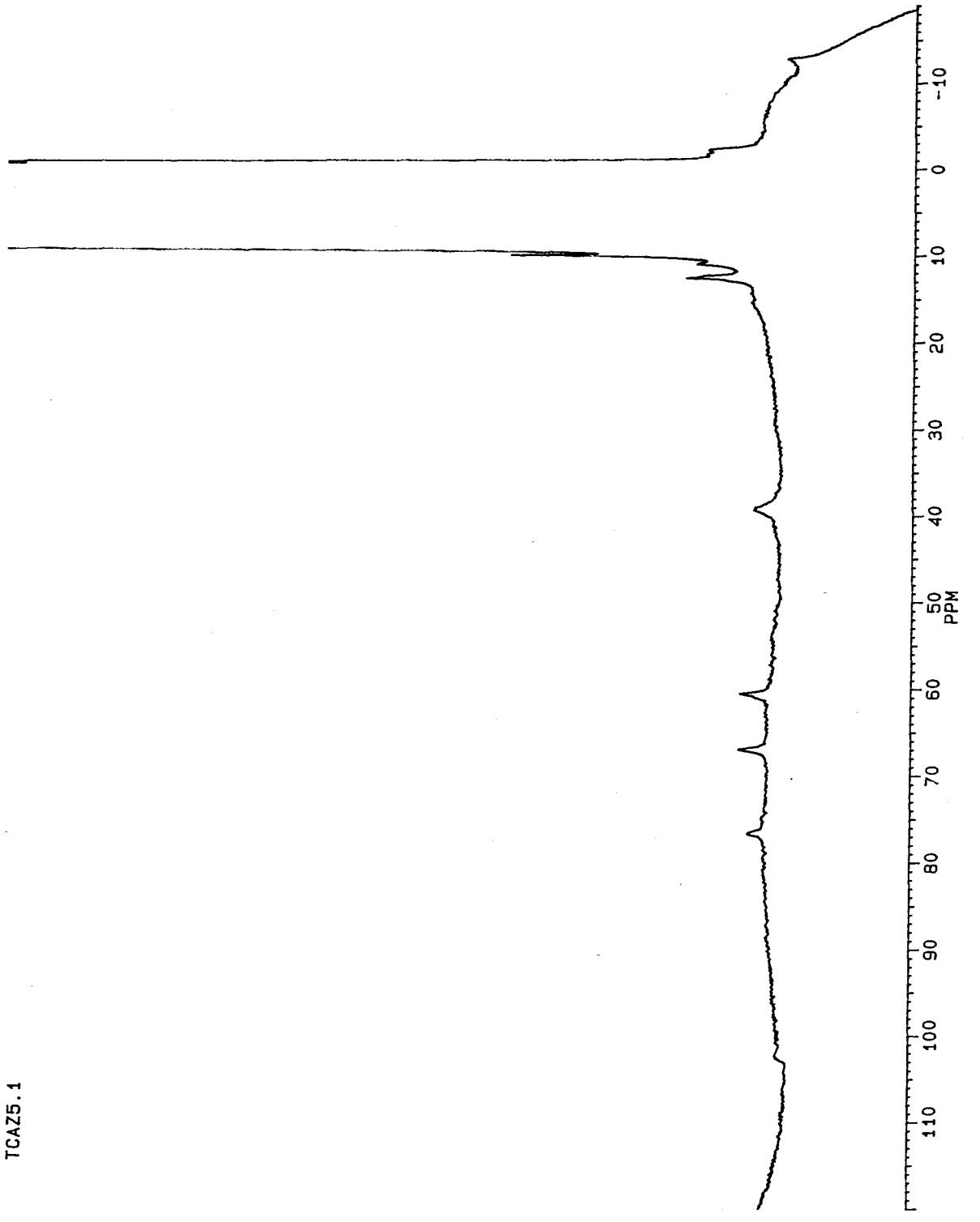


Figure 34. NMR of Nickel F114A azurin



DISCUSSION

Site-saturation cassette mutagenesis was used to create all other 19 amino acid variants as well as the amber-stop codon in one ligation experiment. Surprisingly, all the mutants were stable since they could be stained on Western blot. That all the mutants at Met 121 (Gly, Leu, Asp, Triple, Asn, Ile, His, Val) showed the characteristic blue band was an even bigger surprise since methionine was absolutely conserved in all azurins and plastocyanins. Stellacyanin, however, is one notable exception among the blue copper proteins.[49] Although its structure is not yet known, it lacks methionine in its amino acid composition and therefore cannot use methionine as a ligand. Thus, the methionine does not contribute significantly to the electronic signature of azurin with copper. Interestingly, it was observed during the metal substitution experiments that the metal could be removed or inserted at a much faster rate into the mutants than the wild-type. Therefore, the role of Met 121 must be to give stability to the copper center and to fine-tune the redox potential of azurin.

The nature of the copper center when nickel or cobalt has been substituted is more puzzling. For a long time, the 560 nm peak in the CD spectrum of Ni^{2+}-substituted azurin has been assigned to the πS (Cys)->Cu transition. Yet, the only mutant that retains this band is the one that has methionine (Phe114Ala). Therefore, when nickel is present in the copper center of an azurin, the methionine is most likely coordinating to the metal. Another evidence for this comes from the visible absorbance of the Nickel-substituted azurins. Only the two sample (WT and F114A) show bands at 350 and 550 nm. These absorbances must be from methionine-to-copper charge-transfer. Furthermore, the shape of the entire spectra changes

rather dramatically with nickel. Therefore, it is possible that without Met 121, the copper site is flexible enough to accommodate nickel in a square planar geometry. These conclusions may also explain the radical changes in the NMR spectra upon deletion of methionine. To confirm this, a magnetic susceptibility experiment is awaiting to be performed. This experiment will reveal the geometry adopted by nickel. To further investigate the copper center in the variants, a Resonance Raman experiment is in the plans. These experiments will allow us to reexamine the identities of the charge-transfer bands.

Last but not the least, His83Gln azurin has been constructed. This removes the only surface-accessible histidine from azurin. Consequently, a series of mutants that introduces surface-accessible histidines can be made to investigate, among other variables, the distance dependency of electron transfer. As shown in Figure 35, Lys103His mutant may clarify the role of aromatic residues; Tyr108 may serve as a short cut to the copper center, explaining why this residue is conserved in all azurins. Two separate series of mutants are proposed in Figures 36 and 37, one coupled to Cys 112 and one coupled to Met 121. The former set of mutants are being currently made by another member of the group.[50] These two sets of mutants should prove to be very useful in answering some long-standing questions about the mechanisms of electron transfer.

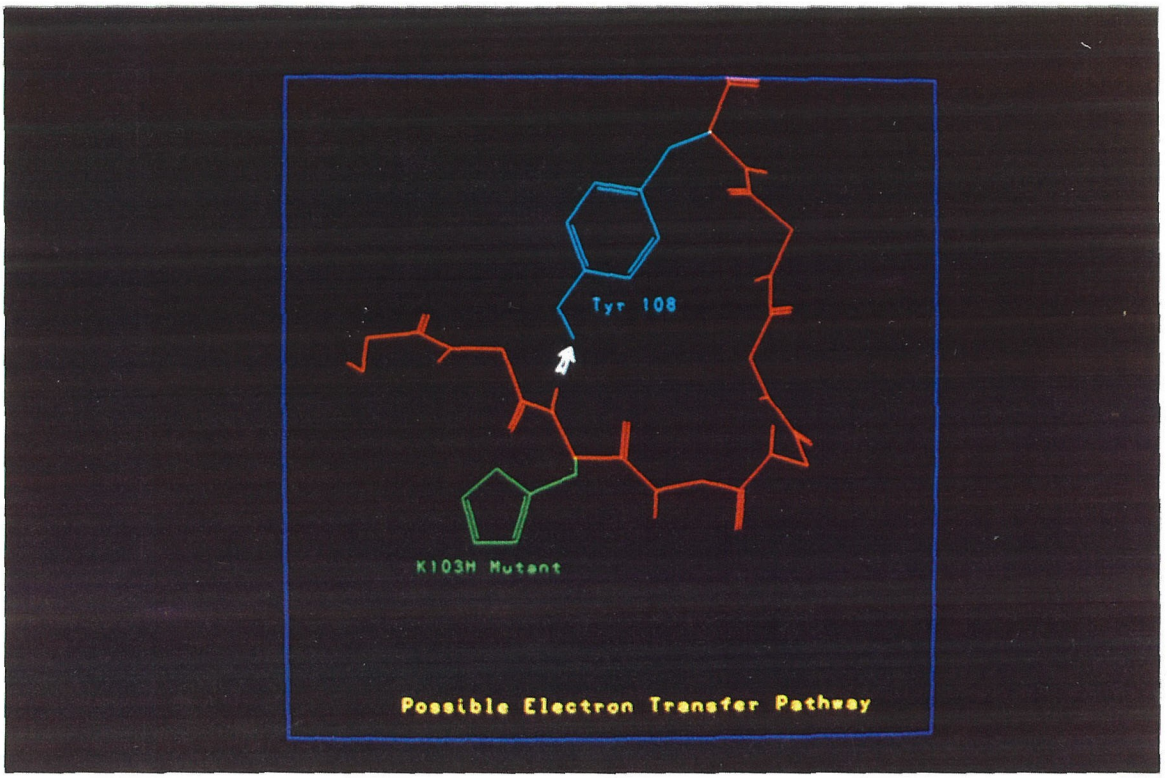


Figure 35. Proposed shortcut pathway of electron from K103H to Cu through Tyr 108.

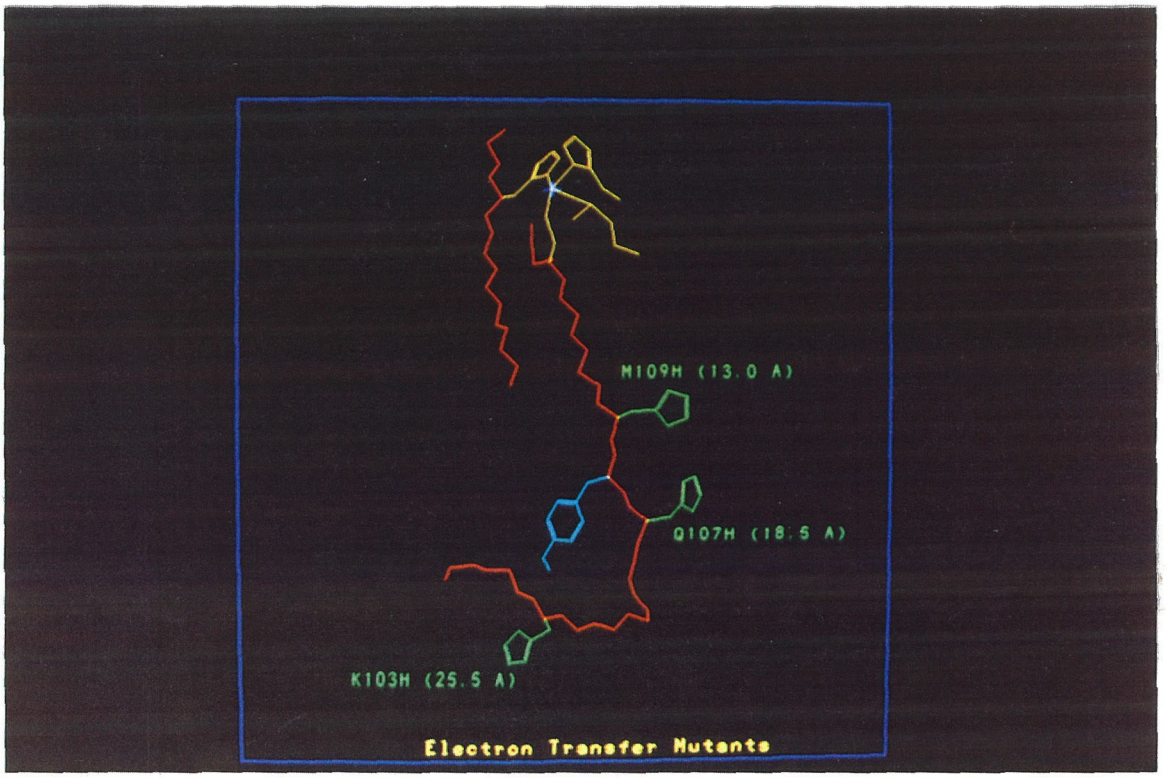


Figure 36. Surface histidine mutants which has histidines at K103, Q107, and M109. The edge to edge distances (imidazole to copper) are shown.

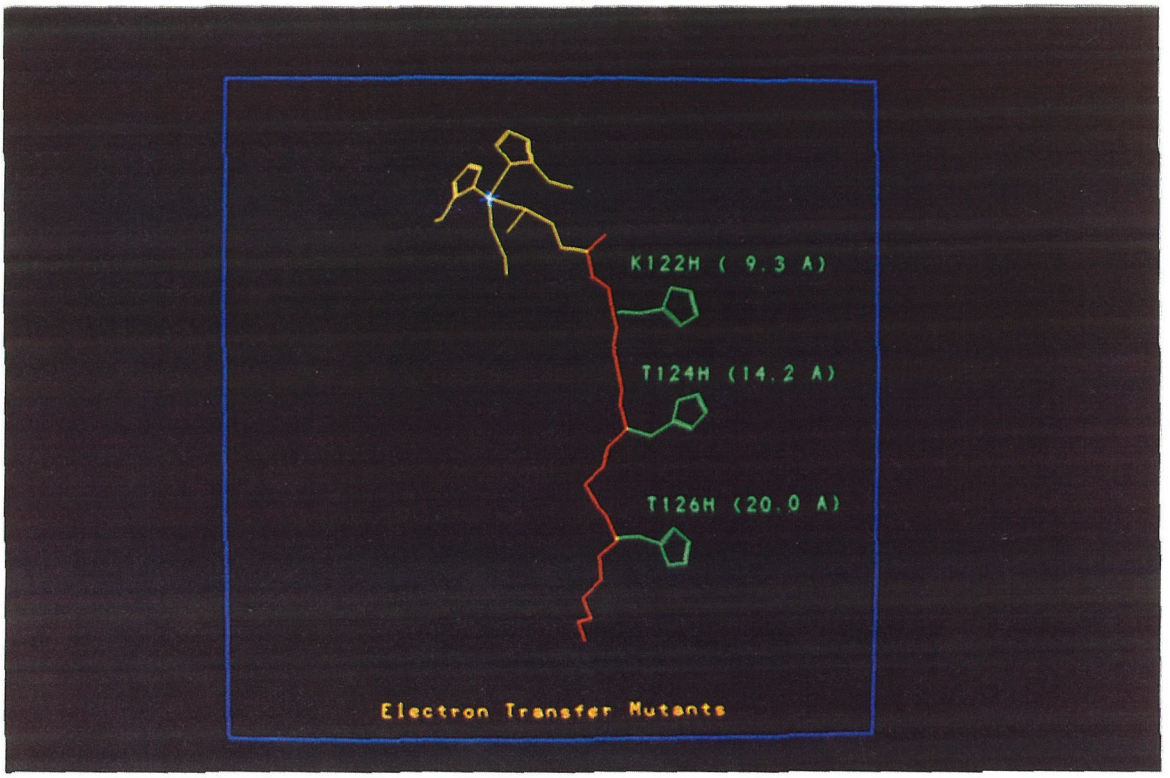


Figure 37. Surface histidine mutants which has histidines at K122, T124, and T126. The edge to edge distances (imidazole to copper) are shown.

REFERENCES

1. Zannoli, D. (1989) *Biochim. Biophys. Acta.* **975**, 299-316.
2. Gray, H. B. and Solomon, E. I. (1981) *In Metal Ions in Biology*, ed. Spiro, T. G. (Wiley, New York), vol. 3, pp. 1-39.
3. Adman, E. T. (1985) in *Topics in Molecular and Structural Biology, Metalloproteins*, ed. Harrison, P. M. (Verlag Chemie, Weinheim), vol. 6, pp. 1-42.
4. Solomon, E. I., Hare, J. W. and Gray, H. B. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 1389-1393.
5. Ugurbil, K., Norton, R. S., Allherhared, A. and Bersohn, R. (1977) *Biochemistry*, **16**, 886-892.
6. Adman, E. T. and Jensen, L. H. (1981) *Isr. J. Chem.* **21**, 8-12.
7. Baker, E. N. (1988) *J. Mol. Biol.* **203**, 1071-1095.
8. Shepard, W. E. B., Anderson, B. F., Lewandaski, D. A., Norris, G. E. and Baker, E. N. (1990) *J. Am. Chem. Soc.* **112**, 7817-7819.
9. Solomon, E. I., Hare, J. W., Dooley, D. M., Dawson, J. H., Stephens, P. J. and Gray, H. B. (1980) *J. Amer. Chem. Soc.* **102**, 168-178.
10. Canters, G. W., Hill, H. A. O., Kitchen, N. A. and Adman, E. T. (1984) *Eur. J. Biochem.* **138**, 141-152.
11. Blair, D. F., Campbell, G. W., Schoonover, J. R., Chan, S. I., Gray, H. B., Malmstrom, B. G., Pecht, I., Swanson, B. I., Woodruff, W. H., Cho, W. K., English, A. M., Fry, H. A., Lum, V. and Norton, K. A. (1985) *J. Am. Chem. Soc.* **107**, 5755-5766.
12. Belhadj, M., Jean, J. M., Friesner, R. A., Schoonover, J. and Woodruff, W. H. (1990) *J. Phys. Chem.* 94, 2160-2166.

13. Loehr, T. M. and Sanders-Loehr, J. 1984) in *Copper Proteins and Copper Enzymes*, ed. Lontie, R. (CRC Press, Florida), Vol. 1, pp. 115-185.

14. Boas, J. F. (1984) 1984) in *Copper Proteins and Copper Enzymes*, ed. Lontie, R., (CRC Press, Florida), Vol. 1, pp. 8-62.

15. St. Clair, C. S. (1989) Ph.D. Thesis, California Institute of Technology.

16. St. Clair and Gray, H. B., manuscript in preparation.

17. Tennent, D. L. and McMillin, D. R. (1979) *J. Am. Chem. Soc.* **101**, 2307-2311.

18. Engeseth, H. R. and McMillin, D. R. (1986) *Biochemistry* **25**, 2448-2455.

19. Engeseth, H. R., McMillin, D. R. and Otvos, J. D. (1984) *J. Biol. Chem.* **259**, 4822-4826.

20. Blaszak, J. A., Ulrich, E. L., Markley, J. L. and McMillin, D. R. (1982) *Biochemistry* **21**, 6253-6263.

21. Blaszak, J. A., McMillin, D. R., Thorton, A. T. and Tennent, D. L. (1983) *J. Biol. Chem.* **258**, 9886-9892.

22. Kostic, N. M., Margalit, R., Che, C.-M. and Gray, H. B. (1983) *J. Am. Chem. Soc.* **105**, 7765-7767.

23. Margalit, R., Kostic, N. M., Che, C.-M., Blair, D. F., Chiang, H.-J., Pecht, I., Shelton, J. B., Shelton, J. R., Schroeder, W. A. and Gray, H. B. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6554-6558.

24. Jackman, M. R., McGinnis, J., Power, R., Salmon, G. A. and Sykes, A. G. (1988) *J. Am. Chem. Soc.* **110**, 5880-5887.

25. Liang, M., Mauk, A. G., Pielak, G. J., Johnson, J. A., Smith, M. and Hoffman, B. M. (1988) *Science* **240**, 311-313.

26. Betts, J. and Beratran, D., Pathway Program, California Institute of Technology.

27. Bowler, B. E., Meade, T. J., Mayo, S. L., Richards, J. H. and Gray, H. B. (1989) *J. Am. Chem. Soc.* **111**, 8757-8759.

28. Farver, O. and Pecht, I. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6968-6972.

29. Gray, H. B. (1985) *Chem. Soc. Rev.* **15**, 17-30.

30. Marcus, R. A. and Sutin, N. (1985) *Biochim. Biophys. Acta* **811**, 265-322.

31. Dalbadie-McFarland, G., Cohen, L. W., Riggs, A. D., Morin, A. D., Itakura, K. and Richards, J. H. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6409-6413.

32. Carter, P. (1987) *Methods in Enzymol.* **154**, 382-403.

33. Richards, J. H. (1986) *Nature (London)* **323**, 187.

34. Wells, J. A., Vasser, M. and Powers, D. B. (1985) *Gene* **34**, 315-323.

35. Hermes, J. D., Blacklow, S. C. and Knowles, J. R. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 696-700.

36. Schultz, S. C. and Richards, J. H. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1588-1593.

37. Schultz, S. C., Carroll, S. S. and Richards, J. H. (1987) in *Proteins*, ed. L'Italien, J. J. (Plenum Publishing Co., New York), pp. 195-224.

38. Carroll, S. C. and Richards, J. H. (1987) *Ann. Rep. Med. Chem.* **22**, 293-301.

39. Healey, W. J., Labgold, M. R. and Richards, J. H. (1989) *Proteins: Structure, Function, and Genetics* **6**, 275-283.

40. Craik, C. S., Roczniak, S., Largman, C. and Rutter, W. J. (1987) *Science* **237**, 909-913.

41. Leatherbarrow, R. J., Fersht, A. R. and Winter, G. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7840-7844.

42. Wright, D. J., King, K. and Modrich, P. (1989) *J. Biol. Chem.* **264**, 11816-11821.

43. Chang, Y.-H., Labgold, M. R. and Richards, J. H. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2823-2927.

44. Ghoush, S. S., Bock, S. C., Rokita, S. E. and Kaiser, E. T. (1986) *Science* **231**, 145-148.

45. Abrahmsen, L. and Wells, J. A. (1991) *Biochemistry*, in press.

46. Imanaka, I., Shibazaki, M. and Takagi, M. (1986) *Nature (London)* **324**, 695-697.

47. Pascher, T., Bergstrom, J., Malmstrom, B. G., Vanngård, T. and Lundberg, L. G. (1989) *FEBS Lett.* **258**, 266-268.

48. van de Kamp, M., Floris, R., Hali, F. C. and Canters, G. W. (1990) *J. Am. Chem. Soc.* **112**, 907-908.

49. Bergman, C., Gandvik, E.-K., Nyman, P. and Strid,, L. (1977) *Biochim. Biophys. Res. Comm.* **77**, 1052-1057.

50. Lew, A. Y.-C., Gray, H. B. and Richards, J.H., work in progress.