

To the Graduate Council:

I am submitting herewith a dissertation written by David Alan Engler entitled "Structure-Function Analysis of Human Epidermal Growth Factor: Site-Directed Mutagenesis of Highly Conserved Residues in the Third Disulfide Loop." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Biomedical Sciences.

Salil K. Niyogi
Salil K. Niyogi, Major Professor

We have read this dissertation
and recommend its acceptance:

Sankar Mitra

Audrey Stevens

Fred C. Hartman

Accepted for the Council:

Cowminkel
Vice Provost
and Dean of the Graduate School

STRUCTURE-FUNCTION ANALYSIS OF HUMAN EPIDERMAL GROWTH FACTOR:
SITE-DIRECTED MUTAGENESIS OF HIGHLY CONSERVED RESIDUES
IN THE THIRD DISULFIDE LOOP

A Dissertation
Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

David Alan Engler

May 1990

Copyright © David Alan Engler, 1990
All rights reserved

DEDICATION

To my children, Alexander and Sabrina

ACKNOWLEDGEMENTS

I would like to express my thanks to my advisor Dr. Salil K. Niyogi, for making available to me the resources of his laboratory and introducing me to the field of protein engineering. I sincerely appreciate his time, patience, and helpful advice during this project. In addition to his guidance, I gratefully acknowledge the independence he afforded me to pursue my own ideas. It was this attribute of Salil that allowed me to develop my own sense of identity as a scientist.

I would also like to thank Dr. Audrey Stevens, a member of my thesis committee, who not only was always generous with helpful advice and materials, but who together with her husband Salil, made me and my family feel like a part of their family and provided us with some very special memories. Together they have made my time in Oak Ridge not only scientifically rewarding, but also very enjoyable.

My thanks are also extended to Drs. Fred Hartman, Sankar Mitra, and Margaret Terzaghi-Howe for serving as members of my thesis committee. I wish to thank them for their time and advice on all those aspects of this work for which their counsel was so helpful.

There have been many other members of the Biology Division that have lent their support to me over the course of my training in Oak Ridge. These include members of my lab

group; Dr. Steve Campion, Margaret Yette, and Doug Tadaki, who together with Marilyn Maupin have helped me technically and otherwise, and been the source of friendship and numerous fruitful discussions. Also, much thanks goes to the members of the Protein Engineering and Molecular Mutagenesis Group, too numerous to list individually here, that have lent their support to me and this project over the years. Special thanks go to Dr. Frank Larimer for the use of his expression vector and Drs. Bob Foote and Richard Mural for their helpful discussions and advice on the early phases of this project including the design and synthesis of the chimeric gene. Special thanks are also extended to Claude Stringer for his technical assistance in preparing the EGF amino acid composition analysis and protein sequence information, and to Dr. Melinda Hauser for testing the various mutant proteins in tissue culture assays.

I would also like to sincerely thank Dr. Gaetano Montelione of the Center for Advanced Biotechnology and Medicine, and the Department of Chemistry, at Rutgers University for allowing me to use his NMR facilities and sharing with me some of his insight into the interpretation of protein structure utilizing NMR spectra.

Finally, I especially thank my wife and fellow graduate student Risë Matsunami. She not only helped immensely in the cloning of the EGF gene, but was always around to listen to my concerns and lend moral as well as scientific support. In

addition, she always found the time to take care of those familial responsibilities that I either neglected or was unable to do.

This research was carried out in the Biology Division of Oak Ridge National Laboratory, and was supported by the National Cancer Institute Training Grant CA09104 to the University of Tennessee - Oak Ridge Graduate School of Biomedical Sciences, and by the Office of Health and Environmental Research, U. S. Department of Energy, under contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc.

ABSTRACT

The polypeptide hormone epidermal growth factor (EGF) stimulates cell division by binding to its specific cell surface receptor and stimulating the receptor's tyrosine kinase activity. A chimeric gene, coding for human EGF fused at its amino terminal end to the signal peptide of bacterial alkaline phosphatase, was synthesized and cloned into an expression vector under the transcriptional control of the *tac* promoter and introduced into *Escherichia coli*. Induction with isopropylthiogalactoside led to the synthesis of the precursor form of the EGF polypeptide containing the signal peptide. Secretion of the correctly processed mature EGF into the periplasm was confirmed by the physicochemical properties and biological activity of the purified protein.

The role(s) of several highly conserved amino acid residues in the third disulfide loop (loop C) of EGF was investigated via site-directed mutagenesis. Single amino acid substitutions of glycine-36 and glycine-39 resulted in mutant proteins with decreased receptor affinities. Preliminary evidence implies that these glycyI residues are conserved primarily for structural reasons, possibly because of their location in turn regions of loop C. Several mutant proteins with single amino acid substitutions of tyrosine-37 were evaluated for their receptor affinity and ability to stimulate the receptor's tyrosine kinase activity. With respect to

these activities, amino acid substitutions were well tolerated at this position, in general favoring aromatic residues above other types, in the following order: Phe > Tyr (wild-type) > His > Ser > Ala > Asp > Arg > Gly. The results indicate that tyrosine-37 is not directly involved in receptor interaction. In contrast, arginine-41 appears to be conserved for functional rather than structural reasons. A series of single amino acid substitutions at this site all resulted in a drastic lowering of receptor affinity, including a conservative lysine substitution which lowered the receptor affinity 2.3 orders of magnitude to 0.43% that of wild-type.

Structural analysis utilizing one-dimensional proton nuclear magnetic resonance spectroscopy of several EGF mutants at positions 37 and 41 was used to ascertain whether the decrease of functional activity was due to altered protein conformations induced by the mutations. The spectra obtained were particularly useful for mutations involving arginine-41, and showed that the dramatic decrease in receptor affinity was not due to drastic changes in overall structure of the mutant EGFs. The conclusions drawn from the structure-function studies of these mutant proteins are correlated with the conservation of these loop C residues in other proteins known to bind to the EGF receptor, and to conservation in EGF-like domains of various proteins with no known EGF-like function.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION	1
II.	CLONING AND MUTAGENESIS OF THE <i>hEGF</i> GENE	12
	Rationale/ Design/ and Construction of the <i>hEGF</i> Gene	12
	Materials	14
	Methods	15
	Design of Chimeric <i>phoA</i> Signal/ <i>hEGF</i> Gene..	15
	Synthesis and Purification of Oligonucleotides	15
	Construction of <i>phoA</i> Signal/ <i>hEGF</i> Gene	16
	Transfer of <i>phoA</i> Signal/ <i>hEGF</i> Genetic Construct into an Expression Vector ...	18
	Mutagenesis of Loop C Residues	19
	Materials	20
	Methods	21
	Design of Mutagenic Primers	21
	Oligonucleotide-directed Mutagenesis of <i>hEGF</i> Gene	21
	Screening and Characterization of Mutant <i>hEGF</i> Clones	23
	Transfer of Mutant <i>hEGF</i> Genes into Expression Vector	23
III.	EXPRESSION, ISOLATION, AND PURIFICATION OF EGFS AND PHYSICOCHEMICAL CHARACTERIZATION OF RECOMBINANT <i>hEGF</i>	25
	Expression, Isolation, and Purification of EGF Proteins	25
	Bacterial Strains and Media Composition ...	25
	Expression of EGF Proteins	25
	Isolation and Purification of EGF Proteins.	31
	Characterization of Wild-type <i>hEGF</i>	32
	Size of Protein Product	32
	Amino Acid Composition and Protein Sequencing	33
	Extinction Coefficient Determination	34
	Biological Activity	34

CHAPTER	PAGE
IV. STRUCTURE-FUNCTION STUDIES OF LOOP C MUTANTS ..	36
Receptor Affinity Studies with Tyr-37 and Arg-41 Analogs	36
Stimulation of EGF Receptor Protein-Tyrosine Kinase Activity	39
Structural Studies	41
NMR Studies of Position 37 Mutants	42
NMR Studies of Position 41 Mutants	45
Conclusions	47
Results of Other Loop C Mutants	50
Summary and Implications	51
LIST OF REFERENCES	56
APPENDICES	66
APPENDIX A. FIGURES AND FIGURE LEGENDS	67
APPENDIX B. TABLES	104
VITA	114

LIST OF FIGURES

FIGURE	PAGE
1. Schematic representation of secondary structural features of hEGF	68
2. Ribbon diagram, derived from nuclear Overhauser effect (NOE) distance constraints, representing the polypeptide folds of EGF	70
3. Alignment of type-I EGF-like protein sequences, comprised of EGF and EGF-like molecules known to bind to the EGF receptor	72
4. Schematic representation of EGF showing highly conserved residues	74
5. Design of the synthetic chimeric <i>phoA</i> Signal/hEGF Gene	76
6. Expression plasmid pEGF-1 showing relative size of insert and position of restriction endonuclease recognition sites	78
7. Comparison of reverse phase HPLC profiles of wild-type hEGF, Tyr ³⁷ →Gly, and Arg ⁴¹ →Gly	80
8. Effect of chloramphenicol addition on hEGF production with respect to hEGF protein induction time	82
9. Growth curve of <i>E. coli</i> /pEGF-1 culture ± chloramphenicol	84
10. Typical elution profile of recombinant hEGF off a reverse phase HPLC column	86
11. SDS-PAGE of recombinant hEGF	88
12. Competition binding curves of [¹²⁵ I]hEGF versus wild-type and tyrosine-37 analogs	90
13. Competition binding curves of [¹²⁵ I]hEGF versus wild-type and arginine-41 analogs	92
14. EGF receptor protein-tyrosine kinase stimulation by wild-type and tyrosine-37 analogs	94

FIGURE	PAGE
15. Comparison of aromatic and downfield C ^α H spectral regions from 500 MHz 1D ¹ H-NMR spectra of wild-type, Tyr ³⁷ →Phe, Tyr ³⁷ →His, and Tyr ³⁷ →Ala	96
16. Comparison of aromatic and downfield C ^α H spectral regions from 500 MHz 1D ¹ H-NMR spectra of wild-type, Arg ⁴¹ →Lys, Arg ⁴¹ →Gln, and Arg ⁴¹ →Ala	98
17. Alignment of non-EGF type-II protein sequences from EGF-like motifs in various proteins with no known EGF receptor binding capability	100
18. Comparison of reverse phase HPLC profiles of wild-type hEGF, Gly ³⁶ →Val, Ile ³⁸ →Leu, and Gly ³⁹ →Leu	102

LIST OF TABLES

TABLE	PAGE
1. OCCURRENCE OF EGF-RELATED SEQUENCES IN NON-EGF TYPE-II PROTEINS	105
2. FEATURES OF MUTAGENIC OLIGONUCLEOTIDES	107
3. COMPARISON OF VARIOUS EGF SPECIES' ABILITY TO BE PRODUCED/SECRETED UPON $\pm$ CHLORAMPHENICOL ADDITION.	108
4. AMINO ACID COMPOSITION OF RECOMBINANT WILD-TYPE hEGF	109
5. FUNCTIONAL CHARACTERIZATION OF RECOMBINANT HUMAN EGF	110
6. COMPARISON OF IC_{50} VALUES AND RELATIVE RECEPTOR AFFINITIES OF hEGF TYROSINE-37 AND ARGININE-41 ANALOGS	111
7. COMPARISON OF EC_{50} VALUES AND RELATIVE RECEPTOR AFFINITIES OF hEGF TYROSINE-37 ANALOGS	112
8. IC_{50} VALUES AND RELATIVE RECEPTOR AFFINITIES OF MUTANT EGF PROTEINS CONTAINING AMINO ACID SUBSTITUTIONS AT OTHER HIGHLY CONSERVED SITES IN LOOP C	113

ABBREVIATIONS

A₆₀₀ = absorbance at 600 nm

Amino acids =	Ala or A, alanine	His or H, histidine
	Arg or R, arginine	Ile or I, isoleucine
	Asn or N, asparagine	Leu or L, leucine
	Asp or D, aspartic acid	Lys or K, lysine
	Asx, asparaginyl or aspartyl	Met or M, methionine
	Cys or C, cysteine	Phe or F, phenylalanine
	Gln or Q, glutamine	Pro or P, proline
	Glu or E, glutamic acid	Ser or S, serine
	Gly or G, glycine	Thr or T, threonine
	Glx, glutaminyl or glutamyl	Trp or W, tryptophan
		Tyr or Y, tyrosine
		Val or V, valine

ATP = adenosine 5'-triphosphate

bp = base pairs

BSA = bovine serum albumin

Cm = Chloramphenicol

cpm = counts per minute

ds = double stranded

DTT = dithiothreitol

ϵ_{280} = molar extinction coefficient at 280 nm

E. coli = *Escherichia coli*

EDTA = ethylenediaminetetraacetic acid

EGF = epidermal growth factor; hEGF, human epidermal growth factor; mEGF, mouse epidermal growth factor; rEGF, rat epidermal growth factor

hEGF = gene for human epidermal growth factor

HEPES = N-2-Hydroxyethylpiperazine-N'-2-ethanesulfonic acid

HPLC = high performance liquid chromatography

IPTG = isopropyl- β -D-thiogalactopyranoside

kD = kilo Dalton

lon⁻ = gene coding for bacterial La protease

M_r = relative molecular weight

MGF = myxoma virus growth factor

MHz = mega Hertz

NMR = nuclear magnetic resonance; 1D ¹H-NMR, one-dimensional proton nuclear magnetic resonance; 2D NMR, two-dimensional nuclear magnetic resonance

PEG₈₀₀₀ = poly(ethylene glycol)

PEI = poly(ethylenimine)

phoA signal = alkaline phosphatase signal peptide

phoA signal = gene for alkaline phosphatase signal peptide

PMSF = phenylmethylsulfonyl fluoride

RF = replicative form

R_f = retardation factor

RPC-18 = reverse phase column with alkyl chain length of 18 carbons

SDS-PAGE = sodium dodecyl sulfate polyacrylamide gel electrophoresis

SGF = Shope fibroma virus growth factor

ss = single stranded

TCA = trichloroacetic acid

T_d = temperature of dissociation

TGF- α = transforming growth factor alpha

TGF- β = transforming growth factor beta

Tris = Tris(hydroxymethyl)aminomethane

VGF = vaccinia virus growth factor

X-Gal = 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside

CHAPTER I

INTRODUCTION

Growth factors are polypeptides which act to regulate the development of tissues by stimulating growth and maintaining viability at the cellular level. Epidermal growth factor (EGF), one of the first such growth promoting factors to be identified, is perhaps the most extensively characterized growth factor to date. However, although three decades have elapsed since its discovery, and many cellular and organismal responses have been ascribed to exogenously added EGF, much still remains to be learned about its mechanism of action. At the organismal level there is still not a clear understanding as to what physiological role endogenous EGF plays in any animal. At the cellular level, where many of the attributes of EGF function have been revealed, more information is needed to ascertain how this growth factor influences the multiple cellular responses observed upon its addition. This type of information will have to be gathered at the very basic level of EGF function, namely, at the molecular level from a chemical and biochemical perspective. Since the study of EGF, as a model for growth factors in general, may lead to a better understanding of the mechanisms of normal cellular growth and differentiation, and possibly of abnormal growth and carcinogenesis, elucidating the

molecular mechanism of EGF action promises to be a very rewarding enterprise.

EGF binds with high affinity to a specific cell surface receptor as first demonstrated by O'Keefe *et al.* in 1974 (1). The EGF receptor, a single polypeptide glycoprotein with an approximate M_r of 170,000, consists of an extracellular ligand binding domain, a transmembrane region, and an intracellular protein-tyrosine kinase domain.* The binding of EGF to its receptor stimulates cellular responses such as changes in intracellular pH and free calcium, increased sugar and amino acid transport, increased glycolysis and protein synthesis, autophosphorylation on receptor tyrosine residues and tyrosine phosphorylation of certain endogenous cellular proteins, and increased transcription of specific genes. This leads ultimately to increased DNA replication and cell division.

The activity that was later to be identified with EGF was first observed in 1960 by Stanley Cohen and Rita Levi-Montalcini during their work on nerve growth factor isolated from the mouse submaxillary gland (4,5). It was noticed that a partially purified carboxymethyl cellulose chromatography fraction that contained only slight nerve growth promoting activity also elicited more general somatic effects when injected into newborn mice.

*The reader is referred to several recent reviews on the EGF receptor (2,3).

These effects included gross anatomical changes such as i) precocious opening of the eyes at 6-7 days after birth instead of the normal 12-14 days, ii) precocious eruption of the incisors at 5-6 days instead of the normal 8-10 days, and iii) a marked stunting of the animals' growth with an inhibition of hair growth. The purification, based on the biological assay of eyelid opening and incisor eruption, of a heat-stable, non-dialyzable, antigenic protein was reported by Cohen in 1962 (6). This protein was called "tooth-lid" factor. When purified "tooth-lid" factor was injected into newborn mice that were subsequently histologically examined, it was noted that the observed eyelid separation and tooth eruption were reflections of a more generalized biological effect, namely, an enhancement of epidermal keratinization and an increase in the overall thickness of the epidermis (7). When the purified factor was tested *in vitro* on embryonic chicken skin or epidermis growing in culture, a direct growth promoting activity as well as increased keratinization was seen (8) and "tooth-lid" factor became known as epidermal growth factor.

In a separate study, initiated by earlier observations that pregnant women had a very low frequency of ulcers, it was shown that the urine of pregnant women had beneficial effects upon peptic ulceration in dogs (9). Several groups showed that this effect was due to a substance in the urine that inhibited gastric acid secretion (10,11) which was named

"urogastrone" because of its possible relationship to, or derivation from, enterogastrone, the acid inhibitory hormone of the duodenum (12). Subsequent to these discoveries, many groups endeavored to purify urogastrone (13-17), with some isolating substances pure enough to reach different conclusions about its physical properties. Some classified it as a low molecular weight protein (14), while others claimed it was a large molecular weight glycoprotein (16,17).

In 1975, Gregory and Willshire (18) purified three anti-gastric acid secretory proteins from urine which they called α -, β -, and γ -urogastrone. Enough of the β - and γ -urogastrone was purified to allow physical studies. These studies showed γ -urogastrone differing from β -urogastrone by one residue. β -urogastrone was shown to be an acidic polypeptide composed of 53 amino acid residues while γ -urogastrone differed only by the absence of the COOH terminal arginine residue. Both were shown to contain three internal disulfide bonds and contained neither threonine nor phenylalanine. Based on studies by Savage *et al.* (19,20) which had elucidated the primary sequence of mouse EGF, and assays attributing EGF-like activity to urogastrone, Gregory (21,22) was able to show that human β -urogastrone and mEGF were homologous proteins. Cohen and Carpenter (23) had concurrently isolated human EGF from urine and re-affirmed the homology between EGF and urogastrone. Subsequent investigations have demonstrated the existence of EGF in other species including rat, guinea-pig, and dog (24,25).

The physical and chemical properties of EGF were described by Taylor *et al.* (26) and Holliday *et al.* (27). All EGF molecules have an average M_r of approximately 6000, contain three internal disulfide bonds (see Figure 1)[§], are heat and acid stable, have acidic isoelectric points, and contain about 25% β structure and no α helicity. Recently, more detailed studies utilizing nuclear magnetic resonance (NMR) spectroscopy have provided tertiary structure information for mEGF (28-30), rEGF (31), and hEGF (32,33). Computer modeling, utilizing data derived from these NMR studies, and energy minimization and distance geometry algorithms, have resulted in detailed solution structures for mEGF (34-36), rEGF (37), and hEGF (38,39). These studies have produced very similar three dimensional representations of the EGF polypeptide backbone (see Figure 2). The distinguishing features of the EGF molecules are as follows. The distal NH_2 - and COOH -terminal regions are the most poorly defined regions based on NMR data and thus show the greatest differences among structures. In addition, rEGF has only 48 amino acid residues compared to mEGF and hEGF which both have 53 residues. Outside of the truncation in the COOH -terminus of rEGF, the tertiary structures of all three molecules are remarkably similar.

[§]All Figures are presented in Appendix A. All Tables are presented in Appendix B.

The EGF molecule contains two structural domains, and folds around Asn-32 and Cys-33 to form a mitten-like arrangement. The NH₂-terminal domain comprises residues 1-33 and encompasses loops A and B (see Figure 1). Loop A, which is delimited by the disulfide bond formed between Cys-6 and Cys-20 (see Figures 1 and 2), forms an irregular helical loop. Loop B of the molecule delimited by the disulfide bond linking Cys-14 to Cys-31, (see Figures 1 and 2) is characterized by a double-stranded, anti-parallel β -sheet linking residues 19-23 with residues 28-32. A triple stranded β -sheet which also links residues 2-4 with residues 21-23 exists approximately 10% of the time (39). The COOH-terminal domain which encompasses loop C, delimited by the disulfide bond linking Cys-33 to Cys-42, comprises residues 33-53 and is characterized by a double-hairpin structure in which there is a short segment of antiparallel β -sheet linking residues 37 and 38 with residues 44 and 45. Residues 34-37 form a type-II β -turn (40), while residues 39-43 form a turn which allows a chain reversal so the third disulfide bridge can close and the minor β -sheet can form. The final COOH-terminal residues (residues 51-53) are thought to exist in an unordered state.

In addition to the various EGF molecules isolated from the aforementioned species, there exist other molecules isolated from various sources that are also capable of binding to the EGF receptor. One such molecule, known as transforming

growth factor- α (TGF- α), was first isolated from the conditioned medium of fibroblasts transformed by murine sarcoma virus (41). The observations that TGF- α expression was often associated with tumor cells (42), and that TGF- α and the structurally unrelated TGF- β could together induce a reversible phenotypic transformation of NRK fibroblasts in culture (43,44), reinforced the belief that TGF- α was associated with malignant transformation. TGF- α has since been found in fetal cells (45,46) and non-transformed adult tissue (47,48). TGF- α and EGF are equipotent in receptor protein-tyrosine kinase stimulation and receptor downregulation (49,50). Marquardt *et al.* (51,52) and Winkler *et al.* (53) showed that TGF- α has a distinct, but EGF-like, sequence with an analogous disulfide arrangement. It has recently been shown by NMR studies (39,54-56) that TGF- α is also structurally very similar to EGF at the tertiary level.

Other EGF-like proteins that bind to the EGF receptor have also been reported. Amphiregulin, a bifunctional cell growth modulator which inhibits the growth of A431 human epidermoid carcinoma and other human tumor cells, but stimulates proliferation of fibroblast cells, has recently been isolated from the medium of MCF-7 human breast carcinoma cells that had been treated with 12-O-tetradecanoylphorbol-13-acetate. It has been sequenced and found to be an EGF-like protein of 78 amino acids capable of binding to the EGF receptor (57). The COOH-terminal sequences from several viral

proteins including the 19kD vaccinia virus protein (58), the Shope fibroma virus growth factor (59), and the myxoma virus growth factor (60) also have been shown to encode EGF-like sequences. Both the vaccinia virus growth factor (VGF) and the Shope fibroma virus growth factor (SGF), have been shown to bind to the EGF receptor (58,59). The myxoma virus growth factor (MGF) is assumed to bind to the EGF receptor based on its 85% sequence homology to SGF, although it has not yet been demonstrated. A sequence alignment of the proteins known to bind to the EGF receptor (hereafter called type-I EGF-like proteins) is presented in Figure 3. The cysteinyl-bounded core sequence of this EGF family has a consensus sequence of $CX_7CX_{4-5}CX_{10-13}CXCX_8C$ and is 36 to 40 residues in length.

EGF-like motifs, based on the general features of six conserved cysteine residues in the above mentioned approximate arrangement and with a unit length of about 40 amino acid residues, have been identified in a number of functionally unrelated proteins (61,62). These EGF-like motifs have the general consensus sequence of $CX_{4-14}CX_{3-9}CX_{4-17}CXCX_{7-12}C$ and occur quite often in tandemly repeating structural units of secreted or transmembrane molecules. These proteins (hereafter called non-EGF type-II proteins) and their functions are listed in Table 1. In the limited number of cases where individual EGF-like domains were generated by proteolytic digestion of the parent type-II molecule, or synthesized chemically, no ability to bind to the EGF receptor was detected (63; J. Tam, personal

communication). However, the disulfide bond arrangement and the overall EGF-like tertiary structure were maintained (63,64; J. Tam, personal communication). The presence of EGF-like sequences in these relatively diverse proteins, besides inter-species homologies among EGFs, suggests evolutionary conservation of amino acid residues for either structural and/or functional reasons.

If one considers exact homology (*i.e.* conservation of identical amino acid residues) among EGF sequences from various species (listed in Figure 3), there is a 55% conservation over the entire protein sequence. If homology is extended to include amino acid residues with side chains exhibiting similar chemical functionality, this value climbs to 68%. If a comparison is made of all type-I EGF-like protein sequences, including TGFs, amphiregulin, and the COOH-terminal portions of the viral growth factors, the overall exact sequence homology falls to 21% with only the 6 cysteine residues, Gly-18, Tyr-37, Gly-39, and Arg-41 absolutely conserved. Consideration of similar chemical functionality only allows the inclusion of two additional residues, an aromatic residue at position #13 (Tyr-13 in hEGF) and a hydrophobic residue at position #23 (Ile-23 in hEGF), for a 25% overall homology. It should also be noted that Leu-47 is absolutely conserved in these proteins except amphiregulin, which terminates at hEGF position #45. Glycine-36 is also relatively highly conserved, lacking only in amphiregulin and

the two pox virus growth factors MGF and SGF. A schematic summarizing these results is shown in Figure 4. Figures 3 and 4 illustrate that the third disulfide loop of EGF (loop C) comprises the most highly conserved region in the EGF family. There is a 70% exact sequence homology in this region among EGFs, and if one considers residues with like chemical functionality, homology is as high as 90% in this region. Even when one considers all type-I EGF-like sequences listed in Figure 3, there is a 50% sequence conservation in this region compared to the 25% overall EGF sequence homology of these molecules.

Over the past several years, EGF and other members of the EGF family have been associated with important therapeutic responses in several areas of biomedical importance. In addition to its historically important activity in decreasing gastric acid secretion, which is currently the basis for an experimental ulcer treatment program, EGF has also been shown to accelerate the processes of mid-dermal wound repair (65,66) and repair of chronic non-healing wounds (67,68). EGF has also been shown to stimulate corneal angiogenesis (69), and to increase the corneal strength after corneal incisions (70). EGF is also being studied as a treatment for psoriasis (71), and has been implicated as playing a role in male reproductive function by stimulating the meiotic phase of spermatogenesis (72). Thus, these clinically important attributes of EGF (besides those of TGF- α , and other growth factors) can be

added to its role in normal/abnormal cell cycle regulation, and other various effects of EGF on different cells and tissues. It is imperative, however, that we try to understand the mechanisms behind such activities if we are to understand the difference between normal versus abnormal cell growth, or to further advance the clinical use of, and/or manipulation of, EGF-mediated responses.

This treatise examines, via site-directed mutagenesis, the initial but pivotal role that EGF plays in all further EGF/EGF-receptor mediated events, namely, the initial binding of EGF to the extracellular domain of the EGF receptor. It describes i) cloning of the EGF gene for use as template in the site-directed mutagenesis studies; ii) specific alterations at the genetic level of the highly conserved amino acid residues in loop C to produce the desired constructs; iii) optimal expression of the mutant proteins; iv) characterization of the functional abilities of the EGF analogs to bind to the receptor and stimulate the protein-tyrosine kinase activity of the receptor; and v) a preliminary structural analysis of wild-type and mutant EGFs using 1D ^1H -NMR.

The following represents one approach to elucidating the mechanism of EGF action at the molecular level. It is the culmination of several years of research and hopefully the beginning of many more fruitful investigations that will ultimately lead to a full understanding of the molecular mechanisms of EGF action.

CHAPTER II

CLONING AND MUTAGENESIS OF THE *hEGF* GENE

Rationale/Design/and Construction of the *hEGF* Gene

An essential prerequisite for a molecular structure-function analysis of a protein utilizing site-directed mutagenesis is the acquisition of a cloned gene for that particular protein. Prior to this work, the gene for hEGF had been cloned in several forms by various groups (73-76). Most of these clones were unsuitable for this study since they produced hEGF as a fusion protein (73,74), or were cloned in yeast secretion vectors (75) that would have required processing of large volumes of medium to isolate the desired product. Oka *et al.* (76) had produced hEGF as a secretory protein in *E. coli*, but proprietary considerations precluded general distribution of their clone.

The initial goal of this study was to produce a cloned gene that would code for an authentic hEGF with the proper NH₂-terminal asparagine residue, and allow the facile construction of mutant EGF molecules via site-directed mutagenesis. All proteins produced in *E. coli* contain initial primary translation products with NH₂-terminal methionine residues. Only further processing of the initial translation product will result in removal of this residue. Therefore, to accomplish these goals, it was decided to design a

synthetic gene that would code for the hEGF molecule fused at its amino terminal end to the signal peptide of bacterial alkaline phosphatase, a normal constituent of the *E. coli* periplasm. This would serve to direct the secretion of hEGF to the periplasmic space of *E. coli*. This synthetic chimeric gene and the bacterial secretory system would accomplish several desirable results. First, as the precursor protein molecule traversed the bacterial inner membrane, the enzymatic cleavage of the signal peptide by *E. coli* signal peptidase would be accomplished *in vivo*, liberating a full length hEGF protein with the proper NH₂-terminal asparagine residue in the periplasm. Second, sequestering hEGF in the periplasmic space would result in an initial purification of the product since only about 4% of the *E. coli* proteins are localized in this space (77). Third, extraction of the hEGF in a small volume, could be accomplished in a simple one-step procedure (see Chapter 3) that would liberate the periplasmic proteins essentially free of contaminating cytoplasmic proteins. Fourth, localization of hEGF in the periplasm would shelter it from cytoplasmic proteases (78-80), and from the high reducing potential in the cytoplasm that could prevent oxidation of the six cysteinyl thiol groups and hence the formation of the three internal disulfide bonds necessary for a properly folded hEGF molecule. It has also been shown that other small molecules such as insulin are approximately ten times more resistant to proteolysis in the periplasm than in the

cytoplasm (81). Lastly, chemical synthesis of the chimeric gene would allow for the placement of several unique restriction sites that would simplify the mutant screening protocol during site-directed mutagenesis procedures. This would also render the *hEGF* gene suitable for cassette mutagenesis (82), in which the mutant codon(s) could be readily introduced utilizing short synthetic double-stranded oligonucleotides between any two successive unique restriction sites.

Materials

Common chemicals and reagents were purchased at the highest level of purity readily available. Other commercial materials and vendors were as follows: T4 polynucleotide kinase, unlabeled nucleotides, and nuclease P1 from Pharmacia LKB Biotechnology Inc.; [γ - ^{32} P]ATP (4500 Ci/mmol) from ICN Radiochemicals; deoxynucleoside phosphoramidites from American Bionetics; T4 DNA ligase from Boehringer Mannheim; Polygram Cel 300 PEI/UV₂₅₄ from Sybron/Brinkmann; and restriction endonucleases from Bethesda Research Laboratories, or New England Biolabs. Plasmid pFL142 DNA was a generous gift from Dr. Frank W. Larimer of the Biology Division, Oak Ridge National Laboratory.

Methods

Design of Chimeric *phoA* Signal/*hEGF* Gene

The nucleotide sequence coding for the chimeric gene is shown in Figure 5. It was designed utilizing *E. coli* preferred codons (83), where possible, for the amino acid sequence of each protein component of the desired construct, as determined for the alkaline phosphatase signal peptide by Inouye *et al.* (84) and for hEGF by Gregory (21). Introduction of restriction endonuclease sites was accomplished by selective codon usage without altering the coding sequence of the final protein product. The initiator codon for the N-formylmethionine of the *phoA* signal peptide is, in this case, ATG as opposed to the naturally occurring GTG codon (84,85). This was necessary for the proper insertion of the *phoA* signal/*hEGF* construct into the *Nco* I site of the "ATG" expression vector pFL142 (86), to form pEGF-1 (see below).

Synthesis and Purification of Oligonucleotides

Twelve ss oligodeoxynucleotides, ranging in length from 34 to 40 bases, were synthesized on a Systec model 1450 automated DNA synthesizer utilizing phosphoramidite chemistry (87). Each oligonucleotide was purified twice by electrophoresis through 16% polyacrylamide gels containing 8M urea (88). 5'-End analysis to check the purity of the final product was as follows: each oligonucleotide (7 pmol) was phosphorylated at the 5' terminus in a 30 μ l reaction mixture

containing 50 mM Tris-Cl (pH 7.6), 10 mM MgCl₂, 0.1 mM spermidine, 0.1 mM EDTA, 5 mM DTT, 0.9 μ M [γ -³²P]ATP (500 Ci/mmol), and 10 units of polynucleotide kinase. After incubation at 37°C for 1 h, the reaction mixture was heated at 65°C for 10 min, and the oligonucleotide was purified by gel filtration on Bio-Gel P-10 equilibrated with 10 mM triethylammonium acetate, pH 7.0. Following lyophilization to concentrate the oligonucleotide to about 1500 cpm/ μ l, 1 μ l of the oligonucleotide was dissolved in 10 mM sodium acetate (pH 5.3) with 20 units of nuclease P1 to give a final volume of 10 μ l. Following complete enzymatic hydrolysis at 37°C for 1 h, the reaction mixture was spotted onto PEI-cellulose thin layer chromatography plates and developed in 1 M LiCl until the solvent front reached 10 cm. Chromatograms were dried and autoradiographed overnight at room temperature and the labeled 5' mononucleotide of the oligonucleotide hydrolysate was determined from its R_f relative to cold 5' dNMP standards.

Construction of *phoA* Signal/ *hEGF* Gene

Each ss oligonucleotide, #3,5,7, or 9 (200 pmol) was annealed with its respective complementary ss oligonucleotide, #4,6,8, or 10 (200 pmol) in a 25 μ l reaction mixture containing 15 mM Tris-Cl (pH 7.5) and 75 mM NaCl to form ds oligonucleotides with 5'-overhangs of 6-9 bases (Figure 5). Annealing was performed in a stepwise fashion, as described below, to preclude the formation of any secondary hairpin

structures. Each reaction mixture was first heated for 10 min at 10-15°C above the empirically calculated T_d (89), then at 20°C below the T_d for 15 min, followed by slow cooling to room temperature over a period of 1 h. The resulting ds complex was phosphorylated at its 5' termini as described above, except that a molar ratio of cold ATP to 5' oligonucleotide ends of 5:1 (2000 pmol cold ATP, 400 pmol 5' oligonucleotide ends) was used in a 50 µl reaction with 50 nCi of [γ -³²P]ATP to follow the efficiency of the phosphorylation reaction. The ss oligonucleotides #2 and #11 (200 pmol each) were phosphorylated at their 5' termini, then annealed to 200 pmol each of their respective complementary ss oligonucleotides (#1 and #12) that were left as the 5'-OH species. Leaving the two 5' termini unphosphorylated on #1 and #12 prevented concatemerization of the final ligated gene product prior to insertion into the expression vector.

The six ds oligonucleotides with complementary ss termini were ligated together (to form the *phoA* signal/ *hEGF* genetic construct) in a 100 µl reaction mixture containing 50 pmol of each ds pair in 20 mM Tris-Cl (pH 7.5), 10 mM MgCl₂, 2 mM DTT, 1 mM ATP and 24 units of T4 DNA ligase. Ligation was carried out at 14°C for 24 h. The ligation efficiency was greater than 85% as determined by agarose gel electrophoresis.

Transfer of *phoA* Signal/*hEGF* Genetic Construct into an Expression Vector

Ten ng of the *phoA* signal/*hEGF* genetic construct was added to 100 ng of the 2731-bp *NcoI*-*EagI* fragment from the expression vector pFL142. Ligation was carried out as above, but for 18 h with 6 units of T4 DNA ligase in a reaction volume of 10 μ l. *E. coli* BNN45 was transformed with various dilutions of the ligation mixture by the procedure of Hanahan (90) and plated by serial dilution onto LB medium containing 25 μ g/ml ampicillin. Plasmid DNA was prepared from ampicillin-resistant colonies by the alkaline lysis method (91) and screened by digestion with *HindIII* followed by agarose gel electrophoresis to determine which plasmid DNA had the *phoA* signal/*hEGF* insert. Plasmid DNA containing the *phoA* signal/*hEGF* insert, called pEGF-1 (see Figure 6), was digested with *Bam*HI to liberate a 300-bp fragment which was subcloned into the *Bam*HI site of M13mp19 RF DNA for DNA sequence analysis. *E. coli* JM107 was transfected (90) with this DNA preparation and plated on LB medium using top agar containing 0.3 mM IPTG and 0.03% X-Gal to select for plaques containing the *Bam*HI fragment. Phage ss DNA was isolated and sequenced by the dideoxynucleotide chain termination protocol (92) to verify the correct sequence of the *phoA* signal/*hEGF* gene construct.

Mutagenesis of Loop C Residues

As previously discussed in Chapter 1, loop C of hEGF is composed of residues 33-42 and comprises the most highly conserved region of the molecule. Of the other eight residues in this loop besides the two conserved cysteines, three are absolutely conserved throughout the entire EGF-like type-I family of molecules that are known to bind to the EGF receptor. These are Tyr-37, Gly-39, and Arg-41 (see Figure 4). The unequivocal conservation of these residues raises the question as to why these particular amino acid residues are so highly conserved. The solution structure of EGF (see Figure 2) shows that Tyr-37 is involved in the minor β -sheet of the molecule, Gly-39 is positioned in a short left-handed helical turn of loop C, and Arg-41 is positioned at the top of the loop at the center position of the chain reversal.

The major objective of this work was to elucidate which residues in EGF might play a role in interaction with the EGF receptor, and several residues in loop C appeared to be likely candidates based on their absolute evolutionary conservation. In addition, work by Nestor *et al.* (93) with synthetic peptide fragments of rat TGF- α had indicated that a peptide fragment corresponding to loop C of TGF- α was slightly active in receptor binding (0.2% that of EGF), whereas fragments corresponding to other regions of TGF- α were inactive. On the other hand, work by Komoriya *et al.* (94) with synthetic peptide fragments corresponding to different loops of mEGF, had shown

no receptor binding activity with a fragment composed of loop C residues, although a fragment corresponding to loop B exhibited 0.01% of EGF affinity. Both of these studies relied on data from relatively low affinity measurements with synthetic peptide fragments and yielded different conclusions on the ability of the loop C region of these molecules to bind to the EGF receptor.

In view of the above, it was decided to produce mutant analogs of hEGF at several of the highly conserved sites in loop C in an effort to elucidate the role of these residues in an otherwise intact molecule of hEGF. Several mutant analogs of hEGF, incorporating various single amino acid substitutions at positions #37 and #41, and single amino acid substitutions at positions #36, #38, and #39, were constructed in an effort to define the role of each of these residues in EGF's function and/or structure.

Materials

Materials were as described above, with the addition of [α - ^{32} P]dATP (3000 Ci/mmol) from ICN Radiochemicals and *E. coli* DNA polymerase I (Klenow large fragment) from Boehringer Mannheim.

Methods

Design of Mutagenic Primers

Oligodeoxynucleotides, degenerate at 1-4 nucleotides corresponding to the targeted codon and/or restriction site (see Table 2) and complementary to the sequence surrounding the targeted sites, were synthesized and purified as described above. The oligonucleotides were phosphorylated at their 5'-termini as described above.

Oligonucleotide-directed Mutagenesis of *hEGF* Gene

Single stranded M13mp19 DNA, with the 300 bp *Bam*HI insert, containing the entire *phoA* signal/*hEGF* chimeric gene, was purified from CsCl density gradient banded phage particles (95) by repeated phenol/chloroform extractions. Primed synthesis for covalently closed RFI DNA utilizing the mutant oligonucleotides as primers and the M13-EGF ssDNA as template was carried out according to modifications of Niyogi *et al.* (96,97) of the "boost protocol" of Nisbet and Beilharz (98). A brief description follows.

A 10 μ l reaction mixture containing solution A (15 mM Tris (pH 7.4), 75 mM NaCl), 1-2 pmol of template DNA, and enough primer to produce a primer-to-template ratio of approximately 30:1, was heated to approximately 20°C above the calculated Td of the primer for 10 min, then quickly placed at a temperature approximately 10°C below the Td for 20 min, followed by incubation at 37°C for 1 h to allow primer-

template annealing. To that reaction mixture was added 30 μ l of solution B (containing of 20 mM Tris (pH 7.5), 15 mM MgCl_2 , 1 mM ATP, and 65 μ M each of dATP, dGTP, dCTP, and dTTP) prewarmed to 37°C, 0.5-1.0 μ Ci [α - ^{32} P]dATP, 8-10 units of T4 DNA ligase, and 5 units of DNA polymerase I (Klenow fragment) and the mixture incubated at 37°C 1 h. Then 10 μ l of solution C (consisting of 200 μ M each of dATP, dGTP, dCTP, and dTTP in 20 mM Tris (pH 7.5)) was added in the pre-warmed state, in addition to 9 units of T4 DNA ligase, and 5 units of DNA polymerase I (Klenow fragment) to accomplish a "boost" of the synthesis reaction. Incubation at 37°C was continued for an additional 1.5 - 2 hours. The synthesis reaction was terminated by the addition of EDTA to a final concentration of 25 mM and synthesis products were precipitated by the addition of 50 μ g yeast tRNA and 7.5% final concentration of PEG_{8000} in 1.6 M NaCl on ice for a minimum of 1 h. The precipitate was collected by centrifugation at 15,000 rpm in a microcentrifuge at 4°C. The pellet was resuspended and washed twice with 150 μ l aliquots of 6.5% PEG_{8000} and 0.8 M NaCl. The final pellet was resuspended in 160 μ l sterile TE (10mM Tris, pH 7.4, containing 1 mM EDTA) plus 40 μ l 1 N NaOH and incubated at room temperature for 15 min. Covalently closed circular RFI DNA was purified by band sedimentation on linear 5-20% alkaline sucrose gradients (containing 0.8 M NaCl, 0.2 N NaOH, and 2 mM EDTA) (99) using a Beckman SW 50.1 rotor at 40,000 RPM for 2 hours at 4°C.

Screening and Characterization of Mutant *hEGF* Clones

The purified and dialyzed RFI DNA was used to transfect *E. coli* JM107 by the method of Hanahan (90). Plaques were transferred to Schleicher and Schuell BA85 nitrocellulose membranes, and the filters were hybridized to ³²P-labeled mutant oligonucleotide probes as described previously (96). Briefly, hybridization was carried out for 1 h at 20°C below the empirically calculated T_d. Filters were then washed under stringent temperature conditions for each mutant and subjected to autoradiography. Candidate mutant phages from individual plaques were then grown, and RF DNA was isolated and digested with the appropriate restriction endonuclease to utilize the gain or loss of the monitored restriction site as a further screening procedure. Phage whose RF DNA either retained or lost the restriction site in question (see Table 2) were propagated for the isolation of ssDNA. DNA sequencing (92) of the *phoA/hEGF* gene was used to confirm the mutations and the absence of secondary mutations in other areas of the gene.

Transfer of Mutant *hEGF* Genes into Expression Vector

*Bam*HI fragments of each mutant clone were removed from the phage RF DNA and purified either by agarose gel electrophoresis, or by ion exchange chromatography utilizing a Waters Gen-Pak FAX column (100) on the HPLC. Purified fragments were ligated into *Bam*HI digested pEGF-1 lacking the

wild-type *hEGF* gene. This ligation was typically carried out in a 20 μ l reaction mixture with 350 ng vector DNA and 8 ng insert DNA (1:5 molar ratio of vector-to-insert) at room temperature for 12 h in the presence of 20 mM Tris (pH 7.5), 10 mM $MgCl_2$, 10 mM DTT, 0.6 mM ATP, and 1 unit T4 DNA ligase. Typically, 10 μ l of the heat-terminated (65°C for 10 min.) ligation mixture was added to 200 μ l of competent (90) *E. coli* JM107 followed by serial dilution onto LB medium plates containing 25 μ g/ml ampicillin. Resultant transformants were screened by colony hybridization protocols (101) to ascertain which transformants contained mutant *hEGF* gene inserts. Since the mutant *hEGF* fragments could insert in either orientation in the expression vector (due to *Bam*HI sites on each end of the fragment) insert-bearing clones were further screened for proper insert orientation by *Eco*RI digestion of purified plasmid DNA and agarose gel electrophoresis. The expression and purification of wild-type and mutant EGFs are discussed in Chapter III

CHAPTER III

EXPRESSION, ISOLATION, AND PURIFICATION OF EGFs AND PHYSICOCHEMICAL CHARACTERIZATION OF RECOMBINANT hEGF

Expression, Isolation, and Purification of EGF Proteins

Bacterial Strains and Media Composition

The following strains of *E. coli* were used in this study: *E. coli* BNN45 (*metB*, *thi*⁻, *hsdR*, *lacY*, *supE44*, *supF*) (102); *E. coli* JM107 ($\Delta(lac-pro)$, *endA1*, *gyrA96*, *thi1*, *hsdR17*, *supE44*, *relA1*, (F' *traD36*, *proAB*⁺, *lacI*^q Δ M15)) (103); *E. coli* C600 (F', *thi1*, *thr1*, *leuB6*, *lacY1*, *tonA21*, *supE44*, λ ⁻) (104); *E. coli* SG1161 (*lon* Δ , *gal* Δ , (F' *lacI*^q)) (105); *E. coli* AB1899 (F', *thr*⁻, *leu*⁻, *pro*⁻, *his*⁻, *thi*⁻, *arg*⁻, *lac*⁻, *gal*⁻, *ara*⁻, *xyl*⁻, *mtl*⁻, *lon*⁻) (106). LB medium consisted of 10 g Bacto-tryptone, 5 g yeast extract, and 10 g NaCl per liter of medium. Modified M9 medium consisted of 6 g Na₂HPO₄, 3 g KH₂PO₄, 0.5 g NaCl, 2 g NH₄Cl, 2 mg thiamine, 2 mM MgSO₄, and 1% glycerol per liter of medium.

Expression of EGF Proteins

Several strains of *E. coli* were tested for their ability to support the expression of EGF using different media and growth conditions. This included a comparison (data not shown) of the above strains during growth at different temperatures (30°C, 37°C, and 42°C) in both nutrient-rich medium, LB, and

minimal medium, modified M9 medium (for those strains that were prototrophic for amino acid biosynthesis). From these experiments a general pattern was observed that influenced future expression studies. It was found that overall net EGF expression was better in those strains that were grown in a nutrient-rich medium at 37°C because of higher cell densities. However, EGF production on a per cell basis was better in cells grown under non-optimal growth conditions, such as lower temperatures and in nutrient-poor medium. This also held true for *E. coli* cells grown under anaerobic conditions as opposed to aerobic conditions. Also, use of the La-protease deficient *lon*⁻ strains of bacteria, such as the *E. coli* SG1161 and AB1899 strains, did not increase the levels of EGF in the periplasm over those of other strains. Since *E. coli* JM107 is a *lacI*^q host, signifying that expression of EGF would be repressed due to higher levels of lac repressor molecules in this host, expression of EGF could be initiated at will by the addition of IPTG. These observations led to the use of *E. coli* JM107 as the standard expression host for this study.

An early observation was made pertaining to the yield of various mutant proteins, including Tyr³⁷→Gly and Arg⁴¹→Gly. It was noticed that certain mutants, the latter two in particular, were produced in *E. coli* in extremely small amounts. Levels of these purified mutant proteins from the periplasm were so small in some cases as to preclude receptor affinity

measurements. It was also shown that the *lon*⁻ strains of *E. coli* had no beneficial effects on production levels of these mutants; therefore instability of these mutants and subsequent protease degradation in the periplasm, at least by the La protease, could be ruled out. In addition, purified preparations of these mutants exhibited multi-peak profiles on reverse phase HPLC chromatography as compared to wild-type EGF profiles (see Figure 7). Purified material from individual HPLC peaks of each of these mutants appeared as single peaks when re-injected on HPLC, and exhibited electrophoretic gel patterns (under denaturing conditions) identical to that of the wild-type (data not shown). This suggested that these mutant proteins did not exhibit multiple peaks on HPLC due to drastic proteolysis, rather, were capable of forming conformationally different, stably folded entities from the same full-length polypeptide backbone. It was assumed that if these mutants were able to exist as conformationally different, stably folded entities in the periplasm, that some higher percentage of these mutant molecules, as compared to wild-type, might also exist as stably folded intermediates in the cytoplasm. This presumably would tend to have a detrimental effect on secretion of the mutant EGF molecules to the periplasm, since previous studies by others (107,108) had suggested that the presence of a stably folded structure inhibits the translocation of proteins across membranes.

It is assumed, for exponentially growing *E. coli* cells in a nutrient-rich medium, that approximately 10 amino acid residues per second are added to an elongating nascent polypeptide chain. This would mean that a new EGF precursor molecule of 74 amino acids (signal + EGF) in length could be liberated from exponentially growing *E. coli* ribosomes approximately every 7.4 seconds. Since the *phoA* signal/*hEGF* gene was located behind the efficient *tac* promoter, it was also assumed that the EGF mRNA would be in constant supply and would not be a limiting factor for EGF production. Although there has been much debate in the literature over the exact nature of secretion sites located in membranes (109-112), it seems fair to conclude that such sites do exist and that there is a finite number of such sites. The concern in this study was, that if the EGF precursor molecule was being translated off the ribosomes with higher efficiency than that of the secretion sites to process and transport them out of the cytoplasm, a buildup of EGF precursor molecules might form in the cytoplasm. This accumulation of precursor EGFs waiting for available secretion sites could afford enough time for the precursor molecules to fold and/or aggregate into some stable conformation or complex that would be non-competent for export. This in turn might lead to the proteolytic destruction of these forms in the cytoplasm.

A series of experiments that addressed the issue of increasing EGF yield in the periplasm by trying to control

protein folding rates *in vivo* was undertaken. In short, the objective was to find conditions under which the cell's environment would not be limiting for any nutrient, yet under which the *in vivo* translation rate might be controlled exogenously. Because of the previously mentioned results that EGF production on a per cell basis was increased in cells grown under non-optimal growth conditions, and hence slower protein translation rates, initial efforts focused on trying to slow protein expression rates. It was hoped that by slowing translation rates of elongating nascent polypeptide chains while they were still associated with the ribosome, sufficient time would be gained to allow the leader sequence of the EGF precursor molecule to associate with a secretion site and initiate the process of translocation while the rest of the nascent precursor molecule was still attached to the ribosome. This would tend to favor a co-translational type of secretion as opposed to a post-translational one, thereby limiting the time any pre-formed precursor molecules would have, while existing free in the cytosolic milieu, to fold or aggregate. Co-translational secretion would also energetically limit the conformational folding process in which the precursor molecule could fold into an entity that was non-competent for transport across the inner membrane. This would be accomplished by associating both ends of the nascent polypeptide with objects other than itself (one end with the membrane secretion site, and the other with the

ribosome). The protein synthesis inhibitor drug chloramphenicol (D(-)threo-2,2-dichloro-N[β -Hydroxy- α (hydroxymethyl)-p-nitrophenethyl]acetamide) is known to bind to the 50S subunit of bacterial ribosomes and inhibit the peptidyl transferase catalyzed peptide bond formation reaction in elongating polypeptide chains (113) by lowering the maximal velocity of this reaction. Therefore the addition of chloramphenicol to cultures of *E. coli* actively growing in LB medium was used for slowing the translational rate of the EGF precursor molecule.

Initial experiments with wild-type EGF showed a 5 fold increase in the amount of EGF isolated from the periplasm after addition of chloramphenicol to actively growing cultures of *E. coli* JM107 (Figure 8). It is also apparent that the beneficial effect of low dosages of chloramphenicol ($\leq 5 \mu\text{g/ml}$ final concentration) on EGF production is temporally related to induction of EGF expression. That is, chloramphenicol must be added to the growing culture at least 1 hour prior to de-repression of the *lac* operator by IPTG to see an increased EGF production. Addition of chloramphenicol concurrent with or post IPTG addition does not stimulate EGF production above the basal level in the absence of chloramphenicol. Figure 9 indicates that the addition of chloramphenicol does slow the overall growth rate of the *E. coli* culture as would be expected if protein translation rates were inhibited. When chloramphenicol was tested for its ability to stimulate EGF

production/secretion in *E. coli* cells producing mutant EGF proteins that normally did not express well, similar beneficial results were observed. These are seen in Table 3.

Based on these observations, a standardized procedure for the expression of wild-type and mutant EGF proteins was established. An overnight culture of *E. coli* JM107 harboring the expression plasmid containing either the wild-type or a mutant EGF gene is grown in LB medium containing 25 $\mu\text{g/ml}$ ampicillin and then diluted 100-fold into the same pre-warmed medium and grown at 37°C and 300 rpm in a rotary shaker. Three hours later ($A_{600}=0.5-0.8$), chloramphenicol is added to a final concentration of 5 $\mu\text{g/ml}$. The culture is allowed to continue growth under the same conditions for an additional two hours, at which time ($A_{600}=1.2-1.4$) expression of EGF protein is induced by the addition of IPTG to a final concentration of 0.1 mM. The culture is allowed to grow until EGF production is maximal, which is typically 10-12 hours later. The cells are then harvested by centrifugation.

Isolation and Purification of EGF Proteins

EGF protein sequestered in the periplasm was isolated from the cell pellet by resuspension in ice-cold 1M Tris-HCl (pH 9.0) until a homogeneous suspension was obtained. EDTA was then added to a final concentration of 2 mM with stirring. The suspension was kept on ice for 20 min with occasional stirring. The Tris supernatant was removed after

centrifugation and precipitated on ice with $(\text{NH}_4)_2\text{SO}_4$ to 80% saturation with stirring for 45-60 min. After centrifugation at 39,000 x g for 50 min. the pellet was resuspended in 25 mM sodium phosphate (pH 7.2) and dialyzed overnight against the same buffer using a Spectrapor #6 (M_r cutoff 2,000) dialysis tubing. The dialyzed fraction was loaded onto a column (2.3 x 40 cm) of Sephadex G-75 equilibrated in 25 mM sodium phosphate (pH 7.2) and eluted with the same buffer. Fractions containing EGF activity were pooled and further purified by reverse phase chromatography through a Vydac 218TPS column (4.6 x 250 mm ; The Separations Group) on a Waters model 600E HPLC system utilizing a 18-34% CH_3CN gradient in 10 mM sodium phosphate (pH 7.2) buffer.

Characterization of Wild-type hEGF

Figure 10 shows a typical elution profile of recombinant hEGF off the RPC-18 column. Protein eluting in this peak was characterized with respect to its physicochemical and biological properties as described below.

Size of Protein Product

The hEGF protein was found to be homogeneous, migrating with the apparent M_r of hEGF on SDS-PAGE (Figure 11).

Amino Acid Composition and Protein Sequencing

Amino acid composition analysis and protein sequencing were carried out under the direction of Mr. Claude D. Stringer of the Biology Division, Oak Ridge National Laboratory. Analysis was carried out according to standard methods (114) after hydrolysis in 6 N HCl and 10 mM 2-mercaptoethanol for 21 h at 110°C, utilizing a single-column procedure and a Beckman model 121-M analyzer. Cysteine determination as cysteic acid was accomplished after dimethyl sulfoxide oxidation (115). Tryptophan determination was accomplished after hydrolysis in 3 N mercaptoethanesulfonic acid (116). Results of the amino acid analysis are shown in Table 4. Sequencing, by Edman degradation, of the carboxymethylated hEGF was performed on an Applied Biosystems 470A gas-phase protein sequenator. Carboxymethylation of the purified hEGF was carried out in a 50 μ l reaction mixture containing 7M guanidine HCl, 10 mM sodium phosphate (pH 8.0), 0.36 mM hEGF, 35 mM recrystallized iodoacetate, 10 mM DTT and 3 mM EDTA for 15 min in the dark at room temperature and stopped by the addition of DTT to a final concentration of 100 mM. Results of this analysis showed that the sequence of the purified recombinant hEGF was identical to that of authentic, full-length, human EGF (22).

Extinction Coefficient Determination

The molar extinction coefficient of wild-type hEGF was calculated based on dry weight measurements. Purified hEGF was exhaustively dialyzed against distilled, deionized H₂O (pH 7.0) to remove any counter ions. Dialyzed protein was dried by lyophilization and kept under vacuum until the dry weight was measured. A 1 mg/ml solution of hEGF in distilled, deionized H₂O (pH 7.0) was spectrophotometrically measured on a Zeiss spectrophotometer. The ϵ_{280} was determined to be 14100 M⁻¹cm⁻¹ with an A_{280/260} of 1.653.

Biological Activity

Previous studies have shown that mEGF and hEGF are equipotent in their receptor binding ability (23). Purified recombinant hEGF was tested for its ability to compete with mEGF (purified from male mouse submaxillary glands and purchased from Collaborative Research) for the EGF receptor. [¹²⁵I]mEGF was prepared by the chloramine-T method (117). Purified membrane fractions containing human EGF receptors were isolated from the human epidermoid carcinoma cell line A431 (kindly provided by Dr. J. E. DeLarco, Monsanto Research Laboratory) as previously described by Akiyama *et al.* (118). Specific binding of hEGF to the EGF receptor was carried out as described by Carpenter (119). Specifically, a 100 μ l reaction mixture containing 20 mM HEPES (pH 7.4), 0.125% BSA, 16 nM [¹²⁵I]mEGF, 20 μ g total membrane protein from the

purified A431 membrane fractions, and known concentrations of hEGF were incubated at 30°C for 30 min. Samples were quickly diluted to 1 ml with 20 mM HEPES (pH 7.4) containing 0.1% BSA and filtered on Millipore GVWP (0.22 μ m pores) filters. Filters were washed four times with 5 ml aliquots of 20 mM HEPES(pH 7.4)/0.1% BSA, dried and quantified by liquid scintillation counting. EGF receptor tyrosine kinase activity was carried out for hEGF and compared to mEGF as described in Chapter IV. The results of these assays are presented in Table 5 and show that the recombinant wild-type hEGF was able to compete with mEGF on an equimolar basis for binding to the EGF receptor, and that it was fully capable of stimulating the EGF receptor's tyrosine kinase activity.

CHAPTER IV

STRUCTURE-FUNCTION STUDIES OF LOOP C MUTANTS

Receptor Affinity Studies With Tyr-37 and Arg-41 Analogs

Mutant EGF proteins with single amino acid substitutions (see Chapter II, Table 2) were characterized with respect to their ability to bind to the EGF receptor. This was accomplished as previously described (see Chapter III) for wild-type EGF. Specifically, known concentrations of mutant protein collected from the HPLC peak which eluted from the RPC-18 column at or around the same retention time as that of wild-type EGF were assayed for their ability to displace a known concentration of wild-type [^{125} I]hEGF, iodinated as previously described for mEGF to an average specific activity of 150,000 cpm/pmol hEGF, in the radioreceptor binding assay. The concentration of competing mutant species of EGF that gave a 50% inhibition of binding of wild-type [^{125} I]hEGF to the EGF receptor was termed the IC_{50} value and was estimated graphically for each mutant species.

Figure 12 shows the competition binding curves for wild-type EGF in relation to the singly substituted mutants of tyrosine 37, namely, Tyr³⁷→Phe, Tyr³⁷→His, Tyr³⁷→Ser, Tyr³⁷→Ala, Tyr³⁷→Asp, Tyr³⁷→Arg, and Tyr³⁷→Gly. Similarly, Figure 13 shows the competition binding curves for wild-type EGF and the singly substituted mutants of arginine 41, namely, Arg⁴¹→Lys,

Arg⁴¹→Gln, Arg⁴¹→Ile, Arg⁴¹→Tyr, Arg⁴¹→Gly, and Arg⁴¹→Ala. No receptor binding was detected for either the Arg⁴¹→Asp or Arg⁴¹→Glu mutants at concentrations of mutant protein below 100 μ M.

The IC₅₀ values for the various mutants are presented in Table 6, along with the relative receptor affinity of each mutant protein as compared to wild-type EGF which was set at 100%. The conservative change of Tyr-37 to Phe-37 produced a mutant that retained the aromatic character of the original phenol ring but lost the polar -OH group. This mutant had a somewhat higher EGF receptor affinity (126%) than did wild-type EGF. A semiconservative change to histidine, in which the imidazole ring retains some of the aromatic character and the approximate size of tyrosine, but again lacks the polar -OH group, resulted in a mutant protein that retained 74% of the receptor affinity as compared to that of wild-type. Substitution of Tyr-37 with the small aliphatic and polar serine, which lacks the benzene ring moiety of the original phenol ring, but retains the polar -OH group, only moderately decreased receptor affinity to 62% of wild-type, while replacement of the tyrosyl side chain with the small hydrophobic methyl group of alanine reduced the affinity to 39% of wild-type. The introduction of charged residues at position 37 produced mutant species that exhibited somewhat different receptor affinities based on the polarity of the charge introduced. Substitution of the non-charged tyrosine

with the acidic aspartic acid resulted in the mutant protein having a receptor affinity of 26% relative to that of wild-type. However, substitution at position 37 with the basic amino acid arginine, resulted in a mutant with a receptor affinity of only 10% relative to wild-type EGF. Substitution of Tyr-37 with a glycyl residue further reduced receptor affinity to only 6.5% that of wild-type. The significance of these values, relative to possible structural influences resulting from each substitution, will be discussed below.

All mutations at amino acid position 41, unlike position 37, resulted in mutant proteins with drastically reduced receptor affinity, regardless of the nature of the substitution. The conservative substitution of Arg-41 with lysine, which conserves the charge properties at that position, lowers the receptor affinity by about 2.5 orders of magnitude. All other substitutions at position 41, whether large and hydrophobic (Arg⁴¹→Ile), small and hydrophobic (Arg⁴¹→Ala), polar and aromatic (Arg⁴¹→Tyr), side chain deficient (Arg⁴¹→Gly), or large, polar and uncharged (Arg⁴¹→Gln) also greatly reduce receptor affinity. In addition, charge transposition (*i.e.* replacing the positively charged arginine residue with acidic residues) is extremely detrimental to binding ability, since both the Arg⁴¹→Asp and the Arg⁴¹→Glu mutants have receptor affinities less than 0.01% that of wild-type EGF.

Stimulation of EGF Receptor Protein-Tyrosine Kinase Activity

It is believed that the mitogenic function of EGF is mediated through the action of the EGF-receptor's ability to phosphorylate certain endogenous cellular components, thereby transferring EGF's initial binding signal to secondary messengers that signal the cell to divide. Phosphorylation on tyrosyl residues is carried out by the tyrosine kinase domain of the EGF receptor.

Determination of the ability to stimulate the EGF receptor's protein-tyrosine kinase activity for mutant protein relative to wild-type EGF was carried out as follows: Five μ l of an EGF receptor preparation [detergent-solubilized and purified on a wheat germ agglutinin agarose column by the method of Akiyama *et al.* (118)], EGF protein, reaction mixture [final concentration of 20 mM HEPES (pH 7.4), 10 mM MgCl_2 , 2 mM MnCl_2 , 1 mM DTT, 250 mM $(\text{NH}_4)_2\text{SO}_4$, 250 mM NaCl, 5% glycerol, 0.1 mM Na_3VO_4 , 0.5 mM PMSF, and 0.05% Triton X-100], and 50 μ g of the synthetic substrate $(\text{Glu4:Tyr1})_n$ (Sigma Chemical Co., avg. $M_r = 35,000$) were incubated for 15 min at room temperature in a 100- μ l assay volume. Reaction was initiated by the addition of ATP (final concentration of 100 μ M ATP and 5 μ Ci [γ - ^{32}P]ATP) and allowed to proceed for an additional 10 min at room temperature. Termination of the reaction was accomplished by the addition of 1 ml of ice-cold 5% TCA + 10 mM sodium pyrophosphate and precipitation on ice for 10 min. Acid-insoluble precipitates were collected by filtration on

Millipore HAWP (0.45 μm pore size) filters, washed four times, dried, and quantified by liquid scintillation counting.

This particular assay was sensitive to EGF stimulation over the range of 0-10 μM final EGF concentration (120). This restricted the use of this assay as a functional characterization to those mutants whose relative receptor affinity was greater than 1% relative to that of wild-type EGF. Therefore, the EGF mutants at position 41 could not be evaluated by this assay. EGF mutants at position 37, however, could be assayed for their ability to stimulate the EGF receptor tyrosine kinase activity in this assay and these results are shown in Figure 14. The effective concentration of each mutant EGF resulting in 50% maximal stimulation of the EGF receptor tyrosine kinase activity was termed the EC_{50} value and was estimated graphically for each mutant species. These values are presented in Table 7 along with the relative receptor affinity, as measured by the tyrosine-kinase stimulation, compared to wild-type EGF which was set at 100%. In general, these values compare favorably with the values for receptor affinity exhibited by each of the mutants as shown in Table 6, implying that mutants involving Tyr-37 are fully capable of stimulating the receptor kinase activity if a sufficient concentration is added to compensate for any lowered receptor affinity caused by the particular amino acid substitution.

Structural Studies

One-dimensional ^1H -NMR studies were undertaken to ascertain whether the decrease in functional activity (*i.e.* receptor affinity) was in fact due to misfolding of the polypeptide chain influenced by the introduced mutation. One-dimensional NMR spectroscopy provides a convenient experimental tool for quickly determining grossly misfolded protein structure from native solution structure. Aromatic residues are spread throughout the hEGF molecule in its 3-dimensional structure (38) and the resonance frequencies of these aromatic ring protons are sensitive to molecular conformation and solvent accessibility. Accordingly, comparisons of proton resonance frequencies from aromatic residues in wild-type and mutant proteins provide a convenient method for distinguishing correctly folded from incorrectly folded mutants. This has been used as a tool for probing the structure of a mEGF mutant by Moy *et al.* (121) and shown to be a reliable indicator of molecular conformation as verified by further analysis by 2D NMR methodology.

The 1D ^1H -NMR spectroscopy of wild-type and mutant proteins was carried out on purified samples of proteins exhibiting single peaks on reverse phase HPLC analysis. Samples were dialyzed against distilled, deionized water at pH 6.5 after purification and lyophilized to dryness. Samples were resuspended in D_2O to a concentration of 1 mM and the pD adjusted to 7.2 ± 0.1 with 0.1 M NaOH in D_2O . Samples were

allowed to exchange all labile protons for deuterons for a minimum of 3 hours before measurements were recorded. ^1H -NMR spectra were recorded at 30°C on a Varian VXR500 (500 MHz) spectrometer, under the direction of Dr. Gaetano T. Montelione of the New Jersey Center for Advanced Biotechnology and Medicine, and the Department of Chemistry, Rutgers University. Figures 15 and 16 show the aromatic and downfield C^αH region (from 5.0 - 8.5 ppm) of the 1D ^1H -NMR spectra for selected mutant proteins at each site under study. The mutants selected for 1D NMR analysis were chosen from the series of mutants produced at each site based on the type of amino acid substitution introduced at the site, the quantity of desired protein able to be readily purified, and from analysis of receptor affinity values, such that typically the best binding, an intermediate binding, and a poorer binding species were represented in the structural comparison analysis.

NMR Studies of Position 37 Mutants

For mutants involving Tyr-37 (see Figure 15), a cautious interpretation of chemical shift differences observed between wild-type and the various mutants was employed, since all substitutions at this site involved aromatic ring substitutions. Alterations in "ring current", and hence local electric field differences, will be involved in any substitution at this site, including substitutions by other aromatic ring structures such as phenylalanine or histidine.

Since it was expected that some conformation-independent changes in chemical shifts would be observed in these mutants, only generalized conclusions about gross structural features of the mutant protein structures were drawn. The Tyr³⁷→Phe and Tyr³⁷→His mutants still retained tertiary structure as indicated by profiles similar to wild-type in the aromatic region (6.0 - 8.5 ppm). The β structure of these two mutants was probably still intact as indicated by near-wild-type profiles in the downfield C ^{α} H region (5.0 - 6.0 ppm), that region of the spectrum known (122) to contain proton resonances from polypeptide backbone α -carbons involved in both the major β -sheet of loop B and the minor β -sheet of loop C. This is perhaps not too surprising given the fact that the Tyr³⁷→Phe mutant actually binds to the EGF receptor with greater affinity (see Table 6) than does wild-type, and that the Tyr³⁷→His mutant has close to 75% of the affinity of wild-type.

The Tyr³⁷→Ala 1D NMR profile (Figure 15) is somewhat more altered. It may only be a reflection of the loss of the tyrosine residue's "ring current" upon substitution with the non-aromatic alanine, or it may truly represent the loss of some structural integrity of the mutant as compared to the wild-type native structure, since the resonance frequencies in its C ^{α} H region also look different from wild-type. Only further 2D NMR analysis will be able to answer this question. However, it is interesting that this mutant still retains a significant (40% relative to wild-type) receptor affinity.

It is also interesting to note, even though no NMR studies have yet been performed on these mutants, that the tyrosine to arginine substitution at this site has a 2.6 fold more deleterious effect on receptor affinity than does a tyrosine to aspartate substitution. This quantitative difference in receptor affinity may also be suggestive of structural perturbations introduced by the localized positively charged arginine at position 37. It is presumed, from previous NMR measurements (32,37) of slowly exchanging amide protons, that the amide proton of residue 37 is hydrogen bonded to the carbonyl oxygen of residue 34 in the native structure of EGF. The amide protons of residues 38 and 44, which lie close in space to residue 37, are also hydrogen bonded to the respective carbonyl oxygens of each other (see Figure 1). It may be that an arginine side chain at position 37 is more destabilizing to these existing hydrogen bonds due to its potential to act as a stronger hydrogen donor (5 potential hydrogen donor sites) than does aspartate as a hydrogen acceptor (2 potential hydrogen acceptor sites at this pH of 7.2). This could attract more water molecules to the area, thereby reducing the potential for the existing hydrogen bonds by making them more solvent-accessible. This might explain the 2.6 fold quantitative difference in receptor affinities between these two mutant proteins. Further NMR measurements looking at exchange rates of amide protons in D₂O would be needed to test this hypothesis.

The Tyr³⁷→Gly mutant has the lowest receptor affinity of the entire series of mutant proteins produced at this site. This particular mutant also displayed the greatest potential to exhibit multi-peaked HPLC profiles of the final purified protein product (Figure 7). This might be explained by different conformational forms of this mutant, since substitution with a glycyl residue would be expected to allow more freedom of rotation around the phi (ϕ) and psi (ψ) bonds at this position, favoring alternate conformations of this mutant compared to wild-type. For mutations at position 37, information gleaned from the functional studies and limited structural studies, seems to indicate that the tyrosine at position 37 is not essential for receptor interaction, and is probably highly conserved in EGF species for reasons regarding the structural integrity of the EGF molecule rather than for reasons of direct chemical interaction with the EGF receptor.

NMR Studies of Position 41 Mutants

Figure 16 shows profiles of some of the position 41 mutants. The 1D NMR spectra of wild-type and the Lys-41 mutant display very similar profiles with extremely few peaks exhibiting significant chemical shifts (≥ 50 ppb as defined by Moy *et al.* (121)). This indicates that there is no major structural changes in the global chain fold of this particular mutant with respect to the wild-type native conformation. The Arg⁴¹→Gln profile exhibits a few peaks with minor chemical

shift differences with respect to wild-type, although no major structural alterations are suspected due to the close relationship to the wild-type profile in the aromatic region and the same $C^{\alpha}H$ resonances as wild-type, suggesting the conservation of β -sheet structure in this mutant. The Arg⁴¹→Ala mutant displays a somewhat more altered profile, with some peak broadening and several peaks with small chemical shift differences.

Although it seems likely that there are no major structural alterations when lysine is conservatively substituted for the arginine at position 41, the extent of structural perturbations with more non-conservative substitutions, such as the alanine substitution, will require further 2D NMR analysis and the assignment of specific resonances to specific residues. The qualitative difference in the spectra between the non-conservative substitution Arg⁴¹→Ala and the conservative substitution Arg⁴¹→Lys, might explain the relative 8 fold lower receptor affinity of the former compared to the latter, irrespective of how low their affinities are relative to wild-type (see Table 6). The important conclusion concerning this site is that although the structural features of the Lys-41 mutant appear very native-like, this particular substitution results in a lowering of receptor affinity by about 2.5 orders of magnitude. This extremely significant lowering of receptor affinity and near native 1D NMR profile strongly suggest that

the guanidino group of arginine at position 41 may be intimately involved in receptor interaction. This may explain the absolute conservation of an arginine residue at position 41 in those protein species capable of binding to the EGF receptor.

Conclusions

In light of the above, it can be inferred that since even non-conservative changes are tolerated quite well at position 37, the naturally occurring tyrosine residue at that position is not essential for direct chemical interaction with the EGF receptor. Likewise, it can be inferred that since all mutations at position 41, even highly conservative ones, drastically reduce receptor affinity, and that there appears to be very little structural perturbations induced by mutations at this site (especially the conservative Lys-41 mutation), the naturally occurring arginine residue at that position is essential for receptor interaction and may serve as a contact point for the receptor.

The roles of Tyr-37 and Arg-41 in hEGF, as discerned from these site-directed mutagenesis studies, are further supported by comparing the evolutionary conservation of chemically similar residues with respect to amino acid position in other EGF-like species. As discussed in Chapter I, there are several EGF-like (type-I) proteins that are able to bind to the EGF receptor. In addition to these proteins, there is

another group of proteins (type-II) that have entirely different functions in the cell and are not known to bind to the receptor, yet contain one or more EGF-like domains within them (see Table I). The sequences of some of these two classes of proteins are shown in Figures 3 and 17, respectively. That tyrosine 37 is conserved for maintaining some structural feature(s) common to *all* EGF-like species and/or domains is supported by the fact that the chemical functionality of an aromatic residue at position 37 is not only highly conserved in those type-I EGF-like species capable of binding the EGF receptor (see Figure 3), but is also highly conserved at those positions corresponding to position 37 in hEGF in the EGF-like domains of type-II proteins that do not bind the EGF receptor (see Figure 17). Conserved chemical functionality of the arginine at position 41 in hEGF however is *not* maintained in both classes of proteins (see Figure 17), rather *only* in the type-I EGF-like protein species that have been shown to bind to the EGF receptor (Figure 3).

Recently, there have been reports by others (123-125) of similar studies utilizing site-directed mutagenesis to examine the role(s) of the analogous residues in hTGF- α , namely Arg-42 and Tyr-38. Defeo-Jones *et al.* (123) showed that substitution of Arg-42 in TGF- α also reduced receptor binding to less than 1% of wild-type TGF- α . This further supports the observation of the essentiality of Arg-41 in hEGF for receptor binding in the present study. However, the similarity between EGF and

TGF- α regarding receptor binding may not completely extend to Tyr-37 in hEGF (Tyr-38 in hTGF- α). Although conservative changes to phenylalanine at position 38 in hTGF- α also produced a mutant with a somewhat higher receptor affinity (124) similar to that demonstrated for hEGF Tyr³⁷→Phe in the present study, the Ala-38 mutant of hTGF- α (unlike the Ala-37 mutant of hEGF discussed in the present study) was only 3% as active as wild-type TGF- α . Since no structural analysis was performed on the TGF- α Ala-38 mutant, it is not yet clear as to whether this is a real difference between EGF and TGF- α in the way they recognize and bind to the same receptor, or simply the result of some structural perturbation in the TGF- α mutant. Lazar *et al.* (125) also attempted to show the essentiality of an aromatic residue at position 38 of TGF- α for receptor binding, although their studies were done with crude samples of mutant analogs, making their specific activity determinations (receptor affinities) suspect, and structural studies of the mutant proteins impossible. The present study, as compared to others (123-125), re-emphasizes the importance of combining biological assays with structural determinations when utilizing site-directed (or random) mutagenesis to draw meaningful relationships between the structure and function of protein molecules.

Results of Other Loop C Mutants

As stated in Chapter I, in addition to the two cysteines, Tyr-37, and Arg-41 in loop C of hEGF, the only other residue absolutely conserved is Gly-39. Reexamination of Figures 3 and 17 show, that with the exception of amphiregulin, and the two pox viruses -- myxoma virus and Shope fibroma virus-- that Gly-36 is also relatively highly conserved. These residues as well as Ile-38, which is not chemically conserved in either the type-I or type-II class of proteins and served as a negative control, were also subjected to preliminary examination by site-directed mutagenesis (see Table 8). Substitution of Ile-38 with leucine produced a mutant protein retaining significant receptor affinity (77% relative to wild-type), and eluted from the RPC-18 column (in the final purification step) with retention time and profile similar to those of wild-type EGF (Figure 18). This suggests that the mutant retained the gross structural characteristics of wild-type EGF. Substitution of the highly conserved Gly-39 with leucine results in a mutant protein that is probably structurally altered from wild-type, as indicated by its elution profile off the RPC-18 column (Figure 18). Since there was no protein peak with a similar retention time as compared to wild-type EGF, the earliest eluting species was assayed for its receptor affinity which was found to be only 6% that of wild-type. This suggests that this mutation might have disrupted the second β turn region of loop C since

glycine would be sterically favored at this position to effect a β turn. A disruption in this β turn could possibly distort or prevent the formation of β -sheet structure in loop C and could explain the lowered receptor affinity of this particular mutant. This is suggestive of a structural role for Gly-39 and may explain its high degree of conservation not only in those type-I EGF-like species able to bind the EGF receptor (see Figure 3), but also in those EGF-like domains in type-II proteins that do not bind to the EGF receptor (see Figure 17). Substitution of Gly-36 with valine resulted in a mutant protein that exhibited many different conformational species distinct from wild-type, as indicated by its reverse phase chromatography profile (Figure 18). Since none of these conformationally distinct species was produced in sufficient quantity to permit binding assays, receptor affinity of this mutant was not determined.

Summary and Implications

The present studies were initiated at a time when relatively little was known about the structure-function relationships in the EGF molecule. Although there is a considerable gap in our complete understanding of the structure/function properties of EGF, some preliminary conclusions can be drawn from these studies. Use of site-directed mutagenesis has allowed the placement of each of the highly conserved amino acid residues located in loop C into

categories based on conservation for structural or functional reasons.

Those residues believed to fall into the category conserved for structural reasons are glycine-36 and glycine-39. Tyrosine-37 is also tentatively placed in this category, based on its non-essential role in receptor interaction and its high degree of conservation throughout both the type-I and type-II family of proteins. Further 2D NMR structural analysis of the wild-type and mutant proteins will be necessary to elucidate the structural role of Tyr-37. While it is acknowledged that conservation for structural reasons is not necessarily mutually exclusive of conservation for some functional property inherent in EGF, this difference is hard to differentiate at present. For residues with both a structural and functional role, the latter can often be obscured by the former. In those cases where a residue is thought to be conserved primarily for structural reasons, it is not clear what the residual receptor affinity reflects. It may represent the ability of some fraction of a dynamic population of mutant molecules to retain sufficient native EGF structure to interact with the receptor, or may suggest that a defined static structural perturbation exists in the mutant protein that causes the molecule to possess a receptor affinity unique to that structural alteration. These two scenarios are indistinguishable from the functional standpoint since present functional studies always rely on an equilibrium

binding between EGF and its receptor, conditions which take several minutes to several hours to occur depending on the assay. However, it is known that native EGF has very limited flexibility, denoted by the very slow rates of exchange of amide protons involved in hydrogen bonds with the solvent. This may lead one to suspect that mutant proteins that do interconvert between two or more forms due to the mutation involved would be detected by NMR as broadened peak resonances. Therefore, this situation may be resolved by NMR structural studies if enough mutants are examined.

It is somewhat easier to place arginine-41 into the category conserved for functional reasons. This is based on mutants at this position retaining native-like structural characteristics yet almost completely abolishing receptor interaction. That the arginine in this position may also be involved in receptor interaction is indicated by its unequivocal conservation in only that group of proteins known to bind to the EGF receptor. Furthermore, the only two proteins in the type-II class of proteins that do have a conserved arginine at position 41 are the laminin B1 and B2 chains. This is significant because it has been shown (at least for the laminin B1 chain) that this arginine residue resides in the segment of the polypeptide chain that is known to be necessary for laminin to bind to its receptor (126). Since it has been shown for laminin that a pentapeptide encoding the residues Y-I-G-S-R (residues #37-41 using the EGF

numbering system) competes for the laminin receptor binding site, it seems highly likely that the analogous residues in EGF from loop C namely, Y-I-G-E-R (residues #37-41), might also be involved in protein-protein interaction. The only other conserved residues in this sequence, namely, Tyr-37 and Gly-39, have been shown in this study either to be nonessential for receptor interaction, as is the case for Tyr-37, or to be conserved for structural reasons, as is the case for Gly-39. This again suggests that Arg-41 functions as an essential residue in receptor interaction. Such observations between otherwise diverse proteins, may eventually lead not only to a better understanding of the EGF/EGF-receptor interaction, but possibly of protein-protein interaction in general.

The knowledge as to which residues in EGF act as functional moieties and which act as structural moieties has far-reaching implications. Once enough information is available as to which residues participate in specific structural or functional roles in EGF, the re-designing of EGF by protein engineering can take place. Such protein engineering could eventually yield super agonists of EGF which might be used for clinical uses that require enhanced cell growth, *e.g.* wound repair, replacement of burned tissue, *etc.* Likewise, antagonists of EGF binding could also be produced which might be able to bind to the receptor without stimulating the mitogenic response. Such antagonists might

find a use in combating EGF-dependent growth of epithelial tumors. The possibilities of the uses of newly designed EGF analogs are limited only by the rate of knowledge gathered on the role of each naturally occurring residue in wild-type EGF. This study has elucidated the role of several of the highly conserved residues in loop C.

LIST OF REFERENCES

LIST OF REFERENCES

1. O'Keefe, E., Hollenberg, M.D., and Cuatrecasas, P.
(1974) *Arch. Biochem. Biophys.* **164**, 518-526
2. Carpenter, G. (1984) *Cell* **37**, 357-358
3. Carpenter, G. (1987) *Ann. Rev. Biochem.* **56**, 881-914
4. Cohen, S. (1960) *Proc. Natl. Acad. Sci. USA* **46**, 302-311
5. Levi-Montalcini, R., and Cohen, S. (1960) *Ann. NY Acad. Sci.* **85**, 324-341
6. Cohen, S. (1962) *J. Biol. Chem.* **237**, 1555-1562
7. Cohen, S., and Elliot, G.A. (1963) *J. Invest. Dermatol.* **40**, 1-5
8. Cohen, S. (1964) in "Metabolic Control Mechanisms in Animal Cells" W.J. Rutter (ed.) *NCI Monograph* **13**, 13-37
9. Sandweiss, D.J., Solzstein, H.C., and Farbman, A. (1938) *Am. J. Digest. Diseases* **5**, 24-30
10. Culmer, C.U., Atkinson, A.J., and Ivy, A.C. (1939) *Endocrinology* **24**, 631-637
11. Gray, J.S., Wieczorowski, E., and Ivy, A.C. (1939) *Science* **89**, 489-490
12. Gray, J.S., Culmer, C.U., Wieczorowski, E., and Adkison, J.L. (1940) *Proc. Soc. Exp. Biol. Med.* **43**, 225-228
13. Friedman, M.H.F. (1951) in "Vitamins and Hormones" Academic Press (New York) Harris, P.S., and Thimann, K.V. (eds.) **Vol. IX**, 313-353
14. Gregory, R.A. (1955) *J. Physiol.* **129**, 528-546
15. Mongar, J.L., and Rosenoer, V.M. (1962) *J. Physiol.* **162**, 163-172
16. Morimoto, T., and Yamamoto, M. (1969) *J. Pharm. Soc. Japan* **89**, 215-221

17. Carrea, G., Casellato, M.M., Manera, E., Pasta, P., and Lugaro, G. (1973) *Biochem. Biophys. Acta.* **295**, 274-282
18. Gregory, H., and Willshire, I.R. (1975) *Hoppe-Seylers Z. Physiol. Chem.* **356**, 1765-1774
19. Savage, C.R., Jr., Inagami, T., and Cohen, S. (1972) *J. Biol. Chem.* **247**, 7612-7621
20. Savage, C.R., Jr., Hash, J.H., and Cohen, S. (1973) *J. Biol. Chem.* **248**, 7669-7672
21. Gregory, H. (1975) *Nature* **257**, 325-327
22. Gregory, H., and Preston, B.M. (1977) *Int. J. Peptide Protein Res.* **9**, 107-118
23. Cohen, S., and Carpenter, G. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1317-1321
24. Simpson, R.J., Smith, J.A., Moritz, R.L., O'Hare, M.J., Rudland, P.S., Morrison, J.R., Lloyd, C.J., Grego, B., Burgess, A.W., and Nice, E.C. (1985) *Eur. J. Biochem.* **153**, 629-637
25. Kobayashi, R., Reeve, J.R., Jr., and Walsh, J.H. (1985) *Biochem. J.* **229**, 611-619
26. Taylor, J.M., Mitchell, W.M., and Cohen, S. (1972) *J. Biol. Chem.* **247**, 5928-5934
27. Holliday, L.A., Savage, C.R., Jr., Cohen, S., and Pruett, D. (1976) *Biochemistry* **15**, 2624-2633
28. Mayo, K.H. (1984) *Biochemistry* **23**, 3960-3973
29. Mayo, K.H. (1985) *Biochemistry* **24**, 3783-3794
30. Kohda, D., Kodama, C., Kase, R., Nomoto, H., Hayashi, K., and Inagaki, F. (1988) *Biochem. Int.* **16**, 647-654
31. Mayo, K.H., Schaudies, P., Savage, C.R., Jr., DeMarco, A., and Kaptein, R. (1986) *Biochem. J.* **239**, 13-18
32. Carver, J.A., Cooke, R.M., Esposito, G., Campbell, I.D., Gregory, H., and Sheard, B. (1986) *FEBS Lett.* **205**, 77-81

33. Makino, K., Morimoto, M., Nishi, M., Sakamoto, S., Tamura, A., Inooka, H., and Akasaka, K. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7841-7845
34. Montelione, G.T., Wüthrich, K., Nice, E.C., Burgess, A.W., and Scheraga, H.A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 5226-5230
35. Kohda, D., and Inagaki, F. (1988) *J. Biochem. (Tokyo)* **103**, 554-571
36. Kohda, D., Go, N., Hayashi, K., and Inagaki, F. (1988) *J. Biochem. (Tokyo)* **103**, 741-743
37. Mayo, K.H., Cavalli, R.C., Peters, A.R., Boelens, R., and Kaptein, R. (1989) *Biochem. J.* **257**, 197-205
38. Cooke, R.M., Wilkinson, A.J., Baron, M., Pastore, A., Tappin, M.J., Campbell, I.D., Gregory, H., and Sheard, B. (1987) *Nature* **327**, 339-341
39. Campbell, I.D., Cooke, R.M., Baron, M., Harvey, T.S., and Tappin, M.J. (1989) *Prog. in Growth Factor Res.* **1**, 13-22
40. Wilmot, C.M., and Thornton, J.M. (1988) *J. Mol. Biol.* **203**, 221-232
41. DeLarco, J.E., and Todaro, G.J., (1978) *Natl. Acad. Sci. USA* **75**, 4001-4005
42. Derynck, R., Goeddel, D.V., Ullrich, A., Gutterman, J.U., Williams, R.D., Bringman, T.S., and Berger, W.H. (1987) *Cancer Res.* **47**, 707-712
43. Anzano, M.A., Roberts, A.B., Meyers, C.A., Komoriya, A., Lamb, L.C., Smith, J.M., and Sporn, M.B. (1982) *Cancer Res.* **42**, 4776-4778
44. Anzano, M.A., Roberts, A.B., Smith, J.M., Sporn, M.B., and DeLarco, J.E. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6264-6268
45. Twardzik, D.R. (1985) *Cancer Res.* **45**, 5413-5416
46. Lee, D.C., Rochford, R.M., Todaro, G.J., and Villareal, L.P. (1985) *Mol. Cell. Biol.* **5**, 3644-3646
47. Coffey, R.J., Jr., Derynck, R., Wilcox, J.N., Bringman, T.S., Goustin, A.S., Moses, H.L., and Pittelkow, M.R. (1987) *Nature* **328**, 817-820

48. Madtes, D.K., Raines, E.W., Sakariassan, K.S., Assoian, R.K., Sporn, M.B., Bell, G.I., and Ross, R. (1988) *Cell* **53**, 285-293
49. Massagué, J. (1983) *J. Biol. Chem.* **258**, 13614-13620
50. Pike, L.J., Marquardt, H., Todaro, G.J., Gallis, B., Casnellie, J., Bornstein, P., and Krebs, E.G. (1982) *J. Biol. Chem.* **257**, 14628-14631
51. Marquardt, H., Hunkapillar, M.W., Hood, L.E., Twardzik, D.R., DeLarco, J.E., and Todaro, G.J. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4684-4688
52. Marquardt, H., Hunkapillar, M.W., Hood, L.E., and Todaro, G.J. (1984) *Science* **223**, 1079-1082
53. Winkler, M.E., Bringman, T.S., and Marks, B.J. (1986) *J. Biol. Chem.* **261**, 13838-13843
54. Kohda, D., Shimada, I., Miyake, T., Fuwa, T., and Inagaki, F. (1989) *Biochemistry* **28**, 953-958
55. Montelione, G.T., Winkler, M.E., Burton, L.E., Rinderknecht, E., Sporn, M.B., and Wagner, G. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 1519-1523
56. Tappin, M.J., Cooke, R.M., Fitton, J.E., and Campbell, I.D. (1989) *Eur. J. Biochem.* **179**, 629-637
57. Shoyab, M., Plowman, G.D., McDonald, V.L., Bradley, J.G., and Todaro, G.J. (1989) *Science* **243**, 1074-1076
58. Stroobant, P., Rice, A.P., Gullick, W.J., Cheng, D.J., Kerr, I.M., and Waterfield, M.D. (1985) *Cell* **42**, 383-393
59. Lin, Y.Z., Caporaso, G., Chang, P.Y., Ke, X.H., and Tam, J.P. (1988) *Biochemistry* **27**, 5640-5645
60. Upton, C., Macen, J., and Mcfadden, G. (1987) *J. Virol.* **61**, 1271-1275
61. Rees, D.J.G., Jones, I.M., Handford, P.A., Walter, S.J., Esnouf, M.P., Smith, K.J., and Brownlee, G.G. (1988) *EMBO J.* **7**, 2053-2061
62. Carpenter, G., and Wahl, M.I. (1990) *Hand. Exptl. Pharmacol.* (in press)

63. Huang, L.H., Ke, X.H., Sweeney, W., and Tam, J.P. (1989) *Biochem. Biophys. Res. Comm.* **160**, 133-139
64. Handford, P.A., Baron, M., Mayhew, M., Willis, A., Beesley, T., Brownlee, G.G., and Campbell, I.D. (1990) *EMBO J.* **9**, 475-480
65. Buckley, A., Davidson, J.M., Kamerath, C.D., Wolt, T.B., and Woodward, S.C. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7340-7344
66. Schultz, G.S., White, M., Mitchell, R., Brown, G., Lynch, J., Twardzik, D.R., and Todaro, G.J. (1987) *Science* **235**, 350-352
67. Knighton, D.R., Fiegel, V.D., Austin, L.L., Ciresi, K.F., and Butler, E.L. (1986) *Ann. Surg.* **204**, 322-330
68. Carter, D.M., Balin, A.K., Gottlieb, A.B., Eisinger, M., Lin, A., Pratt, L., Sherbany, A., and Caldwell, D. (1988) *Prog. in Clin. Res.* **266**, 303-319
69. Gospodarowicz, D., Bialecki, H., and Thakral, T.K. (1979) *Exp. Eye Res.* **28**, 501-514
70. Woost, P.G., Brightwell, J., Eiferman, R.A., Schultz, G.S. (1985) *Exp. Eye Res.* **40**, 47-54
71. King, L.E., Jr., Ellis, D.L., Gates, R.E., Inman, W.H., Stoscheck, C.M., Fava, R.A., and Nanney, L.B. (1988) *Am. J. Med. Sci.* **31**, 154-158
72. Tsutsumi, O., Kurachi, H., and Oka, T. (1986) *Science* **233**, 975-977
73. Smith, J., Cook, E., Fotheringham, I., Pheby, S., Derbyshire, R., Eaton, M.W., Doel, M., Lilley, D. M., Pardon, J.F., Patel, T., Lewis, H. and Bell, L.D. (1982) *Nucleic Acids Res.* **10**, 4467-4482
74. Sumi, S.I., Hasegawa, A., Yagi, S., Miyoshi, K.I., Kanezawa, A., Nakagawa, S., and Suzuki, M. (1985) *J. Biotech.* **2**, 59-74
75. Brake, A.J., Merryweather, J.P., Coit, D.S., Heberlein, U. A. Masiarz, F. R., Mullenback, G. T., Urdea, M. S., Valenzuela, P., and Barr, P.J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 4642-4646

76. 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
77. Neu, H.C., and Heppel, L.A. (1965) *J. Biol. Chem.* **240**, 3685-3692
78. Swamy, K.H.S., and Goldberg, A.L. (1981) *Nature* **292**, 652-654
79. Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heyneker, H., Bolivar, F., and Boyer, H.W. (1977) *Science* **198**, 1056-1063
80. Derynck, R., Roberts, A.B., Winkler, M.E., Chen, E.Y., and Goeddel, D.V. (1984) *Cell* **38**, 287-297
81. Talmadge, K., and Gilbert, W. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1830-1833
82. Wells, J.A., Vasser, M., and Powers, D.B. (1985) *Gene* **34**, 315-323
83. Grantham, R., Gautier, C., and Guoy, M. (1980) *Nucleic Acids Res.* **8**, 1893-1912
84. Inouye, H., Barnes, W., and Beckwith, J. (1982) *J. Bacteriol.* **149**, 434-439
85. Kikuchi, Y., Yoda, K., Yamasaki, M., and Tamura, G. (1981) *Nucleic Acids Res.* **9**, 5671-5678
86. Larimer, F.W., Mural, R.J., and Soper, T.S. (1990) *Protein Engineering* **3**, 227-231
87. Sinha, N.D., Biernat, J., McManus, J., and Koster, H. (1984) *Nucleic Acids Res.* **12**, 4539-4557
88. Atkinson, T., and Smith, M. (1984) in *Oligonucleotide Synthesis* Gait, M.J. (ed.) pp.35-81, IRL Press, Oxford
89. Meinkoth, J., and Wahl, W. (1984) *Anal. Biochem.* **138**, 267-284
90. Hanahan, D. (1985) in *DNA Cloning* (Glover, D.M., ed.) Vol.1, pp. 109-135, IRL Press, Oxford
91. Birnboim, H.C., and Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513-1523

92. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467
93. Nestor, J.J., Newman, S.R., DeLustro, B., Todaro, G.J., and Schreiber, A.B. (1985) *Biochem. Biophys. Res. Comm.* **129**, 226-232
94. Komoriya, A., Hortsch, M., Meyers, C., Smith, M., Kanety, H., and Schlessinger, J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1351-1355
95. Niyogi, S.K., and Mitra, S. (1978) *J. Biol. Chem.* **253**, 5562-5567
96. Niyogi, S.K., Foote, R.S., Mural, R.J., Larimer, F.W., Mitra, S., Soper, T.S., Machanoff, R., and Hartman, F.C. (1986) *J. Biol. Chem.* **261**, 10087-10092
97. Niyogi, S.K., Soper, T.S., Foote, R.S., Larimer, F.W., Mural, R.J., Mitra, S., Lee, E.H., Machanoff, R., and Hartman, F.C. (1987) *J. Biosci* **11**, 203-214
98. Nisbet, J.T., and Beilharz, M.W. (1985) *Gene Anal. Techn.* **2**, 23-40
99. Zoller, M.J., and Smith, M. (1983) *Methods Enzymol.* **100**, 468-500
100. Stowers, D.J., Kein, J.M.B., Paul, P.S., Lyoo, Y.S., Merion, M., and Benbow, R. (1988) *J. Chromatogr.* **444**, 47-56
101. Grunstein, M., and Hogness, D. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3961-3964
102. Murray, N.E., Brammar, W.J., and Murray, K. (1977) *Mol. Gen. Genet.* **150**, 53-61
103. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) *Gene* **33**, 103-119
104. Appleyard, R.K. (1954) *Genetics* **39**, 440-447.
105. Trisler, P., and Gottesman, S. (1984) *J. Bacteriol.* **160**, 184-191
106. Howard-Flanders, P., Simson, E., and Theriot, L. (1964) *Genetics* **49**, 237-246
107. Randall, L.L., and Hardy, S.J.S. (1986) *Cell* **46**, 921-928

108. Maher, P., and Singer, S. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9001-9005
109. Randall, L.L., and Hardy, S.J.S. (1984) *Microbiol. Rev.* **48**, 290-298
110. Oliver, D. (1985) *Ann. Rev. Microbiol.* **39**, 615-648
111. Randall, L.L., Hardy, S.J.S., and Thom, J.R. (1987) *Ann. Rev. Microbiol.* **41**, 507-541
112. Hirst, T.R., and Welch, R.A. (1988) *Trends Biochem. Sci.* **13**, 265-269
113. Pestka, S. (1975) in *Antibiotics: Mechanisms of Action of Antimicrobial and Antitumor Agents* Vol. III S. Corcoran and F. Hahn (eds.) pp. 370-395, Springer-Verlag, New York
114. Moore, S., and Stein, W.H. (1963) *Methods Enzymol.* **6**, 819-831
115. Spencer, R.L., and Wold, F. (1969) *Anal. Biochem.* **32**, 185-190
116. Penke, B., Ferenczi, R., and Kovacs, K. (1974) *Anal. Biochem.* **60**, 45-50
117. Hunter, W.M., and Greenwood, F.C. (1962) *Nature* **194**, 495-496
118. Akiyama, T., Kadooka, T., and Ogawara, H. (1985) *Biochem. Biophys. Res. Comm.* **131**, 442-448
119. Carpenter, G. (1985) *Methods Enzymol.* **109**, 107-108
120. Matsunami, R. (1990) "Structure Function Analysis of the Highly Conserved Leucine 47 of Human Epidermal Growth Factor and EGF Receptor Phosphorylation Studies" Ph.D. Thesis, Univ. of Tennessee
121. Moy, F.J., Scheraga, H.A., Liu, J.F., Wu, R., and Montelione, G.T. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9836-9840
122. Montelione, G.T., Wüthrich, K., Nice, E.C., Burgess, A.W., and Scheraga, H.A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8594-8598

123. Defeo-Jones, D., Tai, J.Y., Vuocolo, G.A., Wegrzyn, R.J., Schofield, T.L., Rieman, M.W., and Oliff, A. (1989) *Mol. Cell. Biol.* **9**, 4083-4086
124. Defeo-Jones, D., Tai, J.Y., Wegrzyn, R.J., Vuocolo, G.A., Baker, A.E., Payne, L.S., Garskey, V.M., Oliff, A., and Rieman, M.W. (1988) *Mol. Cell. Biol.* **8**, 2999-3007
125. Lazar, E., Vicenzi, E., Van Obberghen-Schilling, E., Wolff, B., Dalton, S., Watanabe, S., and Sporn, M.B. (1989) *Mol. Cell. Biol.* **9**, 860-864
126. Graf, J., Iwamoto, Y., Sasaki, M., Martin, G.R., Kleinman, H.K., Robey, F.A., and Yamada, Y (1987) *Cell* **48**, 989-996

APPENDICES

APPENDIX A
FIGURES AND FIGURE LEGENDS

Figure 1. **Schematic representation of secondary structural features of hEGF.** *Solid lines* indicate disulfide bridges, and hydrogen bonds are indicated by *dashed lines*. *Dashed circles* indicate residues with no rigid structural characteristics as established by NMR measurements.

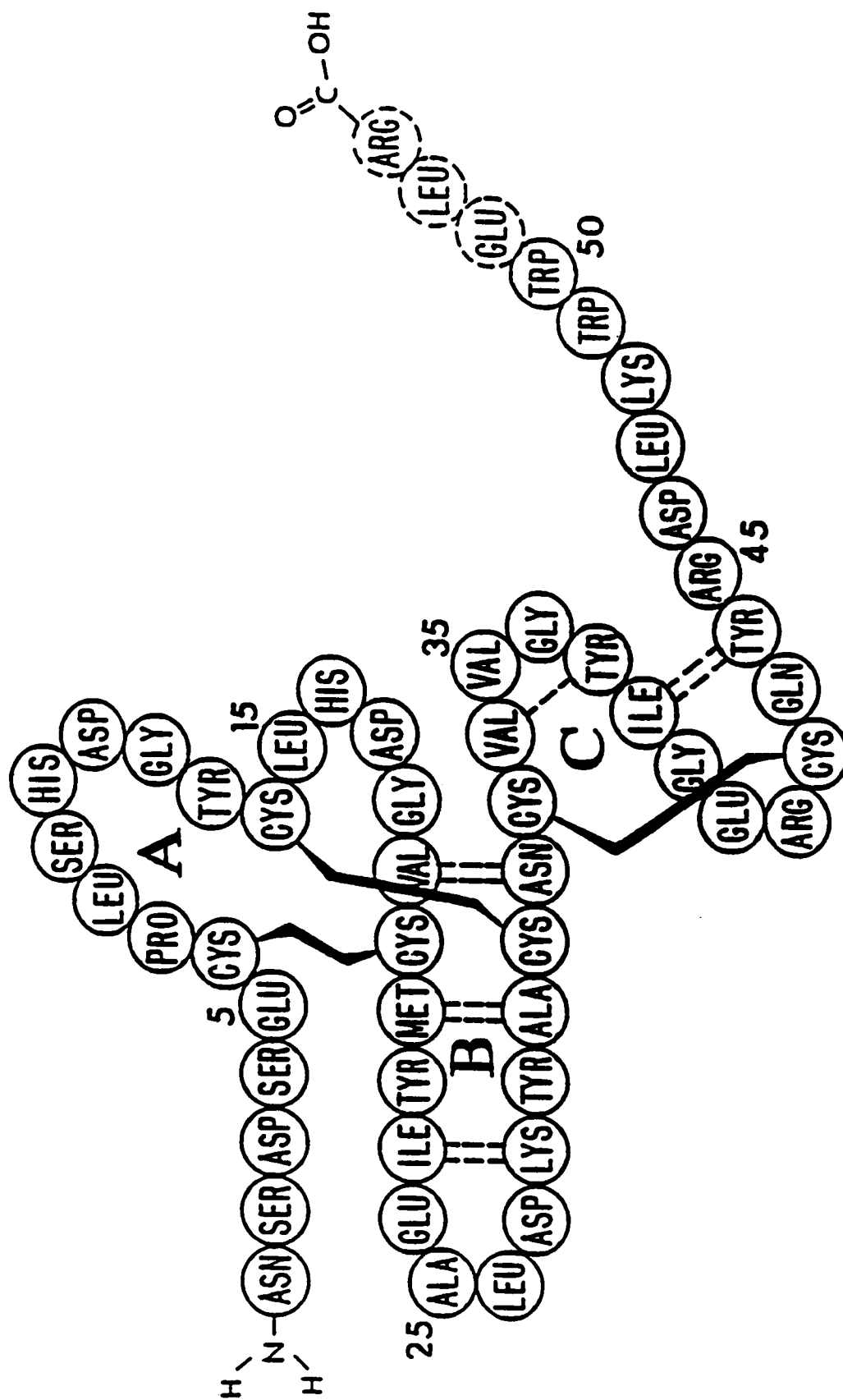


Figure 1

Figure 2. Ribbon diagram, derived from nuclear Overhauser effect (NOE) distance constraints, representing the polypeptide folds of EGF. The polypeptide backbone is represented by the *rope* which broadens into *arrows* which represent the location and direction of β -sheet strands. The *broken rope* indicates the COOH-terminal tail, residues 51-53, which lack NOE tertiary determinants. (Taken from reference #36.)

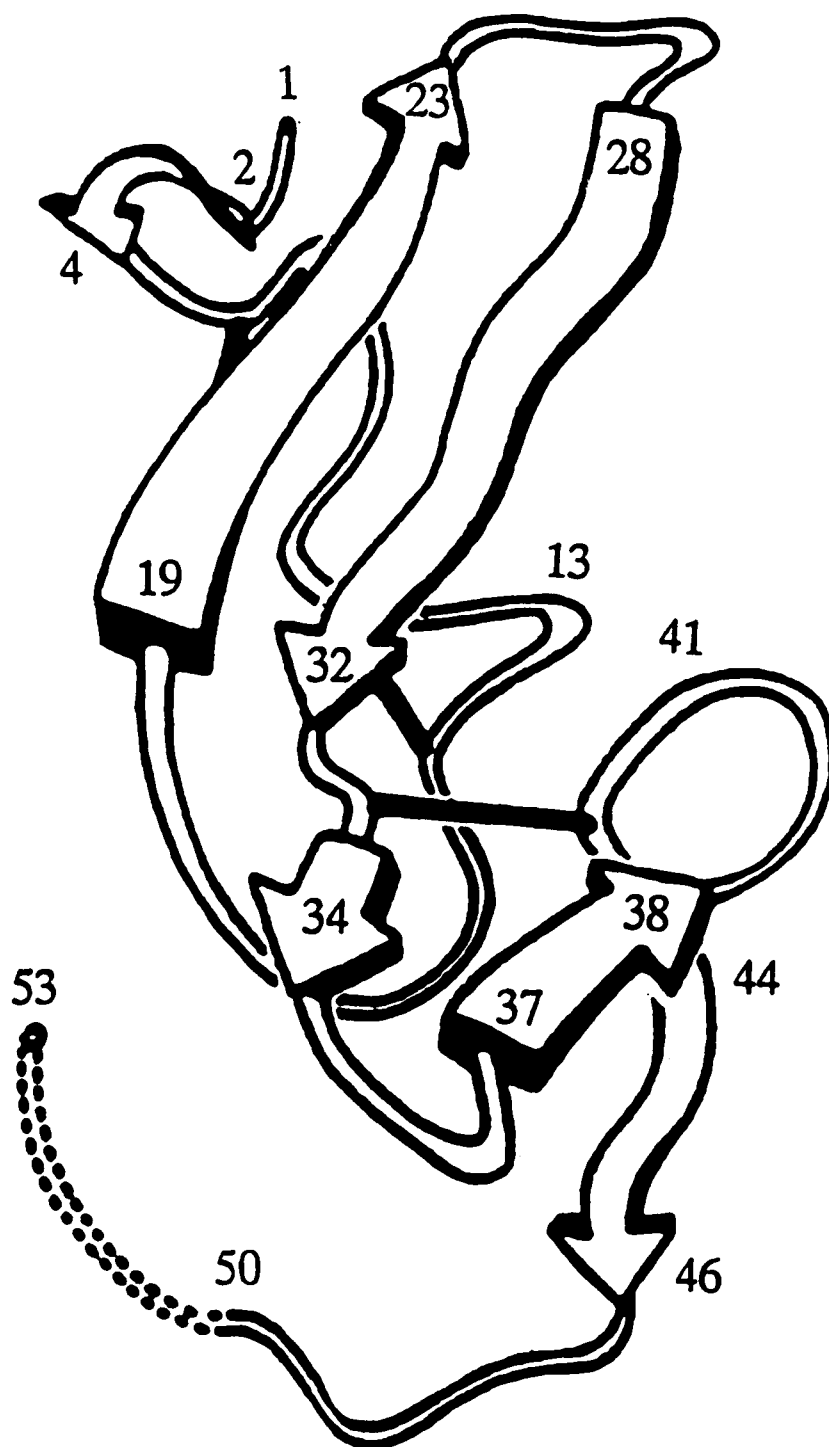


Figure 2

Figure 3. Alignment of type-I EGF-like sequences, comprised of EGF and EGF-like molecules known to bind to the EGF receptor. *Solid boxed* residues indicate unequivocally conserved cysteine residues which were used for alignment. *Dashed boxes* indicate other highly conserved residues. *Numbers* preceding the sequence designate the amino acid residue number the sequence shown initiates with. *Numbers* above the hEGF sequence are for reference with respect to the hEGF sequence.

Human EGF	1	NSDSE	C	PLSHDG	Y	C	LHD	G	V	C	MY	I	EALDKYA	30	C	N	C	VV	G	Y	I	G	E	R	C	QYRD	L	KWWE	50
Rat EGF	1	NSNTG	C	PPSYDG	Y	C	LNG	G	V	C	MY	V	ESVDRYV		C	N	C	VI	G	Y	I	G	E	R	C	QHRD	L	R	
Mouse EGF	1	NSYPG	C	PSSYDG	Y	C	LNG	G	V	C	MH	I	ESLDSYT		C	N	C	VI	G	Y	S	G	D	R	C	QTRD	L	RWWE	
Guinea-pig EGF	1	QDAPG	C	PPSHDG	Y	C	LHG	G	V	C	MH	I	ESLNTYA		C	N	C	VI	G	Y	V	G	E	R	C	EHQD	L	DLWE	
Rat TGF-alpha	1	VVSHFNK	C	PDSHTQ	Y	C	FH	G	T	C	RF	L	VQEEKPA		C	V	C	HS	G	Y	V	G	V	R	C	EHAD	L	LA	
Human TGF-alpha	1	VVSHFND	C	PDSHTQ	F	C	FH	G	T	C	RF	L	VQEDKPA		C	V	C	HS	G	Y	V	G	A	R	C	EHAD	L	LA	
Human Amphiregulin	44	...NP	C	NAEFQ	F	C	IH	G	E	C	KY	I	EHLEAVT		C	K	C	QQ	E	Y	F	G	E	R	C	GEK			
VGF	38	DIPAIRL	C	GPEGDG	Y	C	LH	G	D	C	IH	A	RDIDGMX		C	R	C	SH	G	Y	T	G	I	R	C	QHVV	L	VDYQRS	
MGF	30	IIKRIKL	C	NDDYKN	Y	C	LNN	G	T	C	FT	V	ALNNSLNP		C	A	C	HI	N	Y	V	G	S	R	C	QFIN	L	ITIK	
SFGF	30	IVLHVKV	C	NHDYEN	Y	C	LNN	G	T	C	FT	I	ALDNVSITPF		C	V	C	RI	N	Y	E	G	S	R	C	QFIN	L	VTY	

Figure 3

Figure 4. Schematic representation of EGF showing highly conserved residues. *Bold letters* represent identically conserved residues. *Dashed letters* represent amino acid residues with conserved chemical functionality, and *outlined letters* represent residues conserved in most of the type I EGF-like molecules.

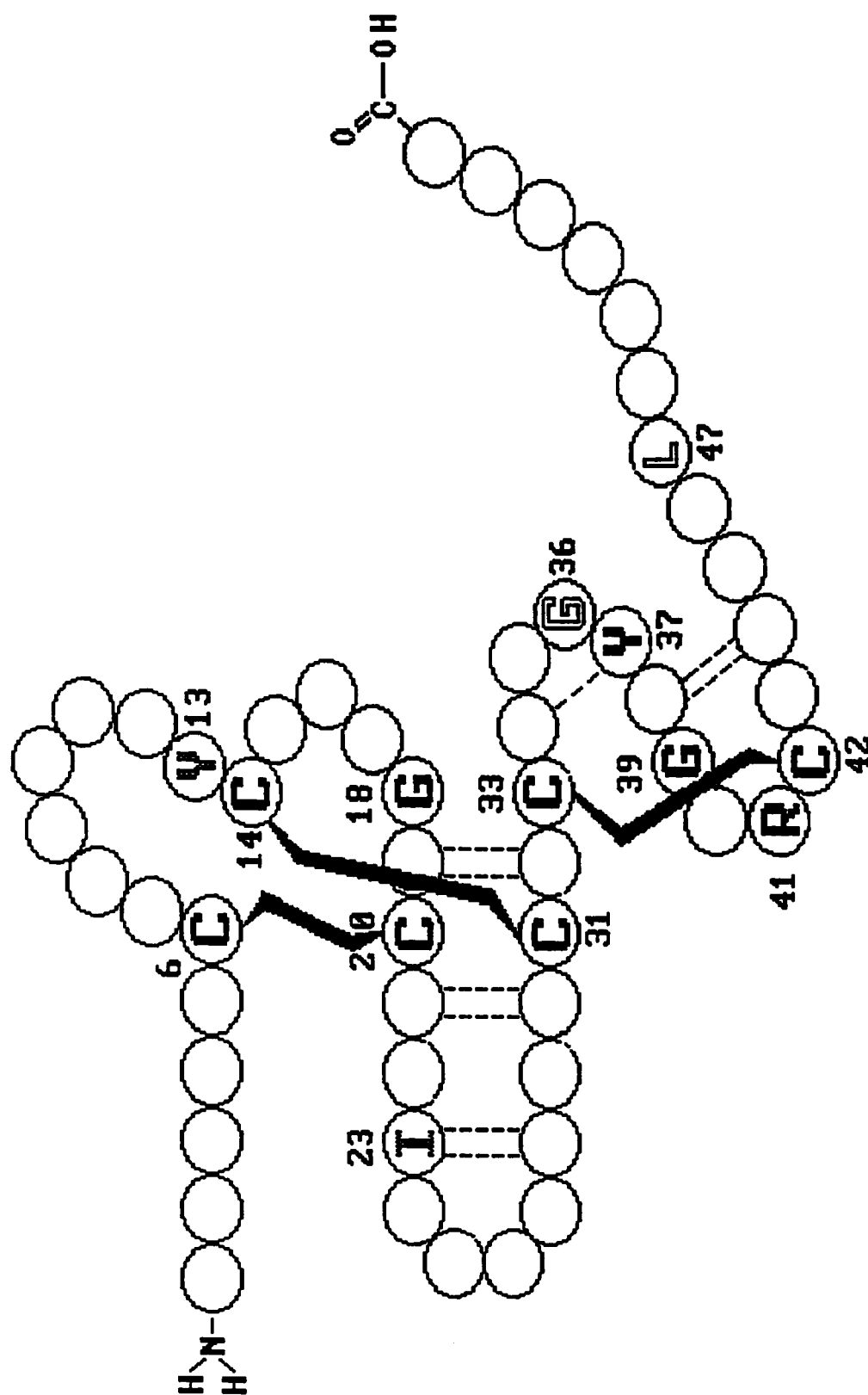


Figure 4

Figure 5. Design of the synthetic chimeric *phoA* Signal/hEGF gene. Numbered lines encompassing the nucleotide sequence designate the individually synthesized single-stranded oligonucleotides. Amino acid sequences for the alkaline phosphatase signal peptide and hEGF are delineated by arrows and are shown above the nucleotide sequence.

Alkaline Phosphatase Signal Peptide →

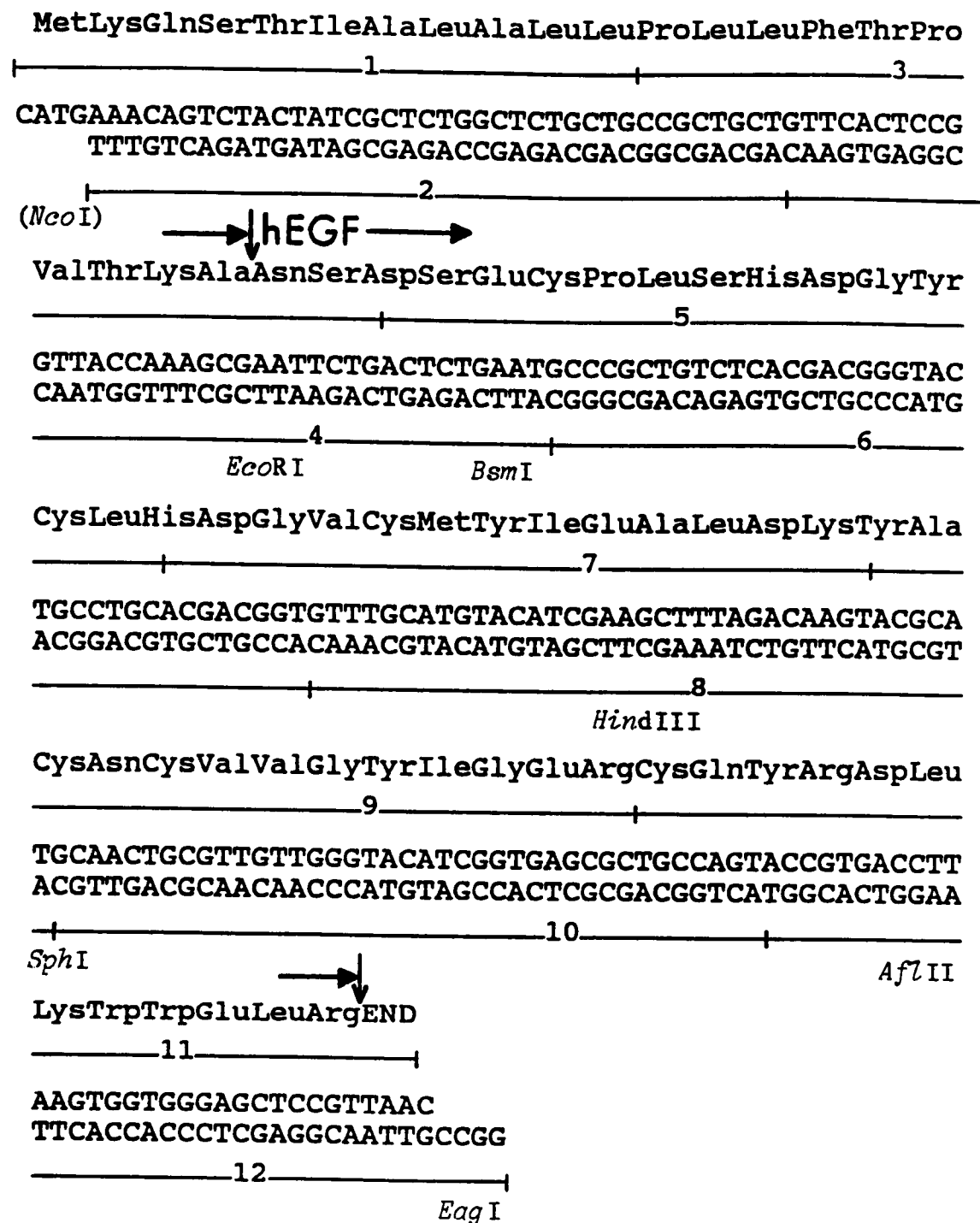


Figure 5

Figure 6. Expression plasmid pEGF-1 showing relative size of insert and position of restriction endonuclease recognition sites. *Parentheses* indicate the loss of the restriction endonuclease recognition site *Nco*I upon insertion into the vector.

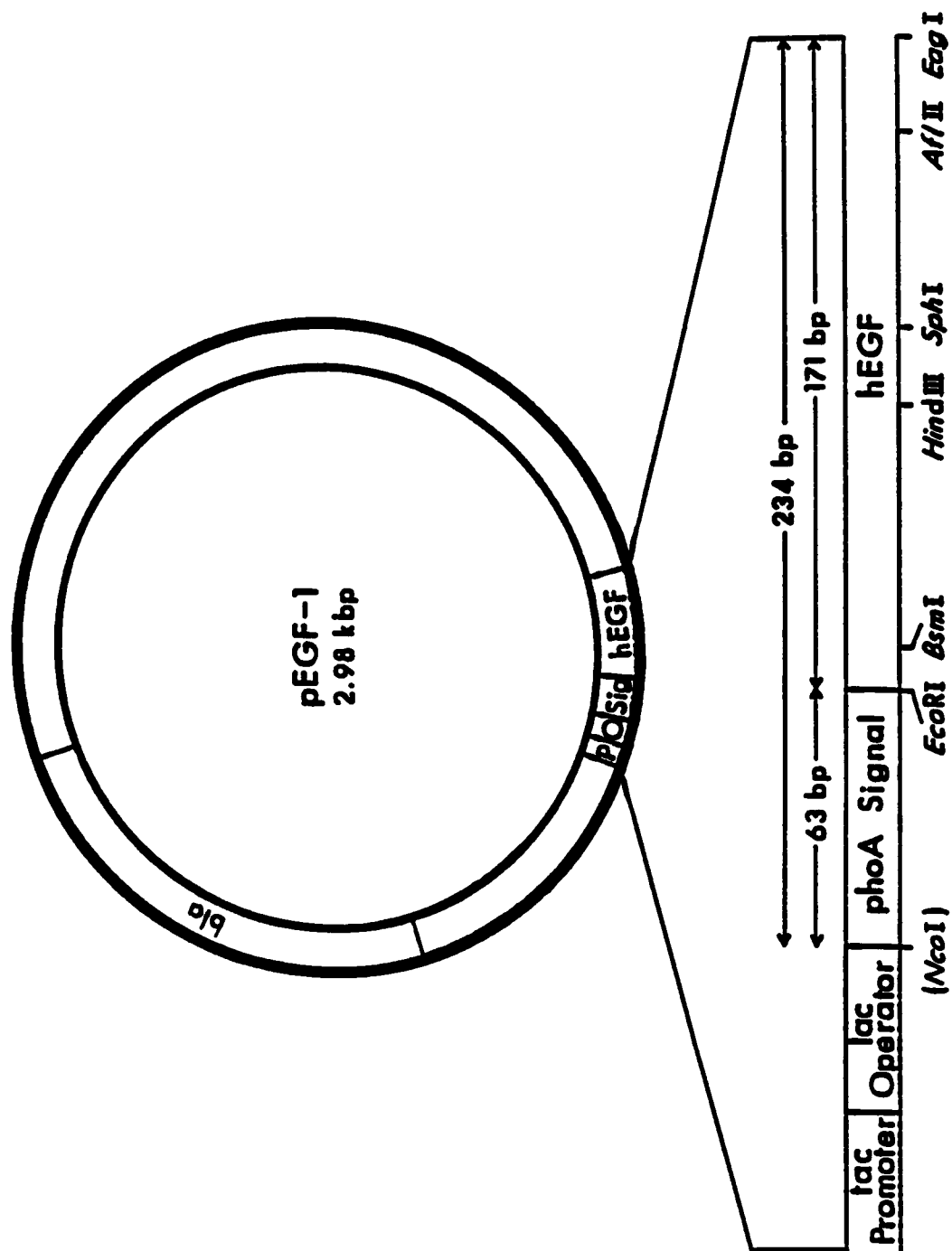


Figure 6

Figure 7. Comparison of reverse phase HPLC profiles of wild-type hEGF, Tyr³⁷→Gly and Arg⁴¹→Gly.

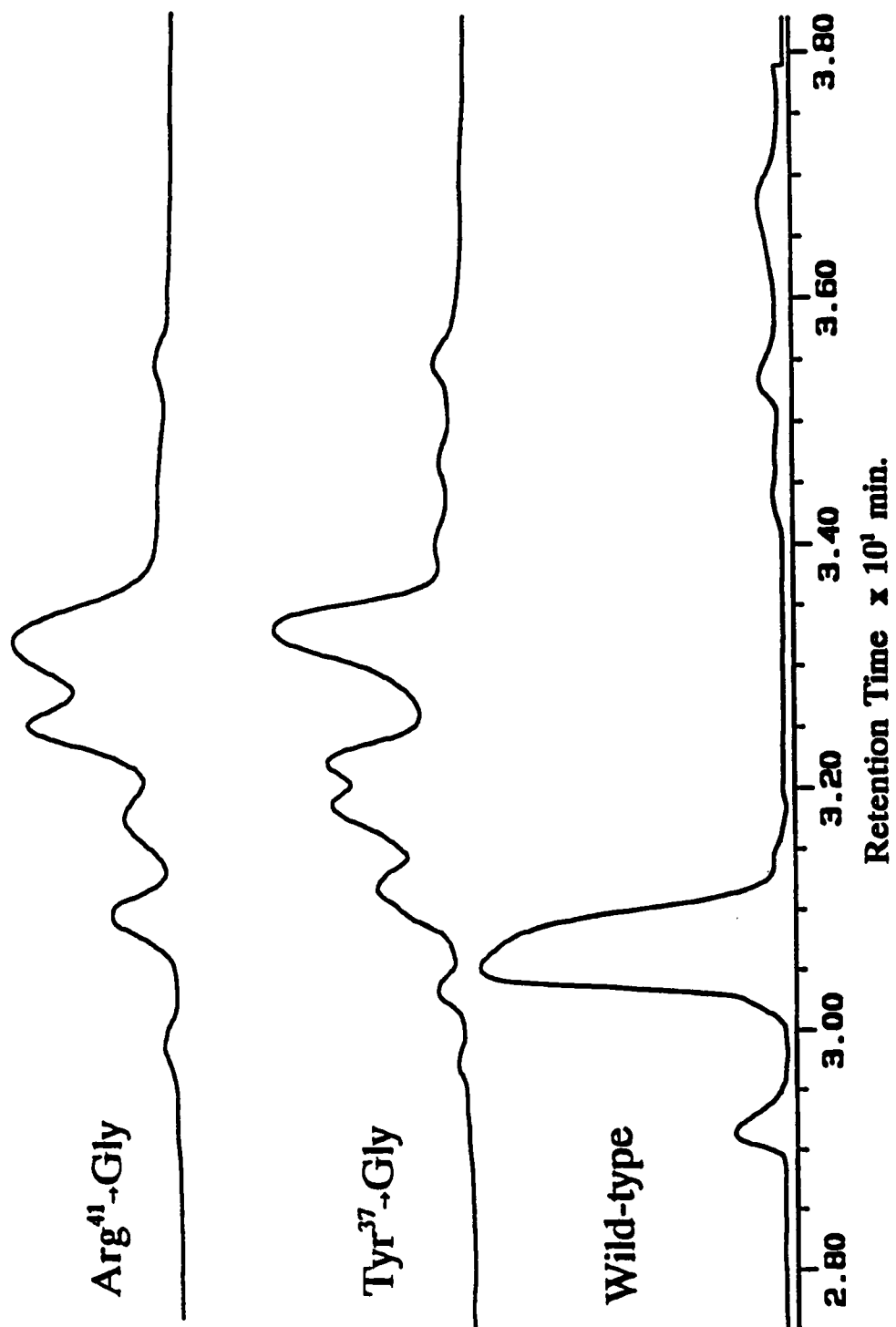


Figure 7

Figure 8. Effect of chloramphenicol addition on hEGF production with respect to hEGF protein induction time.

Time of Cm addition: (Δ) 1 hr pre-induction, ($\circ$) 0 hr pre-induction, ($\bullet$) 1 hr post-induction.

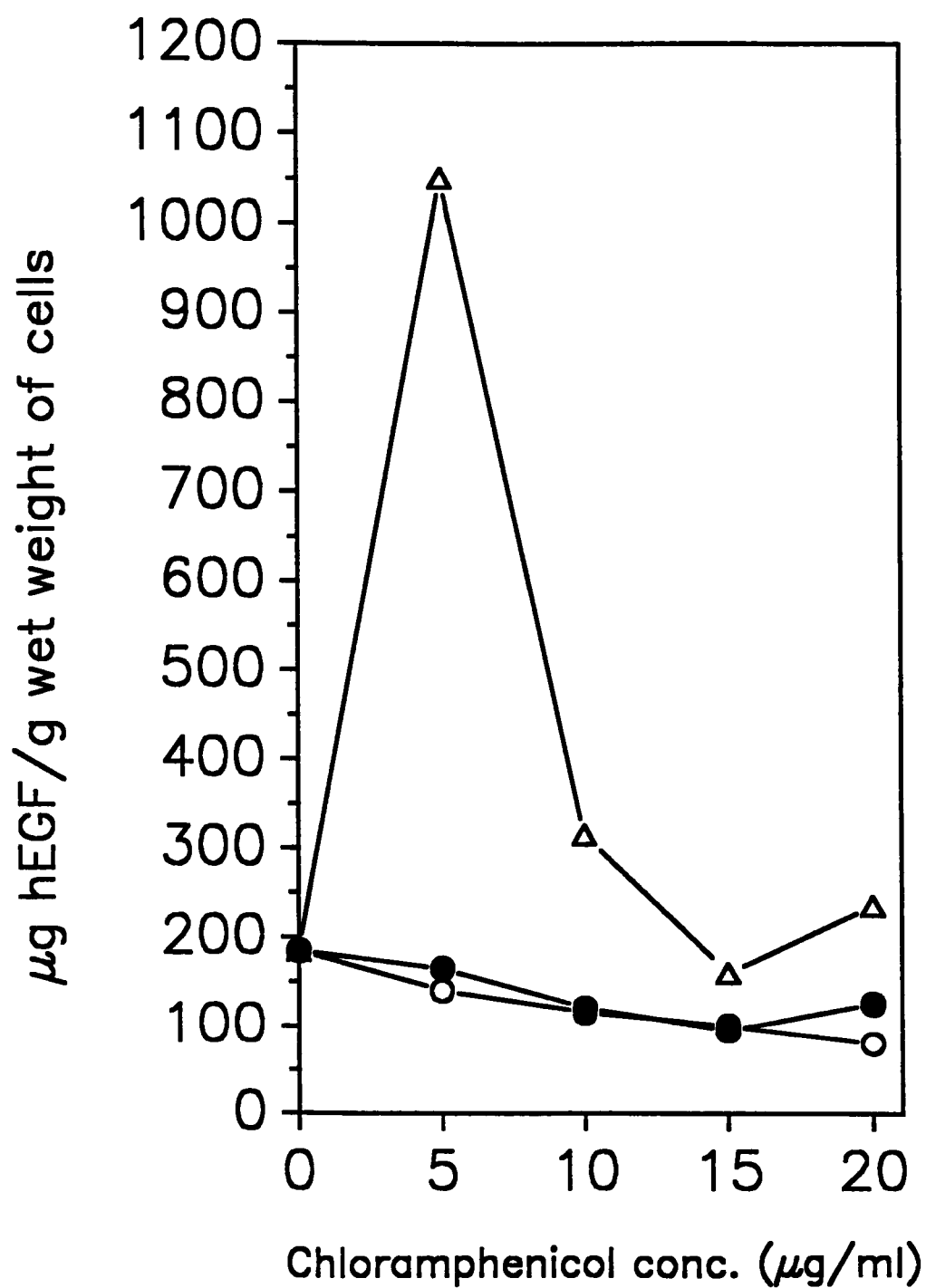


Figure 8

Figure 9. Growth curve of *E. coli*/pEGF-1 culture $\pm$ chloramphenicol. Cm was added at 3 hours after inoculation and IPTG was added at 4 hours after inoculation to induce hEGF production. (○) Cells treated with 5 μ g/ml Cm, (●) cells not treated with Cm.

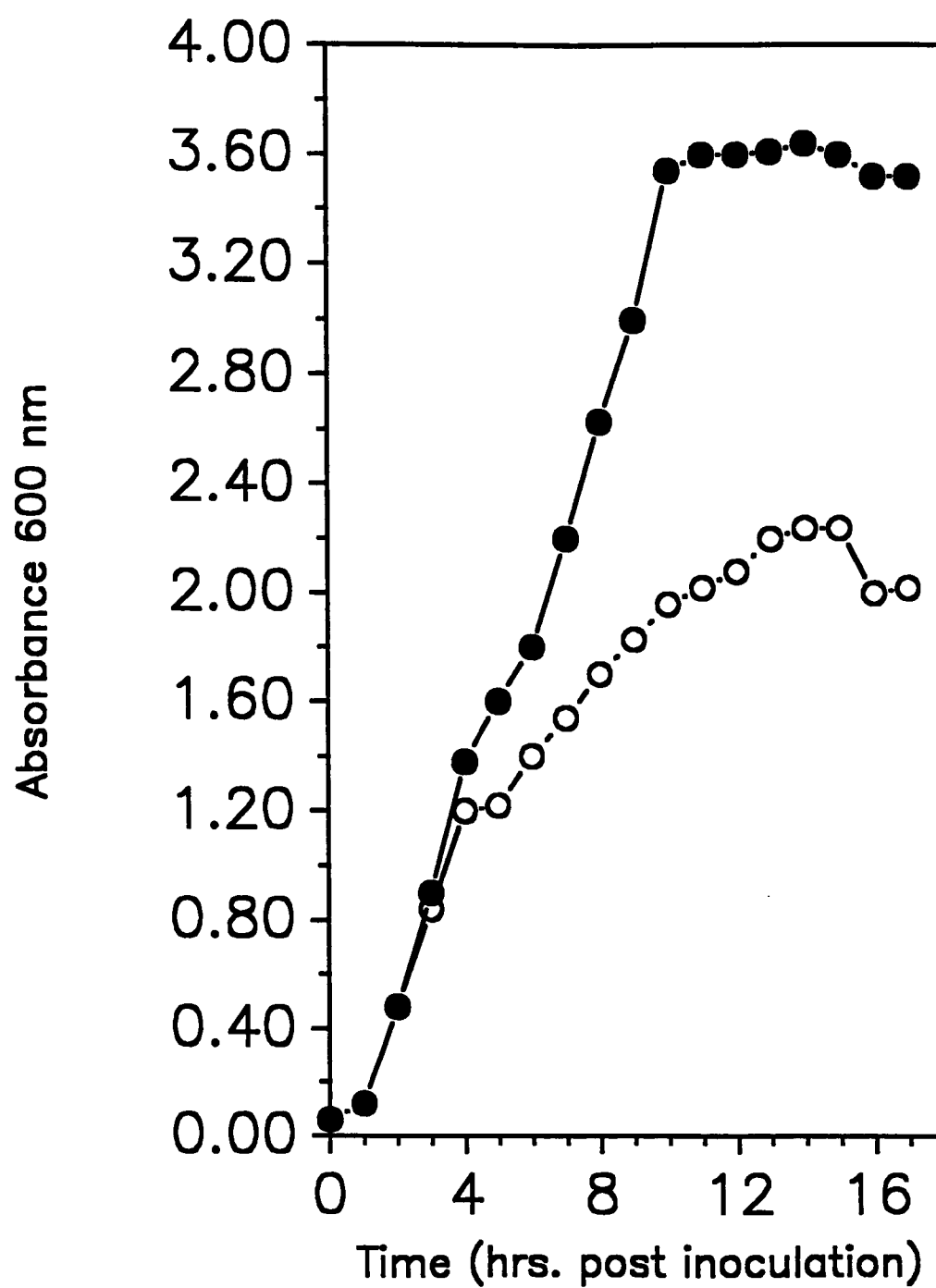


Figure 9

Figure 10. Typical elution profile of recombinant hEGF off a reverse phase HPLC column. *Arrow* indicates peak containing EGF receptor binding activity.

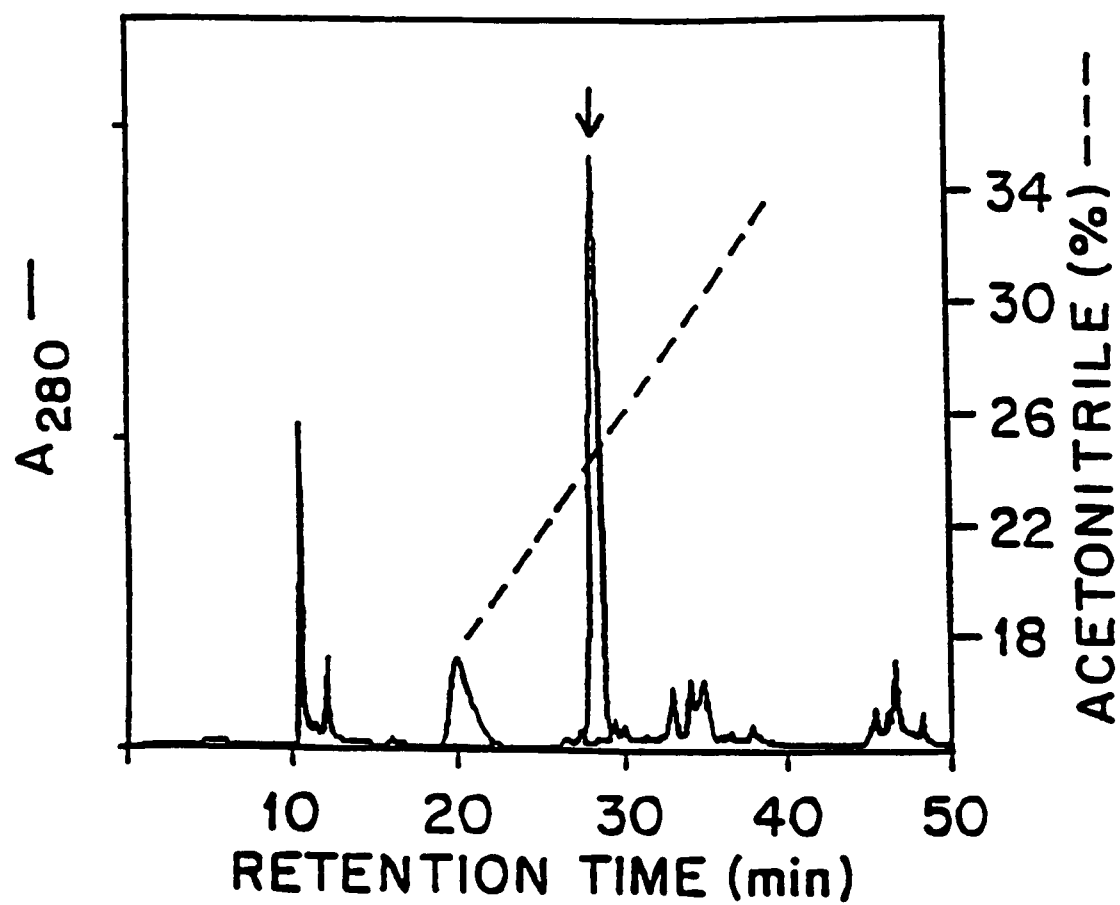


Figure 10

Figure 11. SDS-PAGE of recombinant hEGF. Lane 1 = molecular weight markers, lane 2 = recombinant hEGF.

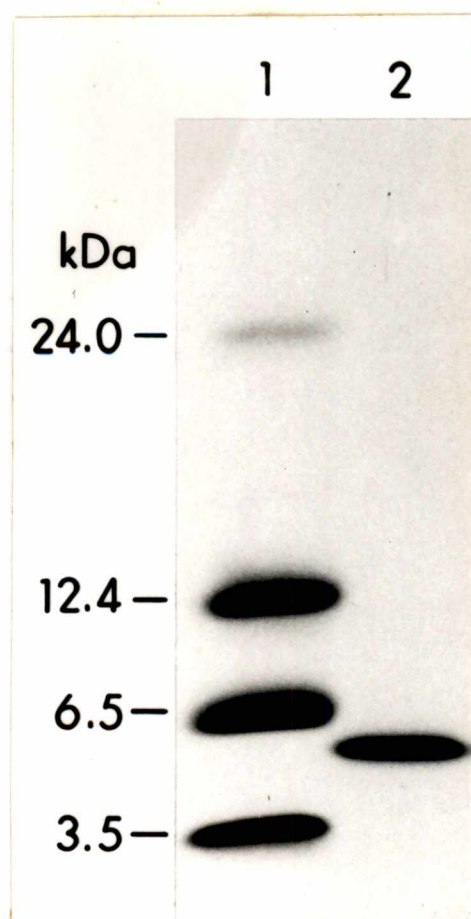


Figure 11

Figure 12. Competition binding curves of [^{125} I]hEGF *versus* wild-type and tyrosine-37 analogs. All curves represent the best fit through the average data points of triplicate experiments. (○) wild-type, (Δ) Tyr³⁷→Phe, (▲) Tyr³⁷→His, (□) Tyr³⁷→Ser, (●) Tyr³⁷→Ala, (■) Tyr³⁷→Asp, (▽) Tyr³⁷→Arg, (▼) Tyr³⁷→Gly.

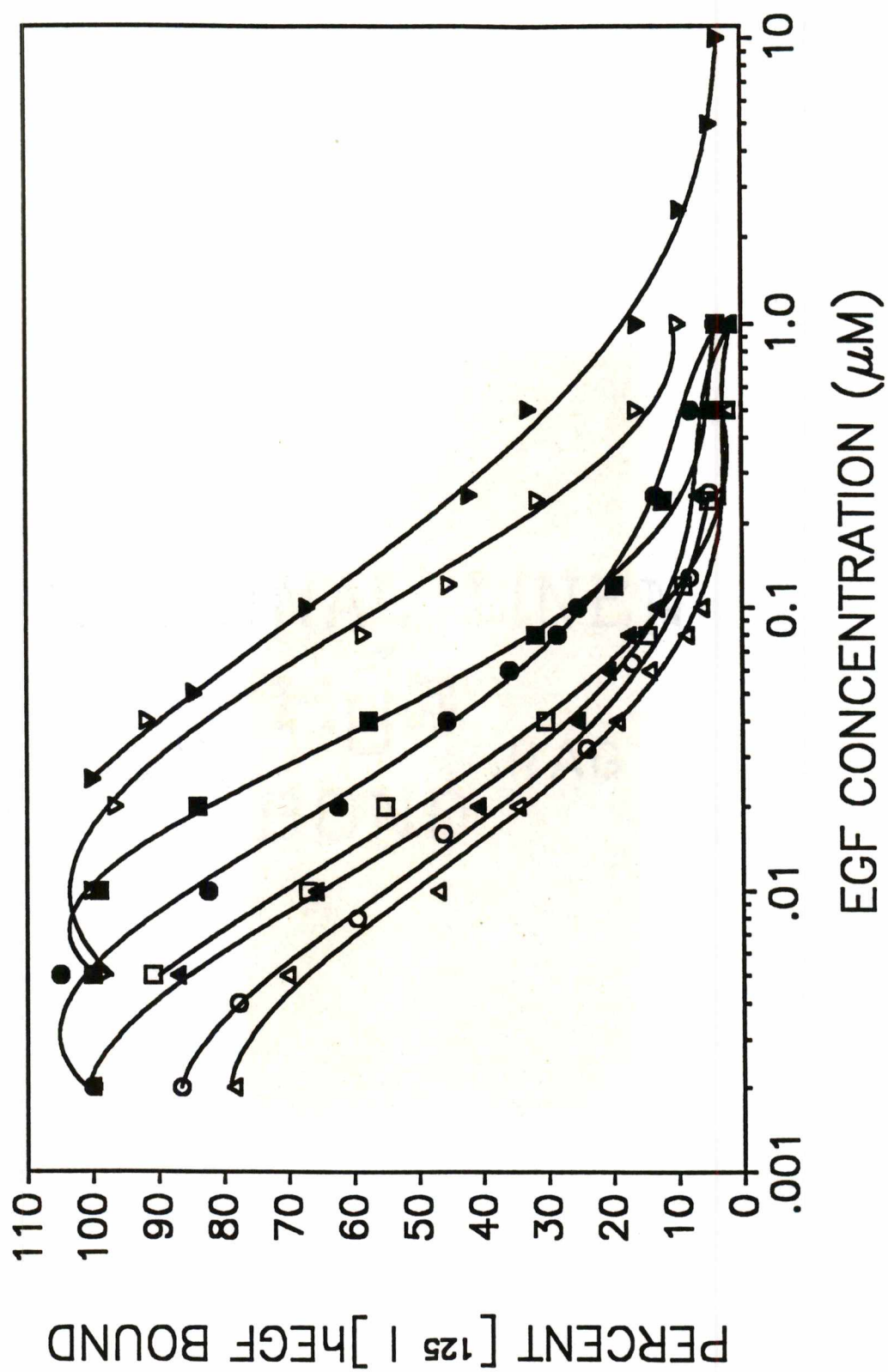


Figure 12

Figure 13. Competition binding curves of [^{125} I]hEGF *versus* wild-type and arginine-41 analogs. All curves represent the best fit through the average data points of triplicate experiments. (○) wild-type, (■) Arg⁴¹→Lys, (Δ) Arg⁴¹→Gln, (●) Arg⁴¹→Ile, (▼) Arg⁴¹→Tyr, (▽) Arg⁴¹→Gly, (□) Arg⁴¹→Ala.

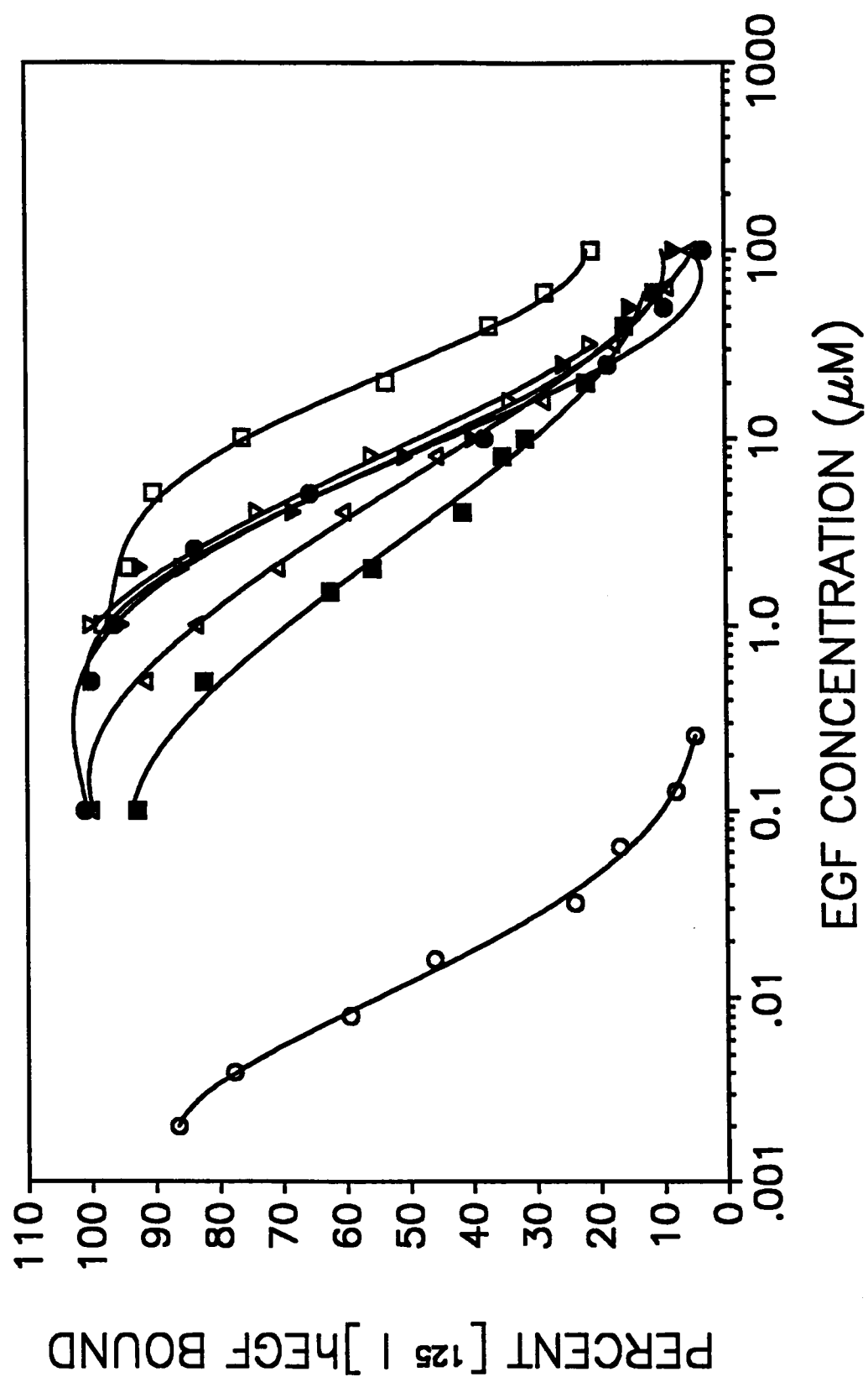


Figure 13

Figure 14. EGF receptor protein-tyrosine kinase stimulation by wild-type and tyrosine-37 analogs. All curves represent the best fit through the average data points of duplicate experiments. (○) wild-type, (Δ) Tyr³⁷→Phe, (▲) Tyr³⁷→His, (□) Tyr³⁷→Ser, (●) Tyr³⁷→Ala, (■) Tyr³⁷→Asp, (▽) Tyr³⁷→Arg, (▼) Tyr³⁷→Gly.

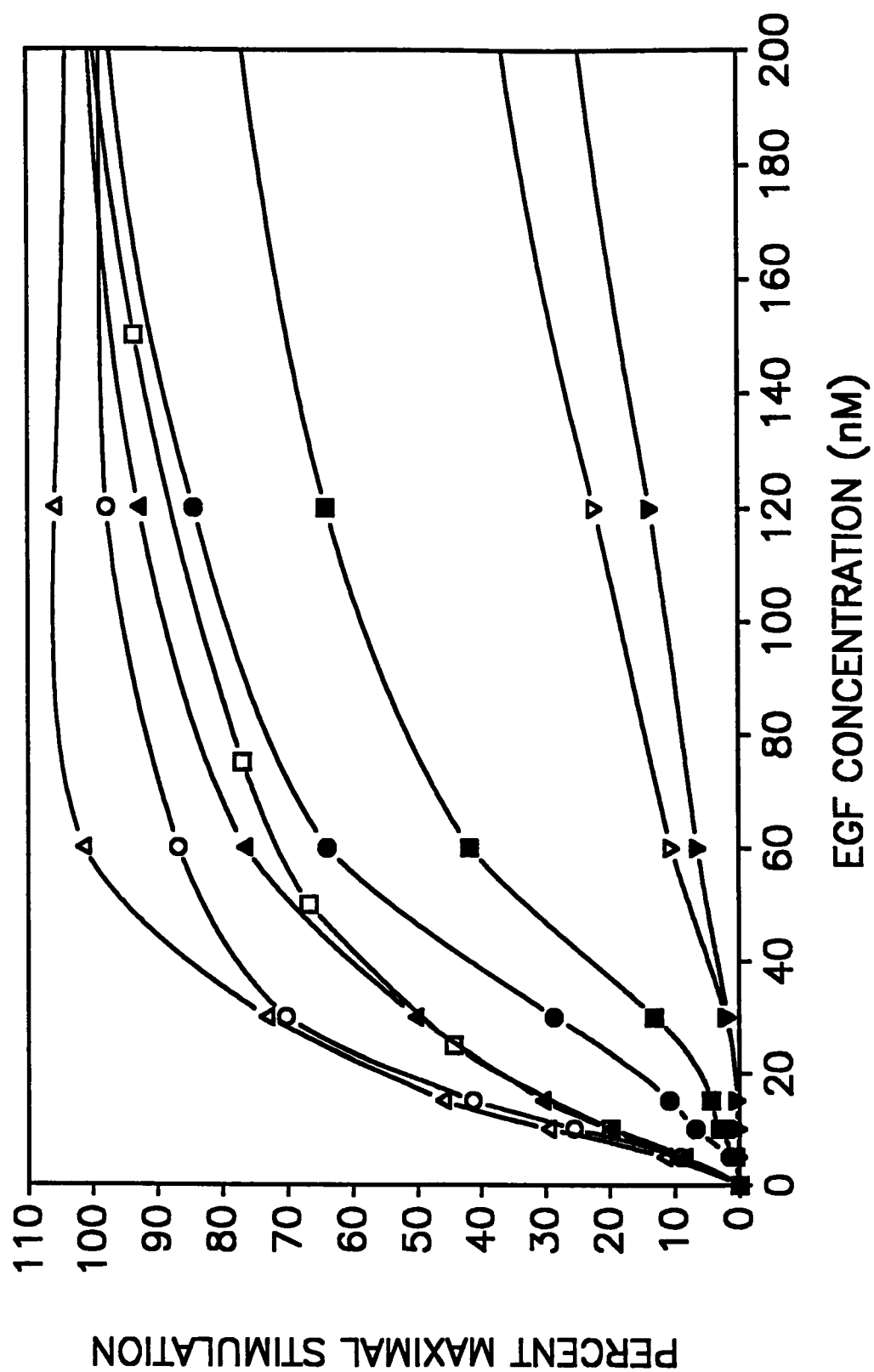


Figure 14

Figure 15. Comparison of aromatic and downfield C^αH spectral regions from 500 MHz 1D ¹H-NMR spectra of wild-type, Tyr³⁷→Phe, Tyr³⁷→His, and Tyr³⁷→Ala. All spectra represent the average of 512 scans on a Varian VXR500 spectrometer under the conditions described in the text.

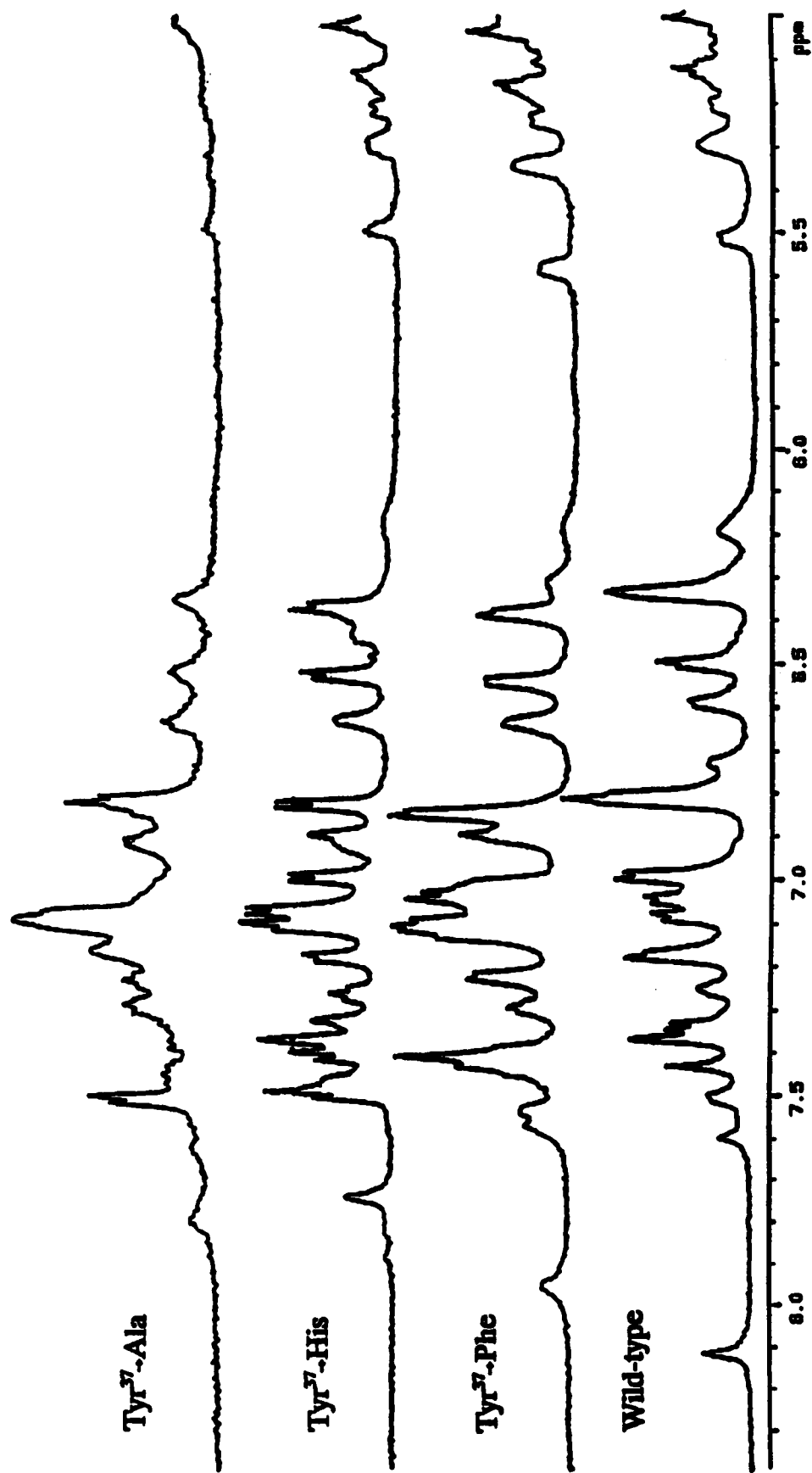


Figure 15

Figure 16. Comparison of aromatic and downfield C^αH spectral regions from 500 MHz 1D ¹H-NMR spectra of wild-type, Arg⁴¹→Lys, Arg⁴¹→Gln, and Arg⁴¹→Ala. All spectra represent the average of 512 scans on a Varian VXR500 spectrometer under the conditions described in the text.

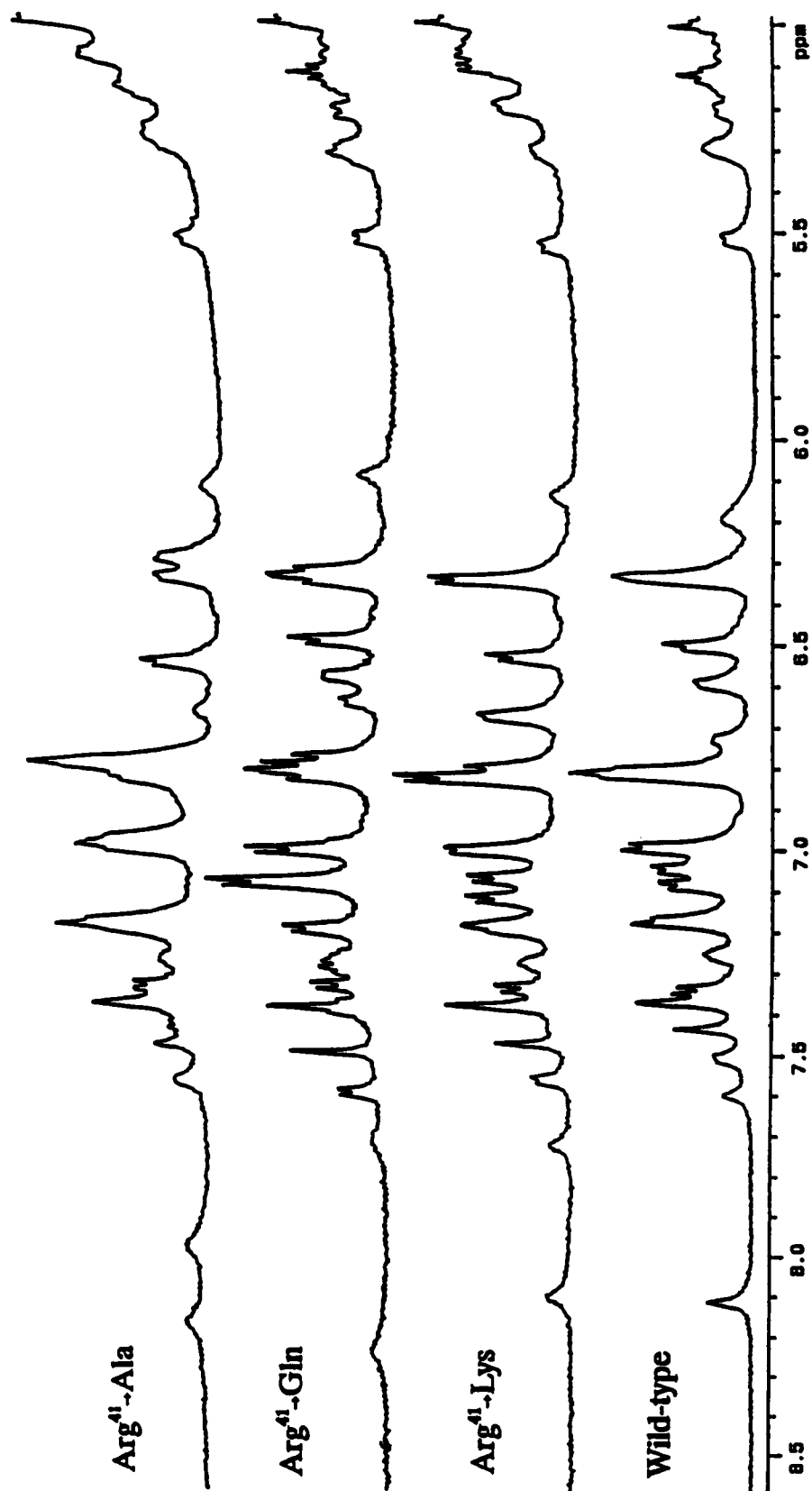


Figure 16

Figure 17. Alignment of non-EGF type-II protein sequences from EGF-like motifs in various proteins with no known EGF receptor binding capability. *Boxed* amino acid residues are those that retain either complete identity or chemical functionality with corresponding residues in hEGF. *Numbers* preceding the sequence designate the amino acid residue number in the parent molecule with which the EGF-like motif listed initiates. *Numbers* above the hEGF sequence are for reference with respect to the hEGF sequence. *Dashes* were placed to maximize homology and account for differing loop sizes in the various molecules.

hEGF	1	NSDSE	C	PL	SHDGY	C	LH	DGV	C	MYIE	ALDKYA	C	N	C	VV	G	Y	I	G	E	40	C	QYR
Factor VII	87	-DQLI	C	VNE	NGG	C	E	QY	C	SDHT	GTGRS	C	R	C	HE	G	Y	S	L	LADGV	S	C	TPT
Factor IX	85	--DVT	C	NIK	NGR	C	E	QF	C	KNSA	DNKVV	C	S	C	TE	G	Y	R	L	AENQK	S	C	EPA
Factor X	85	-TRKL	C	SLD	NGD	C	D	QF	C	HEEQ	NSVV	C	S	C	AR	G	Y	T	L	ADNGK	A	C	IPT
Factor XII	155	-ASQA	C	R	TNP	C	LH	GGR	C	LEVE	GHRL	C	H	C	PV	G	Y	T	P	----	F	C	DVD
Protein C	94	-SFLN	C	SLD	NGG	C	T	HY	C	LEE	GWRR	C	S	C	AP	G	Y	K	L	GDDL	Q	C	HPA
Protein S	160	-DVDE	C	SLK	PSI	C	G	TAV	C	KNIL	GDFE	C	E	C	PE	G	Y	R	Y	NLKS	S	C	E
Protein Z	85	-AESE	C	HPL	RLDG	C	Q	HF	C	YGP	ESYT	C	S	C	AR	G	Y	K	L	GQDRR	S	C	LPH
Complement C1r	125	-DIDE	C	ASRSKSGEEDPQPQ	C	Q	Q	HL	C	HNYV	GGYF	C	S	C	RP	G	Y	E	L	QEDRH	S	C	QA
tPA	61	-PVKS	C	S	EPR	C	FN	GGT	C	QAL	YFSDFV	C	Q	C	PE	G	Y	A	G	----	K	C	----
Urokinase type PA	6	QVPSN	C	D	----	C	LN	GGT	C	VSNK	YFSNIHW	C	N	C	PK	K	F	G	L	----	Q	H	EID
Thrombomodulin	325	-DVDD	C	ILE	PSP	C	P	QR	C	VNTQ	GGFE	C	H	C	YP	N	Y	D	L	VDG	E	C	VEP
Thrombospondin	628	-PRNP	C	TD	GTHD	C	NK	NAK	C	NYLGHY	SDPMYR	C	E	C	KP	G	Y	A	G	NGI	I	C	GED
pg Core Protein	379	--PDR	C	KM	NP	C	LN	GGT	C	YPTE	TSYV	C	T	C	VP	G	Y	S	G	----	D	C	ELD
Cytotactin	40	-GELR	C	PN	D	C	NS	HGR	C	VE	GG	C	V	C	DQ	G	Y	T	G	----	E	D	GEL
Uromodulin	84	-DVDE	C	AEPGL	SH	C	HAL	AT	C	VNVV	GSYL	C	V	C	PA	G	Y	R	G	DGW	H	C	E
EGF precursor	886	-DIDE	C	VLA	RSD	C	PS	TSSR	C	INTE	GGYV	C	R	C	SE	G	Y	E	G	DGI	S	C	F
LDL receptor	333	-DIDE	C	QDP	DT	C	SQ	L	C	VNLE	GGYK	C	Q	C	EE	G	Y	Q	L	DPHTK	A	C	KAV
Laminin A Chain	372	YDDHP	C	RP	DD	C	DPVGSLSV	FARS	C	IKDDRHADLANGKWPQ	LQLA	C	P	C	CK	G	Y	A	G	----	D	K	DR
Mammalian Lamb1	893	CRPCR	C	PDGPDGRQ	CN	C	----	----	C	VQDPVT	ITML	C	V	C	DP	G	Y	I	G	----	S	R	DD
Drosophila Lamb1	907	CRPCR	C	PETVASGLA	----	C	----	HADG	C	SWTHGT	----	C	H	C	QE	G	Y	S	G	----	S	R	EI
Laminin B2 Chain	947	FGPEG	C	KP	CD	C	HPEGSLSLQ	----	C	K	DDGR	C	E	C	RE	G	Y	V	G	----	M	R	DQ
Malarial Antigen	60	EKVLK	C	DEKT	VNKP	C	GD	FSK	C	IKIDGNP	VSYA	C	K	C	NL	G	Y	D	M	VN	N	V	IPN
Sea Urchin Gene	164	-NIDE	C	A	SRP	C	QN	GAV	C	VDGV	NGFV	C	T	C	SA	G	Y	T	G	----	V	L	ET
Notch Gene	217	-DIEE	C	QS	NP	C	KY	GGI	C	VNTH	GSYQ	C	M	C	PT	G	Y	T	G	----	K	X	E
Delta Gene	106	VDIVE	C	SS	GP	C	HN	GGT	C	MNRV	NSFE	C	V	C	AN	G	Y	R	G	----	K	Q	KEE
95K Gene	252	-DIPF	C	EI	TP	C	DN	GGL	C	LTTG	AVPN	C	K	C	SL	G	Y	T	G	----	R	L	EQD
Lin-12 Gene	19	LHIGS	C	LG	LI	C	GR	NGH	C	HAGPVNG	TQTSYW	C	R	C	DE	G	Y	F	G	----	E	Y	EQQ

Figure 17

Figure 18. Comparison of reverse phase HPLC profiles of wild-type hEGF, Gly³⁶→Val, Ile³⁸→Leu, and Gly³⁹→Leu.

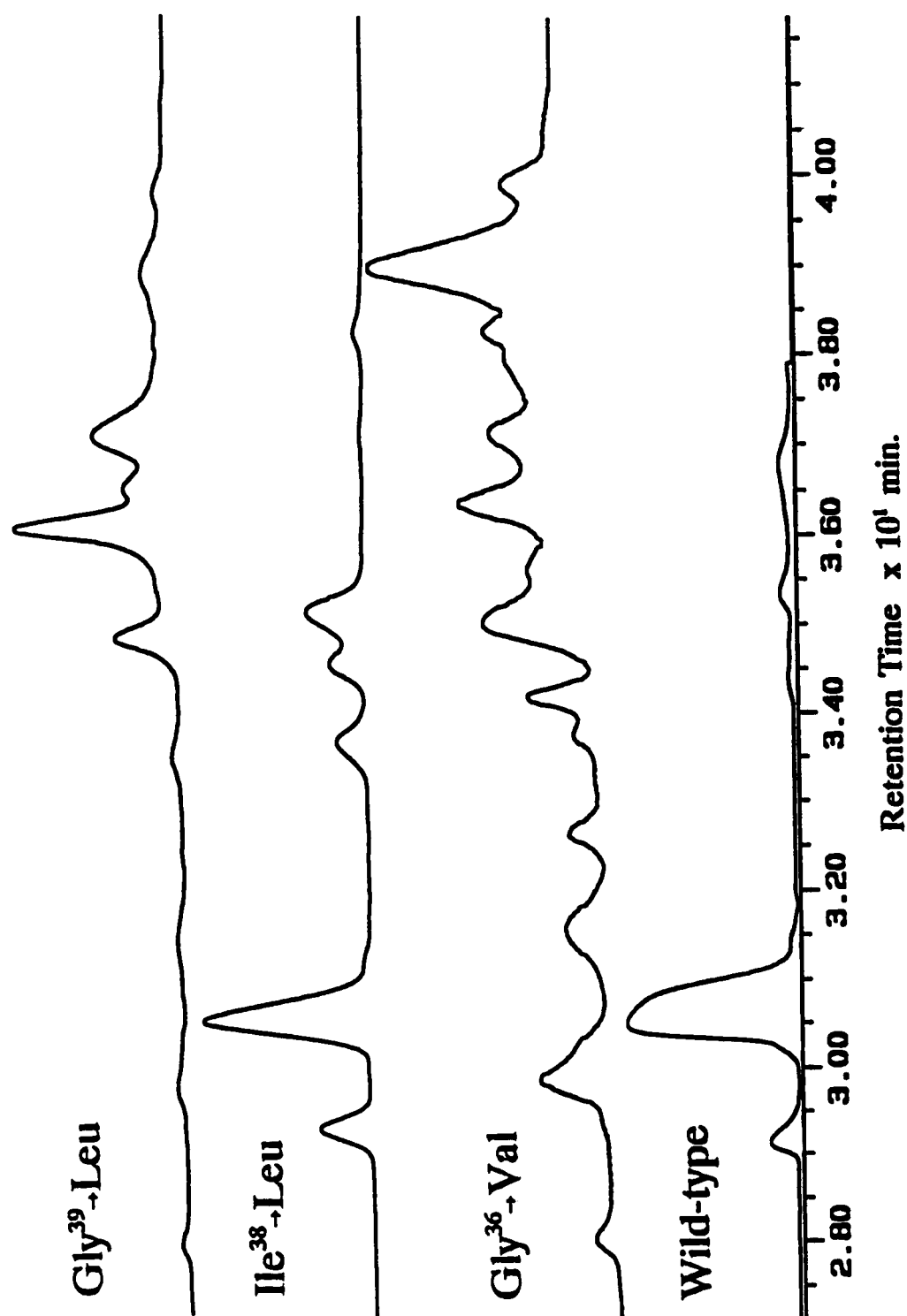


Figure 18

APPENDIX B

TABLES

Table 1

OCCURRENCE OF EGF-RELATED SEQUENCES IN NON-EGF TYPE II PROTEINS

Protein	Function	EGF-Like Units	Protein	Function	EGF-Like Units
Factor VII	serine protease	2	Uromodulin (Tamm-Horsfall urinary glyco- protein)	immunosuppression	2
Factor IX	serine protease	2	EGF precursor	proEGF	9
Factor X	serine protease	2	LDL receptor	cholesterol transport	3
Factor XII	serine protease	2	Laminin A chain	cell adhesion	15
Protein C	serine protease	2	Laminin B1 chain (LamB1) (mammalian)	cell adhesion	13
Protein S	Protein C cofactor	4	Laminin B1 chain (LamB1) (<i>Drosophila</i>)	cell adhesion	13
Protein Z	unknown	1	Laminin B2 chain	cell adhesion	10
Complement Clr	serine protease	1	Malarial antigen	sexual stage surface protein	4

Table 1 (Continued)

Protein	Function	EGF-Like Units	Protein	Function	EGF-Like Units
Tissue-type Plasminogen activator (tPA)	protease	1	Sea urchin gene	development	9
Urokinase-type Plasminogen activator	protease	1	Notch gene (<i>Drosophila</i>)	neurogenesis	36
Thrombomodulin	Thrombin cofactor	6	Delta gene (<i>Drosophila</i>)	neurogenesis	9
Thrombospondin	extracellular interactions	3	95F gene (<i>Drosophila</i>)	neurogenesis	5
Chondroitin SO ₄ proteoglycan (pg Core protein)	extracellular interactions	2	Lin-12 gene (<i>C. elegans</i>)	binary decisions in development	13
Cytotactin	extracellular interactions	4			

Table 2
FEATURES OF MUTAGENIC OLIGONUCLEOTIDES

MUTATION	LENGTH	SEQUENCE ^a	MONITORED RESTRICTION SITE ^b
Gly ³⁶ →Val	21	5'-TGCCTGTTGTAATACATCGGT-3'	<i>AccI</i> (+)
Tyr ³⁷ →Ala	21	5'-GCCGTTGTTGGCGCCATCGGTG-3'	<i>BbeI</i> (+)
Tyr ³⁷ →Arg	26	5'-GCCGTTGTTGGCGCCATCGGGAGCGC-3'	<i>BglI</i> (+)
Tyr ³⁷ →Asp	21	5'-GCCGTTGTTGGGGAATATCGGTG-3'	<i>EcoRV</i> (+)
Tyr ³⁷ →Gly	19	5'-CGTTGTTGGGGGCATCGGT-3'	None
Tyr ³⁷ →His	26	5'-GCCGTTGTTGGCCACATCGGGAGCGC-3'	<i>BglI</i> (+)
Tyr ³⁷ →Phe	26	5'-GTTGTTGGGTTTCATCGGTGAACGCTG-3'	<i>HaeII</i> (+)
Tyr ³⁷ →Ser	21	5'-GCCGTTGTTGGATCGATCGGTG-3'	<i>PvuI</i> (+)
Ile ³⁸ →Leu	21	5'-GTTGGGTACCTAGGTGAGCGC-3'	<i>AvrII</i> (+)
Gly ³⁹ →Leu	21	5'-GGGTACATCCTCGAGCGCTGC-3'	<i>XhoI</i> (+)
Arg ⁴¹ →Ala	23	5'-GTACATCGGTGAGGCTGCCAGT-3'	<i>SmaI</i> (+)
Arg ⁴¹ →Asp	21	5'-ATCGGTGAGGACTGCCAGTAC-3'	<i>HaeII</i> (-)
Arg ⁴¹ →Gln	21	5'-GTACATCGGTGAGCAATGCCA-3'	<i>HaeII</i> (-)
Arg ⁴¹ →Glu	23	5'-GTACATCGGTGAGGAGTGCCAGT-3'	<i>HaeII</i> (-)
Arg ⁴¹ →Gly	19	5'-CATCGGTGAGGGCTGCCAG-3'	<i>HaeII</i> (-)
Arg ⁴¹ →Ile	21	5'-GTACATCGGTGAGATCTGCCA-3'	<i>BglI</i> (+)
Arg ⁴¹ →Lys	23	5'-GTACATCGGTGAGAAATGCCAGT-3'	<i>HaeII</i> (-)
Arg ⁴¹ →Tyr	21	5'-GTACATCGGTGAGTACTGCCA-3'	<i>ScaI</i> (+)

^a Mutated codons, introducing amino acid substitutions, are underlined. Mismatches are boldface.

^b (+) or (-) indicates the gain or loss of a restriction site used in mutant screening.

Table 3
COMPARISON OF VARIOUS EGF SPECIES' ABILITY TO BE PRODUCED/SECRETED
UPON $\pm$ CHLORAMPHENICOL ADDITION

EGF Species	Avg. Yield of EGF Secreted to Periplasm Without Cm Addition (μ g EGF / 1 culture)	Avg. Yield of EGF Secreted to Periplasm With Cm Addition (μ g EGF / 1 culture)	Avg. Fold Increase in EGF Secreted to Periplasm upon Cm Addition
Wild-type	200 - 300	600 - 1000	2 - 5 X
Ile ²³ →Thr	10 - 20	300 - 500	30 - 50 X
Asp ²⁷ →Gly	1 - 3	35 - 50	10 - 50 X
Tyr ²⁹ →Gly	1 - 10	350 - 400	35 - 400 X
Tyr ³⁷ →Gly	1 - 18	600 - 1200	30 - 1000 X
Arg ⁴¹ →Gly	1 - 15	100 - 400	10 - 400 X
Leu ⁴⁷ →His	400 - 600	600 - 1200	1 - 3 X

Table 4

AMINO ACID COMPOSITION OF RECOMBINANT WILD-TYPE hEGF

Amino Acid Residue	Observed ^a residues/mol	Calculated ^b
Asx ^c	7.3	7
Thr	0.3	0
Ser	2.9	3
Glx ^d	4.7	5
Pro ^e	1.3	1
Gly	4.1	4
Ala	2.5	2
Cys ^f	4.7	6
Val	2.8	3
Met	0.6	1
Ile	2.1	2
Leu	5.3	5
Tyr	5.1	5
Phe	0.2	0
His	2.1	2
Lys	2.1	2
Trp ^g	1.8	2
Arg	3.0	3
Total		53

^a Based on a calculated M_r of 6217 after 6 N HCl hydrolysis under reducing conditions, except where noted.

^b Calculated from nucleotide sequence.

^c Five aspartic acid residues and 2 asparagine residues.

^d Four glutamic acid residues and 1 glutamine residue.

^e Determined after oxidation in dimethyl sulfoxide to remove co-eluting cysteine.

^f Determined as cysteic acid after dimethyl sulfoxide oxidation.

^g Determined after hydrolysis in 3 N mercaptoethanesulfonic acid.

Table 5

FUNCTIONAL CHARACTERIZATION OF RECOMBINANT HUMAN EGF

I. EGF Receptor Binding Capacity

nM hEGF	% Competition vs. 16nM [^{125} I]mEGF	nM mEGF	% Competition vs. 16 nM [^{125} I]mEGF
0	0	0	0
7	25	8	25
15	50	16	50
40	75	40	75
70	90	78	90
120	95	132	95

II. EGF Receptor Protein-Tyrosine Kinase Stimulation

nM EGF	% Stimulation	
	hEGF	mEGF
0	0	0
10	13.8	8.9
20	17.4	15.4
50	42.5	36.3
100	94.5	78.8

Table 6

COMPARISON OF IC₅₀ VALUES AND RELATIVE RECEPTOR AFFINITIES
OF hEGF TYROSINE-37 AND ARGININE-41 ANALOGS

EGF Species	IC ₅₀ (μM)	Relative Receptor Affinity (% of Wild-Type)
Wild-type	13.0 x 10 ⁻³	100
Tyr ³⁷ →Phe	10.3 x 10 ⁻³	126
Tyr ³⁷ →His	17.5 x 10 ⁻³	74
Tyr ³⁷ →Ser	21.0 x 10 ⁻³	62
Tyr ³⁷ →Ala	33.0 x 10 ⁻³	39
Tyr ³⁷ →Asp	56.0 x 10 ⁻³	26
Tyr ³⁷ →Arg	130 x 10 ⁻³	10
Tyr ³⁷ →Gly	200 x 10 ⁻³	6.5
Arg ⁴¹ →Lys	3.0	0.43
Arg ⁴¹ →Gln	6.0	0.22
Arg ⁴¹ →Ile	7.5	0.17
Arg ⁴¹ →Tyr	8.0	0.16
Arg ⁴¹ →Gly	8.5	0.15
Arg ⁴¹ →Ala	26.2	0.05
Arg ⁴¹ →Asp	>100	<0.01
Arg ⁴¹ →Glu	>100	<0.01

Table 7
COMPARISON OF EC₅₀ VALUES AND RELATIVE RECEPTOR AFFINITIES
OF hEGF TYROSINE-37 ANALOGS

EGF Species	EC ₅₀ (μM)	Relative Receptor Affinity (% of Wild-Type)
Wild-type	18.9 x 10 ⁻³	100
Tyr ³⁷ →Phe	16.7 x 10 ⁻³	113
Tyr ³⁷ →His	29.1 x 10 ⁻³	65
Tyr ³⁷ →Ser	30.4 x 10 ⁻³	62
Tyr ³⁷ →Ala	46.0 x 10 ⁻³	41
Tyr ³⁷ →Asp	74.4 x 10 ⁻³	25
Tyr ³⁷ →Arg	215 x 10 ⁻³	8.8
Tyr ³⁷ →Gly	330 x 10 ⁻³	5.7

Table 8

IC₅₀ VALUES AND RELATIVE RECEPTOR AFFINITIES OF MUTANT
EGF PROTEINS CONTAINING AMINO ACID SUBSTITUTIONS AT
OTHER HIGHLY CONSERVED SITES IN LOOP C

EGF Species	IC ₅₀ (μM)	Relative Receptor Affinity (% of Wild-Type)
Wild-type	13.0 x 10 ⁻³	100
Gly ³⁶ →Val	not determined	not determined
Ile ³⁸ →Leu	16.9 x 10 ⁻³	77
Gly ³⁹ →Leu	217 x 10 ⁻³	6

VITA

David Alan Engler was born in Petoskey, Michigan on September 14, 1962. In 1963, he moved with his family to Aurora, Illinois where he started his elementary education. In 1970, he moved to Omaha, Nebraska where he completed his primary and secondary education. He graduated from Northwest High School in June of 1980. The following September he entered Florida Institute of Technology in Melbourne, Florida, and in June of 1984 he received a Bachelor of Science degree in Marine Biology. He enrolled in the doctoral program at The University of Tennessee - Oak Ridge Graduate School of Biomedical Sciences at the Biology Division of Oak Ridge National Laboratory. He received his Doctor of Philosophy degree, with a major in Biomedical Sciences with an emphasis in Biochemistry, in May 1990. He is married and has two children. He is a member of the American Association for the Advancement of Science and the International Society of Biorecognition Technology.