Construction of Transforming Growth Factor Alpha (TGF- α) Phage Library and Identification of High Binders of Epidermal Growth Factor Receptor (EGFR) by Phage Display

Xiao-Bo Tang,*.1 Paul Dallaire,† David W. Hoyt,*.2 Brian D. Sykes,* Maureen O'Connor-McCourt,† and Bruce A. Malcolm*

*Protein Engineering Network of Centers of Excellence and Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2H7; *Department of Biochemistry, McGill University, Montreal, Quebec Canada; and *Receptor Group, Biotechnology Research Institute, NRCC, Montreal, Quebec, Canada H4P 2R2

Received for publication, June 30, 1997

 $TGF-\alpha$, a 50 amino acid growth factor containing 3 disulfide bonds, was fused to the N-terminal domain of the pIII protein of fusN, a derivative of phagemid fd-tet, to form a $TGF-\alpha$ phage. The fusion phage showed binding activity to epidermal growth factor receptor (EGFR). A library of approximately 4×10^7 variants of TGF- α was generated with substitutions of total of 10 amino acids located in the C-loop region. This C-loop subdomain of TGF- α consists of a small antiparallel double hairpin structure involving interactions between intra-polypeptide segments. Mutants isolated from the phage library with greatly increased binding affinity were selected through panning with A431 cells (a cell line expressing an elevated number of EGFRs). Following two rounds of stringent selection, variant phages with higher binding affinity than wild type TGF- α were identified and the phage DNAs were sequenced for the alignment analysis. Absolute selection at position 42 as Arg, preferential selection at position 38 and 45 as Tyr or Phe with aromatic side chain and selection at position 41 with acidic residues, were obtained. Although an amino acid residue with smaller side chain at position 35 and one with larger side chain at position 36 were preferred, the steric hindering of the structure in side chains was minimized between these adjacent amino acids.

Key words: epidermal growth factor, epidermal growth factor receptor, phage library.

Transforming growth factor alpha (TGF- α) is a small protein containing 50 amino acids and 3 disulfide bonds. The correct disulfide bond formation of these cysteines has been shown to be essential for receptor binding and biological activity (2, 3). The primary sequence has 35% sequence identity with epidermal growth factor (EGF) to which it is structurally closely related (1). TGF- α shows a variety of biological actions, including in promoting mitogenesis in normal epithelial and mesenchymal cells and stimulating rampant growth in neoplastic cells as well as promoting the wound healing process (5), but it is found to be more potent than EGF in several biological functions (6). TGF- α competes with EGF to bind to the EGFR and then can activate the receptor's cytoplasmic tyrosine kinase which in turn signals growth in epithelial and mesenchymal cells (4, 6-8). Structural (9) and site-directed mutagenesis studies (10) to date have been directed at understanding the growth factor's interaction with the EGFR and its ability to precipitate malignant transformation and growth. It has led to the suggestion that inappropriate expression of TGF- α may have a role in the uncontrolled transformation

and maintenance of several different types of malignant tumors, including breast, renal and squamous cancers.

TGF- α may be divided into three loops (A, B, and C) by the three conserved disulfide bonds (11). Many studies have been performed to clarify the structure-function relationships in TGF- \alpha using recombinant DNA techniques (9), monoclonal antibodies (12), and synthetic peptides derived from the primary sequence (13). Certain residues in TGF- α , such as R42 and L48, appear critical for its biological function (14-16) and it has been suggested that more than one loop of TGF- α is required to trigger the mitogenic response. Structural studies in solution phase on the growth factor and the growth factor-receptor complex (19, 20) suggested features which are important for the binding and the activation of EGFR. NMR spectroscopy data suggest that the C-terminus of TGF- α is probably in close contact with the EGFR in the complex whereas the N-terminus is not (17, 19). It has recently become apparent that a great deal of structure-function and protein engineering information can be gained through the use of phage display technology. This approach allows one to generate a vast array of mutant proteins and, using the appropriate screening procedure, select for variants with the desired characteristics (20, 21). This has been used for the selection of mutants of growth hormone which bind to the growth hormone receptor with higher affinity (22). In addition, identification of protein ligands that modulated function

To whom correspondence should be addressed at the present address: Department of Physiology, University of Alberta, Edmonton, Alberta, Canada T6G 2H7. Tel: +1-403-492-7203, Fax: +1-403-492-8915, E-mail: xiaotang@pop.srv.ualberta.ca

² Present address: Brattelle Pacific Northwest Laboratory, MSIN, P7-55, P.O. Box 999, Richland WA 99352, USA.

was done by thoughtful selection strategies (23, 24). In this report, we used phage display technology to generate and select variants of TGF- α with increased affinity for the EGFR.

The TGF- α gene was retrieved from plasmid pTBI102 (courtesy of R. Harkins, Berlex Biosciences), in a modified form via a polymerase chain reaction (PCR) directed mutagenesis procedure to remove the internal KpnI site and to introduce a unique SphI site without changing amino acid component. The TGF- α was fused to the N terminus of the phage surface protein, pIII, through a glycine-glycineserine spacer to increase the flexibility of the TGF- α , as shown in Fig. 1. The recombinant TGF- α phage clones were screened by restriction enzyme digestion and candidate clones were verified by DNA sequencing.

The TGF- α phage library was constructed by cloning oligonucleotides into the TGF- α phage DNA by replacing a total of 10 amino acid positions loop with a limited number of substitutions at each position, starting at position 33 (Table I), which is located in the TGF- α C loop. An 84bp double stranded, partially degenerate DNA cassette, was synthesized by extension of a partially degenerate synthetic oligonucleotide [5'-ACCGGCATGC (G/A/C)(A/T/ C)C \overrightarrow{TGC} (C/A/G) $(\overrightarrow{A/C/G})T$ $(\overrightarrow{A/C/G})(\overrightarrow{A/C/G})T$ \overrightarrow{GGT} T/A)T GACCTGCTGGCTGGTGGATCTGCGGCCGAAA-C-3' with a primer [5'-GTTTCGGCCGCAGATCCACC-3'] complementary to the 3' constant region using T4 DNA polymerase. The DNA cassette for insertion was isolated from polyacrylamide gel, then digested to completion with SphI and EagI, and ligated into previously digested and purified TGF- α phage DNA (Fig. 1). DNA from a representative aliquot of the library was extracted and sequenced to assess the distribution of bases.

High efficiency transformation (10^7 transfomants/ μ g DNA) was accomplished by electroporation (25). Rather than randomize each position completely (*i.e.* all 20 amino acids), a different subset of amino acids was used for each position changed. The fraction of clones containing a library insert (rather than the wild type $TGF-\alpha$) was determined by comparing the ligation frequency of the vector preparation itself to the ligation frequency with the insert. Based on the vector self-ligation frequency, less than 9% of the clones should have been religated wild-type $TGF-\alpha$. After subtraction of these background clones, the library consists

of approximately 4×10^7 independent recombinants. In order to evaluate the genetic diversity of the TGF- α phage library, Twenty two individual clones isolated from a representative aliquot of the phage library were isolated and sequenced. The pooled sequence gave the expected distribution of bases based on approximate quantitation from a phosphorimager and the individual sequences showed no duplicate clones among those sampled. These data suggest that no gross and inappropriate bias had occurred during the generation of the library (data not shown).

To select EGFR-specific, high affinity TGF- α variants, the phage library was subjected to two consecutive rounds of affinity selection using A431 cells which over-expresses the EGFR approximately 106 receptors/cell. For the first round of panning, 108 cfu of each phage were incubated with 106 A431 cells (approximately 1012 EGFRs) in a 60 mm tissue culture plate with 1 ml EMEM (Gibco BRL) for 1 h at 4°C. The phage solutions were removed, and the plates were rapidly washed 3 times on ice with 1 ml EMEM. Adherent phages were eluted in two stages: first by adding 1 ml of EMEM containing 1 µM EGF for 2.5 h at 4°C to compete wild-type and/or low binders, and second by an elution with 1 ml of elution buffer (0.1 M glycine, pH 2.2, 1 mg BSA/ml, and 0.1 mg/ml of phenol red). After incubating for 10 min at room temperature to release bound phage, this eluate was removed and neutralized with 20 µl of 2 M

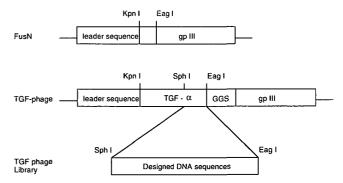


Fig. 1. Construction of TGF- α phage and TGF- α phage library. The FusN-TGF- α plasmids express the fusion proteins in which the gene III signal peptide, the TGF- α protein and its library, and a short spacer sequence (GlyGlySer) were fused to the amino terminus of the mature pIII protein.

TABLE I. Construction of TGF- α phage library with size of 3.4×10^7 variants of amino acid compositions.

Number of possible variations	9	1	9	9	1	2	9	1	8	2	1	2	9	9
Position	33	34	35	36	37	38	39	40	41	42	43	44	45	46
w.t. TGF-α	V	С	H	S	G	Y	V	G	A	R	С		H	A
DNA sequence	GAC AT CC	TGC	CAT AC GG	ATG CA GC	GGT	TAC T	GTT AA TC	GGT	GCA AA CG	CGA A	TGC	CAG G	CAT TT GC	GCT TT CA
Mutations	A D H		A D G	A D G		F	A D F		E G K	Q		Q	A D F	D F H
	I L		N P	H N			I N		P O				L P	L P
	N		R	P			S		2xR				s	s
	T T		S T	к Т			T Y		Т				V Y	V Y

688 X.-B. Tang *et al.*

Tris base. The phages eluted from the each round with low pH glycine buffer were amplified in solution, as described previously (25). An aliquot of DNA representing the two thousand phages eluted with low pH after round two selections was sequenced to estimate the diversity. Comparison of the sequencing gel before and after the two rounds of selection indicates that some positions show stringent selection bias, as the amino acid sequences of the TGF- α mutants encoded by each phage clone listed in Table II. Several genetically identical sequences were found among the 32 TGF- α phage mutants screened. The number of genetically identical siblings obtained is indicated in parentheses after the clone's designation in Table II. These clones are most likely genetic duplicates and not independent selection events and therefore the total number of truly independent clones examined was only 24. Consequently, amino acid frequencies were calculated based on the number of appearances out of 24 independent events.

In the first position varied, Val 33, the wild type amino acid was selected in 14 out of 24 clones analyzed (58%). Ile replaced the Val in five clones (21%). His was found twice (8%). Three amino acids (Asp, Pro, and Thr) were absent from the pool of clones after screening. It appears that this position is highly selective for the original residue or a structurally related amino acid replacement (i.e. an aliphatic side chain).

The second residue varied was His at position 35. None of the sequenced phages had the original amino acid in this position. The residue most often found was Asp (10 clones, 42%). The second most common ones were Ala (5 clones, 21%) and Asn (5 clones, 21%). Arg was also found in two clones (8%). The rest of the substitutions at this position,

such as Ser and Thr, were never found. This suggests that the wild type residue, His, is not favored under the selection conditions and the smaller residues (Asp, Asn, and Ala) may confer improved binding.

The third position randomized was number Ser 36. Fifty percent of the TGF- α mutants sequenced contained an Arg in this position. Both Asn and Ala were present in 13% of the clones (3 clones), respectively. Ser and Asp were each selected in two clones (8%). Thr and His were each found in one clone (4%). The other two possibilities (Gly and Pro) were not found at all. This suggests that this position has some degree of selectivity and that the bigger residues are, not only tolerated, but may improve binding under the selection conditions.

Only two amino acids were introduced into position 38 (Phe and Tyr-wild-type). Equal distribution was found for the wild type (Tyr, 46%) and mutant (Phe, 54%) among the clones. From these data it would appear that the presence or absence of the phenolic OH group has no effect on binding under the conditions of the selection.

The wild type residue Val 39 was the most frequently selected (46%). The other 54% were comprised of by several other amino acids, such as Ser (17%), Tyr (13%), Ile (8%), Thr (8%), Ala (4%), and Phe (4%). Two other available replacements (Asp and Asn) were not seen among the selected clones. It seems that the wild type amino acid (Val) at this position is favored but that virtually all types of replacements can be tolerated under the experimental conditions.

Approximately 66% of the wild type Ala residues were replaced by Glu residues at position 41. Only three clones (13%) retained the wild type $TGF-\alpha$ residue. Other re-

TABLE II. Biopanning of TGF- α phage library for selection of high affinity binders.

Number of														
possible variations	9	1	9	9	1	2	9	1	8	2	1	2	9	9
Position	33	34	35	36	37	38	39	40	41	42	43	44	45	46
Human TGF-α	V	C	H	S	G	Y	V	G	A	R	C	E	H	A
Rat TGF- α	V	C	Н	S	G	Y	V	G	V	R	С	E	H	Α
Human EGF	N	C	V	V	G	Y	I	G	\mathbf{E}	R	C	Q	Y	R
Mouse EGF	N	C	V	I	G	Y	S	G	D	R	C	Q	T	R
Clones														
R16	N		D	Α					E				\mathbf{F}	S
R17 (4 clones)			Α	N					\mathbf{E}				F	S
R18			N	R		\mathbf{F}			\mathbf{E}				F	S F F
R19 (3 clones)			D	\mathbf{T}					\mathbf{E}				Y	\mathbf{F}
R23	H		D	\mathbf{R}			S						Y	V
R25	H		N	R			S						Y	D
R38	L		D						G				F	
R40			Α	Α		F	S		E				Y	V
R44			N	R		\mathbf{F}	I		\mathbf{E}			Q	Y	D
R81 (2 clones)			N				\mathbf{T}		\mathbf{E}				F	V
R84 (2 clones)			D	R		F			Q				F	D
R90			Α	Α			S		E				Y	V
R91 (2 clones)			D	R		\mathbf{F}	F		E				V	F
R103	Α		P	N									Α	Н
R104	I		R	R		\mathbf{F}	Y		Q				F	V
R109			N	R		F	I		\mathbf{E}			Q	Y	D F
R112			D	R		F			E				F	F
R135			D	R		\mathbf{F}	Y		R				F	
R140			Α	N		\mathbf{F}			\mathbf{E}				F	S
R172	I		Α	H			T		\mathbf{E}			Q	F	
R196	I		D	R		\mathbf{F}			P				F	S
R281			G	D			Y		G				Y	F Y
R292	I		D	D		\mathbf{F}	Α		T				Y	Y
R300	I		R	R		F			E				A	Y

placements obtained were: Gln (8%), Gly (8%), Pro (4%), and Thr (4%). Basic amino acids (Lys and Arg) were not found. This suggests that this position strongly favors an acidic residue under these selection conditions.

Absolute selection was seen at position 42. Every clone retained the wild type residue, Arg, after selection rather than the alternative Gln. This was the only case of an absolute selection found in this experiment, suggesting that Gln at this position compromises binding significantly.

Almost 90% of the selectants retained the original amino acid Glu at position 44. The only other amino acid found at this position was the biochemically related Gln (3 clones, 13%). The other radically different residues were not selected. This indicates that this position is highly selective for the wild type amino acid.

At position 45, two amino acids accounted for 87% of the clones: Phe (53%) and Tyr (34%). The original wild type residue, His, was not selected at all. This data suggests the His is not optimal for binding under the selection condition. This may reflect the pK_a of the His imidazole and the impact of protonation (i.e. positive change) at this position upon binding. Finally, at position 46 (Ala) very little selection was seen. The amino acids profile after selection was similar to the original library. This suggests that this position is not selective.

Previous studies on the structure-function activity of TGF- α and EGF strongly suggested that discontinuous regions of TGF- α are involved in receptor binding (9). These observations include: (a) alteration of any of the three disulfide bonds by mutagenesis or treatment with reducing agents and (b) synthesis of various fragments or loops of TGF- α , failed to inhibit binding of TGF- α to its receptor; and (c) site-directed mutants in specific regions of TGF- α , such as Tyr-38 to Ala (15), Arg-42 to Lys (14), and Leu-48 to Ile (16) in the C-loop abolished receptor binding and mitogenic activity. Particularly, the C-loop domain consists of a small antiparallel double hairpin structure (18) involving interaction between polypeptide segments Cys-34/His-35, Tyr-38/Val-39, and His-45/Ala-46. Residues Ser-36/Gly-37 form a type-II β-turn. The polypeptide segment Gly-40/Glu-44 forms a unique multiple-bend structure which has a single turn of a left-handed helix. This kind of structure domain is further stabilized by Cys-34/Cys-43 disulfide bond. The relative orientation of the A, B, C loop subdomains of TGF- α is also determined primarily by long-range distance constrains, such as Phe-17 and Glu-44, His-18 and His-35, His-18 and Tvr-38, Glv-19 and His-35. Some of the residues important for receptor recognition are brought together in space by interactions between the loops and left-handed helical segments.

Amino acid substitutions at ligand-receptor contact points could affect the strength of the binding. Alternatively, substitutions at these residues may also affect the secondary or tertiary structure, resulting in a conformational change that alters binding. The amino acid sequence alignment in the C-loop region of TGF-α and EGF shows some residues are highly conserved (e.g. Cys-34, Gly-37, Tyr-38, Gly-40, Arg-42, and Cys-43; see Table II). Therefore, they might be crucial for the receptor binding. Although the "hinge" residue Val-33 located between two subdomains of B- and C-loops is not conserved across the family of EGFR ligands, all known TGF-α sequences possess Val, while every known EGF protein has an Asn.

Such a structure difference might introduce variation into the conformational flexibility of the two families of proteins.

A study which suggests that EGF and $TGF-\alpha$ appear to bind differently to EGFR was reported (26). The shift of amino acid sequences at certain positions (Val-39 in $TGF-\alpha$ to Ile or Ser in EGF, Glu-44 in $TGF-\alpha$ to Gln in EGF, and His-45 in $TGF-\alpha$ to Tyr or Thr) may also be selected for the protein conformational flexibility. The selections for high binders to EGFR from phage library create these variations (Table II). Many selected clones contain EGF sequences at these three positions. It indicates that the receptor binding property from the phage selected sequences may also be shifted to more favorable for EGF-like binding.

Taken together, our data suggested that these TGF- α mutants obtained after panning may represent variants, which have stronger interactions with the EGFR. Among the 10 positions examined every degree of selectivity was observed. Position 42 (Arg) showed absolute selection for Arg. Positions 33 (Val), 39 (Val), and 44 (Glu) were strongly selective for their wild type residues. In contrast, the original amino acids at positions 35 (His) and 45 (His) were completely replaced by other amino acids. Positions 36 (Ser) and 41 (Ala) showed a partial preference for their original residues. Position 38 (Tyr) had equal distribution between the wild type residue and the closely related replacement (Phe). Only position 46 (Ala) showed little or no selection among the highly diverse amino acids introduced.

REFERENCES

- Marquardt, H., Hunkapiller, M.W., Hood, L.E., and Todaro, G.J. (1984) Rat transforming growth factor type 1: structure and relation to epidermal growth factor. Science 223, 1079-1082
- Todaro, G.J., Fryling, C., and De-Larco, J.E. (1980) Transforming growth factors produced by certain human tumor cells: polypeptides that interact with epidermal growth factor receptors. Proc. Natl. Acad. Sci. USA 77, 5258-5262
- Massague, J. (1983) Epidermal growth factor-like transforming growth factor. II. Interaction with epidermal growth factor receptors in human placenta membranes and A431 cells. J. Biol. Chem. 258, 13614-13620
- Coffey, R.J., Derynck, R., Wilcox, J.N., Bringman, T.S., Goustin, A.S., Moses, H.L., and Pittelkow, M.R. (1987) Production and auto-induction of transforming growth factor-alpha in human keratinocytes. *Nature* 328, 817-820
- Rappolee, D.A., Mark, D., Banda, M.J., and Werb, Z. (1988)
 Wound macrophages express TGF α and other growth factors in vivo: analysis by mRNA phenotyping. Science 241, 708-712
- Carpenter, G. and Wahl, M.I. (1991) Peptide Growth Factors and Their Receptors (Sporn, M.B. and Roberts, A.B., eds.) Vol. I, pp. 69-171, Springer-Verlag New York, New York
- Campbell, I.D., Cooke, R.M., Baron, M., Harvey, T.S., and Tappin, M.J. (1989) The solution structures of epidermal growth factor and transforming growth factor alpha. *Prog. Growth Factor* Res. 1, 13-22
- Reynolds, F.H., Todaro, G.J., Fryling, C., and Stephenson, J.R. (1981) Human transforming growth factors induce tyrosine phosphorylation of EGF receptors. *Nature* 292, 259-262
- Defeo-Jones, D., Tai, J.Y., Wegrzyn, R.J., Vuocolo, G.A., Baker, A.E., Payne, L.S., Garsky, V.M., Oliff, A., and Riemen, M.W. (1988) Structure-function analysis of synthetic and recombinant derivatives of transforming growth factor alpha. *Mol. Cell. Biol.* 8, 2999-3007
- Feild, J.A., Reid, R.H., Rieman, D.J., Kline, T.P., Sathe, G., Greig, R.G., and Anzano, M.A. (1992) Structure-function analy-

690 X.-B. Tang et al.

sis of human transforming growth factor- α by site-directed mutagenesis. *Biochem. J.* 283, 91-98

- Tam, J.P., Lin, Y.Z., Liu, W., Wang, D.X., Ke, X.H., and Zhang, J.W. (1991) Mapping the receptor-recognition site of human transforming growth factor-alpha. *Int. J. Pept. Protein Res.* 38, 204-211
- Hoeprich, P.D., Langton, B.C., Zhang, J.W., and Tam, J.P. (1989) Identification of immunodominant regions of transforming growth factor alpha: implications of structure and function. J. Biol. Chem. 264, 19086-19091
- Eppstein, D.A., Marsh, Y.V., Schryver, B.B., and Bertics, P.J. (1989) Inhibition of epidermal growth factor/transforming growth factor-alpha-stimulated cell growth by a synthetic peptide. J. Cell. Physiol. 142, 420-430
- Defeo-Jones, D., Tai, J.Y., Vuocolo, G.A., Wegrzyn, R.J., Schofield, T.L., Riemen, M.W., and Oliff, A. (1989) Substitution of lysine for arginine at position 42 of human transforming growth factor-alpha eliminates biological activity without changing internal disulfide bonds. Mol. Cell. Biol. 9, 4083-4086
- Lazar, E., Vicenzi, E., Van Obberghen-Schilling, E., Wolff, B., Dalton, S., Watanabe, S., and Sporn, M.B. (1989) Transforming growth factor alpha: an aromatic side chain at position 38 is essential for biological activity. Mol. Cell. Biol. 9, 860-864
- Lazar, E., Watanabe, S., Dalton, S., and Sporn, M.B. (1988)
 Transforming growth factor alpha: Mutation of aspartic acid 47 and leucine 48 results in different biological activities. Mol. Cell. Biol. 8, 1247-1253
- 17. Kohda, D., Shimada, I., Miyake, T., Fuwa, T., and Inagaki, F. (1989) Polypeptide chain fold of human transforming growth factor alpha analogous to those of mouse and human epidermal growth factors as studied by two-dimensional 'H NMR. Biochemistry 28, 953-958

- Montelione, G.T., Winkler, M.E., Burton, L.E., Rinderknecht, E., Sporn, M.B., and Wagner, G. (1989) Sequence-specific ¹H-NMR assignments and identification of two small antiparallel β-sheets in the solution structure of recombinant human transforming growth factor α. Proc. Natl. Acad. Sci. USA 86, 1519-1523
- Hoyt, D.W., Harkins, R.N., Debanne, M.T., O'Connor-McCourt, M., and Sykes, B.D. (1994) Interaction of transforming growth factor α with the epidermal growth factor receptor: binding kinetics and differential mobility within the bound TGF-α. Biochemistry 33, 15283-15292
- Winter, J. (1994) Bacteriophage display: peptide libraries and drug discovery. Drug Dev. Res. 33, 71-89
- Hoess, R.H. (1993) Phage display of peptides and protein domains. Curr. Opin. Struct. Biol. 3, 572-579
- Bass, S., Greene, R., and Wells, J.A. (1990) Hormone phage: an enrichment method for variant proteins with altered binding properties. *Proteins: Struct. Funct. Genet.* 8, 309-314
- Clackson, T. and Wells, J.A. (1994) In vitro selection from protein and peptide libraries. Trends Biotechnol. 12, 173-184
- Fong, S., Doyle, L.V., Devlin, J.J., and Doyle, M.V. (1994) Scanning whole cells with phage-display libraries: identification of peptide ligands that modulate cell function. *Drug Dev. Res.* 33, 64-70
- Christian, R.B., Zuckermann, R.N., Kerr, J.M., Wang, L., and Malcolm, B.A. (1992) Simplified methods for construction, assessment and rapid screening of peptide libraries in bacteriophage. J. Mol. Biol. 227, 711-718
- Winkler, M. E., O'Connor, L., Winget, M., and Fendly, B. (1989)
 Epidermal growth factor and transforming growth factor alpha
 bind differently to the epidermal growth factor receptor. Biochemistry 28, 6373-6378