Epidermal Growth Factor single mutants highlighted by homologs cross-conservation approach differentially affect the EGF Receptor downstream pathway

Dalmira Merzhakupova[1¶], Stefano Pascarelli[1 ¶], Paola Laurino[1*]

[1] Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna-son, Okinawa 904-0412 Japan.

* Corresponding author

E-mail: paola.laurino@oist.jp

¶ These authors contributed equally to this work.

Abstract

Epidermal Growth Factor Receptor is a tyrosine kinase receptor that plays a key role in cells. Its ligands contribute to the system modularity by influencing the stability of the dimerized receptor. A synthetically modified ligand could be a potential anti-cancer drug if able to compete with the endogenous ligands while modulating a different pathway activation. Herein, we developed a cross-conservation approach to identify functionally important EGF residues. Four EGF mutants have been selected and studied. These mutants show binding affinities for EGFR similar to EGF, while the effect at the cellular level changed, increasing growth rate in fibroblasts and reducing apoptosis in skin cancer cells. These results support the theory of biased signaling in the tyrosine kinase receptor system.

Abbreviations

EGF, Epidermal Growth Factor

EGFR, Epidermal Growth Factor Receptor

PPI, Protein-Protein Interaction

MSA, Multiple Sequence Alignment

MSTA, Multiple STructural Alignment

ITC, Isothermal Titration Calorimetry

DMEM, Dulbecco's Modified Eagle Medium

ECD, Extra Cellular Domain

WT, Wild Type

Introduction

Protein-protein interactions (PPIs) regulate many biological processes (1). PPIs are one of the most interesting and well-studied examples of molecular co-evolution in biological systems. These interactions are sometimes defined by one part (receptor) that binds several counterparts (ligands). Receptors and ligands experience different selective constraints, and receptors tend to evolve more slowly due to the necessity of binding multiple ligands. Identification of key residues in a ligand that may affect binding and the resulting cellular phenotype should provide new understanding of PPI co-evolution (2).

In recent years, different experimental techniques have been developed to define the effects of single mutant proteins at the cellular level (3). Often these approaches generate false-positive and false-negative that can be misinterpreted and leading to unclear conclusions (3). Bioinformatic tools can be exploited for a more detailed analysis of PPIs co-evolution. The combination of sequence and structural alignment methods have made possible to identify amino acids essential for ligand-receptor interactions, but the ligand effect at the cellular level still remains unclear (4).

The epidermal growth factor (EGF)-like domain ligand – ErbB1(EGFR) receptor signaling system is involved in many biological events in multicellular organisms (5) and is considered to be ancient (6). Few studies have shown that overexpression of oncogenic receptors and ligands may induce different types of cancers (7). EGF ligands are also involved in ion transport, glycolysis, and synthesis of proteins and nucleic acids (8). They also induce stimulation of fibroblasts in early phases of wound healing (9). As anti-EGFR antibodies often lead to inconsistent outcomes, design of EGF analogues remained an attractive target for biomedical applications (10).

Ligand-induced selective activation of downstream pathway has been observed in G-coupled protein receptor as “biased signaling” (11, 12). Recent reports hints that this mechanism might take place also in EGFR tyrosine kinase (13). Comprehensive analysis of specific residues in EGF ligands from different species and among different EGFR ligands (EGF, HBEGF, EPGN, BTC, EREG, AREG, TGFA) might allow the design of mutants with different or improved functions. Recent studies showed that some EGF residues like R41 and L47 are highly conserved and important for high binding affinity to EGFR in the A431 cancer cell line (14). Another study highlighted Y13, L15 and H16 residues as essential for downstream activity of ErbB1 (15). These outcomes were based on structural analyses of ligands and experimental validation. Although, to the best of our knowledge a study to test a difference in the biased signaling of a single-residue mutated EGF ligand was not performed yet.

In this paper we show a novel approach to study PPIs, through cross-conservation analysis. We combined bioinformatics and experimental tools to study co-evolution of the EGF-EGFR ligand-receptor system at the residue level (Figure 1A). The identified residues have phenotypic implications at the cellular level, influencing protein activation in the EGFR downstream pathway. Overall, this approach has been critical to identify residues that play important role in biased signaling.

Materials and Methods

Sequence and structure analysis

The sequences of EGFR ligands and the multiple sequence alignment of EGF orthologs were obtained from Ensembl database (16). Multiple sequence alignment of all ligands was

performed with MAFFT software (17). X-ray structures were obtained from the PDB database (18). Structure alignments were created using Chimera (19).

Phylogenetic analysis

From the multiple sequence alignment of EGF from different species, nearly identical sequences were removed. The fruit fly EGF sequence was added as an outgroup in the EGF phylogeny tree, while *Caenorhabditis elegans* EGF was used as outgroup in the tree of all ligands. The image of MSA and phylogenetic trees was obtained using unipro UGENE software (20). Three phylogenetic trees were made using Neighbor Joining (NJ), Maximum Likelihood (ML), and Bayesian (MrB) methods.

As additional method, trees were also made with IQTREE (21), using ModelFinder to scan for the most fit evolutionary model and parameters (22).

Calculation of Cross conservation score

From the evolutionary MSA of EGF orthologs and ligands MSA (or MSTA) of EGF paralogs, two conservation measures were obtained. The conservation score was calculated in three different ways: 1) IDENTITY, 2) BLOSUM, and 3) JSDw. The IDENTITY score measures the frequency of appearance of an EGF residue in other ligands. In the BLOSUM score, reference position substitutions are weighted using the BLOSUM62 matrix. The JSDw score is based on Jensen Shannon divergence with a window of residues, and is the method used by Pupko *et al.* in ConSurf (23). The cross-conservation plot and analysis were performed with the Python package SEABORN. The two levels of conservation were plotted, and a cross-conservation score was obtained by computing the distance from the diagonal.

The code used in the analysis of the cross-conservation score and plots and the data used in this paper are shared on Github: https://github.com/oist/CrossConservation.

Mutation rationale

Along with the cross-conservation score, the choice of positions for mutation was influenced by two other factors: the distance from the receptor, and the amino acid variation among ligand types. Each ligand has different binding affinity and activate different pathways; therefore, we designed mutations with the aim of changing pathway activation, by taking into consideration the residues types in the ligands that show a different cellular effect. Finally, we carefully analyzed exceptional cases in the cross conservation scoring. Some of the residues that show high score have intramolecular interactions with another amino acid. These residues, if mutated, will not only change interaction with the receptor but also lose EGF structural stability (namely "residue swapping" behavior showed in Figure S2**)**. The decision of which mutation to introduce was made using EGFR ligands alignment. Overlapping residues at a given position were divided into two groups based on EGF-like and non EGF-like stabilization of the receptor dimer. This separation was previously shown to follow binding affinity (24). Residues that introduced a noticeable shift in amino acid properties in the two groups were selected. The chosen positions have higher-than-median cross conservation scores. Furthermore position 46, 48 and 50 were preferred because, given previous experiments and the overall scores, the EGF C-terminus tail seems to play a critical role in the ligand function, for example (25). N32R was also chosen since it is highly conserved in other ligands in the corresponding evolutionary alignment (e.g. CVC in TGFA, or CRC in EREG), therefore hinting for a high cross-conservation score when computed for other ligands.

Synthetic Peptides

Wild-type EGF (protein sequence:

N'NSDSECPLSHDGYCLHDGVCMYIEALDKYACNCVVGYIGERCQYRDLKWWELR-C')

and EGF variants (See below the list of peptides) with purity >90% and quantity 5 mg/mL were ordered from Scrum Net Co. These peptides were used for ITC measurements, Circular Dichroism (CD) measurements, proliferation studies, and Western Blot (WB) analyses.

The list of Mutations:

Name	Mut1	Mut2	Mut3	Mut4
Position	46	48	50	32
Amino acid substitutions	D46T	K48T	W50Y	N32R

Cell Lines

The *Bj5-tα human normal fibroblast cell line* was purchased from ATCC. Cells were grown in DMEM with 10% fetal bovine serum (FBS), and 5 µg/mL hygromycin B.

The *Swiss Albino 3T3* mouse *normal fibroblast cell line* was obtained from the RIKEN Cell Bank. Cells were grown in DMEM, 10% FBS, 50 ug/mL gentamycin at 37°C in a 5% CO_2 atmosphere with 95% humidity.

The *A431 human epithelial carcinoma adherent cell line* (RIKEN Cell Bank) is a model skin cancer cell line with overexpressed EGFR used for oncogenic pathway studies (G. Carpenter et.al.,1983). Cells were cultured in DMEM supplemented with 10% FBS (Sigma-Aldrich), 50 ug/mL gentamycin antibiotic.

Experiments were conducted at 37°C in a 5% CO_2-enriched air atmosphere with 95% humidity. Cell lines were grown and used for cell ELISA and cell proliferation studies.

Cell Proliferation Assay

We measured cell proliferation using a label-free, non-invasive, cellular confluence assay with IncuCyte Live-Cell Imaging Systems (Essen Bioscience). Human Bj5-tα (2,500 cells / well) and Mouse Swiss Albino 3T3 (1,000 cells/well) were seeded overnight on a 96-well plate (Corning) at 37°C in an incubator. The next day, cells were treated with WT EGF and mutants at 1 nM, 10 nM and 100 nM concentrations and placed in an XL-3 incubation chamber maintained at 37°C. The plate was scanned using a 4x objective at 2-hr intervals over 3 days. Cell confluence was measured using IncuCyte Analysis Software. The IncuCyte Analyzer gives real-time confluence data based on segmentation of high-definition phase-contrast images. Cell proliferation is shown as relative rate increase in confluence to control.

Apoptosis Assay

Experiments were performed with the A431 human cancer cell line. 5,000 cells/well were seeded on a 96-well plate (Corning) and incubated at 37°C for 24 hr. Media were replaced with fresh DMEM containing WT EGF, or EGF mutants at 1, 10, and 100 nM concentrations and fluorescent annexin V green reagent. Plates were pre-warmed prior to data acquisition to avoid condensation and expansion of the plate, which affect autofocus. Images were captured every 2 hrs (4x) for 3 days in the IncuCyte system. Cell proliferation is reported as in the previous assay.

Statistics

Proliferation and apoptosis experiments were performed in duplicates. All results are shown as the mean±s.d. Raw data was analyzed by two-way ANOVA with 95% confidence level. The multiple test was corrected using Bonferroni post hoc test. Prism 8 software was used

for statistical analysis. Asterisks in the pictures show P-values using GraphPad convention: 0.1234 > (ns), 0.0332 > (*), 0.0021 > (**), 0.0002 > (***), 0.0001 > (****).

Isothermal Titration Calorimetry (ITC)

All ITC studies employed a MicroCal PEAQ-ITC System (Malvern). For titration, both EGFR ECD (Sigma-Aldrich) and EGF variants were dialyzed into the same reaction buffer Milli-Q H_2O (22 μm) at 25°C. Each titration involved serial injections of 13×3 μL aliquots of EGF variants (200 μM) into a solution of EGFR ECD (20 μM) in the sample cell. In each case, the reference cell was filled with the same reaction buffer as the control to determine the heat upon binding of the two components. The measured heat constant value was subtracted from the heat per injection prior to analysis of the data. The experiment was replicated twice. Results were analyzed by MicroCal PEAQ-ITC Analysis Software.

Circular Dichroism (CD)

Far UV measurements were taken at a protein concentration of 0.1 μM, using a cuvette with a path length of 0.1 cm. Secondary structure content was calculated from far UV spectra using CAPITO software (26). Five scans in the 190-240-nm wavelength range were taken.

Western Blot Analysis

A431 epidermoid carcinoma cells were harvested using Lysis Buffer (0.4% SDS, 0.2% BETA-ME, 1% Bromophenol Blue, 10% glycerol. Samples were incubated at 65°C for 10 min, sonicated, and centrifuged at 15,000 rpm at 22°C for 10 min. Supernatants were used for further analysis. Sample concentrations were measured with a Pierce™ BCA protein assay kit (ThermoFisher Scientific). Proteins were mixed with 2x Sample Loading Laemmli Buffer and incubated at 65°C for 10 min. Equal amounts of protein were loaded in 4-15% Mini-

PROTEAN® TGX™ SDS-PAGE gel (Bio-Rad) and transferred to PolyVinylidene DiFluoride (PVDF) membranes. Membranes were blocked for 10 min with Turbo Transfer Buffer and probed with monoclonal rabbit anti-EGFR antibody (Santa Cruz Biotechnology, INC), monoclonal rabbit anti-PLCy, and anti-phosphorylated PLCy antibodies (Santa Cruz Biotechnology, INC), monoclonal mouse anti-scr and rabbit anti-phosphorylated src antibodies (Santa Cruz Biotechnology, INC), at dilution 1:1000. Samples were incubated with Goat Anti-Rabbit IgG StarBright Blue 700 at a 1:2000 dilution and Anti-Tubulin hFAB™ Rhodamine Antibody as a loading control at a 1:3,000 dilution for 3 hours and washed with Blocking Buffer and Milli-Q H_2O (22-µm filtration). Immunoreactive fluorescently labeled samples were visualized and analyzed with ImageLab Software.

Results

Initially, a phylogeny of all EGFR ligands was built **(**Figure S3**)**, presenting a high degree of monophyly among the seven paralogs. This monophyly justifies the comparison of different ligands in our further analysis. Then, we developed a method for identifying amino acid residues that are possibly involved in specific functions among paralogs, using what we named cross-conservation approach. We applied the cross-conservation approach on the seven ligands of EGFR to find out which mutations would have been key to switch function in EGF. Cross-conservation analysis highlights functional positions in proteins, based upon previous knowledge of their interacting partners, as explained in the methods section. The combination of two residue conservation measures is used to generate a cross-conservation plot (Figure 1B), by combining the alignment of all the ligands (paralogs that still share binding to same receptor), using either the Multiple Sequence Alignment (MSA, Figure A in Figure S1) or the Multiple Structural Alignment (MSTA, Figure B in Figure S1), and the

alignment of orthologs sequences of EGFs (herein so called evolutionary alignment, Figure C in Figure S1).

Figure 1. Cross-conservation analysis.

(A) Structure of extracellular EGFR-EGF complex (from PDB: 1ivo). EGFR extracellular domain in cyan cartoon. EGF peptide is represented as cartoon in a gradient of blue to red from the N-term to the C-term. I, II and III indicate the three ECD domains. C-tail of EGF is disordered and in close proximity of domain III of EGFR. (B) Cross-conservation plot. The plot is obtained by crossing the conservation measures of ligand alignments (rhombus for MSA conservation, squares for MSTA conservation) and evolutionary alignments (Figure S1). Positions highlighted in purple have been chosen for experimental verification. The color gradient shows the N-/C-end displacement of the amino acid consistently with panel A. Distance from the diagonal (e.g., 'd' in the plot) is used to calculate the cross-conservation score. Interestingly, no point lies in the bottom right half of the plot suggesting that ligand and evolutionary conservation are not independent and differently influenced by evolution pressure. The C-terminus amino acids have higher cross-conservation score on average, highlighting that this part has a functional role. (C) Extract of the paralog ligands alignments focusing on the chosen positions, showing a low degree of conservation. (D) Extract of the logo generated from the orthologs evolutionary alignment (Figure S1 panel C). Positions chosen for mutation are highly conserved, therefore resulting in a high cross-conservation score.

The cross-conservation score is calculated as the distance (d, Figure 1B) from the diagonal. The list of residues sorted by score is shown in S1 Table. According to our analysis, residues with high scores are concentrated on the C-terminal tail. This result is in line with previous

reports on the C-tail of EGF (27-29). Interactions between the interface of the ligand and EGFR have been shown to influence the stability of the dimerized receptor (30), a factor that is related to ligand-specific signaling bias (24, 31).

Based on the factors described in the methods section, we designed EGF mutants with single amino-acid substitutions (N32R, D46T, K48T and W50Y). We characterized the mutants biochemically and at the cellular level. Using circular dichroism (CD) experiments, we confirmed that the secondary structure of these mutants was maintained (Figure S4). Then the K_d for EGFR was determined *in vitro* by Isothermal Titration Calorimetry (ITC). ITC measurements of the binding of all mutants to the ECD of EGFR exhibited similar K_d values to the WT EGF. Only N32R has 2-fold higher affinity for EGFR compared to WT EGF (Figure 2B).

Figure 2. ITC measurements of EGF mutants and the EGFR receptor.

(A**)** ITC analysis of WT EGF ligand and mutants N32R, D46T, and K48T binding to the ECD of the EGFR receptor at 25° C. Measurements were taken by adding WT EGF at 200 μM to the ECD of EGFR at 20 μM. **(**B**)** Four zoom-in of X-Ray structure of the extracellular domain (ECD) of EGFR bound to EGF (PDB 1IVO). In cyan cartoon the ECD of EGFR. Each zoom-in focuses on the mutated residue. Highlighted in yellow stick side chain of the mutated residues and in cyan stick side chain of the residue of ECD in proximity (< 5 Å with the mutated residue. (C) K_d calculated from the ITC measurements using the program Affinimeter KinITC Kintecs Software.

Mutation N32R is on the interface between ligand and receptor (Figure S5). The slightly higher affinity is probably due to the presence of the guanidinium group of R which is positive charged and could interact with Q16 of EGFR ECD.

Surprisingly, although the biochemical parameters are not substantially different, EGF variants differently affected cell growth in cell proliferation studies. Compared to EGF, all the mutants exhibited an intermediate effect on the proliferation of human fibroblasts bj5-tα (Figure 3A). D48T and K48T mutants differed most from the WT (Figure S7). A similar response was observed in Albino Swiss mouse 3T3 cell line (Figure S6), and in both cell lines, the effects are concentration dependent.

In order to validate the differential effect, we performed an apoptosis assay using A431 cell lines. After 24h of growth, the cells were treated with varying concentration of EGF or mutated ligand. 48h after 100 nM treatments, EGF showed the highest decrease in cell proliferation, while the four mutants had an intermediate effect, with W50T being the closest to the WT (Figure 3B). Evidence of apoptosis was found from the globular processes formed in cells treated with WT EGF (Figure 3C).

Figure 3. Results of cell growth assays for cells treated with EGF variants.

(A) Effect of different concentrations of EGF variants on proliferation of the human normal fibroblast bj5-tα cell line. Data represent the growth rate percentage calculated from the confluence of cells relative to the control (mean+- standard deviation) for each concentration of EGF. Percent confluence was estimated 24h after the treatment (two replicates/treatment). (B**)** Effect of EGF mutants on the growth of A431 cells. EGF induced apoptosis is measured as the reduction in cell proliferation compared to the control. (C**)** Comparison of A431 cell growth after treatment with 100 nM WT EGF and EGF variants D46T and K48T. Cells were labeled with fluorescent annexin V green reagent. Plates were pre-warmed prior to data acquisition to avoid condensation and expansion of the plate, which affect autofocus. Images were captured every 2 hrs (4x) for 3 days in the IncuCyte system.

Significantly herein we report the effects of one amino acid change on the C-terminus tail of EGF ligand. To explain this behavior, we proceed analyzing the downstream proteins involved in the EGFR network. Both D46T and K48T reduced the expression of Phospholipase-Cγ1 (PLCγ1), which is a downstream signaling protein required for EGFR-induced squamous cell carcinoma (Figure 4A). The low levels of PLCγ protein lead to a decreased amount of PKCζ protein (Figure 4A). Our findings show that modularity in the downstream kinase cascade could be induced by one single amino acid substitution.

Figure 4. EGF variants D46T and K48T affect the EGFR downstream signaling pathway.

(A**)** Western blot analysis of EGFR-regulated downstream gene expression of EGF variant-treated A431 cancer cells. Expression of Epidermal Growth Factor Receptor (EGFR), Phospholipase-Cγ1 (PLCγ1) and PKCζ protein in A431 cancer cell line after treatment with 100 nM WT, EGF variant D46T or variant K48T. Samples were collected on Day 3 after treatment (two duplicates). Samples were incubated with Goat Anti-Rabbit IgG StarBright Blue 700 at a 1:2000 dilution and Anti-Tubulin hFAB™ Rhodamine Antibody as a loading control at a 1:3,000 dilution for 3 hrs and washed with Blocking Buffer and Milli-Q H_2O (22 µm filtration). Immunoreactive fluorescent labeled samples were visualized and analyzed with ImageLab Software. **(**B**)** A schematic representation of one of the EGF-EGFR-mediated signaling pathways that may be initiated in the A431 epidermoid cancer cell line. Arrows indicate the positive action of downstream gene expression, whereas arrows with flat tips indicate inhibition of gene expression. The “P+” symbol represents phosphorylation of downstream-regulated proteins. The dashed line represents the potential cellular effect regulated by altering gene expression levels involved in the depicted pathway.

Discussion

The prediction of functional residues is a well-developed field (32), where conservation of each residue in a protein is considered a key factor to rely on. Tools like ConSurf (23) and the ET-like methods (33) are able to identify slowly evolving positions that are involved in folding, interaction, or catalytic activity of protein (32). Though, the specific reason why a residue is conserved remains often unclear. In this work, we show that the conservation score in the structural alignment of paralog sequences, combined with the orthologs alignment conservation score is a promising way to identify important residues that affects the downstream pathway and cellular behavior.

Cross-conservation analysis is based on the following assumptions: 1) Orthologs evolutionary alignment conservation shows whether a residue is important for either structural or functional reasons. 2) Ligands alignment conservation scores denote the importance of a residue for receptor binding (the main shared property of all ligands). In our analysis we rely on these two assumptions to conclude that highly conserved residues in the evolutionary alignment (Figure C in Figure S1) that are not conserved in the ligand alignment (Figure A in Figure S1) have ligand-specific relevance related to their function.

It is interesting to note that conserved positions in the paralogs alignments are a subset of those conserved in the orthologs alignment. By subtracting the first ones, positions with a shared function across all ligands are filtered out. Then, Cross-conservation analysis overcomes the limitations of previous methods and highlights ligand-specific functional residues.

In particular, the two tryptophan in positions 49 and 50 are strong outliers on our bioinformatics analysis (Figure 1B). Their score is high even when using conservation measures that do not take amino acid type change into account (data not shown). We

specifically chose W50 for further testing because of its outward facing position, on the hypothesis that it might mediate previously unknown interactions. Mutant W50Y did not bind stronger than WT EGF but it showed a cellular effect. This result further highlights the importance of the disordered C-tail in the signal transduction.

Biochemical characterization of ligand-receptor interaction was done in solution using ITC. The experiment was performed using the soluble extra-cellular domain of the receptor followed by the cellular experiments. Tested mutants have binding affinities similar to that of WT EGF (Figure 2) except mutant N32R which showed slightly higher affinity. N32R was the only position chosen which is not on the C-terminus tail. Interestingly, while retaining the same binding affinity to the receptor, all mutants showed a different cellular effect on Bj-5ta cells, compared to EGF. A delayed proliferation response compared to control might appear as counterintuitive for fibroblast (8). Though, Bj-5ta hTERT immortalized fibroblasts show features that distinguish them from in vivo fibroblast, like differential gene expression and epiregulin-dependent proliferation (34). Competition between endogenously synthesized epiregulin and EGF might generate a different balance in the receptor stabilization, therefore leading to a decreased cell proliferation after EGF treatment, as shown from the cell culture videos in the Supplementary Materials.

A431 constitutionally express EGFR in high levels. Treatment with EGF has been observed to promote STAT1 dependent apoptosis (35). This pathway is sustained by the internalization of the ligand-receptor complex by the endocytosis process, a key factor in the ligand induced biased signaling (13). In cancer cell line A431, the cells treated with mutants show a higher growth rate, therefore a decreased rate in apoptosis compared to WT. Given the setting of the experiment, modularity in the endocytosis pathway is a probable explanation.

Interestingly, K48T treated cells showed an increase in long and thin processes during growth in all replicates (Figure 3). A possible hypothesis could be EGF-induced tubular formation, a further EGF induced pathway reported in intestinal epithelial cells (36).

In both experiments, the growth rate of the mutants is intermediate between the control and EGF, and concentration dependent. This phenomenon, and the cell-line specific response of growth hormones is consistent with previous literature reports (37).

Since the four mutations are located in the disordered C-terminus, we cannot infer whether they disrupt an important contact for EGF high-affinity binding. In fact, we observe the same binding affinity with the isolated ECD as WT EGF. However, we can assume that they might induce a different balance of the receptor dimerization, which then affect the downstream pathway. Previous reports have also identified the importance of the C-terminus for binding specificity (38) and binding affinity (39) as we did. Some studies have examined the effect of individual positions at the C-terminus of EGF (40), although nobody has investigated the residues reported herein. Though, other EGF mutants with a high cross-conservation score have been previously reported to alter the EGF effect (41-43).

We speculate that the mutations might induce a conformational change of the receptor that might affect interactions in the highly modulated endocytosis (44). Cancer treatment has focused on the EGF receptor and deactivation of its intra-cellular tyrosine-kinase (45). As the design of EGFR-based drugs remains complex, our study may support the hypothesis that the D46T and K48T EGF mutants can be used as templates to design anti-cancer drugs.

Author contributions

PL, SP, DM conceived and designed the project; SP developed and performed the computational models and analyses; DM performed the experiments; all authors analyzed the data and wrote the manuscript.

Acknowledgments

We thank Tadashi Yamamoto's Unit and Russel Alenton for technical help with the cell proliferation assay, and Keiko Kono's Unit for the technical help with the WB experiments. We thank Madhuri Gade, Hayato Yanagida, Mary Collins and Mirco Dindo for critical reading.

Funding

Financial support by the Okinawa Institute of Science and Technology to P.L. is gratefully acknowledged.

References

1. Wu JM, Vallenius T, Ovaska K, Westermarck J, Makela TP, Hautaniemi S. Integrated network analysis platform for protein-protein interactions. Nat Methods. 2009;6(1):75-7.
2. Mann M, Jensen ON. Proteomic analysis of post-translational modifications. Nat Biotechnol. 2003;21(3):255-61.
3. Rao VS, Srinivas K, Sujini GN, Kumar GN. Protein-protein interaction detection: methods and analysis. International journal of proteomics. 2014;2014:147648.
4. Laisney J, Braasch I, Walter RB, Meierjohann S, Schartl M. Lineage-specific co-evolution of the Egf receptor/ligand signaling system. BMC Evol Biol. 2010;10:16.
5. Gazdar AF. Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. Oncogene. 2009;28:S24-S31.
6. Cantor AJ, Shah NH, Kuriyan J. Deep mutational analysis reveals functional trade-offs in the sequences of EGFR autophosphorylation sites. Proc Natl Acad Sci U S A. 2018;115(31):E7303-E12.
7. Nicholson RI, Gee JM, Harper ME. EGFR and cancer prognosis. European journal of cancer (Oxford, England : 1990). 2001;37 Suppl 4:S9-15.
8. Carpenter G, Cohen S. Epidermal growth factor. Annual review of biochemistry. 1979;48:193-216.

9. Hardwicke J, Schmaljohann D, Boyce D, Thomas D. Epidermal growth factor therapy and wound healing--past, present and future perspectives. The surgeon : journal of the Royal Colleges of Surgeons of Edinburgh and Ireland. 2008;6(3):172-7.
10. Misale S, Yaeger R, Hobor S, Scala E, Janakiraman M, Liska D, et al. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. Nature. 2012;486(7404):532-6.
11. Evans BA, Sato M, Sarwar M, Hutchinson DS, Summers RJ. Ligand-directed signalling at beta-adrenoceptors. Br J Pharmacol. 2010;159(5):1022-38.
12. Kenakin T. Functional selectivity and biased receptor signaling. J Pharmacol Exp Ther. 2011;336(2):296-302.
13. Ali R, Brown W, Purdy SC, Davisson VJ, Wendt MK. Biased signaling downstream of epidermal growth factor receptor regulates proliferative versus apoptotic response to ligand. Cell Death Dis. 2018;9(10):976.
14. Groenen LC, Nice EC, Burgess AW. STRUCTURE-FUNCTION-RELATIONSHIPS FOR THE EGF/TGF-ALPHA FAMILY OF MITOGENS. Growth Factors. 1994;11(4):235-57.
15. Souriau C, Gracy J, Chiche L, Weill M. Direct selection of EGF mutants displayed on filamentous phage using cells overexpressing EGF receptor. Biol Chem. 1999;380(4):451-8.
16. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. Nucleic Acids Res. 2018;46(D1):D754-D61.
17. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013;30(4):772-80.
18. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000;28(1):235-42.
19. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci. 2018;27(1):14-25.
20. Okonechnikov K, Golosova O, Fursov M, Team U. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics. 2012;28(8):1166-7.
21. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2015;32(1):268-74.
22. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587-+.
23. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 2016;44(W1):W344-W50.
24. Freed DM, Bessman NJ, Kiyatkin A, Salazar-Cavazos E, Byrne PO, Moore JO, et al. EGFR Ligands Differentially Stabilize Receptor Dimers to Specify Signaling Kinetics. Cell. 2017;171(3):683-95.e18.
25. Puddicombe SM, Wood L, Chamberlin SG, Davies DE. The interaction of an epidermal growth factor/transforming growth factor alpha tail chimera with the human epidermal growth factor receptor reveals unexpected complexities. The Journal of biological chemistry. 1996;271(48):30392-7.
26. Wiedemann C, Bellstedt P, Gorlach M. CAPITO--a web server-based analysis and plotting tool for circular dichroism data. Bioinformatics. 2013;29(14):1750-7.
27. Mehrabi M, Mansouri K, Soleymani B, Hoseinkhani Z, Shahlaie M, Khodarahmi R. Development of a human epidermal growth factor derivative with EGFR-blocking and

depleted biological activities: A comparative in vitro study using EGFR-positive breast cancer cells. Int J Biol Macromol. 2017;103:275-85.
28. Panosa C, Tebar F, Ferrer-Batalle M, Fonge H, Seno M, Reilly RM, et al. Development of an epidermal growth factor derivative with EGFR blocking activity. PloS one. 2013;8(7):e69325.
29. Engler DA, Montelione GT, Niyogi SK. Human epidermal growth factor. Distinct roles of tyrosine 37 and arginine 41 in receptor binding as determined by site-directed mutagenesis and nuclear magnetic resonance spectroscopy. FEBS letters. 1990;271(1-2):47-50.
30. Mehrabi M, Mahdiuni H, Rasouli H, Mansouri K, Shahlaei M, Khodarahmi R. Comparative experimental/theoretical studies on the EGFR dimerization under the effect of EGF/EGF analogues binding: Highlighting the importance of EGF/EGFR interactions at site III interface. Int J Biol Macromol. 2018;115:401-17.
31. Knudsen SL, Mac AS, Henriksen L, van Deurs B, Grovdal LM. EGFR signaling patterns are regulated by its different ligands. Growth Factors. 2014;32(5):155-63.
32. Nemoto W, Saito A, Oikawa H. Recent advances in functional region prediction by using structural and evolutionary information - Remaining problems and future extensions. Computational and structural biotechnology journal. 2013;8:e201308007.
33. Lichtarge O, Bourne HR, Cohen FE. An evolutionary trace method defines binding surfaces common to protein families. J Mol Biol. 1996;257(2):342-58.
34. Lindvall C, Hou M, Komurasaki T, Zheng C, Henriksson M, Sedivy JM, et al. Molecular characterization of human telomerase reverse transcriptase-immortalized human fibroblasts by gene expression profiling: activation of the epiregulin gene. Cancer Res. 2003;63(8):1743-7.
35. Grudinkin PS, Zenin VV, Kropotov AV, Dorosh VN, Nikolsky NN. EGF-induced apoptosis in A431 cells is dependent on STAT1, but not on STAT3. Eur J Cell Biol. 2007;86(10):591-603.
36. Ibuka S, Matsumoto S, Fujii S, Kikuchi A. The P2Y(2) receptor promotes Wnt3a- and EGF-induced epithelial tubular formation by IEC6 cells by binding to integrins. J Cell Sci. 2015;128(11):2156-68.
37. Bjorkelund H, Gedda L, Andersson K. Comparing the epidermal growth factor interaction with four different cell lines: intriguing effects imply strong dependency of cellular context. PloS one. 2011;6(1):e16536.
38. Sato K, Nakamura T, Mizuguchi M, Miura K, Tada M, Aizawa T, et al. Solution structure of epiregulin and the effect of its C-terminal domain for receptor binding affinity. FEBS letters. 2003;553(3):232-8.
39. Wingens M, Walma T, van Ingen H, Stortelers C, van Leeuwen JE, van Zoelen EJ, et al. Structural analysis of an epidermal growth factor/transforming growth factor-alpha chimera with unique ErbB binding specificity. The Journal of biological chemistry. 2003;278(40):39114-23.
40. Matsunami RK, Yette ML, Stevens A, Niyogi SK. MUTATIONAL ANALYSIS OF LEUCINE-47 IN HUMAN EPIDERMAL GROWTH-FACTOR. J Cell Biochem. 1991;46(3):242-9.
41. van de Poll ML, Lenferink AE, van Vugt MJ, Jacobs JJ, Janssen JW, Joldersma M, et al. A single amino acid exchange, Arg-45 to Ala, generates an epidermal growth factor (EGF) mutant with high affinity for the chicken EGF receptor. The Journal of biological chemistry. 1995;270(38):22337-43.
42. Lahti JL, Lui BH, Beck SE, Lee SS, Ly DP, Longaker MT, et al. Engineered epidermal growth factor mutants with faster binding on-rates correlate with enhanced receptor activation. FEBS letters. 2011;585(8):1135-9.

43. Lui BH, Cochran JR, Swartz JR. Discovery of improved EGF agonists using a novel in vitro screening platform. J Mol Biol. 2011;413(2):406-15.
44. Bakker J, Spits M, Neefjes J, Berlin I. The EGFR odyssey – from activation to destruction in space and time. Journal of Cell Science. 2017;130(24):4087-96.
45. Sotelo MJ, García-Paredes B, Aguado C, Sastre J, Díaz-Rubio E. Role of cetuximab in first-line treatment of metastatic colorectal cancer. World journal of gastroenterology. 2014;20(15):4208-19.

Supplementary material

Figure S1. EGF Alignments.

(A) Multiple sequence alignment of the mature forms of EGF paralog. Protein were chosen for retaining binding to the EGF receptor in human. Sequences were obtained from ENSEMBL database and aligned with MAFFT software. On top, sequence logo representing the most conserved residue positions, obtained with WebLogo, UC Berkeley. (B) Multiple structure alignment of EGF paralogs. 1IVO pdb file containing crystallized EGFR (teal) and EGF was used as a template to align the structures of the other EGFR ligands. The ligands pdb files are: 2RNL – AREG; 1IP0 – BTC; 5WB8 – EPGN; 5WB7 – EREG; 1MOX – HBEGF; 1XDT – TGFA. All ligands residues are colored by conservation level (red high, blue low) in the MSTA. The superposition and coloring were performed using CHIMERA. (C) EGF orthologs evolutionary alignment. Orthologous sequences were obtained from ENSEMBL and filtered of duplicate sequences (eg. very closely related species). Fruitfly (*Drosophila melanogaster*) was used as outgroup. On top, bar plot representing the conservation value of each position. The conservation measures of these alignments have been used to make the cross-conservation plot.

S1 Table. Cross-conservation score of all EGF positions.

The table reports the cross-conservation score calculated using either the ligands MSA or the ligands MSTA. Highlighted in red, the position with the residues swap issue. In yellow, positions that were initially considered for mutagenesis. In green, the positions used in the further analysis. On the top right, conservation score plot by amino acid position. It is interesting to note how the highest peaks are concentrated in the C-terminus part.

Figure S2. Highlight on the EGF-specific residue swapping problem.

EGF is colored in teal, while HB-EGF is colored in white. The tyrosine in EGF position 29 (29Y) is not found on the corresponding position of HB-EGF. Instead, a proline is found (130P). Though, the interaction between the two residue is probably maintained because, in place of the histidine 10 (10H) of EGF, HB-EGF shows a tyrosine (112Y). When this type of residue swapping is ligand-specific, the lowering of the conservation measures over-estimates the real change happening in that position.

Figure S3. EGFR ligands evolutionary tree.

The tree was obtained using Maximum Likelihood algorithm. *C elegans* EGF was used as outgroup. The EGFR ligands cluster neatly. Interestingly, fish ligands are found to have longer branches on average. The tree was made using iqtree ModelFinder. The best-fit model used is JTT+R6.

Figure S4. The validation of ordered structure of WT EGF and variants.

0.2 mg/ml of the sample was added to 10 µm cuvette and analyzed by Circular Dichroism (CD) spectrometer J-1000. All samples showed the presence of a comparable amount of β-sheet. In contrast, K48T and W50Y had another peak at 215 nm. The CD of EGF was found similar to previously published results. The plot and analysis was made using CAPITO web tool (26).

Figure S5. Crystalized structure of WT EGF with the highlighted position N32, In the interaction with the position Q16 and G18 of the EGFR receptor backbone.

Figure S6. Effect of 1nM, 10nM and 100nM concentration of EGF variants on the proliferation of mouse normal fibroblast Albino Swiss 3T3 cell line.

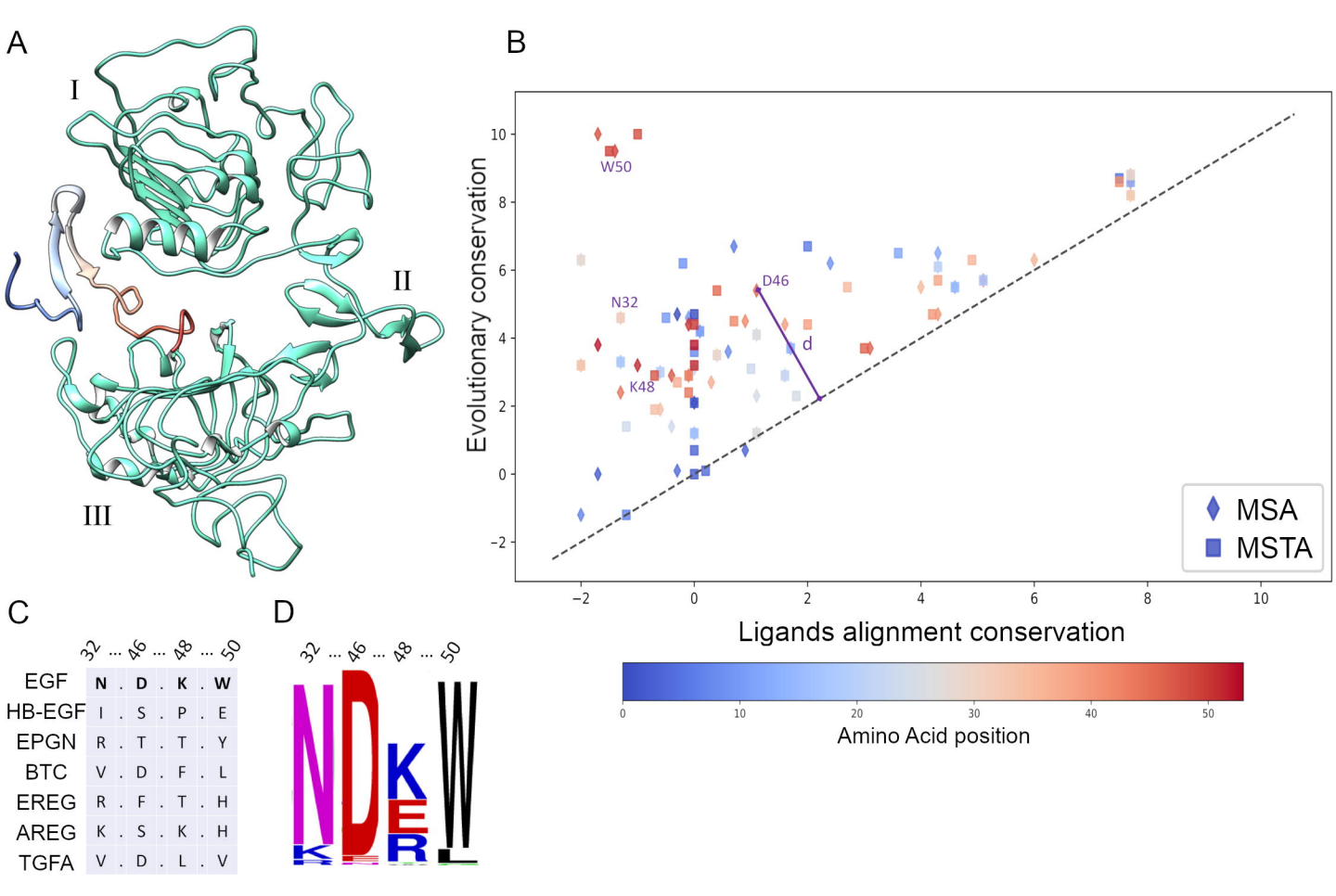
The data represent the relative growth percentage to control based on the confluence of cells (mean ± standard deviation) for each concentration of EGF.

Figure S7. Effect of 1nM, 10nM and 100nM concentration of EGF variants on the proliferation of human normal fibroblast bj5-tα cell line.

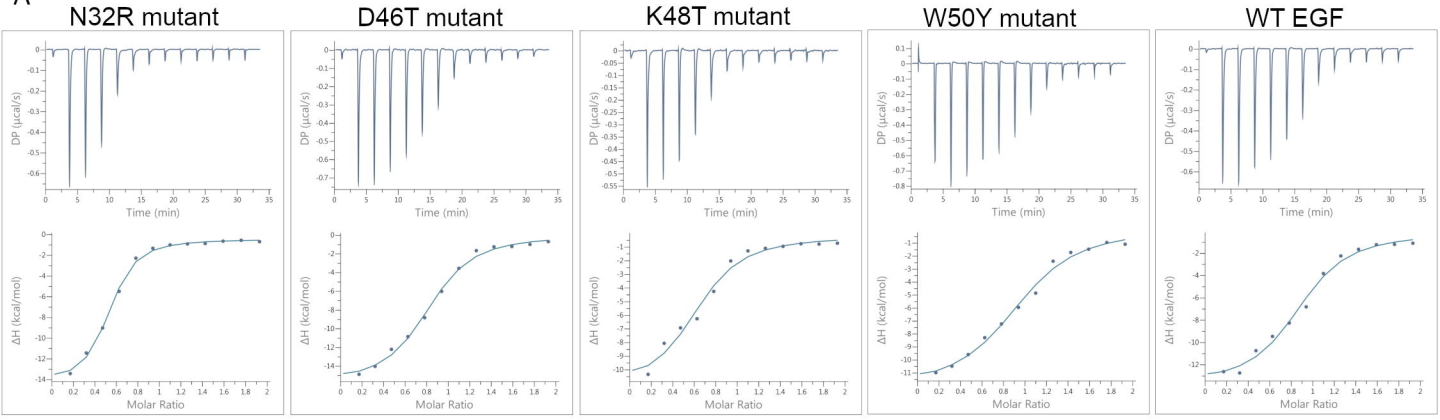
The data represent the relative growth percentage to control based on the confluence of cells (mean ± standard deviation) for each concentration of EGF. The percentage confluence was estimated 48h after the treatment (two replicates/treatment).

Figure S8. Effect of 1nM and 10nM concentration of EGF variants on the proliferation of A431 skin cancer cells.

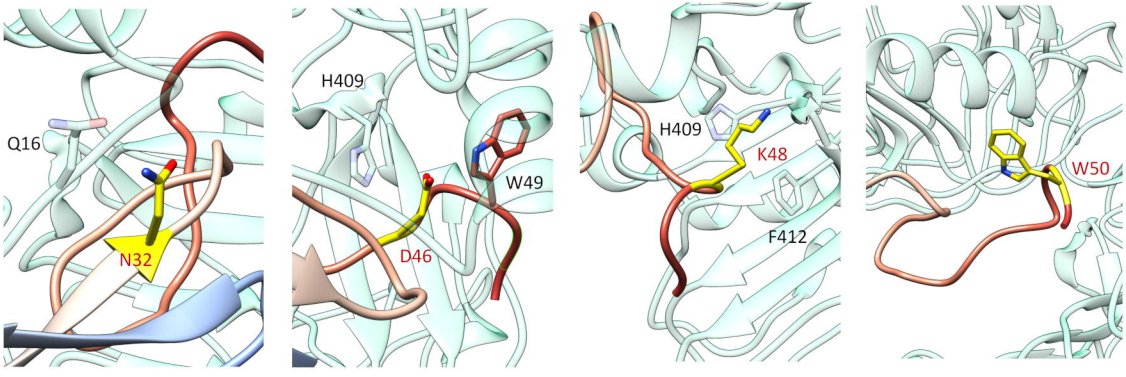
The data represent the relative growth percentage to control based on the confluence of cells (mean ± standard deviation) for each concentration of EGF variants. The percentage confluence was estimated 48h after the treatment (two replicates/treatment).


A
I
II
III
B
Evolutionary conservation
Ligands alignment conservation
W50
D46
N32
K48
d
MSA
MSTA
Amino Acid position
C
32 ... 46 ... 48 ... 50
EGF N . D . K . W
HB-EGF I . S . P . E
EPGN R . T . T . Y
BTC V . D . F . L
EREG R . F . T . H
AREG K . S . K . H
TGFA V . D . L . V
D
32 ... 46 ... 48 ... 50

A



B

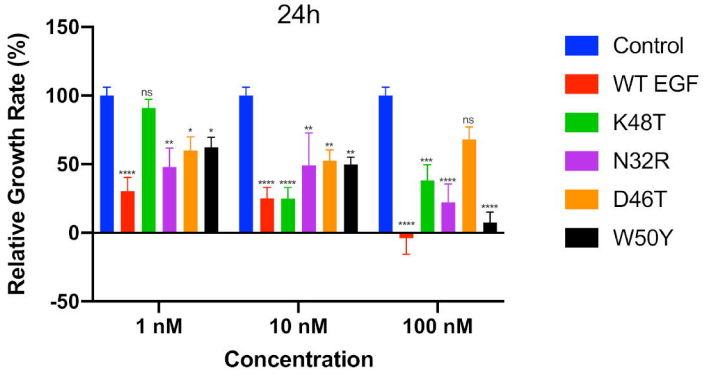


C

	K_d (M)
WT EGF	$1.28e^{-06} \pm 488e^{-09}$
N32R	$6.32e^{-07} \pm 125e^{-09}$
D46T	$1.12e^{-06} \pm 274e^{-09}$
K48T	$1.62e^{-06} \pm 888e^{-09}$
W50Y	$1.86e^{-06} \pm 665e^{-09}$

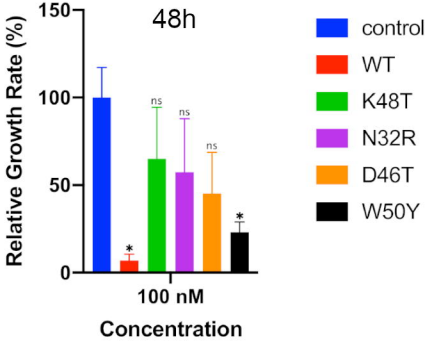
A

Bj5t-α Human Normal Fibroblast Cell Growth



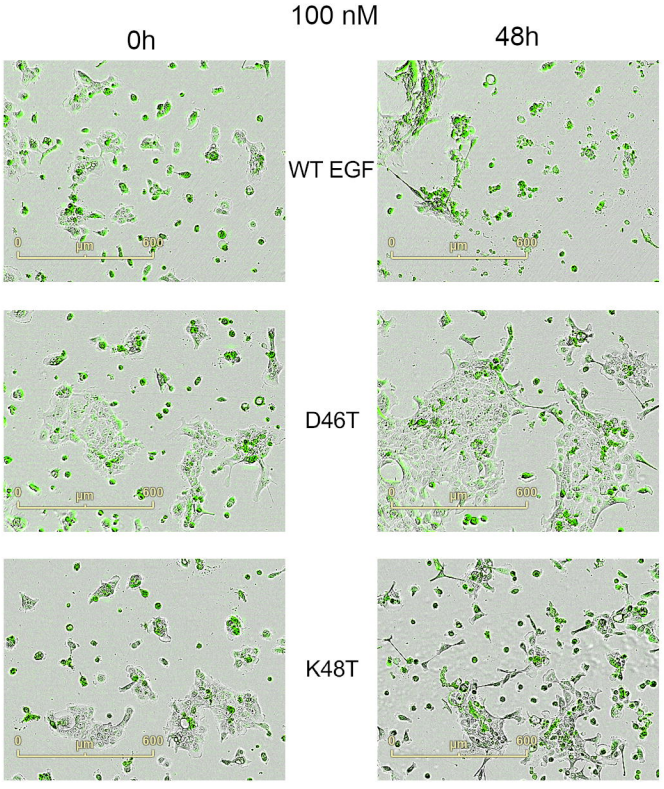
B

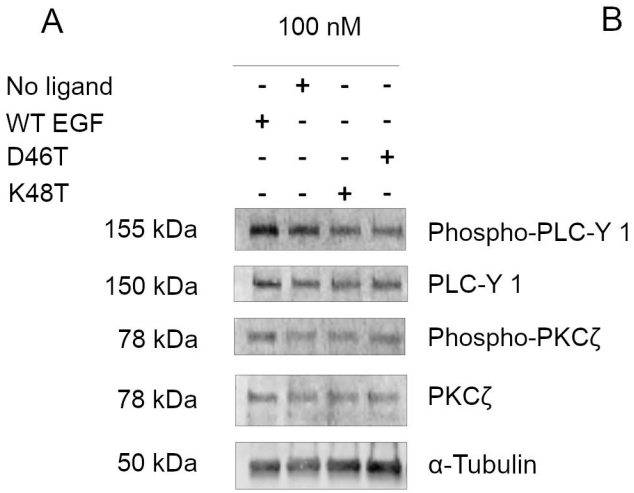
A431 Human Cancer Cell Growth

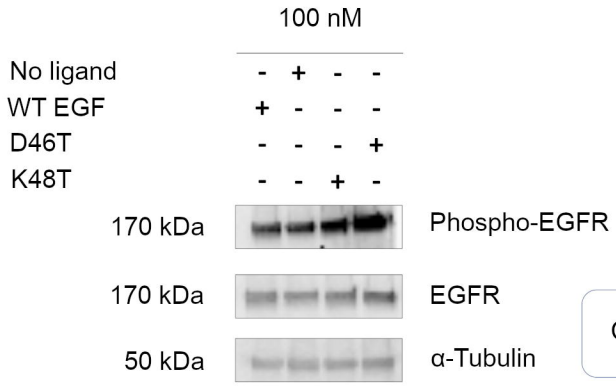


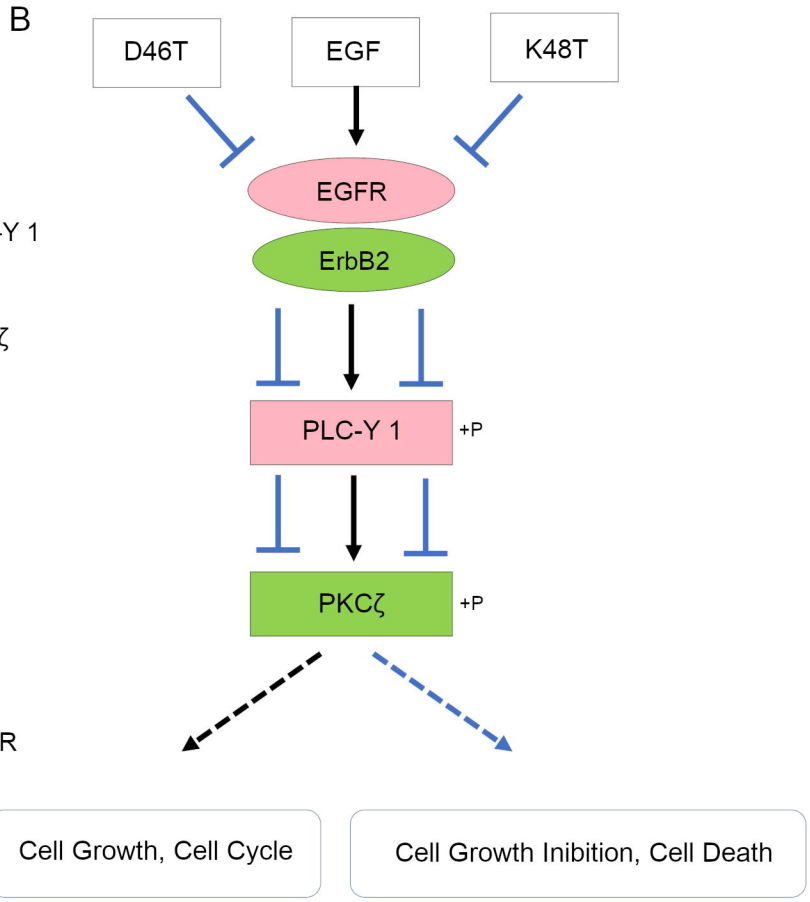
C

A431 Human Cancer Cells




A
100 nM
No ligand - + - -
WT EGF + - - -
D46T - - - +
K48T - - + -
155 kDa
Phospho-PLC-Y 1
150 kDa
PLC-Y 1
78 kDa
Phospho-PKCζ
78 kDa
PKCζ
50 kDa
α-Tubulin


100 nM
No ligand - + - -
WT EGF + - - -
D46T - - - +
K48T - - + -
170 kDa
Phospho-EGFR
170 kDa
EGFR
50 kDa
α-Tubulin


B
D46T
EGF
K48T
EGFR
ErbB2
PLC-Y 1
+P
PKCζ
+P
Cell Growth, Cell Cycle
Cell Growth Inibition, Cell Death