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

or

Differential effect of EGF mutants on EGFR pathway

Merzhakupova, D. [#]**; Pascarelli, S.** [#]**; Laurino P.***

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

[*] Corresponding author (paola.laurino@oist.jp).

[#] D.M. and S.P. equally contributed to this work.

Abstract

Molecular co-evolution is a key feature of biological systems. Molecular interactions (ligand-receptor, protein-protein, etc.) usually evolve simultaneously and independently to optimize binding. Frequently, these interactions involve one receptor that binds multiple ligands. Each ligand often leads to a different pathway activation intra-cellularly. Understanding single amino acid roles in evolving ligands and their contributions to downstream pathways of the receptor is still challenging.

We developed a cross-conservation approach to identify functionally important EGF residues. Four EGF mutants (N32R, D46R, K48T, W50Y) have been selected and studied biochemically and at the cellular level. While these mutants retain binding affinities for EGFR similar to that of EGF, surprisingly the effects of two of them (D46R, K48T) at the cellular level changed, inducing higher proliferation levels in normal fibroblasts and reducing proliferation in skin cancer cells. These results lay the base to understand the basis of EGF signaling.

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 essential amino acids for understanding ligand-receptor interactions but still the ligand effect at the cellular level remain 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).

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 have been shown 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 (11). Another study highlighted Y13, L15 and H16 residues as essential for downstream activity of ErbB1 (12). These outcomes were based on structural analyses of ligands and experimental validation.

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 (Figure 1A). This method allows us to analyze and characterize evolutionarily conserved EGF residues and to determine which residues are conserved among

77 different ligands. Furthermore, the identified residues have phenotypic implications at the
78 cellular level, influencing protein activation in the EGFR downstream pathway. Overall, this
79 approach has been critical to identify residues that play important role in cellular proliferation
80 and cancer cells.
81
82 **Results**
83 Initially, we identified amino acid residues in EGF that are essential for protein-coevolution
84 and mitogenic activity of the ligand using what we named cross-conservation approach.
85 Cross-conservation analysis highlights functional positions in proteins, based upon previous
86 knowledge of their interacting partners. The combination of two residue conservation
87 measures generated a cross-conservation plot (Figure 1B), as combination of the alignment of
88 all the ligands (paralogs that still share binding to same receptor), using either the Multiple
89 Sequence Alignment (MSA, Figure A in S1 Figure) or the Multiple STructural Alignment
90 (MSTA, Figure B in S1 Figure), and from the alignment of orthologs sequences of EGFs
91 (herein so called evolutionary alignment, Figure C in S1 Figure).

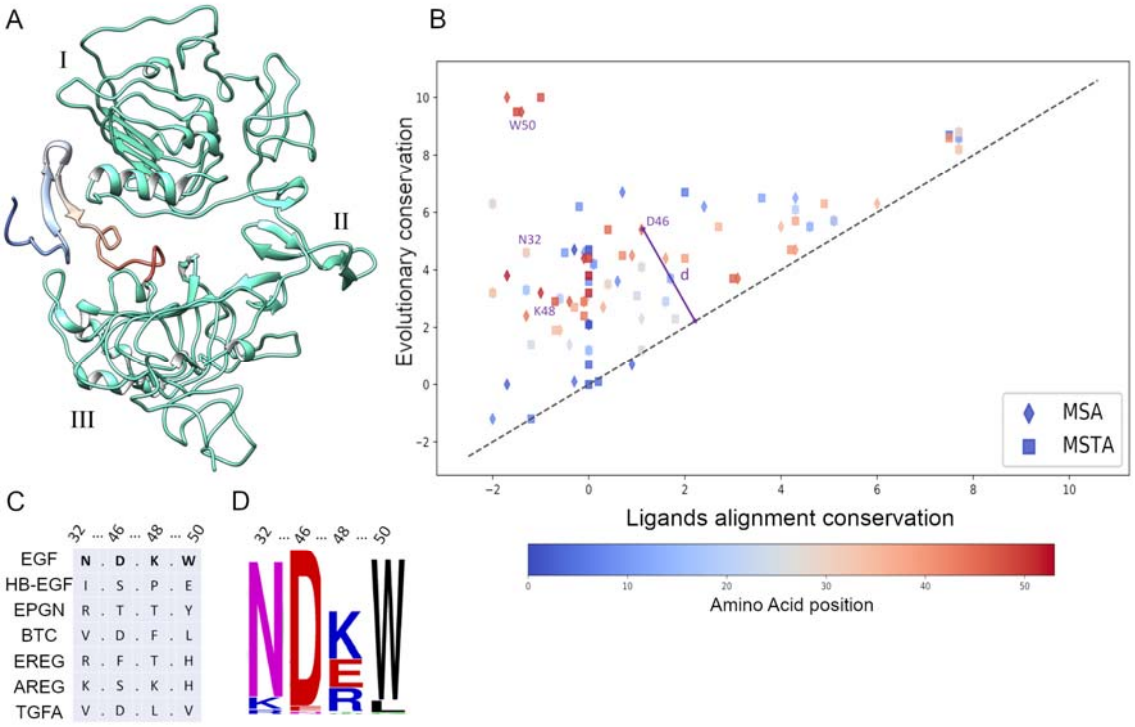


Figure 1. Cross-conservation analysis.

(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 (S1 Figure). 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 C in S1 Figure). 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. Along with the cross-conservation score, the choice of positions for mutation was influenced by three factors: first of all, the distance from the receptor. Secondly, we considered 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 taking into consideration the residues types in the ligands that show a different cellular effect. Finally, some of the residues that show high cross-conservation score have intramolecular interaction with other amino acid and, if mutated, they will not only change interaction with the receptor but also lose EGF structural stability (namely “residue swapping” behavior showed in S2 Figure**)**. A phylogeny of all EGFR ligands was also built **(**S3 Figure**)**, presenting a high degree of monophyly among the seven paralogs. This monophyly justify the comparison of different ligands in our cross-conservation study.

Based on these factors together with cross-conservation analysis, we designed EGF mutants with single amino-acid substitutions (N32R, D46T, K48T and W50Y). All these positions have higher than median cross conservation scores. Furthermore position 46, 48 and 50 were chosen because according to our study the EGF C-terminus tail seems to play a critical role in the ligand function. The amino acid mutation was selected according to its abundance in other ligands having different function as explained above. 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).

We characterized all of them biochemically and at the cellular level. Using circular dichroism (CD) experiments, we confirmed that the secondary structure of these mutants was maintained (S4 Figure). 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).

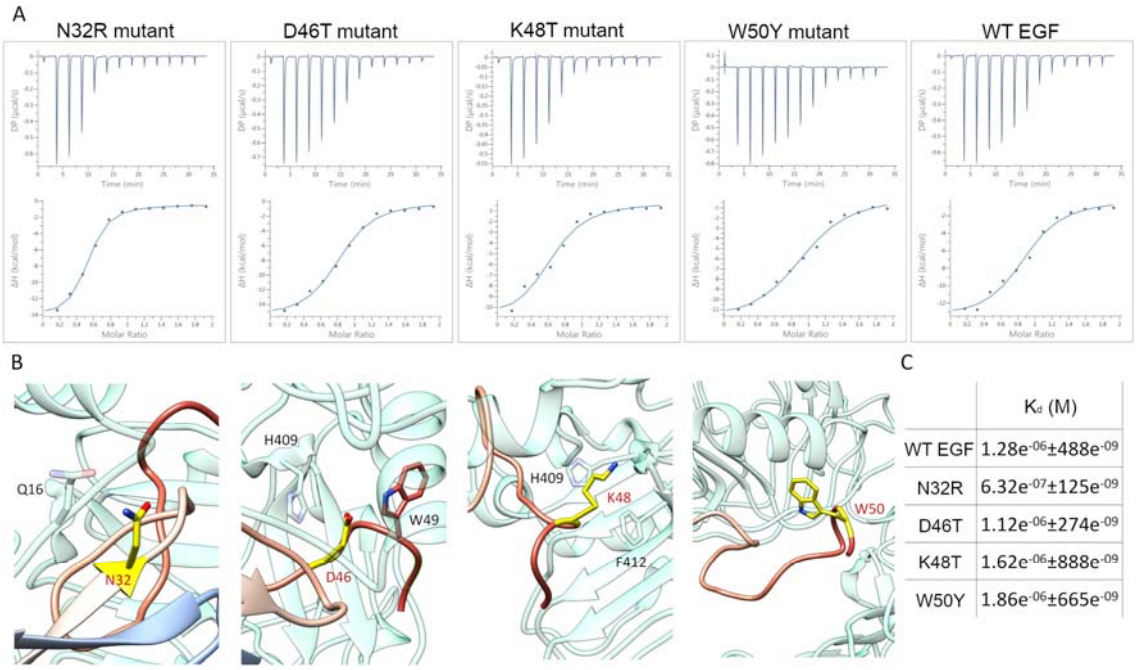
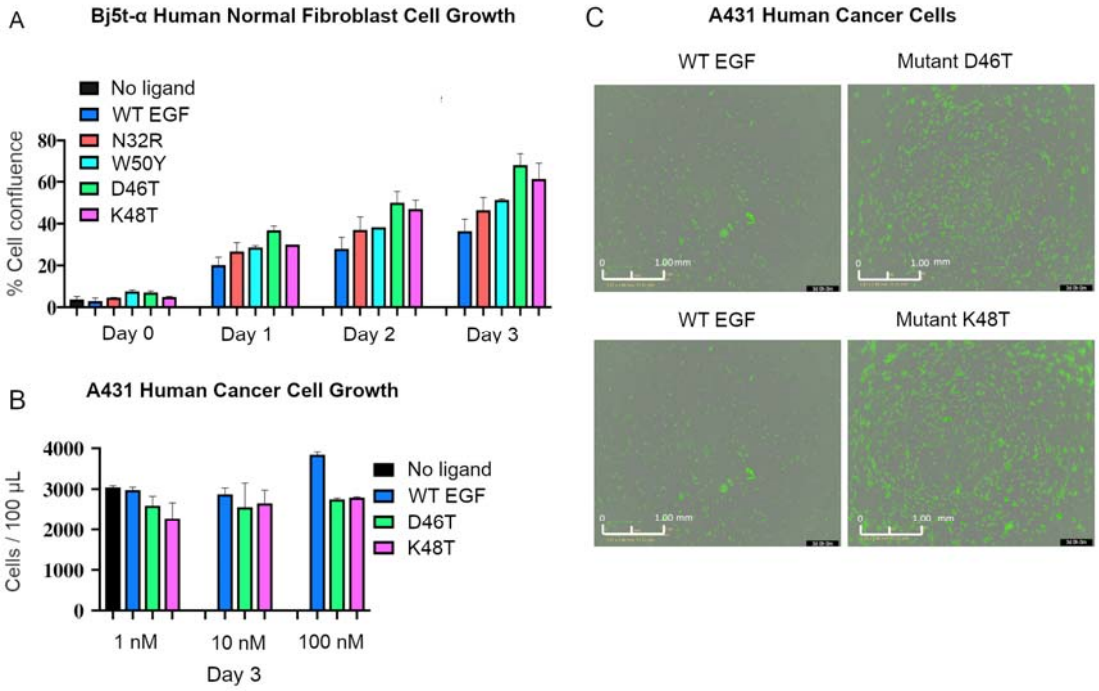


Figure 2. ITC measurements of EGF mutants and the EGFR receptor.

(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

147 mutated residues and in cyan stick side chain of the residue of ECD in proximity (< 5 Å with
148 the mutated residue. (C) K_d calculated from the ITC measurements using the program
149 Affinimeter KinITC Kintecs Software.
150
151 Mutation N32R is on the interface between ligand and receptor (S5 Figure). The slightly
152 higher affinity is probably due to the presence of the guanidinium group of R which is
153 positive charged and could interact with Q16 of EGFR ECD.
154
155 Surprisingly although the biochemical parameters are not substantially different, EGF
156 variants affected cell growth in cell proliferation studies. Human and mouse normal
157 fibroblasts, bj5-tα and Albino swiss 3T3, respectively, and epidermoid carcinoma A431 cell
158 lines, were cultured varying concentrations (1 nM, 10 nM and 100 nM) of wild-type EGF and
159 EGF mutants for three days. EGF mutants D46T and K48T induced cell proliferation in bj5-
160 tα (Figure 3A) more effectively than WT EGF, while no significant effect was detected on an
161 Albino Swiss mouse 3T3 cell line (S6 Figure).
162 We further tested these two mutants and importantly, we found both D46T and K48T
163 increased cell death in the A431 skin cancer cell line, upon 100 nM EGF mutants treatment
164 (Figure 3B). In contrast, 1 nM and 10 nM for both mutants only slightly reduced the number
165 of cancer cells (S7 Figure).



166

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

168 (A) Effect of different concentrations of EGF variants on proliferation of the human normal

169 fibroblast bj5-tα cell line. Data represent the percent confluence of cells (mean+- standard

170 deviation) for each concentration of EGF variants compared to data obtained with WT EGF

171 and negative controls. Percent confluence was estimated on day 3 (three replicates/treatment).

172 (B**)** Different concentrations of EGF mutants D46T and K48T inhibited the growth of A431

173 cells. Data represent the number of cells calculated on Day 3 (three replicates/treatment). (C**)**

174 Comparison of A431 cell growth after treatment with 100 nM WT EGF and EGF variants,

175 D46T and K48T. Dead cells were labeled with fluorescent annexin V green reagent. Plates

176 were pre-warmed prior to data acquisition to avoid condensation and expansion of the plate,

177 which affect autofocus. Images were captured every 2 hrs (4x) for 3 days in the IncuCyte

178 system.

179

Significantly herein we found that one amino acid change on the tail of EGF ligand could affect the downstream pathway. To explain this behavior, we proceed analyzing the downstream proteins involved in the EGFR network. Both D46T and K48T inhibited 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). A cell-line specific response upon interaction between EGFR and ligand (e.g. EGF induce proliferation in normal fibroblasts (8) while apoptosis in cancer ones (13)) is consistent with previous literature as well as a concentration dependent response (14).

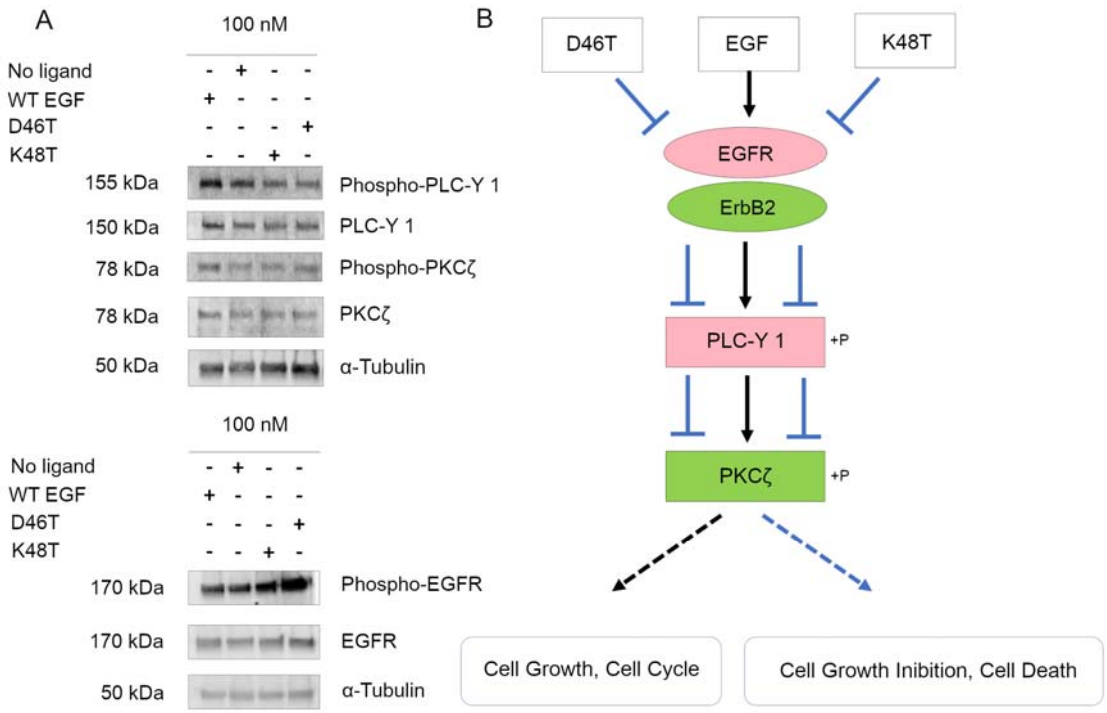


Figure 4. EGF variants D46T and K48T affect the EGFR downstream signaling pathway.

(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 (15), where conservation of each residue in a protein is considered a key factor to rely on. Tools like ConSurf (16) and the ET-like methods (17) are able to identify slowly evolving positions that are involved in folding, interaction, or catalytic activity of protein (15). 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.

Positions conserved in the paralogs alignments are a subset of those conserved in the orthologs alignment. By subtracting the first, 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 neither it showed a cellular effect. The distance of W50 from the receptor might be the reason of this result, from the structure it seems that intramolecular interaction can be favorite rather than interaction with the receptor.

Biochemical characterization of ligand-receptor interaction was done using ITC. 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. Its behavior in the binding affinity is different than other mutants and it has no detected effect at the cellular level, these results confirm the cross-conservation analysis output which highlight the importance of the C-terminus tail rather than the rest of the ligand. Interestingly, cells treated with mutants D46T and K48T show greater proliferation in the normal human fibroblast cell line and increased apoptosis in cancer cell lines (Figure 3A and B). Since these 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 by our data that they might induce some conformational change in the receptor which then affect the downstream pathway. Previous reports have also identified the importance of the C-terminus for binding specificity (18) and binding affinity (19) as we did. Few studies have examined the effect of individual positions at the C-terminus of EGF (20), although nobody has investigated the residues reported herein. We speculate that the mutations might induce a

conformational change of the receptor that might affect interactions in the highly modulated endocytosis (21).

Cancer treatment has focused on the EGF receptor and deactivation of the intra-cellular tyrosine-kinase (22). 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.

Acknowledgments

We thank Tadashi Yamamoto's Unit for technical help with the cell proliferation assay 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. Groenen LC, Nice EC, Burgess AW. STRUCTURE-FUNCTION-RELATIONSHIPS FOR THE EGF/TGF-ALPHA FAMILY OF MITOGENS. Growth Factors. 1994;11(4):235-57.

12. 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.

13. Gill GN, Lazar CS. Increased phosphotyrosine content and inhibition of proliferation in EGF-treated A431 cells. Nature. 1981;293(5830):305-7.

14. 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.

15. 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.

16. 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.

17. 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.

18. 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.

19. 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.

20. 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.

21. 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.

315 22. Sotelo MJ, García-Paredes B, Aguado C, Sastre J, Díaz-Rubio E. Role of cetuximab
316 in first-line treatment of metastatic colorectal cancer. World journal of gastroenterology.
317 2014;20(15):4208-19.
318 23. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl
319 2018. Nucleic Acids Res. 2018;46(D1):D754-D61.
320 24. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7:
321 Improvements in Performance and Usability. Mol Biol Evol. 2013;30(4):772-80.
322 25. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, et al. UCSF
323 ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci.
324 2018;27(1):14-25.
325 26. Okonechnikov K, Golosova O, Fursov M, Team U. Unipro UGENE: a unified
326 bioinformatics toolkit. Bioinformatics. 2012;28(8):1166-7.
327 27. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective
328 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol.
329 2015;32(1):268-74.
330 28. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS.
331 ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods.
332 2017;14(6):587-+.
333 29. Wiedemann C, Bellstedt P, Gorlach M. CAPITO--a web server-based analysis and
334 plotting tool for circular dichroism data. Bioinformatics. 2013;29(14):1750-7.
335
336

Materials and Methods

Sequence and structure analysis

Sequences of all ligands and the multiple sequence alignment of EGF orthologs were obtained using Ensembl (23). Multiple sequence alignment of all ligands was performed using MAFFT software with a built-in scan of optimal parameters (24). Structure images and alignments were created using Chimera (25).

Phylogenetic analysis

From the multiple sequence alignment of the ligand EGF from different species, very similar sequences were removed (mostly from monkeys). 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 were handled using unipro UGENE software (26). Three phylogenetic trees were made using Neighbor Joining (NJ), Maximum Likelihood (ML), and MrBayes (MrB) methods.

As additional method, trees were also made with IQTREE (27), using ModelFinder to scan for the most fit evolutionary model and parameters (28).

Calculation of Cross conservation score

From the evolutionary MSA and ligands MSA (or MSTA), two conservation measures were obtained. The conservation score was calculated in three ways: 1) identity score, 2) BLOSUM62 matrix score, and 3) JSDw score. Identity score measures the frequency of appearance of EGF residue in other ligands. In BLOSUM62, reference position substitutions are weighted using the BLOSUM62 matrix. JSDw is the method used in ConSurf paper (16), and is based on Jensen Shannon divergence, with a window of residues. The cross-

conservation plot and analysis were performed with the Python package SEABORN. The two conservation scores were plotted, and a cross-conservation score was obtained by computing the distance from the diagonal of the plot.

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**.**

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 S1 Figure) that are not conserved in the ligand alignment (Figure A in S1 Figure) have ligand-specific relevance related to their function.

The decision of which mutation to introduce was made using the ligand alignment. Overlapping residues at a given position were divided into two groups based on EGF-like and non EGF-like activation of the receptor. This separation was shown to follow binding affinity. Residues that introduced a noticeable shift in amino acid properties in the two groups were selected.

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 hydromycin 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^{o}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 an increase in percent confluence.

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.

Statistics

Proliferation and apoptosis experiments were replicated three times. All results are shown as the mean±s.d. Raw data was analyzed by multiple t-tests. Prism 8 software was used for statistical analysis.

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 (29). 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 PDFV membranes (gift from Cell Membranology Unit, OIST). 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.