s

Single EGF mutants unravel the mechanism for stabilization of Epidermal Growth Factor Receptor (EGFR) system

Stefano Pascarelli[1], Dalmira Merzhakupova[1], Gen-Ichiro Uechi[1], Paola Laurino[1*]

[1] Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna-son, Okinawa 904-0495 Japan.

* Corresponding author

E-mail: paola.laurino@oist.jp

Abstract

The Epidermal Growth Factor Receptor (EGFR) is a membrane-anchored tyrosine kinase that is able to respond to multiple extra-cellular stimuli in a selective way. Previous studies have pointed out that the modularity of this system could be carried out by ligand-induced differences in the stability of the dimerized receptor. However, nobody has ever explored this hypothesis by observing the effects of single-mutant ligands so far.

Herein, we generated single mutants after we identified positions responsible of inducing functional divergence among paralog ligands using a newly developed approach. Then, we assessed the mutants' effects on the receptor by employing a combination of molecular dynamics and experimental techniques. Although having comparable binding affinities for EGFR to the wild type, the mutants induced different responses at both the receptor and the cellular level. This study shows for the first time that a functional transition in EGFR can be imparted with as little as one targeted mutation on the ligand. These results also 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

43 **CD**, circular dichroism

44

45

46 **Introduction**

47

48 The Epidermal Growth Factor (EGF)-like domain ligand-receptor signaling system is involved
49 in many biological events in multicellular organisms (1). This system is constituted by one
50 receptor (EGFR) and seven distinct peptide ligands (EGF, HBEGF, EPGN, BTC, EREG,
51 AREG, TGFA). Upon binding to the receptor, these ligands can activate multiple intracellular
52 downstream pathways through an interaction network with a bow-tie structure (2). For all
53 ligands, binding induces a transition from monomer or inactive dimer to an active dimer state
54 (3, 4). However, different ligands are able to promote divergent outcomes, even at saturating
55 concentrations; thus, the mechanism responsible of the modular downstream pathway
56 activation is independent of ligand affinity or potency, and likely encompasses intrinsic effects
57 (5). It is well known that the EGFR system play a key role in cancer development. In particular,
58 some studies have shown that overexpression of EGFR or its ligands may induce different
59 types of cancer (6). A better understanding of the interaction between EGFR and its ligands
60 could lead to the development of targeted therapies (7).

61

62 Protein-protein interactions (PPIs), such as that between EGFR and its ligands, are some of the
63 most interesting and well-studied examples of molecular co-evolution in biological systems.
64 These interactions are sometimes defined by one part (receptor) that binds several counterparts
65 (ligands). In these instances, receptors and ligands experience different selective constraints,
66 where receptors tend to evolve more slowly due to the necessity of binding multiple ligands
67 (8). Furthermore, paralogs, proteins related by a duplication event, are less likely to retain the

68 same function as orthologs, proteins related by a speciation event (9). Then, paralogous ligands
69 rather than the receptor are a good candidate to test the functional divergence in the biased
70 signaling system. However, this approach is susceptible to indirect factors that get into play.
71 For example, not all ligands might have the same binding selectivity, thus resulting in the
72 activation of unrelated pathways. In this work, we decided to test single mutants of one ligand,
73 EGF, by identifying and modifying sites responsible for a divergent function among paralog
74 ligands. In recent years, different experimental techniques have been developed to define the
75 effects of single mutant proteins at the cellular level (10). Recent studies have shown that some
76 EGF residues like Arg-41 and Leu-47 are highly conserved and important for high binding
77 affinity to EGFR (11). Another study highlighted Tyr-13, Leu-15 and His-16 in EGF as
78 essential for downstream activity of ErbB1 (12). These outcomes were based on structural
79 analyses of ligands and experimental validation. While bioinformatic tools such as contact
80 prediction (13) or molecular dynamics (14) can give a good overall picture of ligand-receptor
81 interactions, the contribution of single ligand residues to the modularity of the system still
82 remains unclear.
83
84 Ligand-induced selective activation of downstream pathway has been observed in G-protein
85 coupled receptors, a phenomenon known as "biased signaling" (15, 16). A recent report hints
86 that this mechanism might also take place in EGFR tyrosine kinase (17). Initially, the main
87 contributor to the modularity of the system was thought to be the affinity of the ligands to the
88 receptor. However, the discovery of the ligand EPGN, which induces a potent mitogen effect
89 despite low binding affinity (18), and multiple cell-line studies (5) have changed this
90 perspective. The standing theory is that different EGFR ligands induce different receptor dimer
91 stabilities, altering the phosphorylation dynamics; stable receptor dimers show transient
92 activation, reaching the phosphorylation level required to initiate negative feedback signals

93 faster than the unstable counterparts (19). Now, the open question is how these ligands induce
94 different dimer stabilities. Interactions between the interface of the ligand and EGFR have been
95 shown to influence the stability of the dimerized receptor (20), a factor that is related to ligand-
96 specific signaling bias (19, 21). It is known that the EGFR dimer can be observed in a
97 symmetrical, "flush" conformation or an asymmetrical, "staggered" conformation (22)
98 depending on the presence/absence of the membrane, glycosylation, and the number of bound
99 ligands (23). Whether and how this plays a role in biased signaling is still unknown.
100
101 Here we show how single amino acid substitutions alter biased signaling in the EGF-EGFR
102 ligand-receptor system. We developed a new bioinformatic tool to analyze the co-evolution of
103 the ligand-receptor pair and identify candidate function-altering mutations. The identified
104 mutants induced an altered phosphorylation dynamics and different cellular phenotype. This is
105 the first study to explain differences in biased signaling of EGFR using single-residue EGF
106 mutants. Furthermore, our co-evolutionary analysis can be applied readily to other ligand-
107 receptor interactions.
108

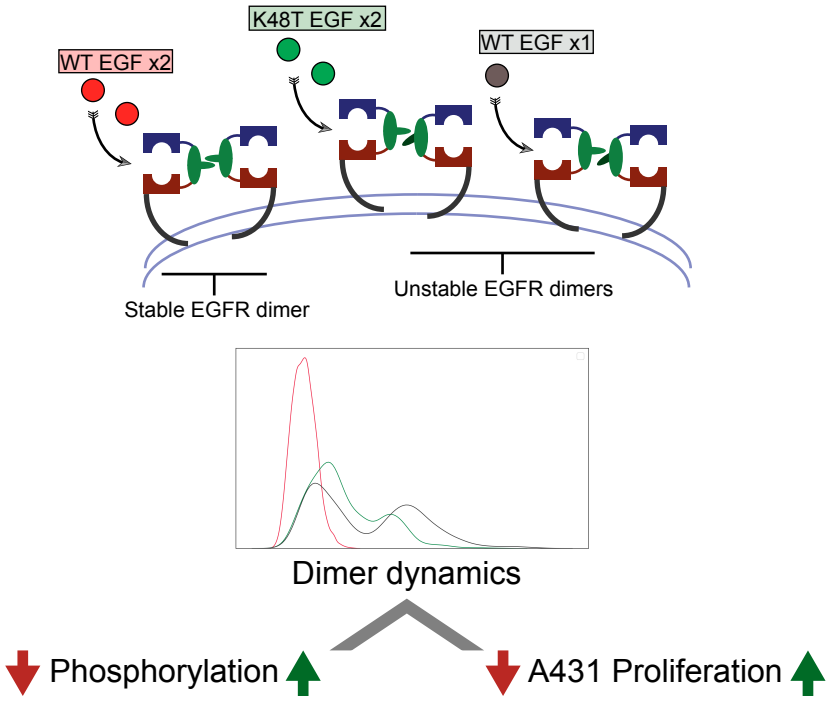


Figure 1. Study rationale. EGF K48T mutant was identified using an *ad hoc* methodology that combines conservation and coevolution measures. Unlike WT EGF, two K48T ligands induce an unstable conformation of the receptor dimer, similarly to a dimer with a single WT EGF. Treating A431 cells with either WT EGF or K48T mutant EGF resulted in a different response both in the phosphorylation dynamics and the cell proliferation phenotype.

Results

DIRpred

Firstly, we developed a method for predicting residues that are likely to be responsible for functional differences induced by paralogous ligands of a given receptor, which we called DIRpred (Divergence Inducing Residues prediction). Our approach combines residue-specific conservation measures to identify positions that are conserved among orthologs while diverging among paralogs. The DIRpred analysis is based on the assumption that conservation of a residue in orthologs of a specific ligand shows whether a residue is important for either

125 structural or functional reasons, while conservation of a residue among paralogous ligands
126 denotes the importance of a residue for receptor binding (the main shared property of all
127 ligands). Thus, residues that are highly conserved in orthologs but not paralogs of a ligand are
128 likely related to the ligand's specific function. Unlike other existing methods for prediction of
129 specificity determining residues (a review can be found in (24)), we included inter-protein
130 coevolution measures to narrow down those residues that are responsible for a specific function.
131 The DIRpred score is calculated as the sum of the four components (I: orthologs conservation,
132 II: complement of the paralog's conservation, III: ligand-receptor co-evolution, IV:
133 complement of ligand internal co-evolution). Optionally, the analysis can be conducted using
134 a structural alignment of the paralogs instead of the sequence alignment.
135
136 **DIRpred analysis of EGF**
137 We applied the DIRpred algorithm to predict single amino acid substitutions in EGF that would
138 alter its function (Figure 2). Since the analysis requires prior knowledge of paralogs and
139 orthologs of the target gene, we first performed a phylogenetic analysis to confirm that the
140 reported paralogous ligands of EGF are monophyletic (Supplementary figure 1). The DIRpred
141 analysis highlighted residue Asn-32, Asp-46, Lys-48 and Trp-50 as potential candidates for
142 paralog functional divergence. Asn-32 and Lys-48 show a very small conservation in both
143 sequence and structure alignments resulting in a high partial score (II). Asp-46 has a relatively
144 high coevolution with the receptor (III), while Trp-50 has high scores overall (Figure 2 —
145 figure Supplement 1). We chose to introduce orthologous substitutions at these four positions,
146 as orthologous have been previously observed as enriched in function (see Materials and
147 Metods) (25). The four designed EGF mutants with a single amino-acid substitution that were
148 selected for functional characterization were N32R, D46T, K48T and W50Y.

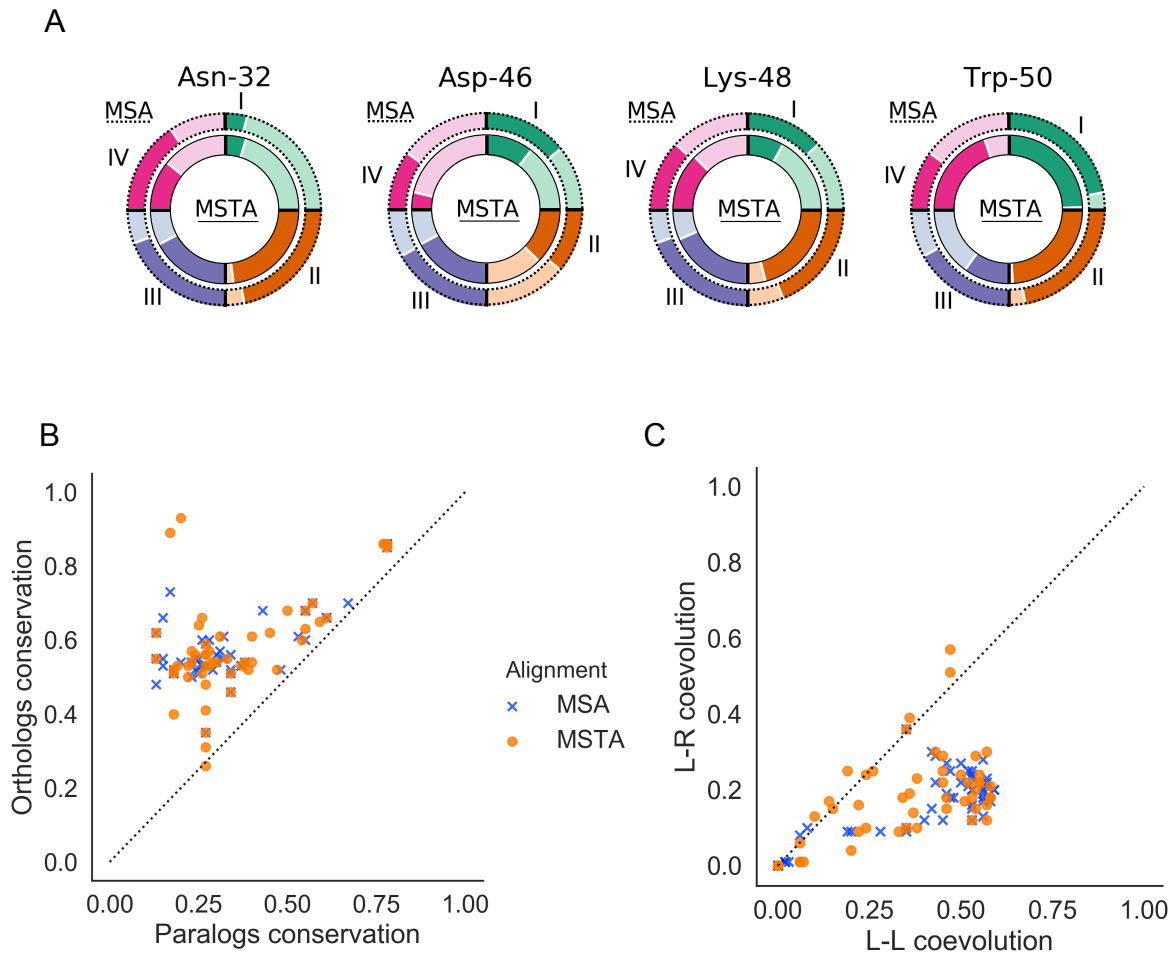


Figure 2. DIRpred analysis of EGF. (A) The relative ranking score of the four mutants. The outer circle represents paralogs MSA-based scores, while the inner circle represents paralogs MSTA-based scores. The darker color indicates a higher ranking in the EGF sites-based scoring. (B) Cross-conservation plot. The plot is obtained by crossing the two conservation scores. Interestingly, no point lies in the bottom right half of the plot (high paralogs conservation and low orthologs conservation), suggesting that paralogs and orthologs conservation are not independent. This observation points out that there is no organism-specific adaptation shared by all ligands at the protein sequence level. (C) Cross-co-evolution score. The plot is obtained by crossing the two co-evolution scores, the ligand-receptor (L-R) co-evolution score (III) on the y axis and the ligand-ligand (L-L) co-evolution score (IV) on the x axis.

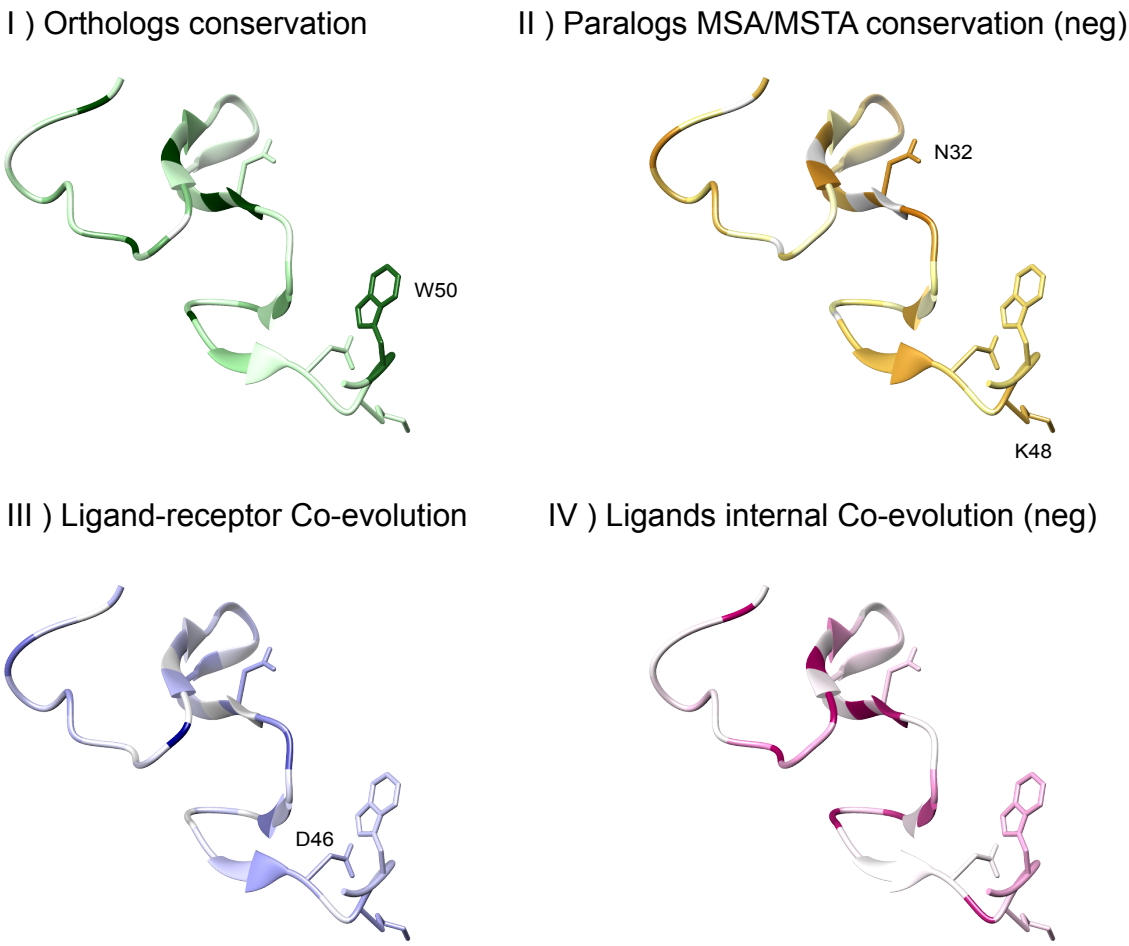


Figure 2—figure Supplement 1. DIRpred score. Structure of EGF (from PDB: 1IVO) colored by the four individual scores. The four sites chosen for mutation are shown in stick representation.

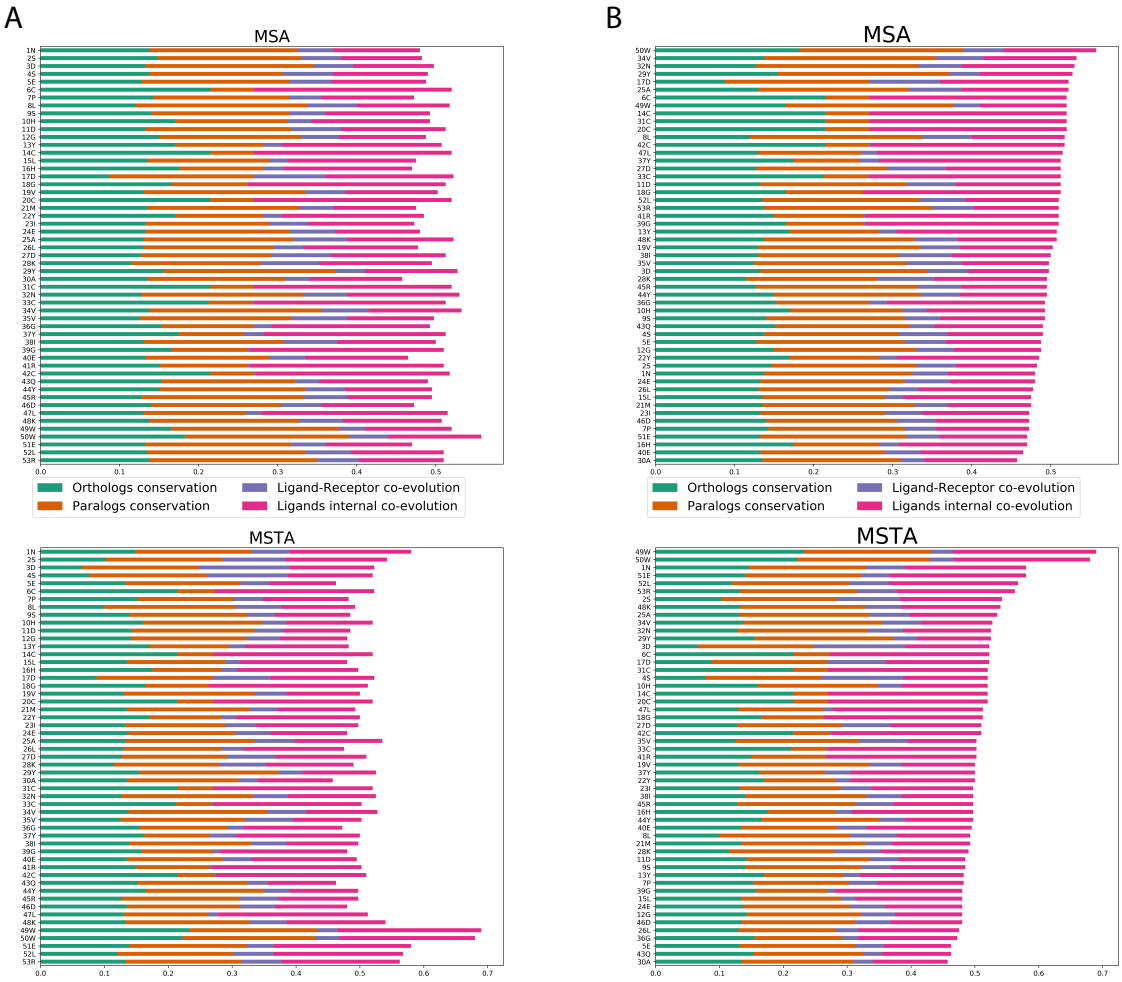
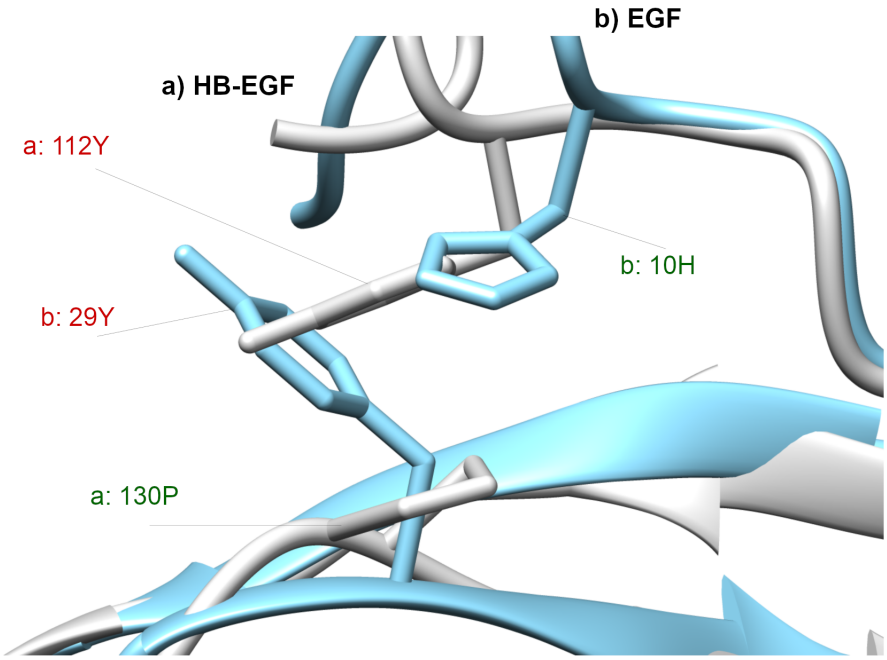


Figure 2—figure Supplement 2. All sites score. This figure shows the four scores for each site of EGF. On the top, the paralogs MSA based scores, while on bottom, scores are based on the paralogs MSTA. In (A) the scores are in order of sequence. In (B), the scores are sorted by highest sum first.



169

170 **Figure 2—figure Supplement 3. Highlight on the EGF-specific residue swapping problem.**
171 EGF is colored in teal, while HB-EGF is colored in white. The tyrosine in EGF position 29
172 (29Y) is not found on the corresponding position of HB-EGF. Instead, a proline is found (130P).
173 Though, the interaction between the two residue is probably maintained because, in place of
174 the histidine 10 (10H) of EGF, HB-EGF shows a tyrosine (112Y). When this type of residue
175 swapping is ligand-specific, the lowering of the conservation measures over-estimates the real
176 change happening in that position.

177

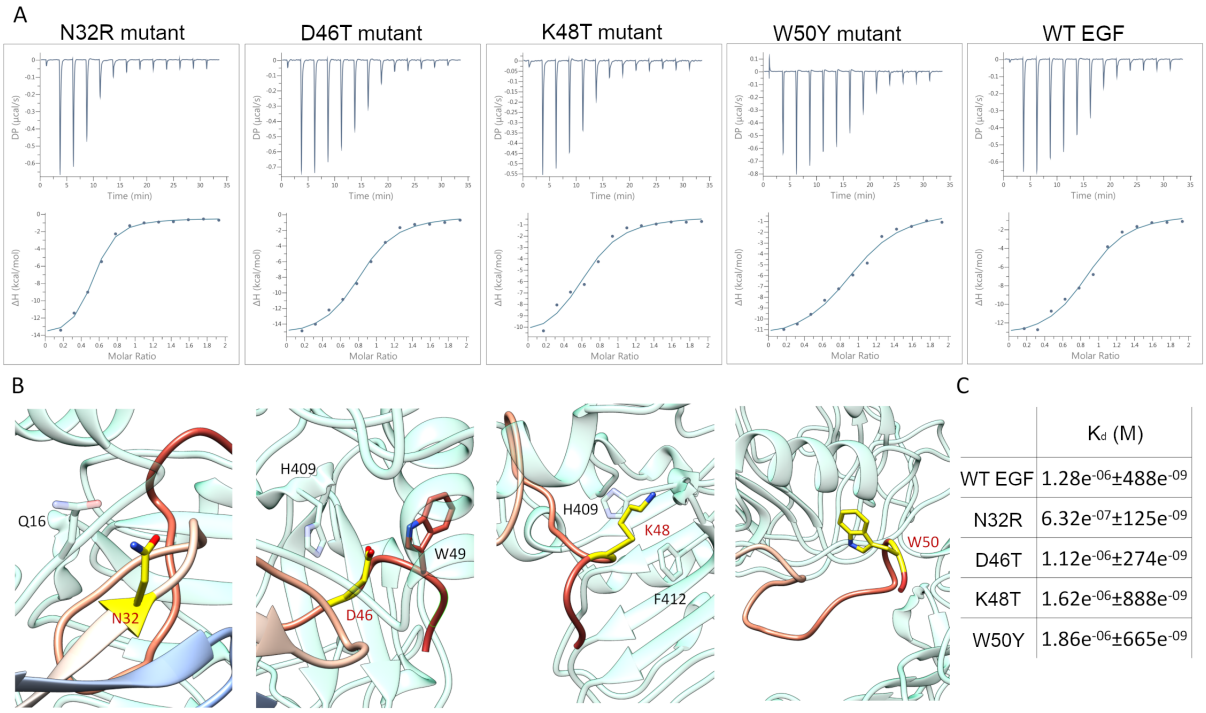
178 **Biochemical properties of the EGF mutants**

179 To determine the functional effects of these amino acid substitutions, we first performed *in-*
180 *vitro* analysis. Initially, Circular Dichroism (CD) spectroscopy was used to confirm that the
181 secondary structure of the mutants was maintained (Figure 3—figure Supplement 1). Then, the
182 K_d of each mutant for EGFR was determined by Isothermal Titration Calorimetry (ITC). ITC
183 measurements of the binding of all mutants to the ECD of EGFR exhibited similar K_d values

184 to the WT EGF. Only N32R has 2-fold higher affinity for EGFR compared to WT EGF (Figure

185 3C). All of the EGF mutants displayed a similar secondary structure and interaction with EGFR

186 in-vitro. Thus, the mutations do not appear to affect the receptor through a change in affinity.



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

187

188 **Figure 3. ITC measurements of EGF mutants and the EGFR receptor.** (A) ITC analysis

189 of WT EGF ligand and mutants N32R, D46T, K48T, and W50Y binding to the ECD of the

190 EGFR receptor at 25˚C. Measurements were taken by adding WT EGF at 200 μM to the ECD

191 of EGFR at 20 μM. (B) Four zoom-in of X-Ray structure of the extracellular domain (ECD) of

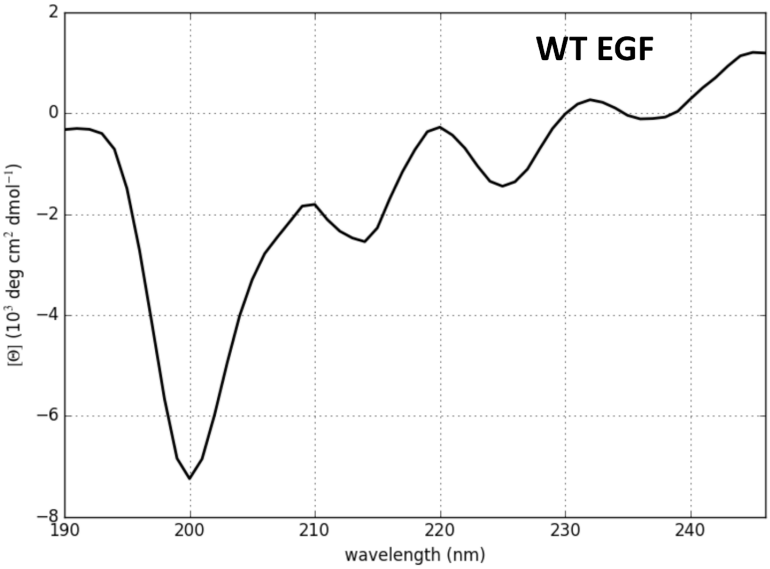
192 EGFR bound to EGF (PDB 1IVO). In cyan cartoon the ECD of EGFR. Each zoom-in focuses

193 on the mutated residue. Highlighted in yellow stick side chain of the mutated residues and in

194 cyan stick side chain of the residue of ECD in proximity (< 5 Å with the mutated residue. (C)

195 K_d calculated from the ITC measurements using the program Affinimeter KinITC Kintecs

196 Software.



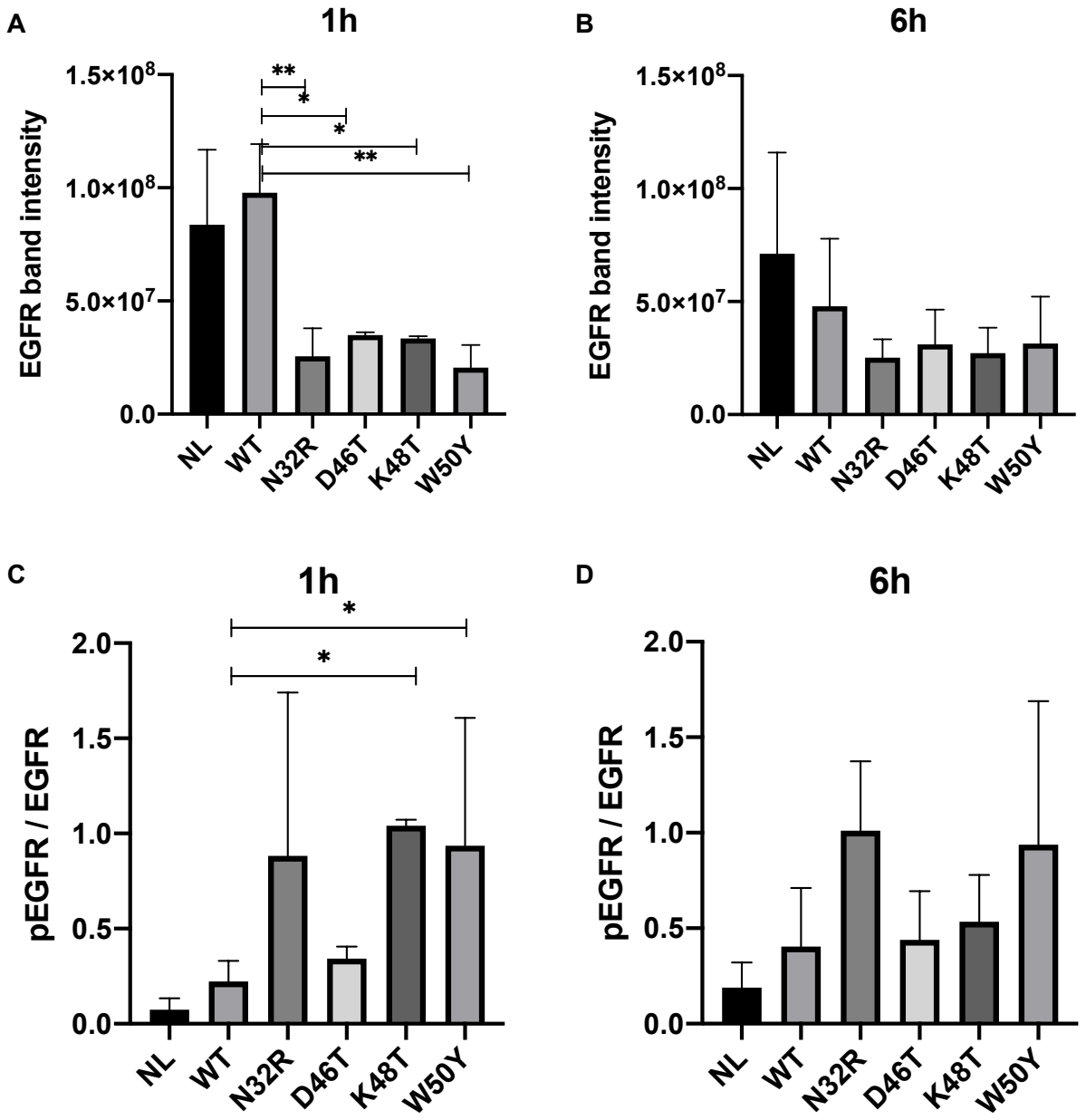
	% β-sheet
WT EGF	0.44
N32R	0.41
D46T	0.47
K48T	0.54
W50Y	0.52

Figure 3—figure Supplement 1. The validation of ordered structure of WT EGF and variants. 0.2 mg/mL of the sample was added to a 10 μm cuvette and analyzed by circular dichroism spectroscopy. All samples showed the presence of a comparable amount of β-sheet. The CD spectrum of EGF was similar to previously published results (26). The plot and analysis was made using CAPITO web tool (27).

Biological effects of the EGF mutants

Next, we sampled the effect of the EGF single mutants on EGFR expression and phosphorylation in A431 cells. A431 is a human epidermoid carcinoma cell line that overexpresses EGFR, thus making it easier to observe differences at the receptor level. Treating this cell line with very low (pM) concentrations of EGF has a small effect on proliferation, while high concentration treatments (>10 nM) induces apoptosis (28). We measured the amount of total and phosphorylated EGFR protein at multiple timesteps by Western Blot after treating with saturating concentrations (100 nM) of WT or mutant ligands (Figure 4).

212 Remarkably, K48T and W50Y showed a greater increase in EGFR phosphorylation 1h after
213 treatment than WT EGF (Figure 4—figure Supplement 1). Treatment with the mutant ligands,
214 but not the WT ligand, also caused a reduction in the amount of EGFR 1h after treatment. At
215 later time points, the effects converge back to the same range of the No Ligand (NL) sample
216 (Figure 4—figure Supplement 2). Thus, while the EGF mutants showed comparable binding
217 affinities to the receptor, at least two of the mutants showed different phosphorylation
218 dynamics compared to the WT ligand. Furthermore, the observed differences in the total
219 amount of EGFR protein indicate a possible alteration in the membrane expression or recycling
220 of the receptor.

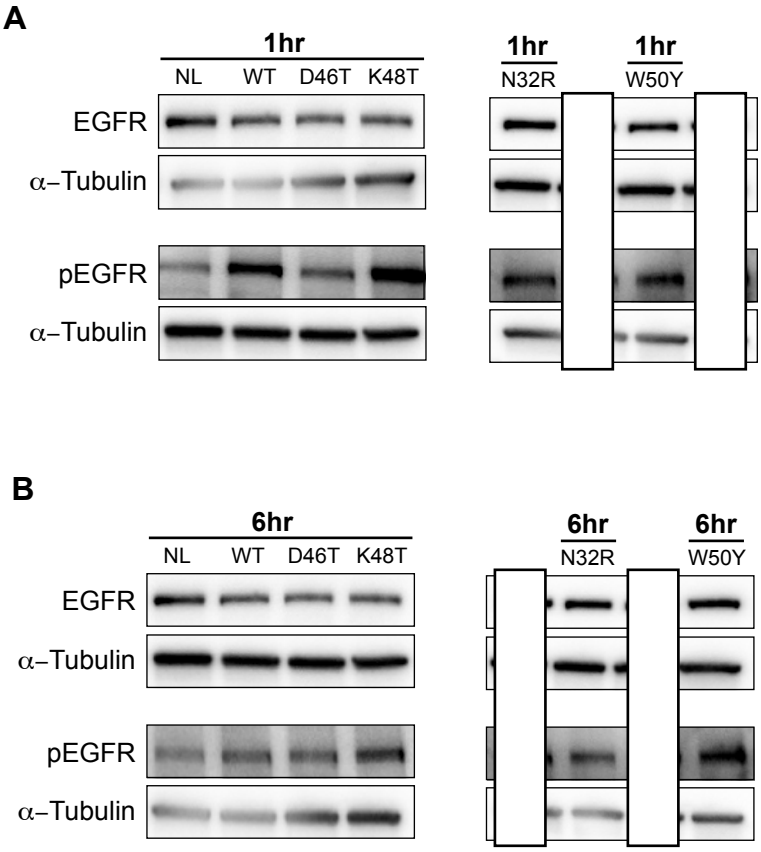


221

222 **Figure 4. Effects of WT and single-mutant EGF on the receptor in A431 cells**. (A-B)
223 Intensity of EGFR band after 1h (A) or 6h (B) treatment of saturating concentrations (100 nM)
224 of the different ligands. (C-D) Relative phosphorylation level, estimated as the ratio of pEGFR
225 (Tyr-1173) band intensity to EGFR band intensity, at 1 h (C) or 6 h (D). All intensities were
226 normalized using the respective alpha-tubulin bands. The bars represent mean ± s.d. (n = 2,

227 technical replicates). Band intensity estimates were calculated using BIO-RAD ImageLab

228 software (BioRad). Plots and statistics were performed using PRISM software (GraphPad).



229

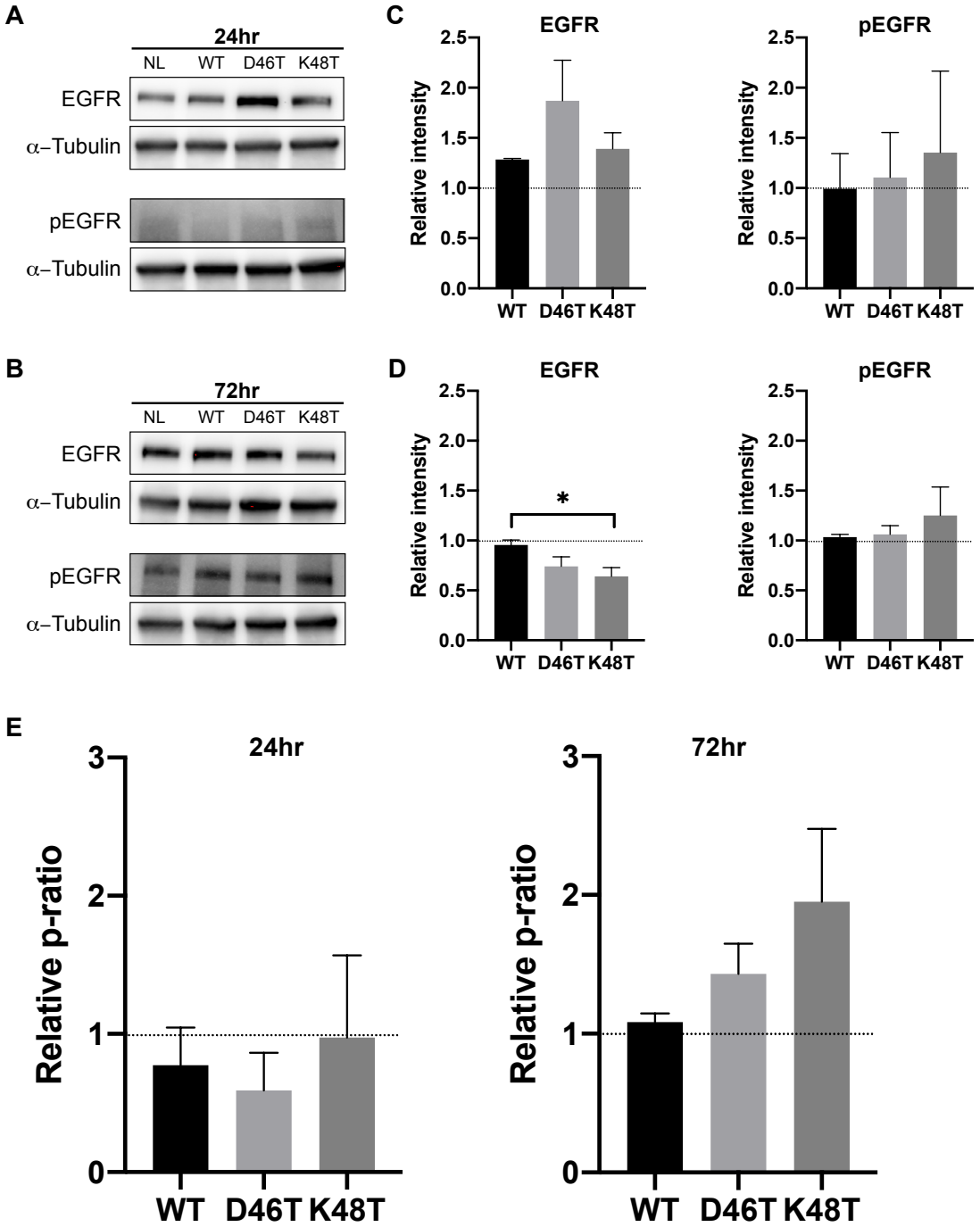
230 **Figure 4—figure Supplement 1. 1h and 6h post-treatment Western Blots**. Image of one

231 duplicate of A431 treated cells western blot for 1h (A) and 6h (B). The membrane containing

232 EGFR and tubulin was separated after the transfer. N32R and W50Y gel is the same in both

233 rows, but its lanes have been obscured to facilitate comparison at the same time steps with

234 other samples.



235

236 **Figure 4—figure Supplement 2. Effects on the receptor at 24h and 72h**. (A) Western blot
237 of A431 cells lysate after treatment with No Ligand (NL), WT EGF, D46T or K48T mutants
238 24h after treatment. (B) Image of the same experiment collected at 72h after treatment. (C)
239 EGFR and pEGFR (Tyr-1173) band intensities, relative to the NL sample at 24h and (D) at 72h
240 after treatment. In (E) and (F), the relative phosphorylation ratio compared to NL respectively
241 at 24h and 72h. A value of 1.0 represent the same phosphorylation ratio as observed in NL

242 lanes. All intensities were normalized using the respective alpha-tubulin bands. The plots were
243 obtained from duplicate of technical repetitions. Band intensity estimates were calculated using
244 BIO-RAD ImageLab (BioRad). Plots and statistics were performed using PRISM software
245 (GraphPad).
246
247 In order to observe the effects of the mutants on the cellular phenotype, we performed a cell
248 growth assay using an IncuCyte® live-cell analysis system. After one day of initial incubation,
249 we subjected Bj5-tα and A431 cells to 1, 10 or 100 nM concentration of growth factors.
250 Treating Bj5-tα fibroblasts with 100 nM WT EGF resulted in a reduction in the proliferation
251 rate compared to the control. The reduction in proliferation was accompanied by a change in
252 cellular morphology that could be a signal of differentiation (Supplementary data 1). However,
253 treatment with any of the mutants decreased the ability of EGF to suppress growth (Figure 5A).
254 The D46T and K48T mutants differed most from the WT (Figure 5—figure Supplement 1).
255 We observed a similar trend in the Albino Swiss mouse 3T3 cell line (Figure 5—figure
256 Supplement 2). In both cell lines, the effects are concentration dependent (Supplementary
257 figure 3). Next, we further tested the mutants in an apoptosis assay using the A431 cell lines.
258 This cell line is known to exhibit apoptosis and reduced growth rate when treated with EGF
259 peptide (29). 48h after treatment with 100 nM WT or mutant EGF, the WT showed the highest
260 decrease in cell proliferation, while the four mutants had an intermediate effect. Cells treated
261 with the W50T mutant had the closest proliferation level to the WT (Figure 5B). In WT EGF
262 cells, we observed evidence of apoptosis in globular processes. Meanwhile, K48T treated cells
263 displayed signs of cellular differentiation (Figure 5C). Significantly, these results show that
264 single amino acid changes in the EGF ligand display differential effects on the EGFR
265 transduction mechanism.

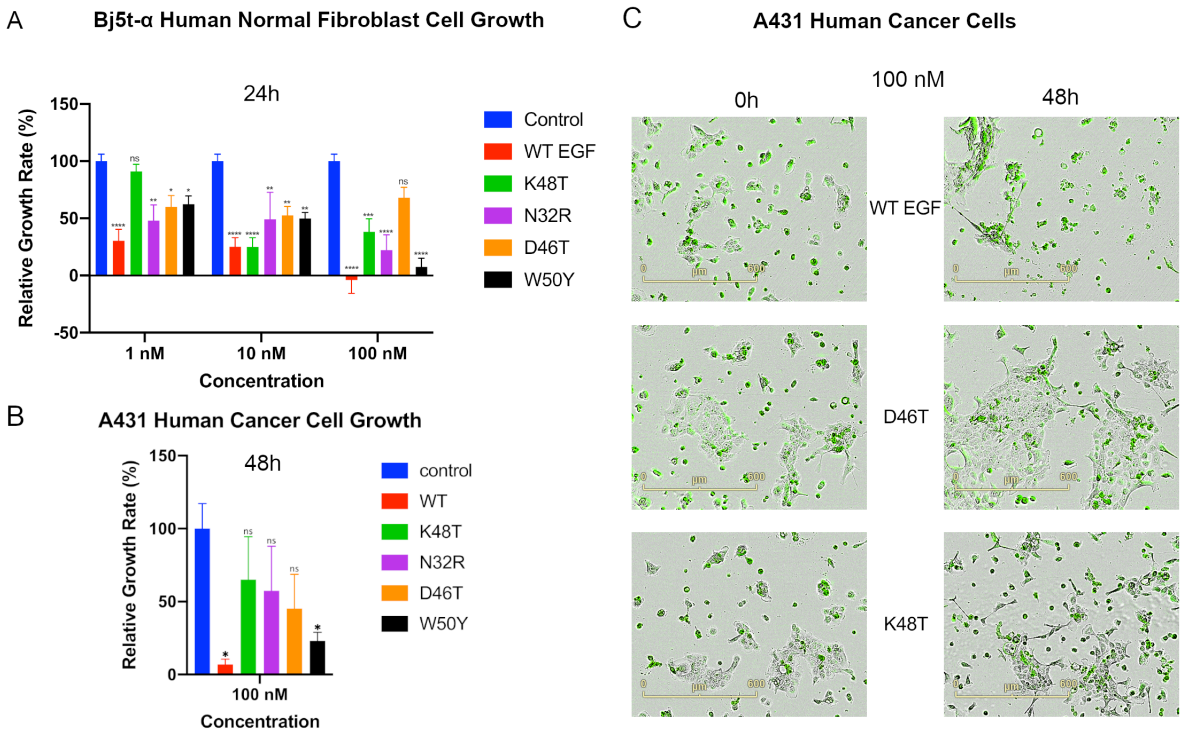


Figure 5. Growth assays of cells treated with EGF variants. (A) Effect of different concentrations of EGF variants on proliferation of the human normal fibroblast Bj5-tα cell line. Data represent the growth rate percentage calculated from the confluence of cells relative to the control (mean ± standard deviation) for each concentration of EGF. Percent confluence was estimated 24 h after the treatment (two replicates per 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 h (4x magnification) for 3 days in the IncuCyte system.

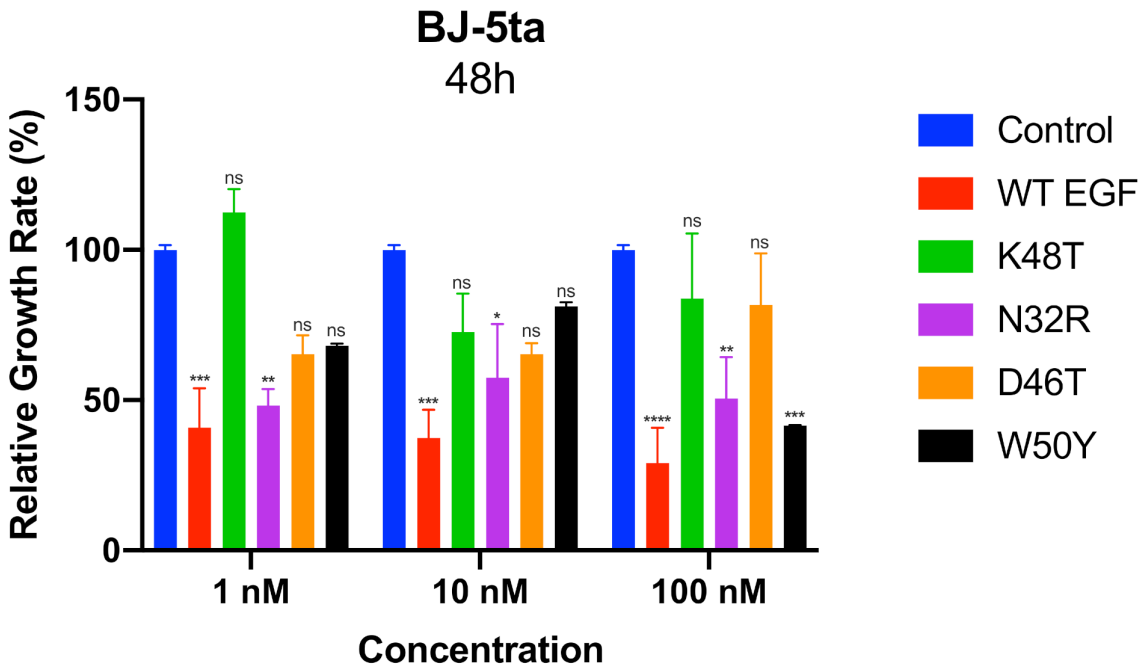


Figure 5—figure Supplement 1. Effect of 1nM, 10nM and 100nM concentrations of EGF variants on the proliferation of the human normal fibroblast Bj5-tα cell line. The data represent growth as a percentage of the control based on the confluence of cells (mean ± standard deviation) for each concentration of EGF. The percentage confluence was estimated 48 h after the treatment (two replicates per treatment).

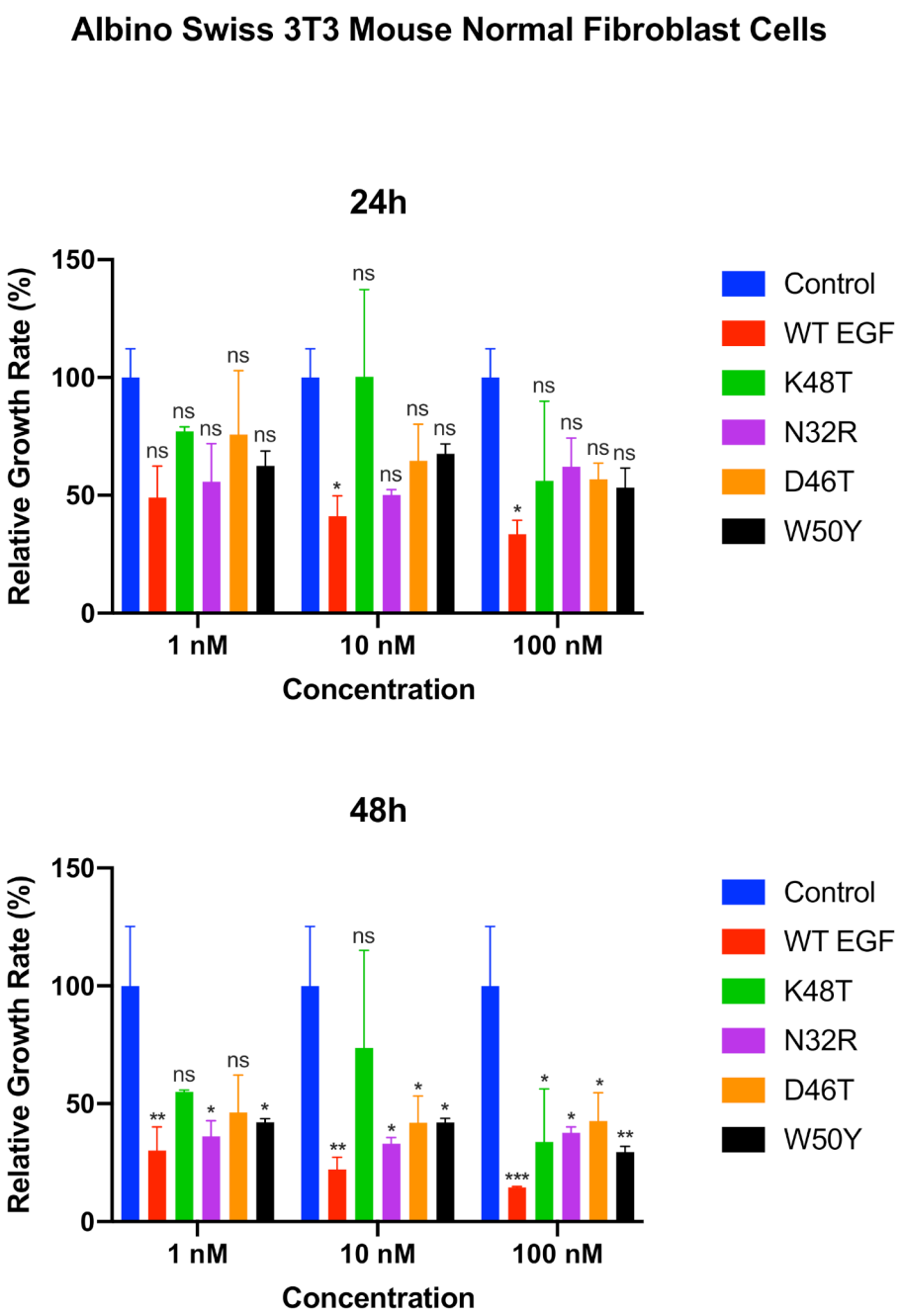


Figure 5—figure Supplement 2. Effect of 1nM, 10nM and 100nM concentrations of EGF variants on the proliferation of the mouse normal fibroblast Albino Swiss 3T3 cell line. The data represent the relative growth percentage to control based on the confluence of cells (mean ± standard deviation) for each concentration of EGF.

291 **Molecular dynamics**

292 To understand how EGF single mutants affect the receptor signaling transduction, we
293 performed full atomistic molecular dynamic (MD) simulations of the extracellular EGFR in
294 complex with WT or mutant ligands (100 ns each). We modeled the receptor starting from the
295 asymmetric, "unstable" conformation of 5WB7 (22). In this way, we expected to observe a fast
296 rearrangement for those simulations where a ligand more favorably induce a stable dimer. In
297 previous literature, an unstable conformation was observed when removing one of the EGF
298 ligands (23). We also performed one simulation of the EGFR dimer in complex with only one
299 WT ligand as a comparison (1ligEGF). The MD simulations quickly converged to a stable
300 RMSD (Figure 6A). For each simulation, we calculated the number of H-bonds between the
301 two receptor dimers, and between ligand and receptor. In the WT EGF simulation, the EGFR
302 dimer had an average of 15 H-bonds, higher than any other simulation (Figure 6—figure
303 Supplement 1A). However, the number of bonds between receptor and ligand was not altered
304 (Figure 6—figure Supplement 1B). In addition, during the course of our simulations we also
305 noted differences in the RMSF specifically located at the dimerization arm domain (Figure 6B).
306 The conformational space sampled by the dimerization arm of K48T simulation was much
307 wider than the WT simulation (Figure 6C). To analyze the temporal distribution of these
308 motion, we measured the distance between Pro-272 and Gly-288 of different EGFR monomers.
309 This distance was chosen because it was able to discriminate between the EGFR dimer in
310 complex with EREG (PDB ID 5WB7) and EGF (PDB ID 1IVO). In 5WB7, one of the two
311 distances is much bigger (~ 1.10 nm vs 0.4 nm) compared to 1IVO (Figure 6—figure
312 Supplement 1C). In our simulations, we observe WT showing a sharp peak at 1 nm, while the
313 mutants and 1ligEGF have a secondary peak at higher distances (Figure 6—figure Supplement
314 1D). Thus, the mutations in EGF appear to affect the stability of the EGFR dimer without
315 affecting the stability of the EGF-EGFR interaction.

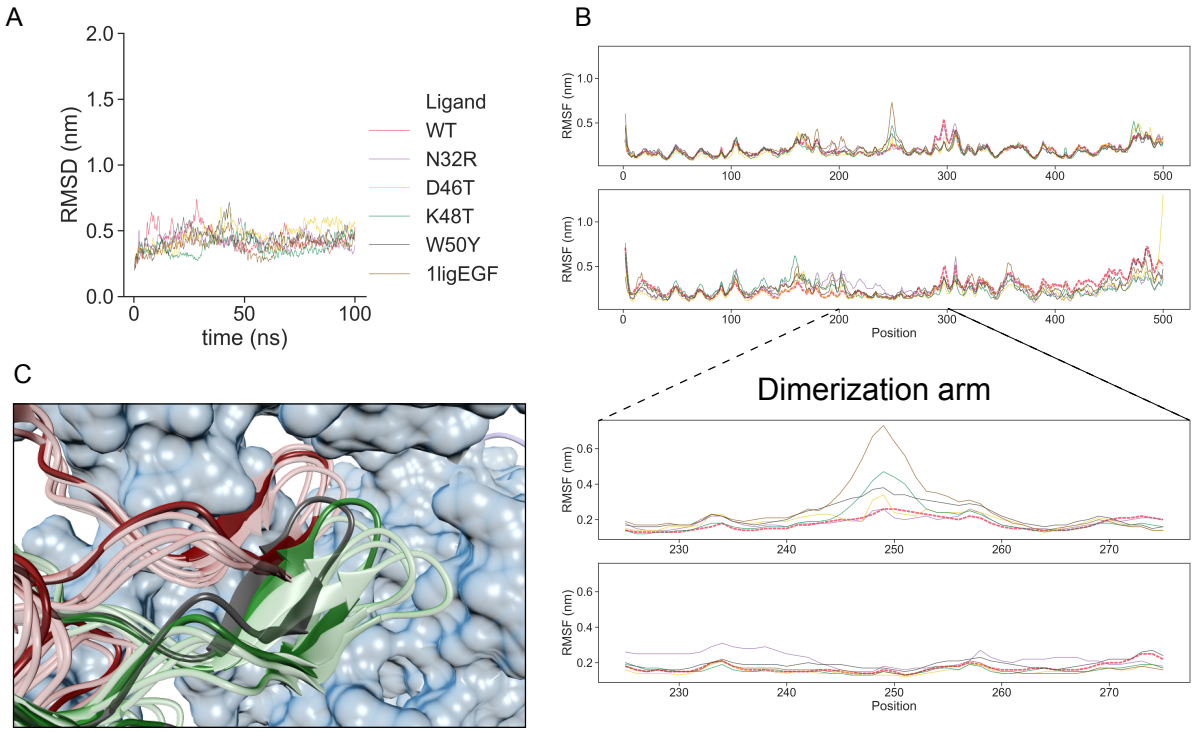


Figure 6. Molecular Dynamics of the EGFR extracellular domain. (A) Root Mean Square Deviation from the initial structure. (B) Top, Root Mean Square Fluctuations of the 100 ns simulations. The WT level is marked in bold and dashed red line. Bottom, focus on the RMSF of the dimerization arm domain. The two plots represent the two EGFR in the receptor dimer. (C) Dynamics of the dimerization arm domain in WT (red) and K48T (green) simulations. A structure every 20 ns of simulation was taken and aligned to 1IVO reference structure (blue surface and black ribbon). The structure at 20 ns is represented in solid colors. K48T simulation shows a more dynamic dimerization arm.

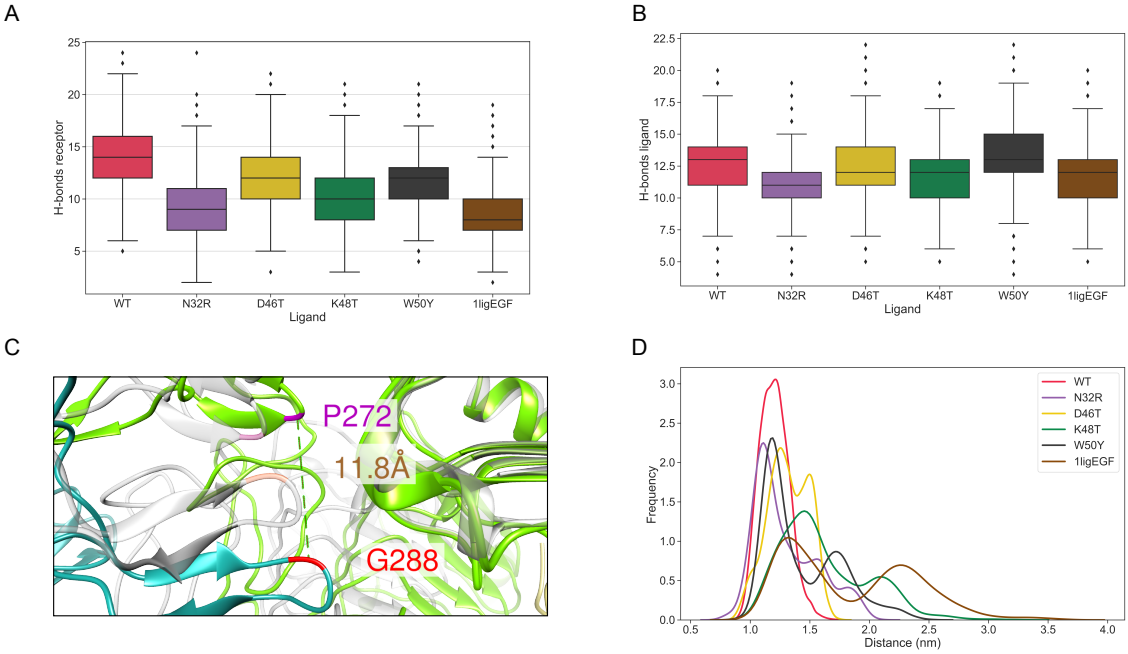


Figure 6—figure Supplement 1. RMSF and distance between Pro-272 and Gly-288. (A) Number of H-bonds formed between the two receptors during the 100 ns simulation time. (B) Number of H-bonds formed between the receptor that shows a dynamic dimerization arm and the corresponding ligand during the 100 ns simulation time. (C) Distance used to investigate the fluctuations of the dimerization domain highlighted on the structure. Pro 272 (purple) and Gly 288 (red) are highlighted both on 5WB7 (green and teal) and in 1IVO (grey). The distance between the alpha carbons of the 5WB7 couple is shown in brown and is notably longer than 1IVO. (D) Distribution of the distance between Pro272 and Gly 288 in all the simulations. WT simulation shows a single peak at ~1nm, while in all other ligands the distance has two peaks.

Discussion

The prediction of functional residues is a well-developed field (30), where conservation measures are considered a key factor to rely on. Tools such as ConSurf (31) and the ET-like methods (32) are able to identify slowly evolving positions that are involved in folding, interaction, or catalytic activity (30). However, the specific reason why a residue is conserved often remains unclear. In this work, we show a new method to identify residues that affects

343 specific functions in a system of interest. Our approach combines a conservation score
344 calculated from the structural alignment of paralogs and among orthologs, with an intra- and
345 inter-molecular co-evolution score with previously known interactors. Conservation and
346 coevolution give a complementary signal, thus improving the overall predictive performance
347 (33). The coevolution score was introduced to DIRpred to highlight residues that are directly
348 responsible of an interaction with the receptor, at the expenses of those that interact
349 intramolecularly within the ligand. While the first part efficiently identifies Asn-32 and Lys-
350 48 as putative interactors, the second part does not properly give a penalty to residue Tyr-29.
351 The interaction of this position with His-10 is also conserved in EGF but with reversed
352 positioning (Figure 2 – figure supplement 3), resulting in a low paralog conservation (then high
353 contribution to DIRpred score). While it is still possible to optimize the co-evolution scoring
354 function, integrating conservation and coevolution measures is a promising way to recall
355 information of specific functions involving protein-protein interactions.
356
357 Three out of the four positions were highlighted by previous reports, however none of them
358 were individually mutated. The tryptophan in position 50 is a strong outlier in our
359 bioinformatics analysis, along with Trp-49 (Figure 2). Their score is high even when using
360 conservation measures that do not take amino acid type change into account (data not shown).
361 However, Trp-50 was a better candidate for testing because of its outward facing position, and
362 since Trp-49 is involved in buried protein contacts (34). Trp-49 and Trp-50 could be
363 responsible in facilitating the interaction of EGF with the membrane phospholipids, as it
364 happens for Pro-7 and Leu-8 (35). In 1992, a biophysics work showed how Trp-49 and Trp-50
365 are not burying inside the bilayer when EGF and a membrane are in solution alone (36), though
366 this might be different when in complex with the receptor. For example, this could be achieved
367 by Trp-50 through the formation of a helix, clustering together with Val-34 and Arg-45 around

368 the conserved Leu-47 (37). Mutation N32R is on the interface between ligand and receptor.
369 The slightly higher affinity is probably due to the presence of the guanidinium group of R
370 which is positively charged and could interact with Gln-16 of EGFR ECD (Supplementary
371 figure 2). Interestingly, mutations of the corresponding position in chicken TGFA were able to
372 alter the mitogenicity without strongly affecting the binding affinity (38). While no previous
373 literature reported about Asp-46 before, Lys-48 was found in two mutants that showed higher
374 affinity (39). Our results give an explanation of the previous reports, while shedding new light
375 on the role of the previously unreported Asp-46.
376
377 With the exception of N32R, all mutant ligands do not change in binding affinity to EGFR of
378 compared to the wild type ligand. However, the mutants showed a different cellular effect on
379 Bj-5ta cells (Figure 5). A delayed proliferation response compared to control might seem
380 counterintuitive for fibroblast (40). Though, Bj-5ta hTERT immortalized fibroblasts show
381 features that distinguish them from *in vivo* fibroblast, like differential gene expression and
382 epiregulin-dependent proliferation (41). A decrease in cell proliferation after EGF treatment
383 could be the result of competition between the endogenously synthesized epiregulin and EGF,
384 thus altering the balance between proliferation and differentiation.
385 A431 cells constitutionally express EGFR at high levels. Treatment with EGF has been
386 observed to promote STAT1 dependent apoptosis (42). This pathway is dependent on the
387 internalization of the ligand-receptor complex by the endocytosis process, a key factor in the
388 ligand-induced biased signaling (17). In our experiments, A431 cells treated with mutant EGF
389 ligands show a higher growth rate, therefore a decreased rate of apoptosis compared to WT.
390 Given the setting of the experiment, modularity in the endocytosis pathway is a straightforward
391 explanation to the observed differences in growth rate and in protein levels and
392 phosphorylation. Interestingly, K48T treated cells showed an increase in long and thin cellular

393 processes in all replicates (Figure 5). This mutant was also observed with a markedly higher
394 EGFR phosphorylation 1 h after treatment (Figure 3). A possible explanation could be EGF-
395 induced tubular formation, an alternative EGF-induced pathway reported in intestinal epithelial
396 cells (43). In both the cellular growth experiments, the growth rate of the mutants is
397 intermediate between the control and EGF, and concentration dependent. This phenomenon,
398 and the cell-line specific response of growth hormones is consistent with previous literature
399 reports (44).
400
401 The observation that the four mutations in EGF alter signal transduction without disrupting any
402 contacts between the ligand and the receptor can be explained by the “loss of symmetry” model
403 of EGFR signaling (19). Freed *et al*. observed how stable symmetrical dimers show a short-
404 lived nature of signaling, while asymmetric dimers conduct a sustained activation, lasting
405 longer than 24 h. Such a model provides an explanation on how a low level of phosphorylation
406 after EGF treatment causes the apoptotic behavior observed in the cellular growth experiments.
407 Compared with WT EGF, the K48T and W50Y variants show a higher phosphorylation signal,
408 suggesting a more sustained activation of EGFR, possibly through the formation of a less stable,
409 asymmetric dimer. Thus, the observed effects of the EGF mutations on EGFR phosphorylation
410 and dimer stability are consistent with the "loss of symmetry" mechanism.
411
412 The analysis of the fluctuations in the dimerization arm reveals an underlying bias in ligand-
413 induced receptor dimerization, originally not visible from the static images of structure
414 comparison. K48T is the mutant that induce the biggest deviation from the WT, in accordance
415 to the phosphorylation data. Although, all mutants show a transition. While giving an initial
416 outlook on the effect of the mutants, our MD simulations do not take into account other factors
417 that could be important in the mechanism of action of EGFR. Differential multimerization (45),

418 receptor glycosylation, and the interaction with the membrane (23) are factors where the
419 underlying bias induced by the mutant ligands could also have an effect in.
420
421 In this work, we showed how a single mutation of EGF is able to alter the specific functional
422 relationship with the receptor. For functional divergence to arise, it could take as little as
423 mutating 15% of the sequence (47). However, the EGFR ligands divergence date back to the
424 vertebrate ancestor of R1/R2 whole genome duplication, up to 500 million years ago (48), as
425 hinted by the low (~25%) sequence identity. From our results, the sequence distance does not
426 reflect the distance in function. In fact, the functional divergence of EGF was altered with just
427 a single targeted mutation.
428 In conclusion our data suggests that a single mutant ligand induces a conformational change of
429 the receptor that then affects receptor dimer stability, with plausible effects on phosphorylation
430 level and downstream pathway activation (49). This shows how the persistence of biased
431 signaling in EGFR is in an unstable equilibrium, where the observed conservation of diverging
432 sites among paralogs is naturally reinforced to maintain the functional divergence. To
433 determine whether a short distance in function space among paralogs is a consequence or a
434 necessity for living systems, further studies will be required.
435
436 **Materials and Methods**
437
438 **Sequence and structure analysis**
439 The sequences of EGFR ligands and the multiple sequence alignment of EGF orthologs were
440 obtained from the Ensembl database (51). Multiple sequence alignment (MSA) of all ligands
441 was performed with MAFFT software (52). X-ray structures were obtained from the PDB
442 database (53). Structural alignments were created using Chimera (54).

443

444 **Phylogenetic analysis**

445 From the multiple sequence alignment of EGF from different species, nearly identical
446 sequences were removed. The *Drosophila melanogaster* EGF sequence was added as an
447 outgroup in the EGF phylogeny, while *Caenorhabditis elegans* EGF was used as outgroup for
448 the ligand phylogeny. MSA and phylogenetic tree images were created using Unipro UGENE
449 software (55). Three phylogenetic trees were made using the Neighbor Joining (NJ), Maximum
450 Likelihood (ML), and Bayesian (MrB) methods in IQTREE (56), using ModelFinder to scan
451 for the best-fit evolutionary model and parameters (57).

452

453 **Divergence Inducing Residue PREDiction**

454 We identified sites in EGF responsible for functional divergence using a method that combines
455 evolutionary and co-evolutionary data, called DIRpred (Divergence Inducing Residue
456 PREDiction). The DIRpred scoring function combines four components to evaluate each
457 residue: I) The combined conservation scores in the ortholog alignments. This score is
458 calculated by averaging the conservation of a reference ligand site with the conservation in the
459 respective positions of the other ligands' orthologs alignments; II) The complementary value
460 (1 – x) of the conservation score in the paralogs alignment. This score can be obtained either
461 from sequence alignment (paralogs MSA) or structural alignment (paralogs MSTA). III) The
462 highest co-evolution score between a reference ligand site and all of the receptor sites. IV) The
463 complementary value (1 – x) of the highest co-evolution score between a reference ligand site
464 and all of the other ligand sites in the joint orthologs alignment of all ligands (Figure 7).

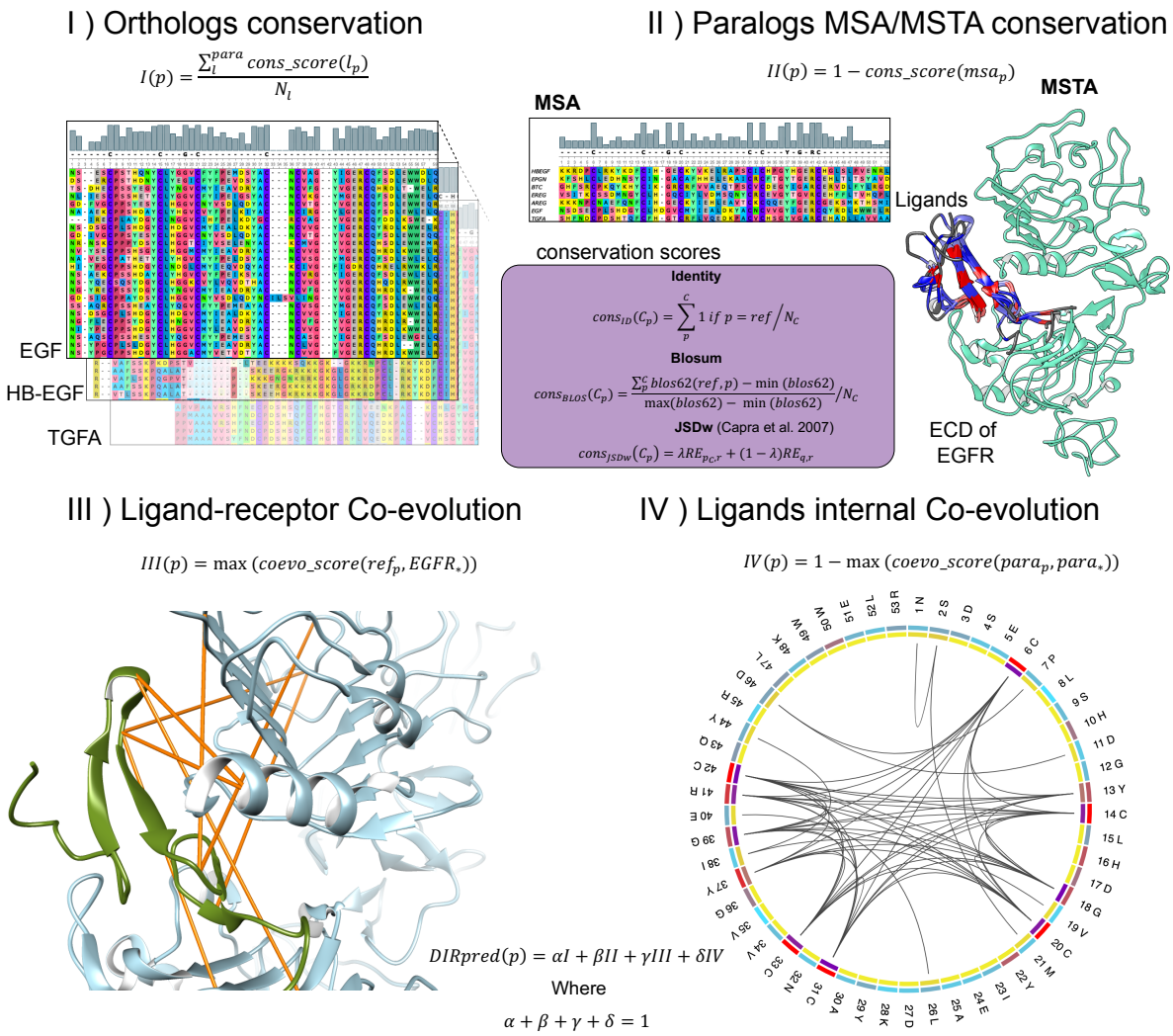


Figure 7. DIRpred rationale. The prediction of Divergence Inducing Residues consists of a linear combination of four site-specific scores, in roman numerals. In I, the orthologs MSA of each paralog is used to compute a conservation score (as exemplified from the histograms on the alignments). Alignment images were produced using Unipro software (58). In II, the human paralog sequences or structures were aligned and the same conservation score function was used. In the purple box, the conservation score functions are represented. In III, the co-evolution score between EGF and each receptor site was computed using Mutual Information (MI). The yellow lines on the structure connect the highest scoring co-evolving residues between ligand and receptor. In IV, MI co-evolution score was computed from the combined ligands orthologs alignment. The circular plot was performed using MISTIC2 (59).

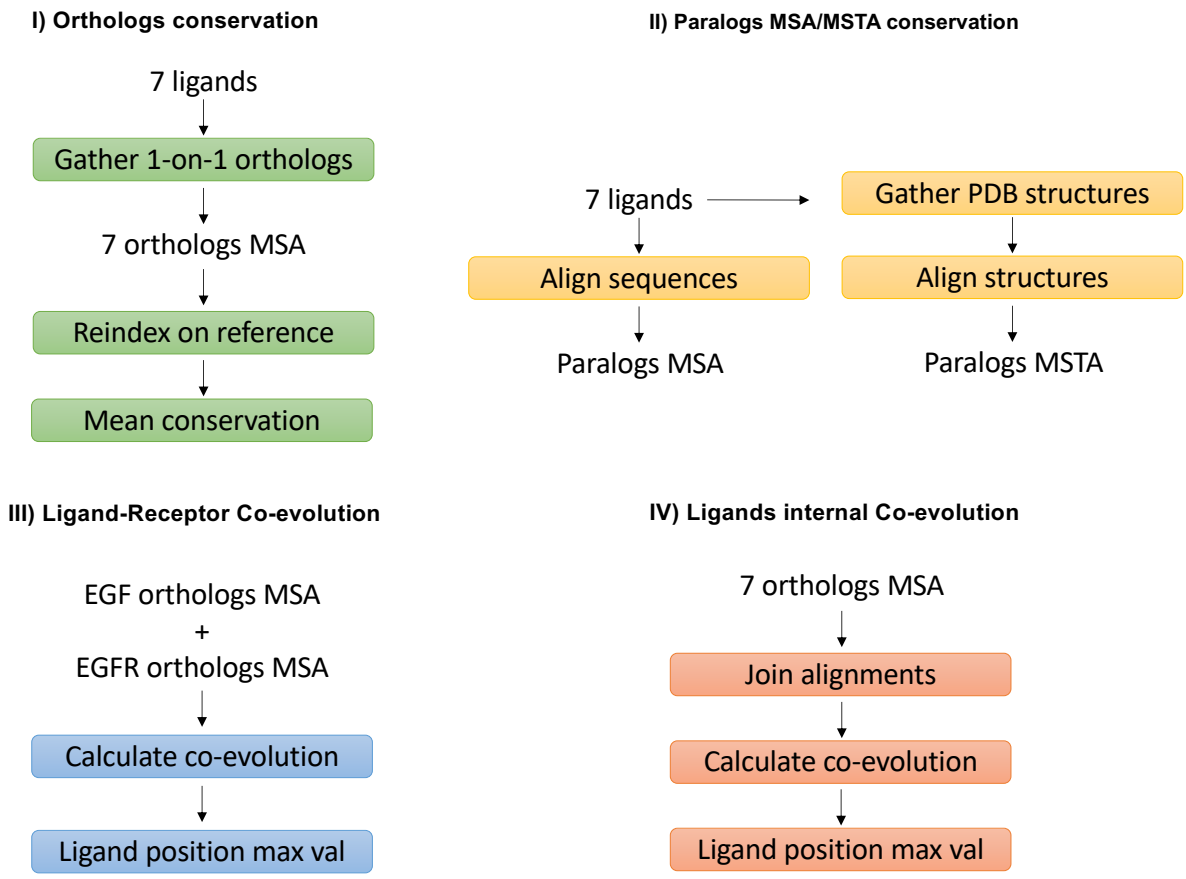


Figure 7—figure Supplement 1. DIRpred schema. A schematic representation of the previous figure, showing the workflow to produce each of the four scores.

Conservation scores were calculated using three different formulations: 1) IDENTITY, 2) BLOSUM, and 3) JSDw. The IDENTITY score measures the frequency of appearance of a reference residue in the MSA. The BLOSUM score takes the amino acid substitution frequency into account using the BLOSUM62 matrix, normalized by the maximum and minimum score for the BLOSUM matrix. The JSDw score is based on Jensen Shannon divergence with a window of residues (60). The analysis and plots were performed with Python3 package SEABORN. EGF index positions from the paralogs MSA and MSTA were respectively used to align the MSA and MSTA based scores.

488 The code used in the analysis of the DIRpred score and plots and the data used in this paper
489 are shared on Github: https://github.com/oist/DIRpred.
490
491 **Selection of the mutations**
492 Along with the DIRpred score, the choice of positions for mutation was influenced by two
493 manually curated factors: the distance from the receptor and the amino acid variation among
494 ligand types. We designed mutations with the aim of inferring a transition to the amino acid
495 properties in sites where the ligands show a different pattern. Overlapping residues at a given
496 position were divided into two groups, based on an EGF-like and non EGF-like stabilization
497 of the receptor dimer. This property was previously shown to follow binding affinity (19).
498 Residues that introduced a noticeable shift in amino acid properties between the two groups
499 were selected. For example, position N32 is hydrophobic in the high affinity ligands group,
500 positively charged in the low affinity group and negatively charged in EGF. Finally, we
501 carefully analyzed exceptional cases in the DIRpred scoring. Some of the residues that show
502 high score have intramolecular interactions with another amino acid in the ligand. These
503 residues, if mutated, will lose EGF structural stability (namely "residue swapping" behavior
504 showed in Figure 2—figure Supplement 3). The decision of which mutation to introduce was
505 made using the paralogs alignment. Positions 32, 48 and 50 have high DIRpred score. Position
506 46 was included although having a lower score, because the substitution pattern matches the
507 two ligands groupings. Furthermore position 46, 48 and 50 were preferred because, given
508 previous experiments and the overall scores, the EGF C-terminus tail seems to play a critical
509 role in the ligand function, for example (61).
510
511 **Synthetic Peptides**

512 The wild-type, N32R, D46T, K48T, and W50Y variants of EGF were ordered from Scrum
513 Net Co with purity >90% and quantity 5 mg/mL. These peptides were used for ITC
514 measurements, Circular Dichroism (CD) measurements, proliferation studies, and Western
515 Blot (WB) analyses.
516
517 **Cell Lines**
518 The *Bj5-tα human normal fibroblast cell line* was purchased from ATCC. Cells were grown in
519 DMEM with 10% fetal bovine serum (FBS), and 5 μg/mL hygromycin B. The *Swiss Albino*
520 *3T3* mouse *normal fibroblast cell line* was obtained from the RIKEN Cell Bank. Cells were
521 grown in DMEM, 10% FBS, 50 ug/mL gentamycin at 37°C in a 5% CO_2 atmosphere with 95%
522 humidity. The *A431 human epithelial carcinoma adherent cell line* (RIKEN Cell Bank) is a
523 model skin cancer cell line with overexpressed EGFR used for oncogenic pathway studies
524 (62).Cells were cultured in DMEM supplemented with 10% FBS (Sigma-Aldrich), 50 ug/mL
525 gentamycin antibiotic or a combination of 100 unit/ml Penicillin G (Nacalai Tesque) and 100
526 µg/ml streptomycin sulfate (Nacalai Tesque). Experiments were conducted at 37°C in a 5%
527 CO_2-enriched air atmosphere with 95% humidity. Cell lines were grown and used for Western
528 Blot and cell proliferation studies.
529
530 **Cell Proliferation Assay**
531 We measured cell proliferation using a label-free, non-invasive, cellular confluence assay with
532 IncuCyte Live-Cell Imaging Systems (Essen Bioscience). Human Bj5-tα (2,500 cells / well)
533 and Mouse Swiss Albino 3T3 (1,000 cells/well) were seeded overnight on a 96-well plate
534 (Corning) at 37°C in an incubator. The next day, cells were treated with WT EGF and mutants
535 at 1 nM, 10 nM and 100 nM concentrations and placed in an XL-3 incubation chamber
536 maintained at 37°C. The plate was scanned using a 4x objective at 2-hr intervals over 3 days.

537 Cell confluence was measured using IncuCyte Analysis Software. The IncuCyte Analyzer
538 gives real-time confluence data based on segmentation of high-definition phase-contrast
539 images. Cell proliferation is shown as an increase in confluence rate relative to control.
540

541 **Apoptosis Assay**

542 Experiments were performed with the A431 human cancer cell line. 5,000 cells/well were
543 seeded on a 96-well plate (Corning) and incubated at 37°C for 24 hr. Media were replaced with
544 fresh DMEM containing WT EGF, or EGF mutants at 1, 10, and 100 nM concentrations and
545 fluorescent annexin V green reagent. Plates were pre-warmed prior to data acquisition to avoid
546 condensation and expansion of the plate, which affect autofocus. Images were captured every
547 2 hrs (4x) for 3 days in the IncuCyte system. Cell proliferation is reported as in the previous
548 assay.
549

550 **Statistics**

551 Proliferation and apoptosis experiments were performed in duplicates. All results are shown
552 as the mean±s.d. Raw data was analyzed by two-way ANOVA with 95% confidence level.
553 The multiple test was corrected using Bonferroni post hoc test. Prism 8 software was used
554 for statistical analysis. Asterisks in the pictures show P-values using GraphPad
555 convention: 0.1234 > (ns), 0.0332 > (*), 0.0021 > (**), 0.0002 > (***), 0.0001 > (****).
556

557 **Isothermal Titration Calorimetry (ITC)**

558 All ITC studies employed a MicroCal PEAQ-ITC System (Malvern). For titration, both EGFR
559 ECD (Sigma-Aldrich) and EGF variants were dialyzed into the same reaction buffer Milli-Q
560 H_2O (22 μm) at 25°C. Each titration involved serial injections of 13 × 3 μL aliquots of EGF
561 variants (200 μM) into a solution of EGFR ECD (20 μM) in the sample cell. In each case, the

562 reference cell was filled with the same reaction buffer as the control to determine the heat upon
563 binding of the two components. The measured heat constant value was subtracted from the heat
564 per injection prior to analysis of the data. The experiment was replicated twice. Results were
565 analyzed by MicroCal PEAQ-ITC Analysis Software.
566
567 **Circular Dichroism (CD)**
568 Far UV measurements were taken at a protein concentration of 0.1 μM, using a cuvette with a
569 path length of 0.1 cm. Secondary structure content was calculated from far UV spectra using
570 CAPITO software (27). Five scans in the 190-240-nm wavelength range were taken.
571
572 **Western Blot Analysis**
573 A431 epidermoid carcinoma cells were harvested using Lysis Buffer (0.4% SDS, 1 mM DTT,
574 1%). Samples were incubated at 37°C for 10 min with Benzonase and centrifuged at 15,000
575 rpm at 22°C for 10 min. Supernatants were used for further analysis. Sample concentrations
576 were measured with a BCA protein assay kit (ThermoFisher Scientific). Lysate were mixed
577 with 4x Sample Loading Laemmli Buffer and incubated at 90°C for 5 min. Equal amounts of
578 protein were loaded in 12% Mini PROTEAN® TGX™ SDS-PAGE gel (Bio-Rad) and
579 transferred to PVDF membranes (Trans-Blot Turbo RTA Mini 0.2 μm PVDF Transfer Kit).
580 Membranes were blocked for 10 min with Bullet Blocking One (Nacalai) and reacted with
581 monoclonal rabbit anti-EGFR antibody (Cell Signaling Technology, Inc.), Phospho-EGF
582 Receptor (Tyr1173) (Cell Signaling Technology, Inc.), monoclonal rabbit anti-PLCy (Cell
583 Signaling Technology, Inc.), and anti-phosphorylated PLCy antibodies (Cell Signaling
584 Technology, Inc.), rabbit anti-PKCζ (Cell Signaling Technology, Inc.), rabbit anti-phospho-
585 PKCζ/λ (Thr410/403) (Cell Signaling Technology, Inc.) and rabbit anti-α-tubulin pAb (MBL)
586 at dilution of 1:1000. Samples were incubated with Goat Anti-Rabbit IgG HRP at a 1:5000

587 dilution and chemiluminescent signals were detected by CDP Plus (ThermoFisher Scientific)
588 and ChemiDoc touch MP (Bio-rad).
589
590 **Molecular Dynamics**
591 All the simulations were performed with Gromacs version 2020 (63) using charmm36-
592 mar2019 force field (64) and SPC216 for water. PDB ID 5wb7 was used as model and template.
593 EGF structure model was extracted from PDB ID 1ivo and superimposed to the ligand using
594 UCSF Chimera Match-Maker algorithm (65). Mutants were generated using Swissmodel
595 webserver (66), while missing atoms were compensated using Scwrl4 (67). A system was
596 composed of EGFR dimer in complex with one or two EGF wild type (WT), N32R, D46T,
597 K48T or W50Y were solvated and neutralized using NaCl ions in a dodecahedral box. Then,
598 energy temperature and pressure equilibrations were performed to the system following the
599 guidelines in Lemkul 2019 (68).
600
601 A 100 ns production simulation was run using the Verlet cut-off scheme (69) for non-bonded
602 interactions and LINCS as constraint algorithm (70). All the simulations reached convergence
603 of RMSD. Long range electrostatic interactions were computed using the Particle Mesh Ewald
604 method (71) using a dedicated GPU. The parameter files for all runs are included in the
605 supplementary data [#ADD#]. We checked for differences in relative motions between the
606 three simulations by extracting and concatenating the backbone trajectories using catDCD
607 plugin of VMD (72), then performing a PCA using Bio3d R package (73).
608
609 **Author contributions**

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

Acknowledgments

We thank Tadashi Yamamoto, Ichiro Maruyama for providing the antibodies, Keiko Kono for the technical help with the WB experiments. We are grateful for the help and support provided by the Scientific Computing section of Research Support Division and in particular Jan Moren at OIST. We thank Madhuri Gade, Mirco Dindo, Harry Wilson and Benjamin Clifton for critical reading. We thank Pierre Matricon for the discussion on the molecular dynamics data, Russel Alenton on the cell proliferation assay and Christine Guzman on the western blot analyses.

Funding

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

References

1. 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.
2. Oda K, Matsuoka Y, Funahashi A, Kitano H. A comprehensive pathway map of epidermal growth factor receptor signaling. Mol Syst Biol. 2005;1:2005.0010.
3. Landau M, Fleishman SJ, Ben-Tal N. A putative mechanism for downregulation of the catalytic activity of the EGF receptor via direct contact between its kinase and C-terminal domains. Structure. 2004;12(12):2265-75.
4. Purba ER, Saita EI, Maruyama IN. Activation of the EGF Receptor by Ligand Binding and Oncogenic Mutations: The "Rotation Model". Cells. 2017;6(2).
5. Wilson KJ, Gilmore JL, Foley J, Lemmon MA, Riese DJ, 2nd. Functional selectivity of EGF family peptide growth factors: implications for cancer. Pharmacol Ther. 2009;122(1):1-8.

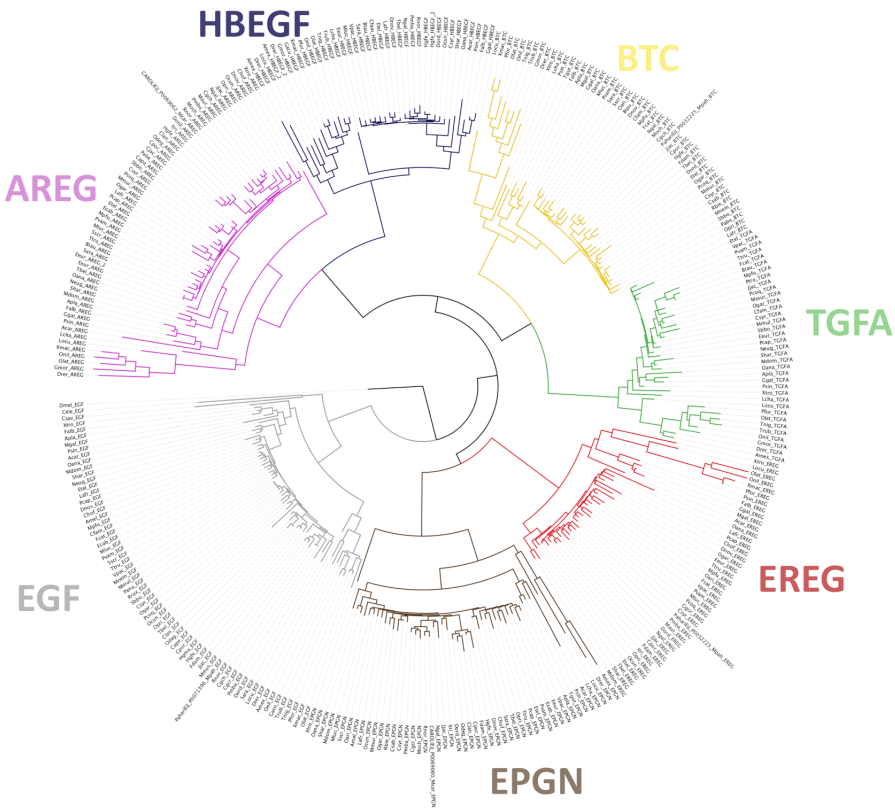
641 6. Nicholson RI, Gee JM, Harper ME. EGFR and cancer prognosis. European journal of
642 cancer (Oxford, England : 1990). 2001;37 Suppl 4:S9-15.
643 7. Xu MJ, Johnson DE, Grandis JR. EGFR-targeted therapies in the post-genomic era.
644 Cancer Metastasis Rev. 2017;36(3):463-73.
645 8. Alvarez-Ponce D, Feyertag F, Chakraborty S. Position Matters: Network Centrality
646 Considerably Impacts Rates of Protein Evolution in the Human Protein-Protein Interaction
647 Network. Genome Biol Evol. 2017;9(6):1742-56.
648 9. Jensen RA. Orthologs and paralogs - we need to get it right. Genome Biol.
649 2001;2(8):Interactions1002.
650 10. Rao VS, Srinivas K, Sujini GN, Kumar GN. Protein-protein interaction detection:
651 methods and analysis. International journal of proteomics. 2014;2014:147648.
652 11. Groenen LC, Nice EC, Burgess AW. STRUCTURE-FUNCTION-RELATIONSHIPS
653 FOR THE EGF/TGF-ALPHA FAMILY OF MITOGENS. Growth Factors. 1994;11(4):235-
654 57.
655 12. Souriau C, Gracy J, Chiche L, Weill M. Direct selection of EGF mutants displayed on
656 filamentous phage using cells overexpressing EGF receptor. Biol Chem. 1999;380(4):451-8.
657 13. Skwark MJ, Raimondi D, Michel M, Elofsson A. Improved Contact Predictions Using
658 the Recognition of Protein Like Contact Patterns. PLOS Computational Biology.
659 2014;10(11):e1003889.
660 14. Gocheva G, Ivanova A. A Look at Receptor-Ligand Pairs for Active-Targeting Drug
661 Delivery from Crystallographic and Molecular Dynamics Perspectives. Mol Pharm.
662 2019;16(8):3293-321.
663 15. Evans BA, Sato M, Sarwar M, Hutchinson DS, Summers RJ. Ligand-directed
664 signalling at beta-adrenoceptors. Br J Pharmacol. 2010;159(5):1022-38.
665 16. Kenakin T. Functional selectivity and biased receptor signaling. J Pharmacol Exp Ther.
666 2011;336(2):296-302.
667 17. Ali R, Brown W, Purdy SC, Davisson VJ, Wendt MK. Biased signaling downstream of
668 epidermal growth factor receptor regulates proliferative versus apoptotic response to ligand.
669 Cell Death Dis. 2018;9(10):976.
670 18. Strachan L, Murison JG, Prestidge RL, Sleeman MA, Watson JD, Kumble KD. Cloning
671 and biological activity of epigen, a novel member of the epidermal growth factor superfamily.
672 The Journal of biological chemistry. 2001;276(21):18265-71.
673 19. Freed DM, Bessman NJ, Kiyatkin A, Salazar-Cavazos E, Byrne PO, Moore JO, et al.
674 EGFR Ligands Differentially Stabilize Receptor Dimers to Specify Signaling Kinetics. Cell.
675 2017;171(3):683-95.e18.
676 20. Mehrabi M, Mahdiuni H, Rasouli H, Mansouri K, Shahlaei M, Khodarahmi R.
677 Comparative experimental/theoretical studies on the EGFR dimerization under the effect of
678 EGF/EGF analogues binding: Highlighting the importance of EGF/EGFR interactions at site
679 III interface. Int J Biol Macromol. 2018;115:401-17.
680 21. Knudsen SL, Mac AS, Henriksen L, van Deurs B, Grovdal LM. EGFR signaling
681 patterns are regulated by its different ligands. Growth Factors. 2014;32(5):155-63.
682 22. Liu P, Cleveland TEt, Bouyain S, Byrne PO, Longo PA, Leahy DJ. A single ligand is
683 sufficient to activate EGFR dimers. Proc Natl Acad Sci U S A. 2012;109(27):10861-6.
684 23. Arkhipov A, Shan Y, Kim ET, Shaw DE. Membrane interaction of bound ligands
685 contributes to the negative binding cooperativity of the EGF receptor. PLoS Comput Biol.
686 2014;10(7):e1003742.
687 24. Chagoyen M, García-Martín JA, Pazos F. Practical analysis of specificity-determining
688 residues in protein families. Briefings in Bioinformatics. 2015;17(2):255-61.

689 25. Cochran JR, Kim YS, Lippow SM, Rao B, Wittrup KD. Improved mutants from
690 directed evolution are biased to orthologous substitutions. Protein Eng Des Sel.
691 2006;19(6):245-53.
692 26. Prestrelski SJ, Arakawa T, Wu CS, O'Neal KD, Westcott KR, Narhi LO. Solution
693 structure and dynamics of epidermal growth factor and transforming growth factor alpha. The
694 Journal of biological chemistry. 1992;267(1):319-22.
695 27. Wiedemann C, Bellstedt P, Gorlach M. CAPITO--a web server-based analysis and
696 plotting tool for circular dichroism data. Bioinformatics. 2013;29(14):1750-7.
697 28. Gill GN, Lazar CS. Increased phosphotyrosine content and inhibition of proliferation
698 in EGF-treated A431 cells. Nature. 1981;293(5830):305-7.
699 29. Gulli LF, Palmer KC, Chen YQ, Reddy KB. Epidermal growth factor-induced
700 apoptosis in A431 cells can be reversed by reducing the tyrosine kinase activity. Cell Growth
701 Differ. 1996;7(2):173-8.
702 30. Nemoto W, Saito A, Oikawa H. Recent advances in functional region prediction by
703 using structural and evolutionary information - Remaining problems and future extensions.
704 Computational and structural biotechnology journal. 2013;8:e201308007.
705 31. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an
706 improved methodology to estimate and visualize evolutionary conservation in macromolecules.
707 Nucleic Acids Res. 2016;44(W1):W344-W50.
708 32. Lichtarge O, Bourne HR, Cohen FE. An evolutionary trace method defines binding
709 surfaces common to protein families. J Mol Biol. 1996;257(2):342-58.
710 33. Teppa E, Wilkins AD, Nielsen M, Buslje CM. Disentangling evolutionary signals:
711 conservation, specificity determining positions and coevolution. Implication for catalytic
712 residue prediction. BMC Bioinformatics. 2012;13(1):235.
713 34. Gallay J, Vincent M, Li de la Sierra IM, Alvarez J, Ubieta R, Madrazo J, et al. Protein
714 flexibility and aggregation state of human epidermal growth factor. A time-resolved
715 fluorescence study of the native protein and engineered single-tryptophan mutants. Eur J
716 Biochem. 1993;211(1-2):213-9.
717 35. Tynan CJ, Roberts SK, Rolfe DJ, Clarke DT, Loeffler HH, Kastner J, et al. Human
718 epidermal growth factor receptor (EGFR) aligned on the plasma membrane adopts key features
719 of Drosophila EGFR asymmetry. Mol Cell Biol. 2011;31(11):2241-52.
720 36. Li De La Sierra IM, Vincent M, Padron G, Gallay J. Interaction of recombinant human
721 epidermal growth factor with phospholipid vesicles. A steady-state and time-resolved
722 fluorescence study of the bis-tryptophan sequence (TRP49-TRP50). European Biophysics
723 Journal. 1992;21(5):337-44.
724 37. Hommel U, Harvey TS, Driscoll PC, Campbell ID. Human epidermal growth factor.
725 High resolution solution structure and comparison with human transforming growth factor
726 alpha. J Mol Biol. 1992;227(1):271-82.
727 38. Puddicombe SM, Chamberlin SG, MacGarvie J, Richter A, Drummond DR, Collins J,
728 et al. The significance of valine 33 as a ligand-specific epitope of transforming growth factor
729 alpha. The Journal of biological chemistry. 1996;271(26):15367-72.
730 39. Lahti JL, Lui BH, Beck SE, Lee SS, Ly DP, Longaker MT, et al. Engineered epidermal
731 growth factor mutants with faster binding on-rates correlate with enhanced receptor activation.
732 FEBS letters. 2011;585(8):1135-9.
733 40. Carpenter G, Cohen S. Epidermal growth factor. Annual review of biochemistry.
734 1979;48:193-216.
735 41. Lindvall C, Hou M, Komurasaki T, Zheng C, Henriksson M, Sedivy JM, et al.
736 Molecular characterization of human telomerase reverse transcriptase-immortalized human
737 fibroblasts by gene expression profiling: activation of the epiregulin gene. Cancer Res.
738 2003;63(8):1743-7.

739 42. Grudinkin PS, Zenin VV, Kropotov AV, Dorosh VN, Nikolsky NN. EGF-induced
740 apoptosis in A431 cells is dependent on STAT1, but not on STAT3. Eur J Cell Biol.
741 2007;86(10):591-603.
742 43. Ibuka S, Matsumoto S, Fujii S, Kikuchi A. The P2Y(2) receptor promotes Wnt3a- and
743 EGF-induced epithelial tubular formation by IEC6 cells by binding to integrins. J Cell Sci.
744 2015;128(11):2156-68.
745 44. Bjorkelund H, Gedda L, Andersson K. Comparing the epidermal growth factor
746 interaction with four different cell lines: intriguing effects imply strong dependency of cellular
747 context. PloS one. 2011;6(1):e16536.
748 45. Macdonald-Obermann JL, Pike LJ. Different epidermal growth factor (EGF) receptor
749 ligands show distinct kinetics and biased or partial agonism for homodimer and heterodimer
750 formation. The Journal of biological chemistry. 2014;289(38):26178-88.
751 46. Innan H, Kondrashov F. The evolution of gene duplications: classifying and
752 distinguishing between models. Nature Reviews Genetics. 2010;11(2):97-108.
753 47. Zhou J, Liu D, Sa Z, Huang W, Zou Y, Gu X. Effective estimation of the minimum
754 number of amino acid residues required for functional divergence between duplicate genes.
755 Mol Phylogenet Evol. 2017;113:126-38.
756 48. Laisney J, Braasch I, Walter RB, Meierjohann S, Schartl M. Lineage-specific co-
757 evolution of the Egf receptor/ligand signaling system. BMC Evol Biol. 2010;10:16.
758 49. Bakker J, Spits M, Neefjes J, Berlin I. The EGFR odyssey – from activation to
759 destruction in space and time. Journal of Cell Science. 2017;130(24):4087-96.
760 50. Sotelo MJ, García-Paredes B, Aguado C, Sastre J, Díaz-Rubio E. Role of cetuximab in
761 first-line treatment of metastatic colorectal cancer. World journal of gastroenterology.
762 2014;20(15):4208-19.
763 51. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl
764 2018. Nucleic Acids Res. 2018;46(D1):D754-D61.
765 52. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7:
766 Improvements in Performance and Usability. Mol Biol Evol. 2013;30(4):772-80.
767 53. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein
768 Data Bank. Nucleic Acids Res. 2000;28(1):235-42.
769 54. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, et al. UCSF
770 ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci.
771 2018;27(1):14-25.
772 55. Okonechnikov K, Golosova O, Fursov M, Team U. Unipro UGENE: a unified
773 bioinformatics toolkit. Bioinformatics. 2012;28(8):1166-7.
774 56. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective
775 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol.
776 2015;32(1):268-74.
777 57. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder:
778 fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587-+.
779 58. Okonechnikov K, Golosova O, Fursov M, team tU. Unipro UGENE: a unified
780 bioinformatics toolkit. Bioinformatics. 2012;28(8):1166-7.
781 59. Colell EA, Iserte JA, Simonetti FL, Marino-Buslje C. MISTIC2: comprehensive server
782 to study coevolution in protein families. Nucleic Acids Res. 2018;46(W1):W323-w8.
783 60. Capra JA, Singh M. Predicting functionally important residues from sequence
784 conservation. Bioinformatics. 2007;23(15):1875-82.
785 61. Puddicombe SM, Wood L, Chamberlin SG, Davies DE. The interaction of an epidermal
786 growth factor/transforming growth factor alpha tail chimera with the human epidermal growth
787 factor receptor reveals unexpected complexities. The Journal of biological chemistry.
788 1996;271(48):30392-7.

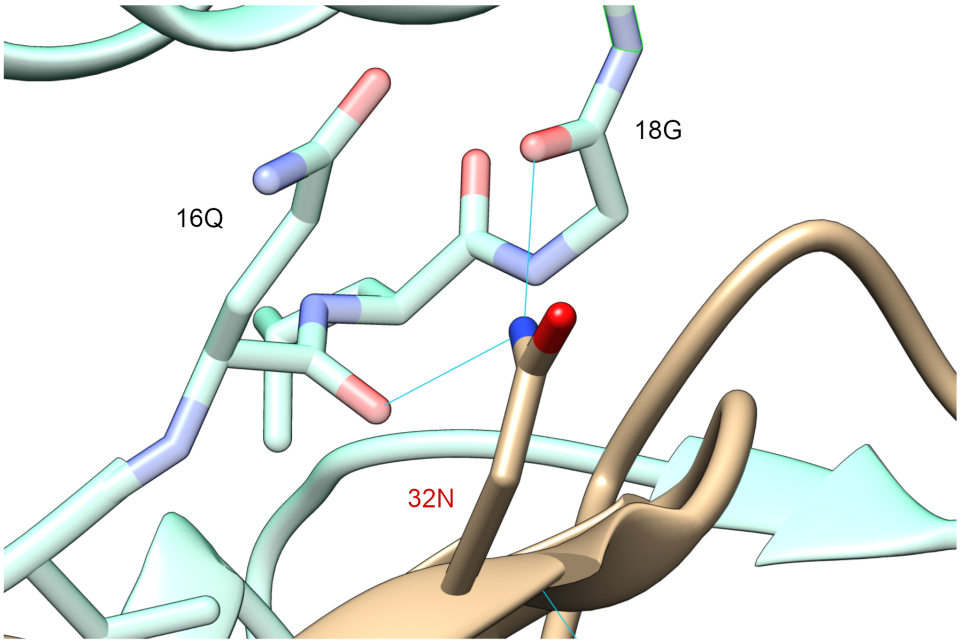
62. Giard DJ, Aaronson SA, Todaro GJ, Arnstein P, Kersey JH, Dosik H, et al. In vitro cultivation of human tumors: establishment of cell lines derived from a series of solid tumors. J Natl Cancer Inst. 1973;51(5):1417-23.
63. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1-2:19-25.
64. Huang J, MacKerell AD, Jr. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. J Comput Chem. 2013;34(25):2135-45.
65. Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE. Tools for integrated sequence-structure analysis with UCSF Chimera. BMC Bioinformatics. 2006;7:339.
66. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46(W1):W296-W303.
67. Krivov GG, Shapovalov MV, Dunbrack RL, Jr. Improved prediction of protein side-chain conformations with SCWRL4. Proteins. 2009;77(4):778-95.
68. Lemkul JA. From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0]. Living J Comp Mol Sci. 2019; 1(1):5068.
69. Páll S, Hess B. A flexible algorithm for calculating pair interactions on SIMD architectures. Computer Physics Communications. 2013;184(12):2641-50.
70. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. Journal of Computational Chemistry. 1997;18(12):1463-72.
71. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. The Journal of Chemical Physics. 1995;103(19):8577-93.
72. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14(1):33-8, 27-8.
73. Grant BJ, Rodrigues AP, ElSawy KM, McCammon JA, Caves LS. Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics. 2006;22(21):2695-6.

818 **Supplementary material**



819

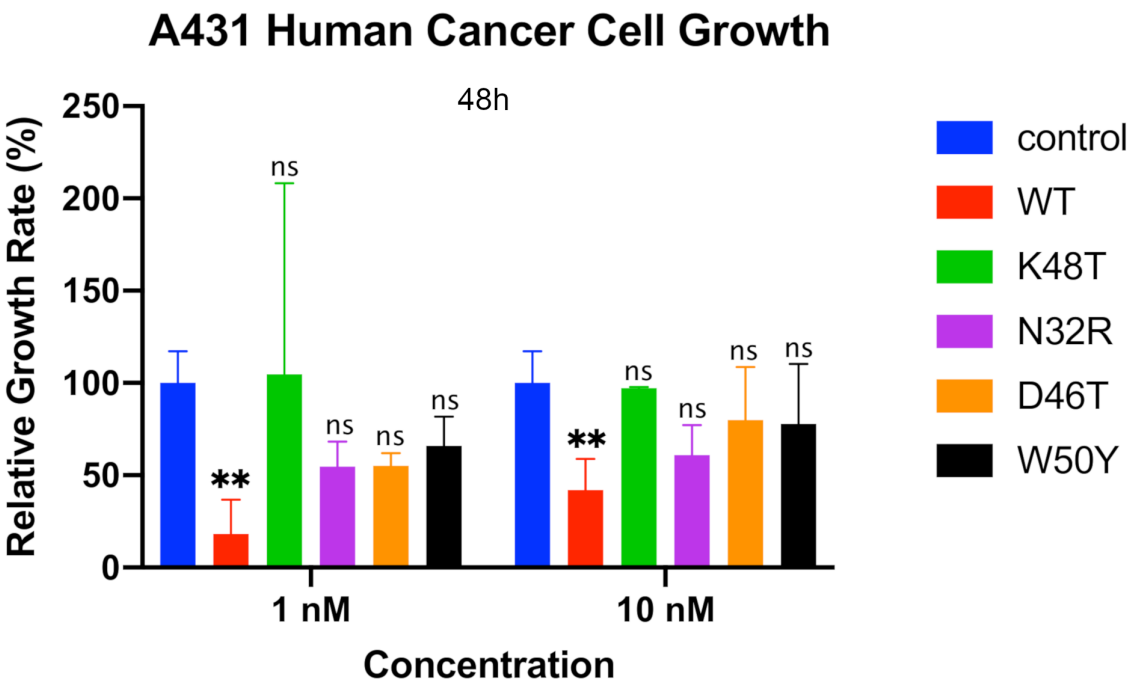
820 **Supplementary figure 1. EGFR ligands orthologs tree.** The tree was obtained using
821 Maximum Likelihood algorithm. *C elegans* EGF was used as outgroup. The EGFR ligands
822 cluster neatly. Interestingly, fish ligands are found to have longer branches on average. The
823 tree was made using iqtree ModelFinder. The best-fit model used is JTT+R6.



824

825 **Supplementary figure 2. Crystalized structure of WT EGF with the highlighted position**

826 **N32, In the interaction with the position Q16 and G18 of the EGFR receptor backbone.**



827

828 **Supplementary figure 3. Effect of 1nM and 10nM concentration of EGF variants on the**
829 **proliferation of A431 skin cancer cells.** The data represent the relative growth percentage to
830 control based on the confluence of cells (mean ± standard deviation) for each concentration of
831 EGF variants. The percentage confluence was estimated 48 h after the treatment (two
832 replicates/treatment).
833
834 **Supplementary data 1. Time-lapse video of the proliferation of A431 and Bj-5tα cells.**
835 Treatment with 100 nM ligand was performed at the 24h mark. Evidences of a balance between
836 differentiation and proliferation are found in the fibroblast, where a reduced proliferation is
837 associated to cellular morphology differences in WT and K48T.
838