Supplementary Information (SI) for Physical Chemistry Chemical Physics.
This journal is © the Owner Societies 2024

Probing Conformational Dynamics of EGFR Mutants via SEIRA Spectroscopy: Potential Implications for Tyrosine Kinase Inhibitor Design

Emiliano Laudadio**[a], Federica Piccirilli**[b,c], Henrick Vondracek[b,d], Giovanna Mobbili[e], Marta S. Semrau[b], Paola Storici [b], Galeazzi Roberta[e], Elena Romagnoli[e], Leonardo Sorci[a], Andrea Toma[f], Vincenzo Aglieri[f], Giovanni Birarda*[b], Cristina Minnelli*[e]

**These authors contributed equally

*Corresponding authors
Email: giovanni.birarda@elettra.eu; c.minnelli@staff.univpm.it

Table of Contents

1. Experimental

1.1. SEIRA spectroscopic analysis

1.2. Model building and molecular dynamics simulations

2. Hydration degree of αC-helix in WT and L858R/T790M EGFR

3. References

1. Experimental

1.1. SEIRA spectroscopic analysis

Nanoantenna arrays are promising platforms for Surface-Enhanced Infrared Absorption (SEIRA) microscopy, enabling enhanced signal detection in selected spectral regions. We exploited here the protocol for the design, fabrication, and functionalization of ad-hoc substrates tailored for SEIRA microscopy developed by Zucchiatti et al[1]. The substrates used consist of a matrix of gold nanoantennas with a size of 50x50 μm^2, patterned on CaF_2 optical windows (thickness 1 mm), optimized for a pronounced signal enhancement in the Region of Interest (ROI) centred around 1600 cm^{-1}. Each CaF_2 window was endowed with 20 antenna arrays for IR analysis. After the nanofabrication process, the devices underwent a thorough cleaning step employing O_2 plasma treatment (100 W for 2 minutes) to clean and activate the gold surface of the antenna arrays. The subsequent functionalization step involved the utilization of the Hexahistidine (His6)-Ni^{2+}: nitriloacetic acid (NTA) methodology for the immobilization of proteins onto the SEIRA devices. The human EGFR WT and L858R/ T790M constructs were produced at the Protein Facility of Elettra Sincrotrone Trieste, as described in Zucchiatti et al[1]. Briefly, the EGFR kinase domain (residues 696–1022) containing an N-terminal His6-tag followed by the TeV cleavage site was expressed in baculovirus infected Sf9 insect cells and purified in 3 chromatographic steps (IMAC, anion exchange and size-exclusion), without removing the His6-tag. The final protein constructs were frozen in 25 mM Hepes pH 8, 250 mM NaCl, 5 % glycerol, 2 mM TCEP buffer and stored in aliquots at -80 ◦C until further use. The T790M/ L858R mutations were generated using Q5 site-directed mutagenesis (New England Biolabs). The SEIRA devices were subjected to an overnight incubation with a 15 µM NTA-EG6-(CH_2)16-SH (SH-NTA) solution at 4 °C, facilitating the formation of a Self-Assembled Monolayer (SAM) on the gold surface of the nanoantennas. The substrate was subsequently rinsed with 95% ethanol and dried under a nitrogen stream to remove any excess reagents. To promote subsequent protein immobilization, the SH-NTA SAM-coated chips were equilibrated with a 10 mM $CoCl_2 \cdot H_2O$ solution for 30 minutes. These SH-NTA functionalized chips were then incubated with a solution containing 500 nM of EGFR WT or EGFR L858R-T790M in HEPES for 1 hour and washed with fresh buffer solution comprising 25 mM HEPES buffer, to remove any non-specifically bound proteins. SEIRA microscopy measurements were conducted at the Chemical and Life Sciences branch of the SISSI beamline (Synchrotron Infrared Source for Spectroscopy and Imaging) at Elettra Sincrotrone Trieste. In all the PIR-SEIRA measurements, the chips with the immobilized protein monolayer were loaded in top-down configuration into a microfluidic device containing fresh HEPES buffer solution.

Spectral acquisition was performed using the Bruker Hyperion 3000 Vis-IR microscope coupled to a Bruker Vertex 70v interferometer. For spectral acquisition, continuous measurements were taken, with each spectrum being acquired in 1 minute, resulting in a total measurement time on each device of 20-30 minutes. The setup included a mid-band HgCdTe detector, enabling FTIR micro-spectroscopy. Point-spectra were captured in reflection mode using a 15× Schwarzschild objective and knife-edge apertures set at 50 × 50 µm^2 to target individual antenna arrays. We covered a spectral range from 4000 to 800 cm^{-1} with a spectral resolution of 4 cm^{-1} in a dual-side, forward/backward acquisition mode at a scanner speed of 40 kHz. Fourier transform analysis was executed with Mertz phase correction and a Blackman-Harris 3-term apodization function. Background spectra were acquired in reflection mode using a gold patch present in the device, using the same instrumental parameters. Raw spectra were corrected for carbon dioxide and water vapor contributions using the routines in the software OPUS 8.5 from Bruker Optics GmbH. A buffer spectrum from a clean device was acquired and used to subtract both the contributions of the antenna resonance profile and of water molecules near the anchored protein monolayer from the protein spectra. Subtraction was performed after the application of a multiplicative factor to the buffer spectrum ensuring an Amide II/Amide I band intensity, calculated at the maxima on baseline corrected spectra, between 0.62 and 0.68. 2^{nd}-derivative were calculated by applying a Savitzky-Golay filter (with 4th-order polynomial and 24 smoothing points). Spectra were thus fitted (with a Gaussian shape) in the region 1640-1670 cm^{-1} by using Quasar 1.7.0 software.

1.2. Model building and molecular dynamics simulations

The inactive tyrosine kinase domain of EGFR WT and EGFR L858R/T790M were prepared starting from PDB files 4HJO [2] and 3W2O [3]. The missing loops were obtained and modelled using Modloop server [4]. The junctions between carboxyl and amino terminal residues were modelled using GalaxyWEB [5], while the DISULFIND online server [6] was used to predict the disulfide bonding state of cysteines and their disulfide connectivity. As expected, no disulfide bridges were found in the EGFR catalytic domain (CD).

A simulation box of 120 nm^3 was settled including 54867 TIP3P water molecules added to solve each system. Na^+ and Cl^- ions were included explicitly to reach the physiological conditions (0.15 M NaCl) and to neutralize the net charge of the different EGFR proteins. Solvation and ions were added using GROMACS 5.1.1 [7,8] . A minimization phase composed of 10,000 cycles of steepest descent followed by 5,000 cycles of conjugate gradient minimization was used to converge to the energy threshold of 1,000 kJ/mol/nm. Then, both EGFR WT and LT proteins were gradually accommodated in their salt-

aqueous environment through the following equilibration step of 2 ns, in which the reference temperature of 310 K was gradually reached. Atom velocities were generated in the NVT ensemble using the Maxwell distribution function with a generated random seed and a weak temperature coupling using the Berendsen thermostat. A time constant of 1 ps was applied to maintain the reference temperature (310 K) for the whole run. Verlet cutoff [9] was used in combination with Particle Mesh Ewald (PME) for electrostatics [10]. The cutoff for the calculation of the van der Waals force was set to 1.2 nm, while the force smoothly was switched to zero between 1.0 and 1.2 nm. After the equilibration step, each EGFR system underwent to 200 ns of MD simulation in NPT ensemble implementing an accurate leapfrog algorithm or interacting Newton's equations of motion with a time step of 0.002 ps. The weak coupling was maintained also for pressure control (i.e., Berendsen barostat). For all simulation runs, the isotropic conditions were set with a reference pressure of 1 atm and a time constant for coupling of 5 ps. Position restraints were applied to EGFR proteins. A shift to the Nose-Hoover [11] and Parrinello-Rahman algorithm for pressure coupling [12] was operated for the production phase in NPT ensemble. To be sure about the simulation results, each system underwent to 200 ns of MD simulations for three times maintaining the described conditions. The root means square deviation (RMSD) and the root means square fluctuation (RMSF) of EGFR forms were calculated to remark structural differences between systems. The analysis of the simulation trajectories was performed by means of the VMD [13] and CHIMERA software [14].

2. Hydration degree of αC-helix in WT and L858R/T790M EGFR

In order to validate the MD simulation results, the WT and the L858R/T790M EGFR were subjected to MD simulations in triplicate, and the RMDS for each replicate was calculated through the respective trajectory. After 100 ns of MD simulation, it is noted that in all simulations the same dynamic stabilization is achieved, with the same RMSD values including oscillations (**Fig S1**). This reinforces the *in silico* data as truthful and representative of the *in vitro* behavior of the observed systems.

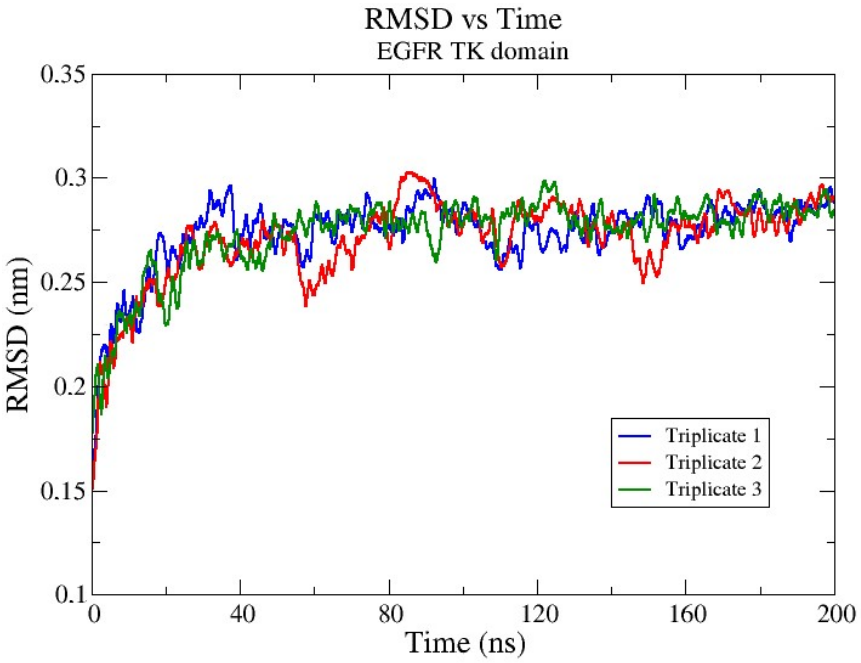

Figure S1. RMSD vs Time of WT EGFR in triplicate

Analysing the MD simulations of the WT form, a clear variation in positioning between the N and C lobes is observed, as can also be seen from an unwinding of the activation loop. The result is a preserved three-dimensionality for the amino acids of the N-lobe, but overall, a greater exposure of the helices to the solvent, compatible with an increased hydration of the helices in the WT form found by spectroscopy. In the case of the LT form, the activation loop follows a different spatial rearrangement, which determines a rotation of the αC-helix and a greater exposure of this helix to the solvent, in line with what is already known in the literature about the activation mechanism of this mutated form of EGFR (**Fig S2**).

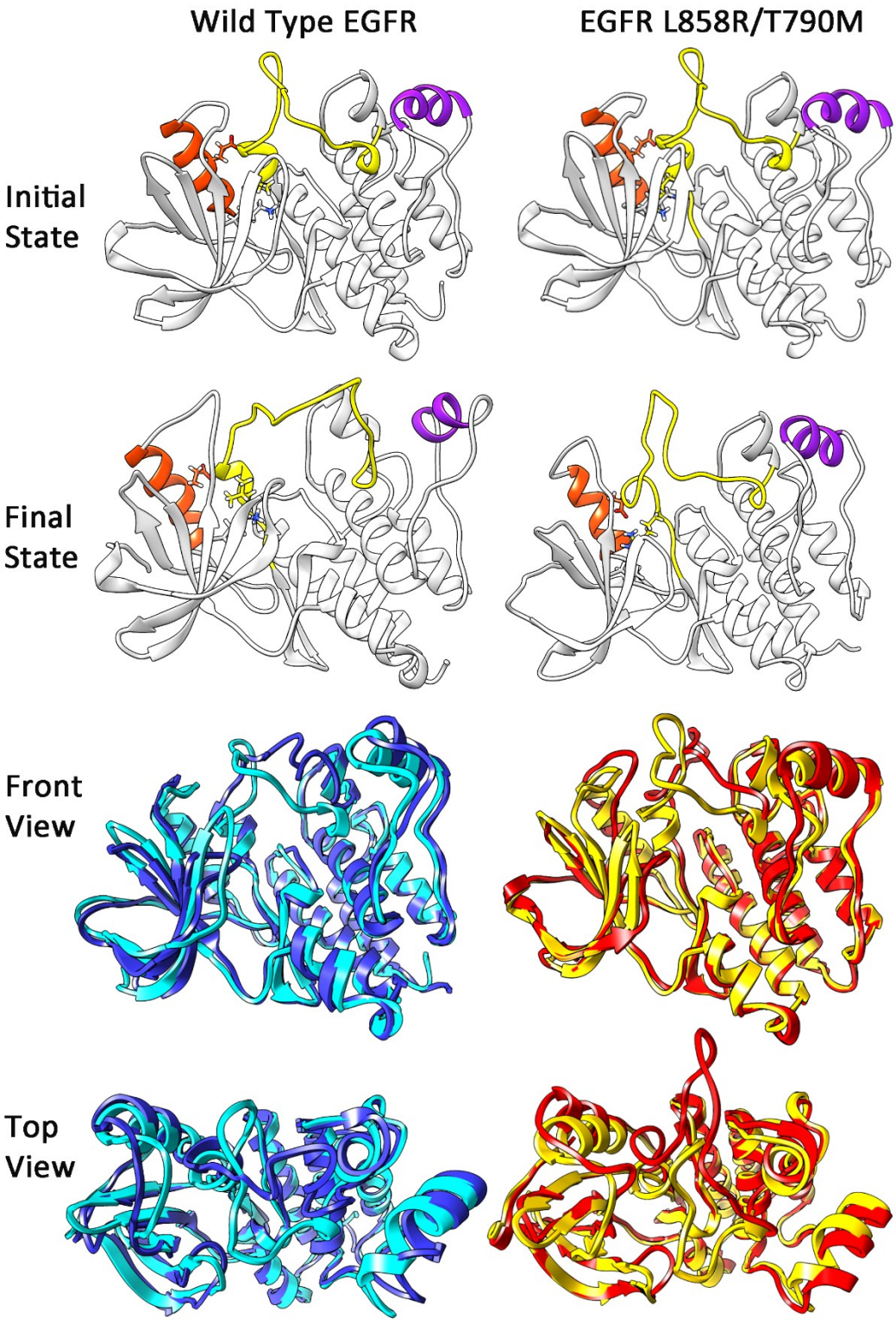

Figure S2. Time evolution dynamics of WT and L858R/T790M EGFR along the 200 ns of MD. The αC-helix, activation loop, and helix of C-lobe are reported in red, yellow, and purple respectively. More, residues K745, E762, and L/R858 are also highlighted in grey, red, and yellow respectively. Superimposition of the models. In cyan and gold are reported the starting structures of WT and L858R/T790M respectively, while in blue and red are remarked the final structures of WT and L858R/T790M, respectively.

By monitoring the progress of αC-helix, it remains more hydrated in the LT form (**Fig. S3a**). We observe 10 hydrogen bonds in the WT system, 2 of which with side groups and 8 in the backbone. In the upper half of the helix up to the Glu762 residue, the helix pace is 4, after Glu762 a pace of 3 is observed.

In the mutated form, the movement of Glu762 seems to be the driving force for the change in the pace of the helix, since this residue makes an ionic bond with Arg858, i.e. with one of the two mutated residues. This leads to a total of 7 H-bonds in the C helix, 6 of them in the backbone and 1 with a side chain, also observed in the wild type. In the mutated form, helix C shows a distorted displacement, since the Ile579 makes 2 H-bonds in the backbone, both with Glu762 and with Ala763. The same is observed for Glu762, which makes two H bonds in the backbone and one, precisely with Arg858, which is ionic. We also reported the change in curvature of helix C in the mutated form compared to the wild type (**Fig. S3b**).

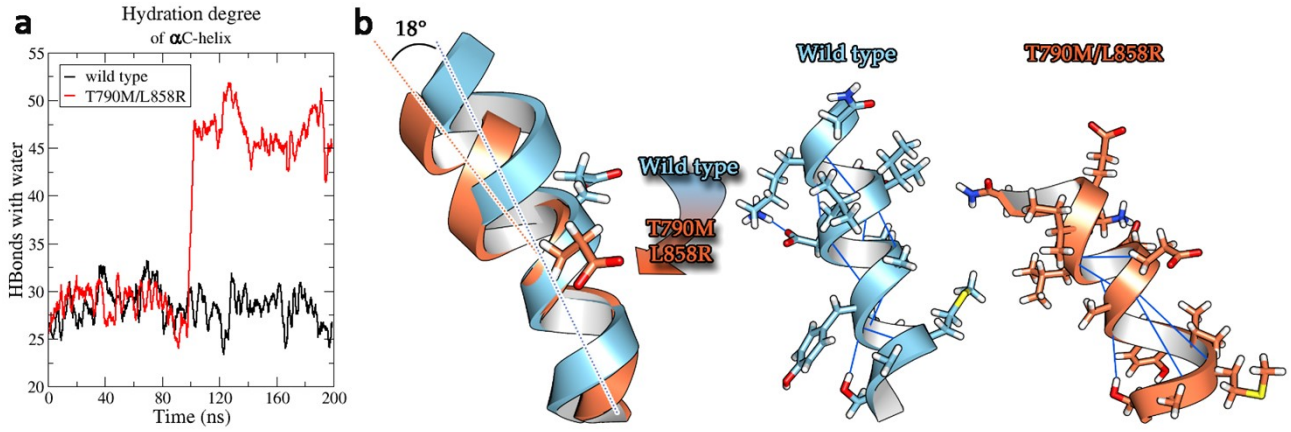

Figure S3. Hydration degree (**a**) and comparison of the displacement (**b**) of αC-helix in WT and L858R/T790M EGFR.

Finally, monitoring the Root Mean Square Deviation (RMSD) of the entire EGFR forms (**Fig. S4**), two different stabilization pathways are observed. While the WT form shows an RMSD value of 0.281 ± 0.022 after 100 ns, the LT system shows larger oscillations and different values at different simulation times, reaching the final value of 0.312 ± 0.047 in the last 40 ns. The dynamic stabilization of the LT shows the different steps that the protein goes through to reach the active conformation, while the WT form, having reached a steady state, will maintain this conformation over time.

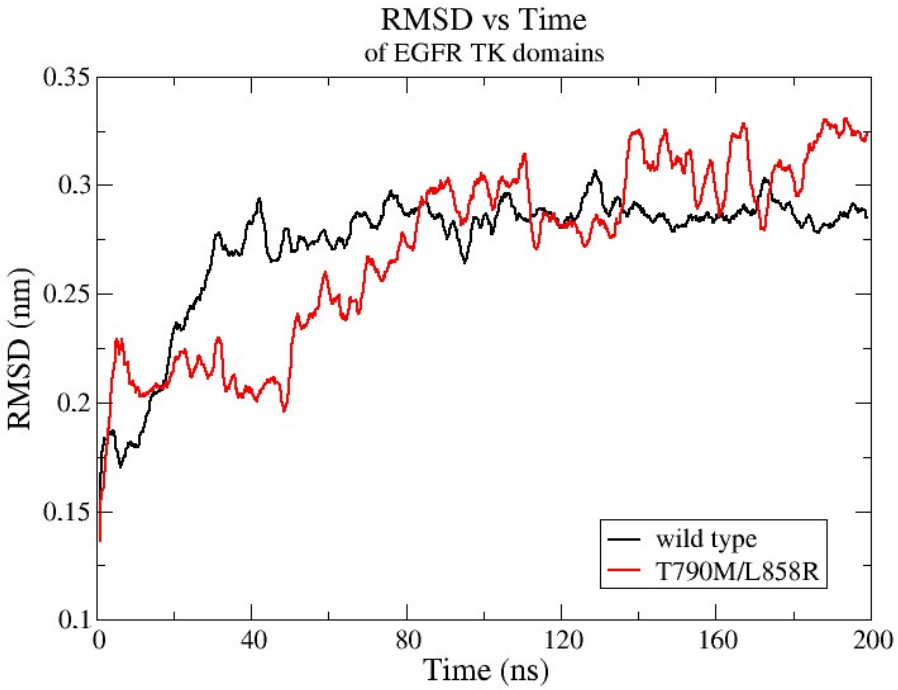

Figure S4. Root mean square deviations of WT and L858R/T790M EGFR.

3. References

1 P. Zucchiatti, G. Birarda, A. Cerea, M. S. Semrau, A. Hubarevich, P. Storici, F. De Angelis, A. Toma and L. Vaccari, *Nanoscale*, 2021, **13**, 7667–7677.

2 J. H. Park, Y. Liu, M. A. Lemmon and R. Radhakrishnan, *Biochemical Journal*, 2012, **448**, 417–423.

3 S. Sogabe, Y. Kawakita, S. Igaki, H. Iwata, H. Miki, D. R. Cary, T. Takagi, S. Takagi, Y. Ohta and T. Ishikawa, *ACS Med Chem Lett*, 2013, **4**, 201–205.

4 A. Fiser, R. K. G. Do and A. Šali, *Protein Science*, 2000, **9**, 1753–1773.

5 J. Ko, H. Park, L. Heo and C. Seok, *Nucleic Acids Res*, 2012, **40**, W294-7.

6 A. Ceroni, A. Passerini, A. Vullo and P. Frasconi, *Nucleic Acids Res*, 2006, **34**, W177-81.

7 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindah, *SoftwareX*, 2015, **1–2**, 19–25.

8 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J Comput Chem*, 2005, **26**, 1701–1718.

9 S. Páll and B. Hess, *Comput Phys Commun*, 2013, **184**, 2641–2650.

10 T. Darden, D. York and L. Pedersen, *J Chem Phys*, 1993, **98**, 10089–10092.

11 S. Nosé, *J Chem Phys*, 1984, **81**, 511–519.

12 M. Parrinello and A. Rahman, *J Appl Phys*, 1981, **52**, 7182–7190.

13 W. Humphrey, A. Dalke and K. Schulten, *VMD: Visual Molecular Dynamics*, 1996.

14 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J Comput Chem*, 2004, **25**, 1605–1612.