PCCP

COMMUNICATION

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2024, **26**, 22853

Received 30th May 2024,
Accepted 18th August 2024

DOI: 10.1039/d4cp02232g

rsc.li/pccp

Probing conformational dynamics of EGFR mutants *via* SEIRA spectroscopy: potential implications for tyrosine kinase inhibitor design†

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

Missense mutations in EGFR's catalytic domain alter its function, promoting cancer. SEIRA spectroscopy, supported by MD simulations, reveals structural differences in the compactness and hydration of helical motifs between active and inactive EGFR conformations models. These findings provide novel insights into the biophysical mechanisms driving EGFR activation and drug resistance, offering a robust method for studying emerging EGFR mutations and their structural impacts on TKIs' efficacy.

Missense mutations in the catalytic domain (CD) of tyrosine kinases (TKs) significantly impact their function, often contributing to cancer development and progression. Understanding these mutations through structural conformational dynamics is crucial for designing selective tyrosine kinase inhibitors (TKIs). Epidermal growth factor receptor (EGFR), a potent oncogene kinase commonly altered in cancers such as non-small cell lung cancer (NSCLC), can adopt inactive and active conformations.[1] In physiological conditions, EGFR is usually inactive, becoming active upon growth factor binding, which induces conformational changes in three conserved motifs: the activation loop (A loop), the αC-helix, and the Asp-Phe-Gly (DFG) motif.[2,3] Mutations within the CD can stabilize the active form or destabilize the inactive one, enhancing EGFR activation. TKIs, which mimic ATP, block tyrosine residue phosphorylation and downstream signaling. Studying these structural dynamics helps in the rational design of effective EGFR inhibitors, improving treatment outcomes for cancer patients.[4] They represent the standard of care for patients with NSCLC harbouring activating mutations of the EGFR gene.[5,6] There are several types of TKIs showing different conformation dependent mechanisms of binding. Inhibitors of type I bind to active conformations, type $I\frac{1}{2}$ and II are known to bind inactive ones while others are bivalent or allosteric.[7,8] Therefore, the interaction behaviour of TKIs strictly depends on the conformational features of the protein[9] and the knowledge of which conformer of protein is most present could provide opportunities for a rational design of effective EGFR inhibitors. Several approaches are used to investigate how the mutations changed the conformation landscape of EGFR. A number of crystal structures of EGFR in the wild type (WT) and mutated forms (L858R, T790M and L858R/T790M) were retrieved from the protein data bank (PDB) revealing to adopt either active or inactive conformation. Notably, most crystallographic structures lack well-defined density for the A-loop which is one of the EGFR subdomains involved in the receptor activation. Therefore, computational approaches have been used to model A-loop and to ascertain where the conformational equilibrium tends, whether towards active or inactive conformer.

In the last decade, SEIRA (surface enhanced infrared absorption) spectroscopy has emerged as a highly sensitive tool addressing the study of protein monolayers at the level of their secondary structure.[10–12] In SEIRA spectroscopy, IR microscopy is coupled to functionalized plasmonic devices. Their resonance frequency is designed to match a spectral range specific to proteins, allowing to depict subtle secondary structure modifications produced by the interaction of the protein with a ligand or a change in the environmental chemo-physical conditions.[13] Herein, we present an ultrasensitive and label-free detection of the subtle structural differences between the active and inactive conformations of the EGFR kinase domain by using a plasmonic platform based on ordered micro-arrays of gold nanoantennas.[13] In this context, we selected the wild type

[a] *Department of Science and Engineering of Matter, Environment and Urban Planning, Marche Polytechnic University, 60131, Ancona, Italy*
[b] *Elettra Sincrotrone Trieste S.C.p.A, 34149 Basovizza, Trieste, Italy*
[c] *Area Science Park, Padriciano 99, 34149 Trieste, Italy*
[d] *Diamond Light Source, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0DE, UK*
[e] *Department of Life and Environmental Sciences, Marche Polytechnic University, 60131 Ancona, Italy. E-mail: c.minnelli@staff.univpm.it*
[f] *Istituto Italiano di Tecnologia, Via Morego 30, 16163 Genova, Italy*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4cp02232g

‡ These authors contributed equally.

(WT) EGFR, as the inactive model, and the drug-resistant double mutant (L858R/T790M, LT EGFR) as the active one, since the co-presence of L858R and T790M mutations shifts the conformers equilibrium toward the active conformation.[1,14] The experimental setup allows the quantity of recombinant proteins required for the study to be minimized, probing a molecular monolayer anchored to the gold antennas patterns. Moreover, thanks to the adoption of the PIR (plasmonic internal reflection) geometry, the protein analysis can be performed in a buffered environment, close to the physiological conditions, minimizing the contributions of surrounding molecules to the IR proteins' spectra that typically hinder a proper evaluation of their conformation. The SEIRA-observed spectral shifts, corroborated by Molecular Dynamics (MD) simulations and literature data, give the first experimental evidence of the relation between the different hydration degree of the helix motifs and EGFR conformations.

Molecular dynamics simulation

MD simulation has been extensively used to theorize the transitions from inactive to catalytically active conformation in EGFR kinase. Overall, the destabilization, unfolding, solvent-exposure and fluctuation of the helix seem to be the main drivers of the EGFR's dynamic activation. It is well known that the WT EGFR, in the absence of growth factors, is in its autoinhibited conformation while the presence of double mutations L858R/T790M (LT) stabilizes the CD in its active state.[1] Therefore, we simulated the conformational changes of the CD of WT and LT EGFR kinases transitioning from the inactive to the active state. To confirm the reliability of the MD results, each system has been investigated for three to have a triplicate. After 200 ns, a univocal dynamic stabilization has been observed (Fig. S1, ESI†). The MD simulation shows that WT EGFR retains the same conformation of the inactive one. As reported in Fig. 1a, the short helix in the N-terminal of the activation loop (residues G857–G863) is folded and the E762 residue of the αC-helix is rotated out of the catalytic site (αC-out). In agreement with previous *in silico* studies,[1,15–17] the L858 sidechain projects into the hydrophobic cavity created by the outward movement of αC and prevents the inward movement of the αC helix (αC-in), which represents the main characteristic of the active conformation. The LT EGFR conformation evolves instead from the inactive to the active one along MD simulation (Fig. S2, ESI†). These data clearly confirm the recently reported results,[1] showing how the substitution of the hydrophobic leucine with a more hydrophilic arginine can strongly stabilize the active conformation. In particular, the R858 (Fig. 1b) induces an unfolding of the short helix in the A loop leading the αC-helix to assume the αC-in conformation, which is then kept in place by a salt bridge formed by E762, K745 and R858.

As shown in Fig. 1, the effect of these structural differences is reflected in a different organization of the helix motifs including the four parallel helices coloured in grey. Since this

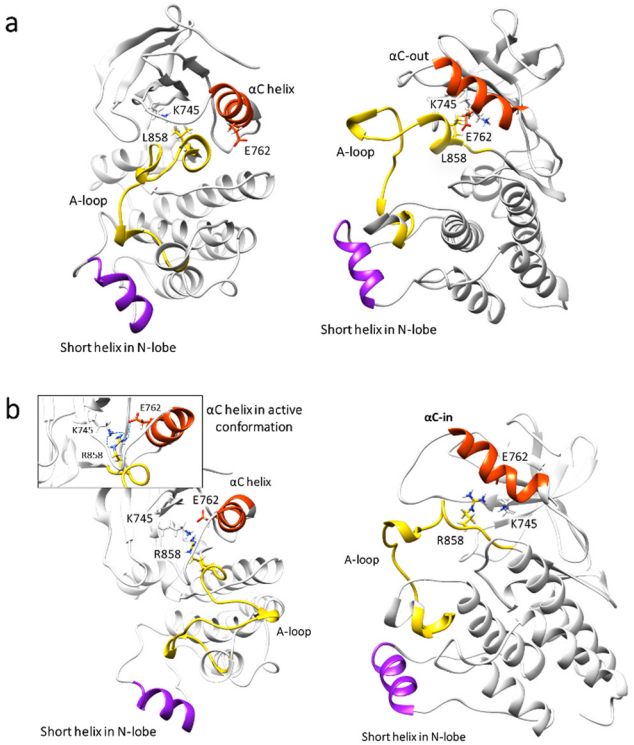

Fig. 1 Conformation of WT (a) and L858R/T790M EGFR (b) after 200 ns of MD. Front (left) and side view (right) of EGFR proteins.

portion of the protein is rarely investigated, we examined how the mutations affect the degree of helical compactness in the EGFR active conformer with respect to the inactive one by assessing the radius of gyration (R_g) of the all helix motifs (Fig. 2a). According to the analysis, a smaller R_g, which is associated with a higher degree of compactness, is observed

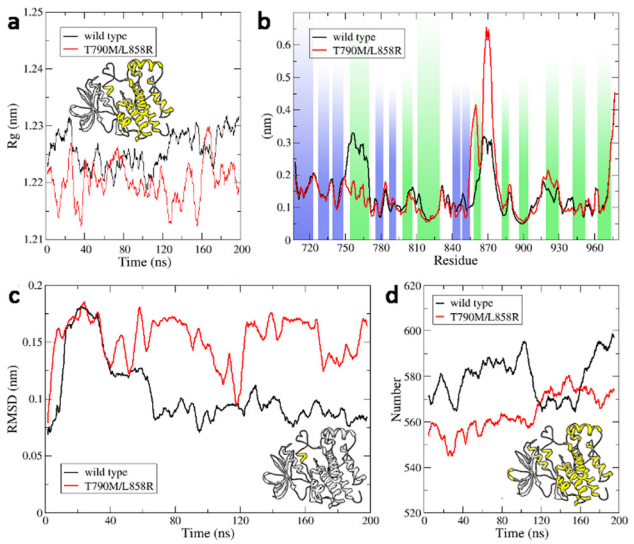

Fig. 2 Radius of gyration (R_g) of helices in WT and L858R/T790M EGFR (a). Root-mean square fluctuations (RMSF) of WT and L858R/T790M EGFR (b). Root-mean square deviation (RMSD) of the helix in the A loop (c). Hydration degree of all helices (d).

for the LT compared to WT EGFR. This could be in part correlated to the inward movement of the αC helix (αC-in) in the LT (Fig. 1).

To better describe these evident differences, we calculated root-means square fluctuations (RMSF) from MD trajectories for both WT and LT EGFR. The analysis shows a conformer-type specific degrees of fluctuation (Fig. 2b). A decreased fluctuation of the αC helix (aa 758 to 767) in the LT EGFR is observed compared with the WT form. This closely matches the αC-in conformation of LT, where the αC helix exhibits limited mobility due to a salt bridge formed between E762, R858, and K745, leading to a more compact structure. A different behaviour is instead detected for the aa 855–857 (DFG motif), 857–863 (short helix of the A-loop) and 864–875 fragment, in which the RMSF value is higher in LT EGFR than in WT form, confirming the increased mobility of the A-loop in the presence of L858R/T790M mutations. To better evaluate the stability and dynamics of the protein along MD simulation time, we also calculated the root mean square deviation (RMSD) of the αC short helix in the A-loop along MD simulation time (Fig. 2c). Results showed that, the RMSD value detected for the WT EGFR, after 100 ns of simulation, is 0.08 ± 0.002 nm, while for the EGFR LT is of about two-times higher (0.15 ± 0.04 nm) without any defined steady state along the whole MD time (Fig. 2c). This is attributed to a higher movement of the unfolded αC short helix along MD simulation. On the other hand, the WT EGFR showed lower RMSD values, in line with a more organized αC short helix in the A-loop.

Fluctuations of the protein surface groups drive and are driven by the surrounding network of water molecules. Therefore, the increase or decrease of domain movements in proteins translates in a change of the hydration dynamics. Kannan *et al.* showed that the unfolding of the short helix of the A-loop leads it to be more exposed to the solvent in the LT EGFR than to WT EGFR.[17] Therefore, with the aim of identifying the hydration degree variations linked to the only αC helix movements, we measured the number of hydrogen bonds between the aa of only this helix and water molecules in the two EGFR forms (Fig. S3a, ESI†). The αC helix in the WT form of EGFR showed the same degree of hydration from the beginning to the end of the simulation, with an average of 28 ± 4 H-bonds. Interestingly, the behaviour of the same helix in the LT EGFR was like that of the WT one during the first 100 ns, with 29 ± 3 H-bonds, but it increased drastically after 100 ns towards the end of the MD simulation, reaching 46 ± 5 H-bonds. This change appeared to be rapid and stable over time, suggesting that the helix increases its hydration degree. By analysing the pitch of the αC helix (Fig. S2b, ESI†), we note how this actually changes in the LT system at the end of the simulation. In this instance, the helix appears more distorted, showing an 18° deviation from its main axis compared to the WT. This distortion leads to a reduction in hydrogen bonds between residues, which is counterbalanced by an increase in interactions with water molecules due to a greater exposure. This phenomenon also influences the stabilization pathway of the LT form since it appears different in RMSD analysis (Fig. S4, ESI†) and supports previous MD simulations showing the presence of intermediate conformers in the EGFR conformational transitions.[18]

With the aim of finding out a relationship between the overall solvent hydration degree and EGFR tyrosine kinase domain conformations, we considered the number of hydrogen bonds (H-bonds) between the aa of all helical motifs and the water molecules during the MD simulation (Fig. 2d). From this analysis, a time-dependent increase in the hydration degree of the helices is observed in WT and LT systems. More in details, the number of H-bond increases from 572 to 595 in WT, while the LT model shows an increase from 556 to 573. These data suggest a significant change in the spatial arrangement of the domains in the mutated form after 200 ns of simulation with the inactive EGFR conformation model showing higher hydration degree (WT system, HB_{final} = 595 ± 3) than to the active one (LT EGFR system, HB_{final} = 573 ± 2).

Therefore, as overall effects, LT showed a higher compactness and lower hydration degree at helical moieties level than to WT EGFR.

SEIRA spectroscopy

In order to experimentally verify the results of the *in silico* study, SEIRAS measurements in the region 1490–1710 cm^{-1} of WT and LT EGFR monolayers, respectively, were acquired in PIR configuration and compared after processing. Fig. 3 summarizes the SEIRAS experiment (panel a), and the results obtained on both WT and LT EGFR (panels b–d).

The buffer-subtracted absorbance spectra in the region of amide I and amide II bands are shown at Fig. 3b. By the observation of the absorbance spectra it is possible to appreciate a relative broadening of the bands, therefore for more details an analysis of the 2nd derivative is necessary. After differentiation in the amide I region, it is possible to notice that the band is composed of three main components (Fig. 3c): at 1633 cm^{-1}, 1652–1660 cm^{-1}, and at 1681 cm^{-1}. Being EGFR-

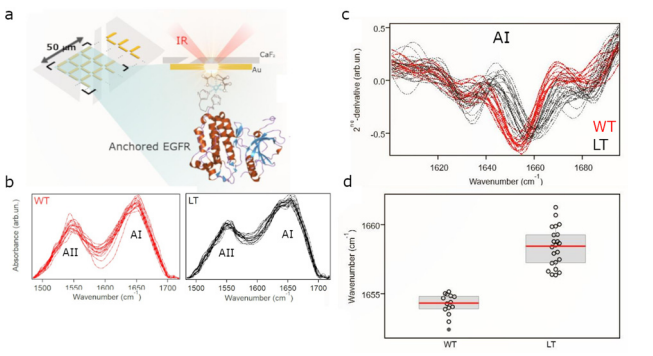

Fig. 3 Scheme of the experimental setup: kinase domain of EGFR proteins was anchored to gold antenna arrays, each having an overall width of 50 μm × 50 μm and measured with the IR//VIS microscope in PIR configuration (a). Absorbance spectra in the 1480–1710 cm^{-1} range of WT and LT, after buffer subtraction (b) 2nd-derivative of buffer subtracted spectra acquired for WT (red) and LT (black) on single antennas (c). Spectral position of the main minimum, associated to α-helix, observed in 2nd-derivative spectra calculated through Gaussian fitting (d).

kinase domain mainly constituted by α-helices, the most intense of the three is associated with these structures. The frequency of this vibration undergoes a shift between the two proteins: it is centred at an average frequency of 1652 cm^{-1} for WT, while in the LT it is at 1660 cm^{-1}. This difference observed in the spectral position reflects changes in the environment experienced by helices in WT compared to LT.[19,20] A shift of this signal towards higher energies, can be due to a variation in the compactness of the helical motifs including the four parallel helices rarely investigated until now (highlighted in grey in Fig. 1). Karjalainen and Barth,[21] demonstrate that as the helices bundle together, there is a coupling of their vibrational signals, called inter-helical vibrational coupling, which shifts the amide signal to higher wavenumbers. By MD simulations (Fig. 2a), it is possible to see that the helical motifs of the LT protein are indeed more compact in LT than in WT. Therefore, SEIRA spectroscopy experimental data correlate well with MD simulations, and it is possible to differentiate WT and LT and therefore active and inactive EGFR conformations. At 1633 cm^{-1}, there is the signal of the short β-sheets present in the EGFR-kinase domain. In LT, this signal is stronger due to the overlapping with one associated with increasing helix bending in coiled coils,[22] modification experienced by the short helix of the A-loop, as can be seen in MD data. Then, the third minimum at 1681 cm^{-1}, is associated with turn/loop structures, and it does not show relevant differences in spectral position between WT and LT, even though it is broader for LT.

In conclusion, in studying protein dynamics and their interaction with other molecules, it is fundamental to work in hydrated conditions. This approach allows to appreciate the protein flexibility in function of the variation of intramolecular interactions and can be used to compare WT and mutated proteins, for example. Several studies have demonstrated that the SEIRA platform enables protein characterization under physiological conditions, and MD simulations can aid in interpreting spectral data. In our specific case, we believe that emerging EGFR mutated forms can be thoroughly studied using a combined SEIRA and *in silico* approach with potential implications for TKIs design. Moreover, the quality of the spectral data acquired on a monolayer of proteins is enough to analyse them in second derivative, demonstrating that SEIRA spectroscopy is sensitive to subtle variations in protein structure, specifically the different packing of the helices. This structural difference may correspond to variations in protein hydration, which could lead to a different drug accessibility and a possible resistance to treatments.

Author contributions

Conceptualization: G. B., C. M., E. L., F. P. data curation: G. B., C. M., E. L., F. P., H. V. formal analyses: E. L., F. P., E. R., H. V. Methodology: A. T., V. A., E. L., F. P., P. S., M. S., G. B., H. V. visualization: G. M., R. G., C. M., E. L., F. P., A. T., V. A., L. S., G. B. investigations: G. B., C. M., E. L., F. P., H. V. writing – original draft: G. B., C. M., E. L., F. P. writing – review & editing: G. B., C. M., E. L., F. P. G. M., A. T., V. A., R. G., P. S., M. S., H. V. funding acquisition: C. M.

Data availability

The data that support the findings of this study are available on request from the corresponding author, [CM and GB].

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported by FIRC-AIRC (Fondazione Italiana per la Ricerca Sul Cancro), providing a postdoctoral research grant no. 25234 (two years fellowship) for the year 2021–2022 (C. M.). The authors would like to acknowledge CINECA-HPC ISCRA MARCONI-100 computer system (ATOM-HMV project no. HP10CRO8W4) for the calculations on GROMACS and Elettra beamline SISSI-Chem – Life Sci (project no. 20210191). F. P. contribution was supported by the European Union – NextGeneration EU within the project PNRR ''PRP@CERIC'' IR0000028 – Mission 4 Component 2 Investment 3.1 Action.

References

1 E. Laudadio, G. Mobbili, L. Sorci, R. Galeazzi and C. Minnelli, *J. Biomol. Struct. Dyn.*, 2023, **41**, 6492–6501.
2 Y. Shan, A. Arkhipov, E. T. Kim, A. C. Pan and D. E. Shawa, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 7270–7275.
3 N. Jura, X. Zhang, N. F. Endres, M. A. Seeliger, T. Schindler and J. Kuriyan, *Mol. Cell*, 2011, **42**, 9–22.
4 E. Laudadio, L. Mangano and C. Minnelli, *ACS Chem. Biol.*, 2024, **19**(4), 785–1024.
5 F. Ciardiello, F. R. Hirsch, R. Pirker, E. Felip, C. Valencia and E. F. Smit, *Cancer Treat. Rev.*, 2024, 122.
6 E. E. Ke and Y. L. Wu, *Trends Pharmacol. Sci.*, 2016, **37**, 887–903.
7 R. Roskoski, *Pharmacol. Res.*, 2016, **103**, 26–48.
8 Q. Wang, J. A. Zorn and J. Kuriyan, *Methods in Enzymology*, Academic Press Inc., 2014, vol. 548, pp. 23–67.
9 C. J. Tsai and R. Nussinov, *PLoS Comput. Biol.*, 2014, **10**, e1003394.
10 R. Adato and H. Altug, *Nat. Commun.*, 2013, **4**, 2125.
11 D. Etezadi, J. B. Warner, F. S. Ruggeri, G. Dietler, H. A. Lashuel and H. Altug, *Light: Sci. Appl.*, 2017, **6**, e17029.
12 F. Piccirilli, H. Vondracek, L. Silvestrini, P. Parisse, F. Spinozzi, L. Vaccari, A. Toma, V. Aglieri, L. Casalis, A. P. Piccionello, P. Mariani, G. Birarda and M. G. Ortore, *Spectrochim. Acta, Part A*, 2024, **322**, 124772.
13 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.
14 C. Minnelli, E. Laudadio, G. Mobbili and R. Galeazzi, *Int. J. Mol. Sci.*, 2020, **21**, 1721.

15 C. H. Yun, T. J. Boggon, Y. Li, M. S. Woo, H. Greulich, M. Meyerson and M. J. Eck, *Cancer Cell*, 2007, **11**, 217–227.
16 M. A. Hasenahuer, G. P. Barletta, S. Fernandez-Alberti, G. Parisi and M. S. Fornasari, *PLoS One*, 2017, **12**, e0189147.
17 S. Kannan, G. Venkatachalam, H. H. Lim, U. Surana and C. Verma, *Chem. Sci.*, 2018, **9**, 5212–5222.
18 Y. Shan, A. Arkhipov, E. T. Kim, A. C. Pan and D. E. Shawa, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 7270–7275.
19 C. M. Davis, A. K. Cooper and R. B. Dyer, *Biochemistry*, 2015, **54**, 1758–1766.
20 A. Barth, *Biochim. Biophys. Acta, Bioenerg.*, 2007, **1767**, 1073–1101.
21 E.-L. Karjalainen and A. Barth, *Phys. Chem. B*, 2012, **116**(15), 4448–4456.
22 W. C. Reisdorf and S. Krimm, *Biochemistry*, 1996 **35**, 1383.