JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

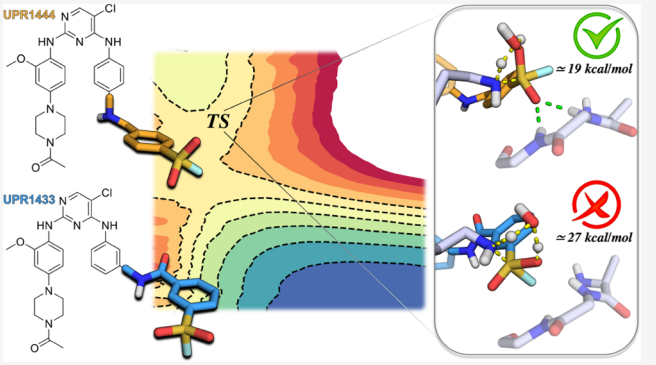
Mechanistic Modeling of Lys745 Sulfonylation in EGFR C797S Reveals Chemical Determinants for Inhibitor Activity and Discriminates Reversible from Irreversible Agents

Kemel Arafet, Laura Scalvini, Francesca Galvani, Sergio Martí, Vicent Moliner, Marco Mor, and Alessio Lodola*

Cite This: J. Chem. Inf. Model. 2023, 63, 1301−1312 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Targeted covalent inhibitors hold promise for drug discovery, particularly for kinases. Targeting the catalytic lysine of epidermal growth factor receptor (EGFR) has attracted attention as a new strategy to overcome resistance due to the emergence of C797S mutation. Sulfonyl fluoride derivatives able to inhibit $EGFR^{L858R/T790M/C797S}$ by sulfonylation of Lys745 have been reported. However, atomistic details of this process are still poorly understood. Here, we describe the mechanism of inhibition of an innovative class of compounds that covalently engage the catalytic lysine of EGFR, through a sulfur(VI) fluoride exchange (SuFEx) process, with the help of hybrid quantum mechanics/molecular mechanics (QM/MM) and path collective variables (PCVs) approaches. Our simulations identify the chemical determinants accounting for the irreversible activity of agents targeting Lys745 and provide hints for the further optimization of sulfonyl fluoride agents.



INTRODUCTION

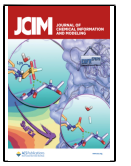
Epidermal growth factor receptor (EGFR) is a transmembrane protein, featuring an extracellular EGF binding domain and an intracellular tyrosine kinase domain, responsible for the transduction of signals promoting cell proliferation.[1] Hyper-activating mutations in the kinase domain of EGFR, at the level of exon 19 (del19) or exon 21 (L858R mutation), have been observed in at least 50% of non-small-cell lung cancer (NSCLC).[2] These alterations promote NSCLC insurgence and progression as their presence results in the constitutive activation of EGFR.[3]

Osimertinib (**1**, Figure 1) is a targeted covalent inhibitor (TCI) bearing an acrylamide warhead[4] used in first line for patients affected by EGFR mutation-positive NSCLC and in second line for patients harboring the acquired mutation T790M, emerged after treatment with first-generation inhibitors.[5] Despite its initial efficacy, the use of osimertinib is hampered by the acquired resistance driven by C797S mutation.[6] The lack of a thiol sidechain at position 797 prevents the formation of a covalent bond between the acrylamide group of osimertinib and EGFR,[7,8] allowing EGFR itself to escape irreversible inhibition.[9]

Medicinal chemistry efforts have recently demonstrated that the detrimental effect of C797S mutation can be over-come.[10,11] Compounds able to occupy accessory pockets in the kinase active site of EGFR, forming an extensive network of polar interactions with the catalytic lysine of EGFR (Lys745) and neighboring residues, have been reported to display an exceptional affinity for EGFR C797S variants.[12] Different compounds exploiting this approach have been described lately,[13] including the clinical candidate BLU-945, reported to block $EGFR^{L858R/T790M/C797S}$ in animal models.[14]

An alternative strategy to target EGFR variants insensitive to osimertinib is represented by the development of covalent agents targeting a residue distinct from Cys797.[15] The catalytic lysine of EGFR has attracted growing attention in the discovery of next-generation TCIs in light of its unique chemical and biological properties.[16,17] Indeed, this lysine is (i) part of the kinase catalytic machinery, and its modification would prevent ATP accommodation leading to enzyme inhibition; (ii) it is surrounded by an environment that can increase its reactivity versus electrophilic warheads; and (iii) it is vital for the phosphorylation activity resulting less prone to mutation than other residues. On the other hand, the

Received: December 16, 2022
Published: February 10, 2023



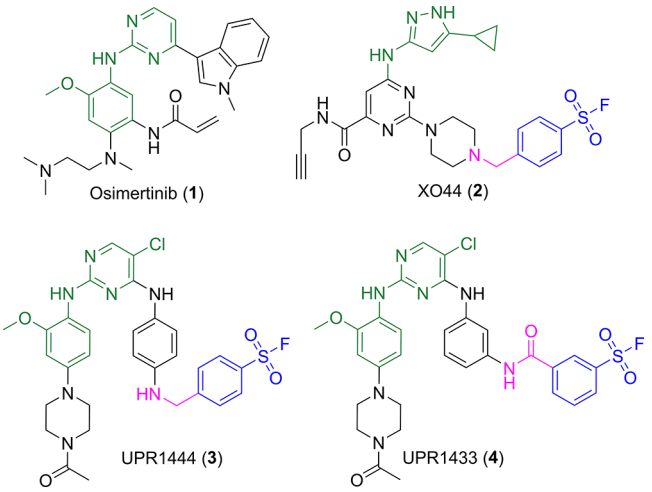


Figure 1. Relevant EGFR inhibitors cited in this work. The reactive warhead of the inhibitors is represented in blue, the linker connecting the warhead and the hinge region binding scaffold is colored in pink, and the hinge region binding scaffold is highlighted in green.

conservation of this residue across the whole kinome anticipates that selectivity might be a severe issue.[18]

Sulfonyl fluorides and related S^{VI}–F groups are emerging as effective warheads for the covalent targeting of nucleophilic residues other than cysteines and serines for chemical biology applications.[19,20] Sulfonyl fluorides are well suited to target the lysine residues in proteins as the resulting reaction product (a sulfonamide) is stable to solvolysis, and it is thus expected to ensure irreversible inhibition of the target of interest.[21] Moreover, S^{VI}–F groups react with nucleophiles only in well-organized environments, such as protein binding sites, in which stable dipoles and residues with hydrogen bonding properties can enhance their electrophilicity and favor the expulsion of the fluorine leaving group (LG).[22] The lack of a well-organized environment prevents sulfur fluoride exchange (SuFEx) process (i.e., sulfonylation) to occur.[23] This feature is expected to dramatically reduce the cross-reactivity of S^{VI}–F groups versus bionucleophiles promoting their use in drug discovery research.

In the context of EGFR inhibition, sulfonyl fluorides have attracted attention thanks to XO44 (**2**, Figure 1).[24] This compound has been recently reported to sulfonylate Lys745 of $EGFR^{L858R/T790M/V948R}$[25] and to inhibit $EGFR^{L858R/T790M/C797S}$ more potently than osimertinib.[26] The lack of selectivity displayed by XO44, inhibiting more than 100 kinases,[25] hampers covalent targeting of Lys745 as a strategy to search for fourth-generation EGFR inhibitors. On the other hand, the use of an EGFR-selective 2-anilinopirimidine hinge binding scaffold has led to the discovery of UPR1444 (**3**, Figure 1), a second-generation sulfonyl fluoride inhibitor able to block $EGFR^{L858R/T790M/C797S}$ activity through the irreversible sulfonylation of Lys745 that spares the activity of other kinases.[26] Thus, UPR1444 has emerged as the first covalent inhibitor of $EGFR^{L858R/T790M/C797S}$ endowed with good selectivity, able to inhibit proliferation of cells resistant to osimertinib.

Despite its promising profile, UPR1444 remains a tool compound featuring suboptimal potency in cells. A deep understanding of its mechanism of action, including structure–activity relationships (SARs), is fundamental for a future optimization of the class and for an unbiased evaluation of the catalytic lysine targeting approach. Available data suggest that the chemical requirements to achieve an efficient and irreversible inhibition of $EGFR^{L858R/T790M/C797S}$ through lysine sulfonylation are rather strict. The simple modification of the linker connecting the warhead to the 2-aminopirimidine scaffold (as for UPR1433, **4**, Figure 1) dramatically reduces the rate of EGFR inactivation and the stability of the reaction product. Differently to what was observed for UPR1444, $EGFR^{L858R/T790M/C797S}$ inhibition by UPR1433 is less dependent on the time of preincubation and can be fully reversed by dilution.[26]

In the present work, we applied a hybrid quantum mechanics/molecular mechanics (QM/MM) approach,[27,28] coupled with the path collective variables (PCVs) method,[29] to investigate at the atomic level the mechanism of Lys745 sulfonylation in $EGFR^{L858R/T790M/C797S}$. Computational strategies of this kind have allowed to clarify mechanisms of action of enzymes[30,31] and relevant covalent inhibitors.[32−34]

Once a likely mechanism was identified for XO44, for which X-ray coordinates of the covalent adduct with EGFR are available,[25] reaction energetics were computed also for the recently reported inhibitors UPR1444 and UPR1433. Minimum free-energy paths of EGFR sulfonylation were thoroughly analyzed and compared to identify the chemical determinants and key stereo-electronic factors at the basis of an efficient sulfonylation of EGFR catalytic lysine.

RESULTS AND DISCUSSION

$EGFR^{L858R/T790M/C797S}$–XO44 Reactant Complex Model. We started our investigation by preparing a Michaelis complex of $EGFR^{L858R/T790M/C797S}$ and XO44 suitable for QM/MM mechanistic simulations by conveniently modifying the X-ray

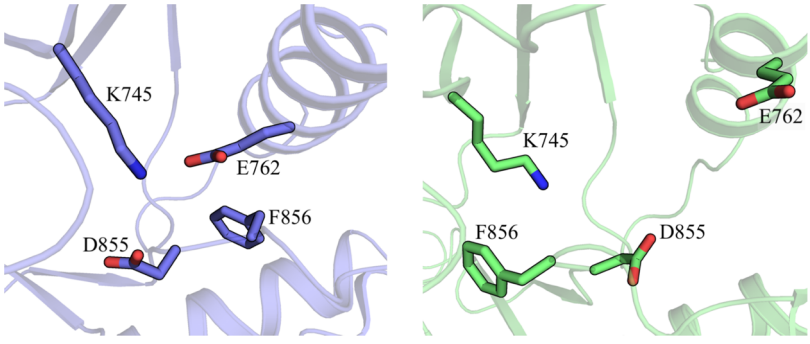
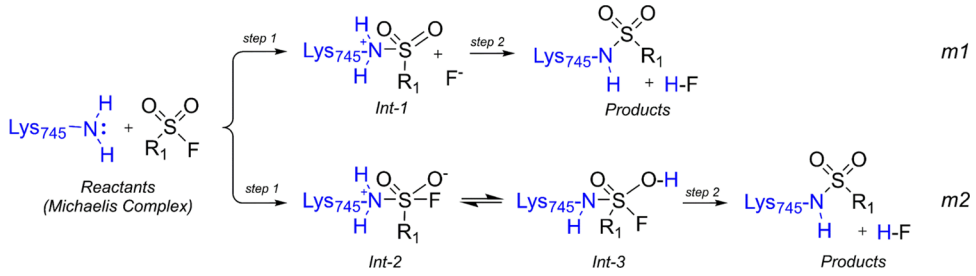


Figure 2. Lys745 environment in EGFR active state represented with violet carbon atoms (left, 1M17.pdb) and in EGFR inactive state colored with green carbon atoms (right, 5HG5.pdb).

https://doi.org/10.1021/acs.jcim.2c01586
J. Chem. Inf. Model. 2023, 63, 1301−1312

Scheme 1. Alternative Mechanisms for Lys745 Sulfonylation by Aromatic Sulfonyl Fluorides, Based on What Have Been Previously Reported in Solution[40−42]



coordinates of the covalent adduct, in which EGFR Lys745 is sulfonylated by XO44 (see the Supporting Information, SI).[25] During the preparation of the EGFR-XO44 noncovalent complex, special attention was given to the protonation state of Lys745.

According to structural data, XO44 recognizes an inactive state of EGFR, in which both the conserved Asp-Phe-Gly (DFG) motif and the αC-helix of the N-lobe of the kinase domain assume an "out" conformation.[35] In this form, known as DFG-out/αC-out (DOCO) inactive state, the basicity of the catalytic lysine is significantly reduced (i.e., by nearly 2 pK_a units)[36] compared to the DFG-in/αC-in (DICI) active state of the kinase. This pK_a downshift is due to the DICI/DOCO transition, which modifies the environment of Lys745. In the DICI state, the catalytic lysine forms salt bridges with Asp855 of the DFG motif and with Glu762 of αC-helix (Figure 2, left). Conversely, in the DOCO state, these interactions are missing since Lys745 is surrounded by hydrophobic residues, including Phe856 of the DFG motif (Figure 2, right). Calculations performed with the empirical scoring function PROPKA3 (which has been tested on 51 distinct lysine residues with an RMSE of 0.65 pK_a units)[37] were consistent with the finding reported by Tsai and colleagues[36] that, in the DOCO inactive state, the basicity of Lys745 is significantly reduced due to a perturbation of the environment. As a result, in the DOCO state, we estimated a downshift of the pK_a of nearly 2.6 log units, compared to the value estimated for the DICI state (predicted pK_a values = ∼9.0 (DOCO) vs 11.6 (DICI), Table S1).

The presence of a micro-environment, destabilizing the protonated form of Lys745 in the DOCO inactive state, along with kinetic data in solution showing that sulfonylation of N-acetyl lysine by aromatic sulfonyl fluorides is pH-dependent with the highest rate reached in basic conditions,[21] indicates that Lys745 likely reacts with XO44 as a free base. Therefore, in the EGFR-XO44 Michaelis complex, the terminal amino group of Lys745 was modeled in its neutral form. The complex was equilibrated by classic molecular dynamics (MD) and, once the QM and MM regions were defined (Figure S1), a QM/MM MD simulation was performed. Then, the resulting structure was employed to evaluate likely mechanisms of sulfonylation by applying the adiabatic mapping (AM) approach.[38,39]

Possible Mechanism of Lys745 Sulfonylation by XO44. Nucleophilic substitution at the sulfonyl fluorides may occur by direct substitution of the fluorine atom (mechanism *m1*, Scheme 1)[40] or by an elimination-addition pathway involving the generation of a trigonal bipyramidal intermediate (mechanism *m2*, Scheme 1).[41,42]

We started our investigation by simulating the first step of mechanism *m1* by means of an adiabatic mapping approach at the PM6/AMBER level,[43,44] which emerged as a suitable method to investigate this kind of reactions, as suggested by simulations performed on cluster models using a set of convenient reaction coordinates (Figure S2 and Table S2).

Step 1 was simulated reconstructing (i) a monodimensional (1D) potential energy surface (PES) using [d(S_{SO2F},F_{SO2F})-d(S_{SO2F},N_{Lys745})] as reaction coordinate (RC1), in which d(S_{SO2F},N_{Lys745}) distance describes the nucleophilic attack at the sulfur center by Lys745 nitrogen, and d(S_{SO2F},F_{SO2F}) distance accounts for the expulsion of the fluorine LG; (ii) a bidimensional (2D) PES in which d(S_{SO2F},N_{Lys745}) and d(S_{SO2F},F_{SO2F}) were explored as independent RCs (RC2 and RC3, respectively). Calculations showed that Int-1 was not a stationary point on both 1D and 2D PESs (Figure S3). Visual inspection of Int-1 geometry showed that the sulfonamide group displayed an arrangement resembling the one observed in the X-ray structure of EGFR-XO44 covalent adduct. The fluorine anion, while still close to the sulfur center, could form hydrogen-bond (H-bond) interactions with p-loop residues of EGFR active site (Figure 3). Despite these polar interactions,

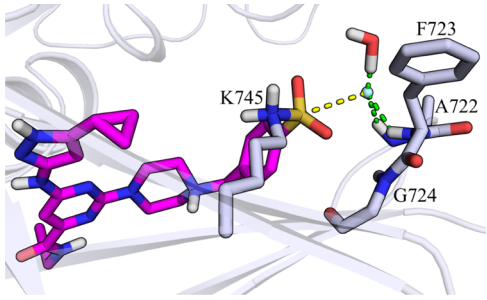


Figure 3. Structure of Int-1. Catalytic lysine Lys745 and p-loop residues (Ala722, Phe723, and Gly724) of EGFR interacting with the sulfonyl fluoride warhead of XO44 (**2**, pink carbon atoms) are represented with gray carbon atoms. The secondary structure of the kinase is displayed in gray cartoon. Polar interactions undertaken by XO44 with p-loop residues and a water molecule are highlighted as dashed lines colored in green. Yellow dashed lines represent the breakage of the S−F bond.

the high energy content of this configuration (∼28 kcal/mol over the level of the reactants) suggested that the enzymatic environment did not sufficiently stabilize the negative charge on the fluorine atom. This led us to speculate that sulfonylation of Lys745 does not occur following a direct substitution of the fluorine atom.

We continued our investigation simulating the first step of mechanism *m2* using two independent RCs. The first one,

https://doi.org/10.1021/acs.jcim.2c01586
J. Chem. Inf. Model. 2023, 63, 1301−1312

$d(S_{SO2F},N_{Lys745})$ reported as RC2, accounting for the nucleophilic attack, and the second one (RC4), defined by the difference of two distances $[d(N_{Lys745},H_{Lys745})\text{-}d(O_{SO2F},H_{Lys745})]$, describing the movement of one of the $-NH_2$ hydrogens of Lys745 on the sulfonyl fluoride oxygen O_{SO2F} of XO44. Topology of the resulting 2D PES (Figure S4a) showed the presence of only two minima, corresponding to the geometries of the Michaelis complex and of the Int-3 intermediate (Figure 4). Int-2 was not a stationary point on the QM/MM surface and a simple unrestrained minimization was sufficient to conduct the system back to the reactants. The energy barrier separating the Michaelis complex and Int-3 was rather high (~28 kcal/mol, Figure S4a), while the energy content of the reaction intermediate was fairly lower (~16 kcal/mol, Figure S4a). Visual inspection of the resulting geometries confirmed that Int-3 was characterized by a trigonal bipyramidal arrangement (Figure 4).

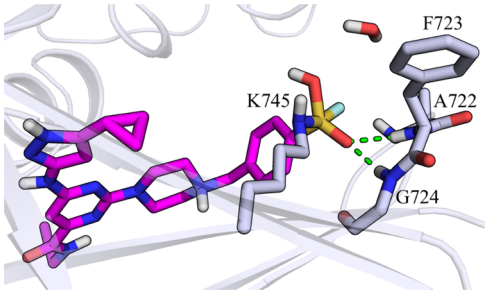


Figure 4. Structure of Int-3. Catalytic lysine Lys745 and p-loop residues (Ala722, Phe723, and Gly724) of EGFR interacting with the sulfonyl fluoride warhead of XO44 (**2**, pink carbon atoms) are represented with gray carbon atoms. The secondary structure of the kinase is displayed in gray cartoon. Green dashed lines represent polar interaction undertaken by XO44 with p-loop residues.

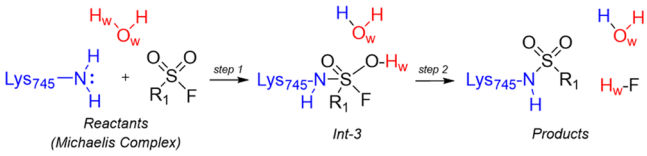
One of the sulfonyl fluoride oxygens O_{SO2F} was engaged in a set of H-bonds with p-loop residues of the kinase. Furthermore, the former hydrogen of Lys745, now bound to the other sulfonyl oxygen of XO44, appeared well oriented to be transferred to the fluorine atom to generate the final product of the reaction.

The subsequent adiabatic mapping performed employing the combination of distances $[d(S_{SO2F},F_{SO2F}) + d(O_{SO2F},H_{Lys745}) - d(F_{SO2F},H_{Lys745})]$ as RC (RC5, Figure S4b), which described the simultaneous expulsion of the fluorine atom and its direct protonation, showed that the generation of the products from Int-3 requires only 5.0 kcal/mol to occur (Figure S4b). Analysis of the geometries along the path showed that S−F breakage occured once the protonation at the fluorine atom is completed. Mechanism *m2* emerged as a sounding mechanism for Lys745 sulfonylation, although the first step featured a rather high energy barrier (Figure S5).

To search for an explanation for the higher barrier of step 1, the geometry of TS1, which resulted from the synchronous nucleophilic attack by Lys745 nitrogen at the sulfur center and by the proton transfer from Lys745 to O_{SO2F}, was thoroughly analyzed. The N−H−O atoms involved in this proton transfer were poorly aligned forming an angle of ~106° (far from the ideal value of 180°), somehow accounting for the high energy barrier computed for mechanism *m2*.

We looked at the EGFR-XO44 X-ray adduct searching for residues that could assist the reaction. A crystallographic water molecule (*wat1*) was identified in proximity of Lys745 nitrogen, prompting us to hypothesize that such water molecule could act as proton shuttle for step 1. Calculations were thus repeated including this water molecule in the definition of the RCs, according to the mechanism *m3* reported in Scheme 2.

Scheme 2. Mechanism *m3* for Lys745 Sulfonylation



The $d(S_{SO2F},N_{Lys745})$ distance, accounting for the nucleophilic attack, and the polynomial expression $[d(N_{Lys745},H_{Lys745}) - d(O_w,H_{Lys745}) + d(O_w,H_w) - d(O_{SO2F},H_w)]$, describing the movement of one of the $-NH_2$ hydrogens of Lys745 on the oxygen atom of the water molecule O_w and the transfer of the one the hydrogens of the water molecule H_w on the oxygen atom O_{SO2F} of the sulfonyl fluoride warhead, were used as RCs (RC2 and RC6, respectively) to reconstruct a 2D PES (Figure 5a).

The 2D PES reported in Figure 5a showed that the first step of the inhibition mechanism *m3* of Lys745 sulfonylation took place with a significantly lower barrier (~16 kcal/mol), consistent with a fast process of sulfonylation. The resulting intermediate Int-3 was employed to simulate step 2 using the combination of distances $[d(S_{SO2F},F_{SO2F}) + d(O_{SO2F},H_w) - d(F_{SO2F},H_w)]$ as RC. The 1D PES reported in Figure 5b revealed that LG expulsion was prompted by protonation of the fluorine atom, requiring to overcome a barrier of ~7.5 kcal/mol to occur. The formation of a neutral product was also evidenced by a charge analysis performed on the reactive atoms showing that, differently from mechanism *m1*, the fluorine atom did not assume a net negative charge (Figures S6−S8).

According to the identified minimum potential energy paths, mechanism *m3* (Figure 6, red line) emerged as the most likely mechanism for Lys745 sulfonylation by XO44 in EGFR. This result pointed out that the architecture of EGFR binding site favors, among reasonable sulfonylation mechanisms that may occur in solution,[40−42] a specific process in which proton transfers assisted by a water molecule concur to lower the energy barriers of the key TSs of the SuFEx reaction.

Potential of Mean Force (PMF) for Lys745 Sulfonylation. As the next step of our investigation, we employed umbrella sampling (US)[45] simulations coupled to WHAM analysis[46,47] to reconstruct the free-energy surface (FES) of Lys745 sulfonylation according to mechanism *m3*, which emerged as the one featured by the lowest potential energy barrier at PM6/AMBER level.[43,44] The reaction was simulated in two steps, starting from geometries generated with the adiabatic mapping approach and equilibrated by QM/MM MD simulations (see the SI for details).

The FES of step 1 (formation of Int-3), reported in Figure 7a, was to some extent similar to the corresponding PES (Figure 5a). The minimum free-energy path connecting reactants and the trigonal bipyramidal intermediate Int-3 indicated that the presence of a concerted process in which nucleophilic attack and protonation of the O_{SO2F}, via the crystallographic water molecule *wat1*, were tightly coupled

https://doi.org/10.1021/acs.jcim.2c01586
J. Chem. Inf. Model. 2023, 63, 1301−1312

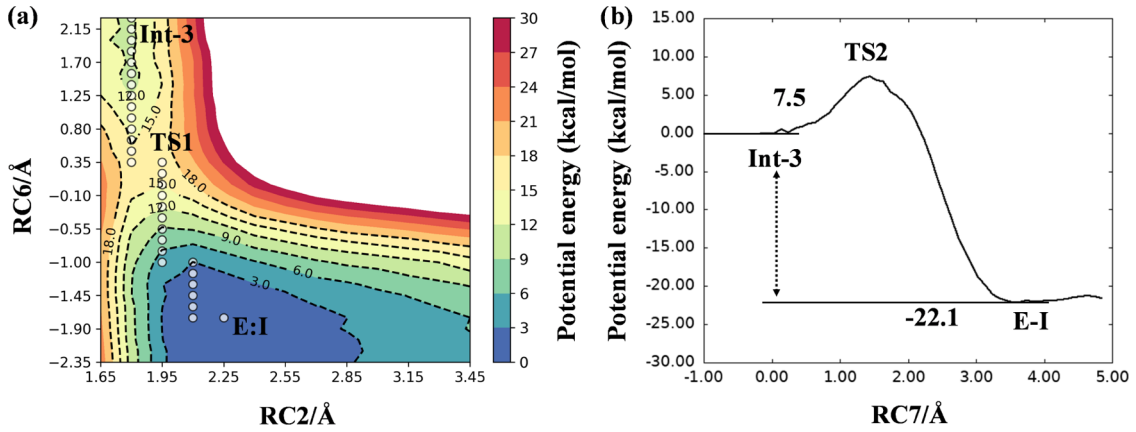


Figure 5. PM6/AMBER PESs for the inhibition mechanism *m3* by XO44. (a) Step 1 of the inhibition mechanism. RC2 corresponds to $d(S_{SO2F},N_{Lys745})$. RC6 corresponds to $[d(N_{Lys745},H_{Lys745}) - d(O_w,H_{Lys745}) + d(O_w,H_w) - d(O_{SO2F},H_w)]$. White dots mark the minimum potential energy path. (b) Step 2 of the inhibition mechanism. RC7 corresponds to $[d(S_{SO2F},F_{SO2F}) + d(O_{SO2F},H_w) - d(F_{SO2F},H_w)]$.

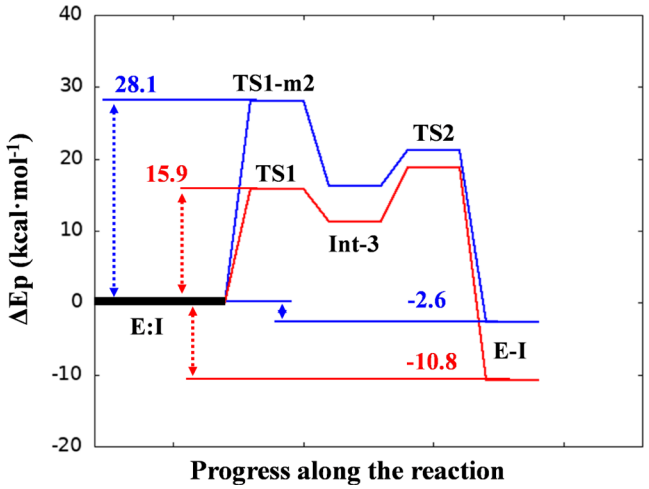


Figure 6. PM6/AMBER potential energy profile describing the inhibition mechanisms *m2* (blue line) and *m3* (red line) by XO44.

events. The barrier for this chemical process was ∼17 kcal/mol, while the energy content of the newly formed intermediate was ∼11 kcal/mol.

Analysis of configurations along the minimum free-energy path showed that the transition state region TS1 was populated by geometries resembling six-membered-ring structures, with the water oxygen of *wat1* (O_w) involved in two proton transfer reactions and the sulfur center at the two opposite endings of the "ring" (Figure 8, TS1 configuration; Figure S10). The FES of the second step showed the presence of a nearly spontaneous process (Figure 7b). These structures were well stabilized by H-bond interactions involving the sulfonyl fluoride warhead and the −NH backbone groups of the p-loop (Figure 8), similarly to what observed in the preliminary exploration of the PES. LG protonation and its subsequent expulsion generated the final product of the reaction overcoming a free-energy barrier lower than ∼1.7 kcal/mol. The second step was highly exergonic with the reaction product significantly more stable than the reactant by ∼11 kcal/mol. With an activation energy (E_{act}) of ∼17 kcal/mol, formation of Int-3 emerged as the rate-limiting step of the whole reaction. This barrier was consistent with experimental

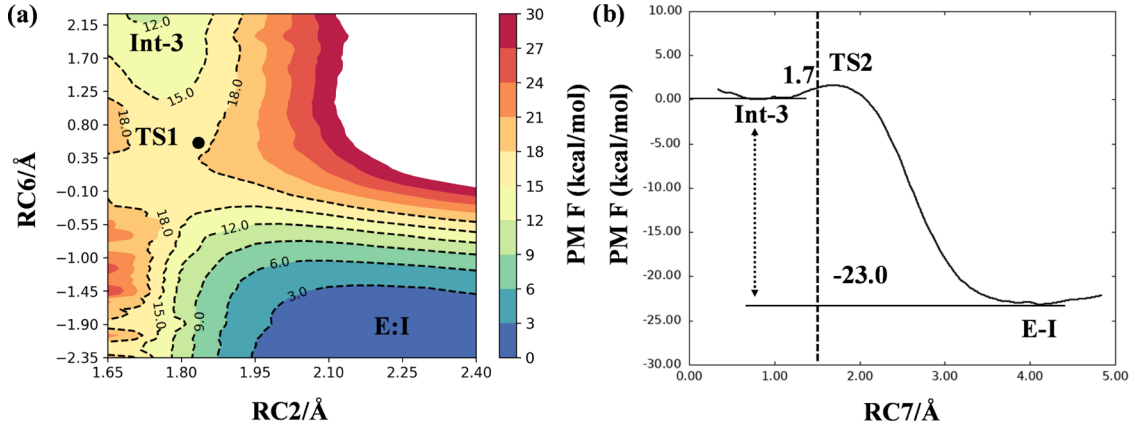


Figure 7. PM6/AMBER FESs for the inhibition mechanism *m3* by XO44. (a) Step 1 of the inhibition mechanism. RC2 corresponds to $d(S_{SO2F},N_{Lys745})$ distance, and RC6 corresponds to the polynomial expression $[d(N_{Lys745},H_{Lys745}) - d(O_w,H_{Lys745}) + d(O_w,H_w) - d(O_{SO2F},H_w)]$. (b) Step 2 of the inhibition mechanism. RC7 corresponds to $[d(S_{SO2F},F_{SO2F}) + d(O_{SO2F},H_w) - d(F_{SO2F},H_w)]$. The position of the optimized TSs at the M06-2X:6-31+G(d,p)/AMBER level is indicated as a black dot (a) and a dashed vertical line (b). Convergence was checked by assessing the evolution of free-energy profiles for each US window (Figure S9).

https://doi.org/10.1021/acs.jcim.2c01586
J. Chem. Inf. Model. 2023, 63, 1301−1312

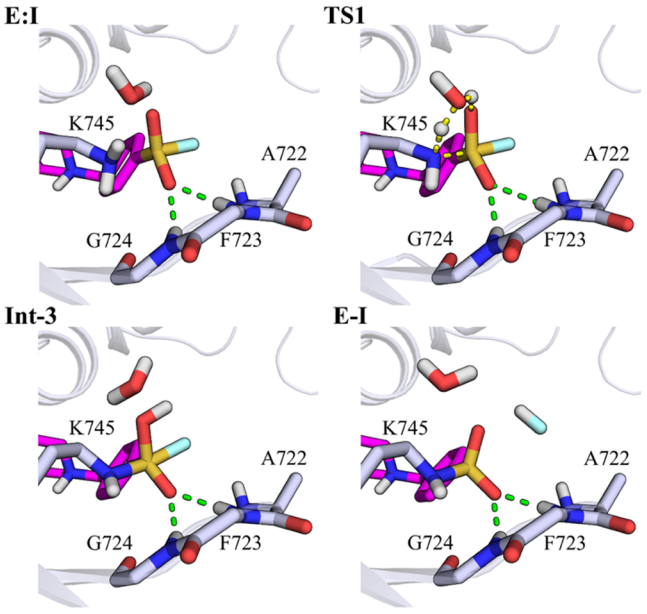


Figure 8. Representative snapshots of the key states of the proposed inhibition mechanism by XO44. Catalytic lysine Lys745 and p-loop residues of EGFR interacting with the sulfonyl fluoride warhead of XO44 (**2**, pink carbon atoms) are represented with gray carbon atoms. The secondary structure of the kinase is displayed in gray cartoon. Green dashed lines represent polar interactions undertaken by XO44 with p-loop residues, and yellow dashed lines underline bond forming and breaking.

data on kinase domain modification, showing that catalytic lysine sulfonylation by XO44 is a rapid process (estimated barrier significantly lower than 20 kcal/mol).[48]

TS structures, optimized at the M06-2X:6-31+G(d,p)/AMBER level, laid in the quadratic region of the PM6/AMBER FESs (Figure 7). Subsequent IRC calculations from optimized TSs led to geometries consistent with E:I, Int-3, and E-I minima, confirming the reaction path deduced from the PM6/AMBER FESs (see the Methods section for details).

Unrestrained QM/MM MD simulations of Int-3 showed that in the time scale of hundreds of picoseconds, this intermediate spontaneously converted in the reaction product, confirming that Int-3 was a transient configuration along the path.

Path Collective Variables (PCVs) Simulations. To further support mechanism *m3* while describing the energetics of the overall reaction without separating it into discrete chemical steps (i.e., step 1, from reactants to Int-3, and step 2, from Int-3 to the products), we applied the path collective variables (PCVs) approach.[49−52] This method allowed to connect reactants and products in a single simulation exploring the possible concerted nature of chemical reactions.

The PCVs method requires the definition of solely two descriptors, representing the progress along a reference reaction path (S), and the distance from it (Z; see the Methods section for details). Geometries corresponding to configurations connecting the reactants to Int-3 and Int-3 to the products (Figure 5) were used to build an initial guess-path that was optimized through a set of consecutive QM/MM steered-MD (SMD) simulations.[53] The variable S assured the progress of the reaction starting from the reactants ($S = 0$) up to the formation of the products ($S = 1$), and the variable Z allowed the system to explore geometries alternative to those defining the guess-path. Analysis of the geometries collected from consecutive SMD simulations (Figures S11 and S12) showed that the nodes of the reference path were minorly affected by the optimization procedure, suggesting that the starting guess-path was constituted by low-energy configurations. The optimization was thus stopped after six consecutive runs, and we performed five independent SMD simulations on the converged path to obtain information on reaction energetics.

The resulting free-energy profile, calculated according to Jarzynski's equality,[54,55] and the geometries of relevant configurations are reported in Figure 9. Consistently with

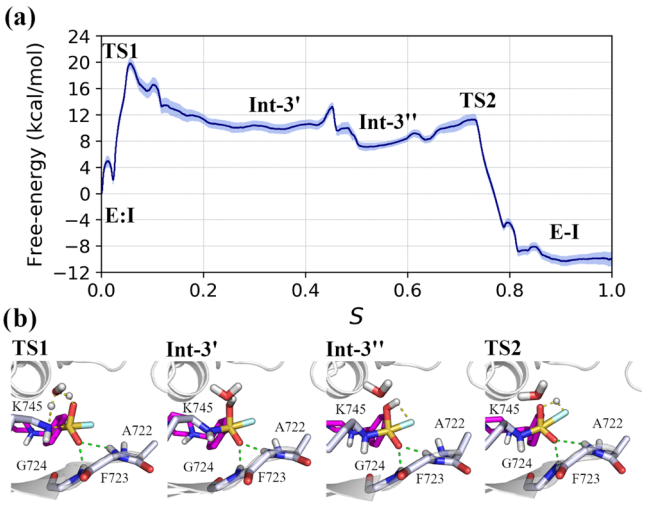


Figure 9. Lys745 sulfonylation by XO44. (a) Averaged free-energy profile of the reaction from reactants (E:I) to products (E-I); light-blue shadow represents the standard deviation. (b) Geometries of relevant steps (TS1 to TS2), taken from one of the five SMD simulations performed on the converged path, are represented. Catalytic lysine Lys745 and p-loop residues of EGFR interacting with the sulfonyl fluoride warhead of XO44 (**2**, pink carbon atoms) are represented with gray carbon atoms. The secondary structure of the kinase is displayed in gray cartoon. Green dashed lines represent polar interactions undertaken by XO44 with p-loop residues, and yellow dashed lines underline bond forming and breaking.

simulations based on the US with classical RCs, the reaction initiated with the concerted nucleophilic attack by Lys745 nitrogen atom at the sulfur center, and the proton transfer from Lys745 to *wat1* and from *wat1* to O_{SO2F} (Figure 9b, TS1). Configurations of the system corresponding to this event featured the same geometries that populated the TS1 region identified by US simulations, with the peculiar six-membered ring structure in which the oxygen atom of *wat1* and the sulfur center were placed at the opposite ends of the ring (Figure S10). The barrier of this event was of ∼20 kcal/mol (a value 3 kcal/mol higher compared to the one from US simulations) and led to the formation of a metastable intermediate (Int-3′) with an energy content of ∼11 kcal/mol, that evolved to a lower-energy intermediate (Int-3″, ∼8 kcal/mol) in which the $O_{SO2F}H_w$ group attached to sulfur center pointed toward the F_{SO2F} atom. This intermediate rapidly progressed to the formation of the product of the reaction, passing an energy barrier of only ∼4 kcal/mol in which protonation and expulsion of the LG occured (TS2). In accordance with the results of US simulations, the water-assisted nucleophilic attack

https://doi.org/10.1021/acs.jcim.2c01586
J. Chem. Inf. Model. 2023, 63, 1301−1312

performed by Lys745 thus emerged as the rate-limiting step of the entire reaction.

Reaction Energetics for Other Sulfonyl Fluoride Derivatives. As a final step of our investigation, we reconstructed the FESs of Lys745 sulfonylation also for UPR1444 and UPR1433 inhibitors. To this end, we applied the same supervised computational protocol, initially employed for XO44, which, being based on reaction coordinates coupled with US simulations, allowed a finer description of the reaction energetics.[56] Analysis of the resulting PESs and FESs showed that the three inhibitors share the same mechanism (Figures S13–S18) with the formation of Int-3 as the rate-limiting step of the overall process.

Activation energy barriers (E_{act}) for XO44, UPR1444, and UPR1433 (~17, ~19, and ~27 kcal/mol, respectively, Figure 10) were qualitatively consistent with inhibition potencies

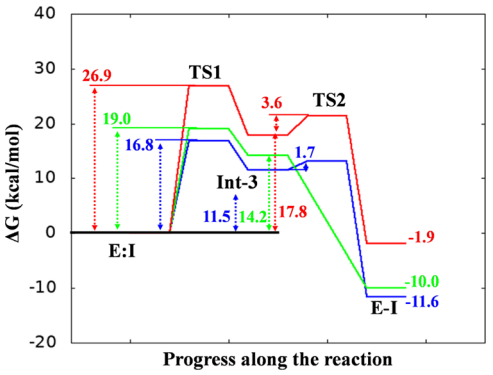


Figure 10. PM6/AMBER free-energy profiles derived from the PMFs for the inhibition mechanism of EGFR by XO44 (blue line), UPR1444 (green line), and UPR1433 (red line).

reported for these compounds (6, 1.6, and 146 nM, respectively).[26] Calculations also provided a mechanistic rationale for the different kinetic behaviors displayed by these inhibitors. While XO44 and UPR1444 efficiently sulfonylated Lys745, leading to a stable covalent adduct, UPR1433 failed to covalently engage the same lysine when tested in similar experimental conditions.[26] Moreover, the reaction-free energies (calculated as the difference between the energies of the product and the reactant) indicated that EGFR sulfonylation by both XO44 and UPR1444 was highly exergonic, with a computed ΔG lower than or equal to −10 kcal/mol (Figure 10). This computational finding agreed with the irreversible mechanism of action displayed by both UPR1444 and XO44. Recent findings on sulfur(VI) fluoride agents suggest that the accessibility of the lowest unoccupied molecular orbital (LUMO) controls the reaction rate of the SuFEx process.[48] This seems to not be the case for the class of aromatic sulfonyl fluorides considered here. UPR1433, while possessing a LUMO energy value significantly lower than that of XO44 and UPR1444 (Table S3), resulted less efficient in the sulfonylation of Lys745. This suggests that the enzymatic environment is at least as important as the intrinsic reactivity of the warhead in the context of aliphatic amine sulfonylation by aromatic sulfonyl fluorides.

We thus analyzed the geometries along the path connecting E:I and Int-3 for the three inhibitors to search for structural determinants accounting for the different computed E_{act}. From this geometry analysis, the role of H-bond interactions, involving the sulfonyl fluoride warhead and the −NH backbone group of the p-loop residues, emerged. The average interatomic distances obtained from the PM6/AMBER FESs calculated for the first step of the reaction (Tables S4–S6) indicated that strong interactions were established between the O_{SO2F} atom, not involved in the proton transfer, and the p-loop residues Phe723 and Gly724 during the reaction of EGFR with both XO44 (O_{SO2F}–H_{Phe723} distance 2.04 ± 0.16 Å and O_{SO2F}–H_{Gly724} distance 2.16 ± 0.19 Å in the TS1, Table S4) and UPR1444 (O_{SO2F}–H_{Phe723} distance 2.21 ± 0.21 Å and

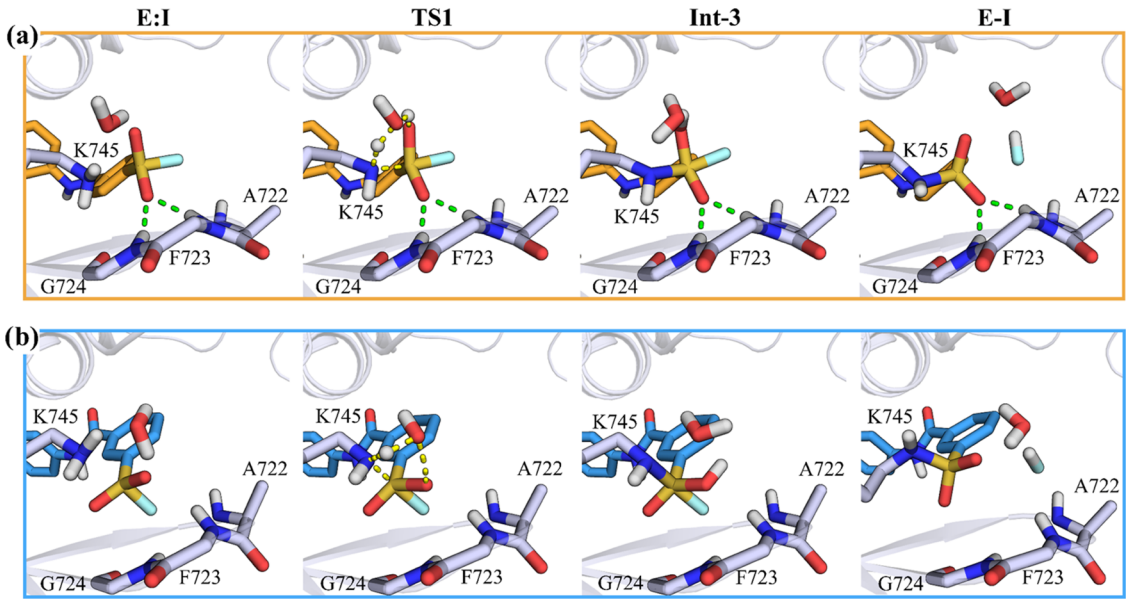


Figure 11. Representative snapshots of the key states of the proposed inhibition mechanism by UPR1444 (**3**, orange carbon atoms, a) and UPR1433 (**4**, blue carbon atoms, b). Catalytic lysine Lys745 and p-loop residues of EGFR interacting with the sulfonyl fluoride warhead of UPR1444/UPR1433 are represented with gray carbon atoms. The secondary structure of the kinase is displayed in gray cartoon. Green dashed lines represent polar interactions undertaken by UPR1444 with p-loop residues, and yellow dashed lines underline bond forming and breaking.

https://doi.org/10.1021/acs.jcim.2c01586
J. Chem. Inf. Model. 2023, 63, 1301−1312

O_{SO2F}−H_{Gly724} distance 3.06 ± 025 Å in the TS1, Table S5), as also shown in Figures 8 and 11a. These interactions were missing in the case of UPR1433 (O_{SO2F}−H_{Phe723} distance 4.12 ± 0.22 Å and O_{SO2F}−H_{Gly724} distance 3.32 ± 0.30 Å in the TS1, Table S6), which instead placed the fluorine atom under the p-loop (Table S6 and Figure 11b). The different arrangement assumed by the sulfonyl fluoride warhead of UPR1433 within the active site of $EGFR^{L858R/T790M/C797S}$ can be attributed to the rigid carbamoyl linker connecting the recognition scaffold targeting the hinge region of the kinase and the warhead itself. In the case of XO44 and UPR1444, the presence of a flexible methylamino linker allowed to place the warhead in a region of the EGFR active site suitable for the attack by Lys745. Overall, these results suggest that the high energy of TS1 for UPR1433 is related to the lack of electrostatic stabilization provided by the environment.

CONCLUSIONS

Even though the third-generation inhibitor osimertinib represents a valid therapeutic option for NSCLC patients, the occurrence of new mutations, including C797S, is responsible for the loss of effectiveness of osimertinib and other EGFR inhibitors in lung cancer therapies. For this reason, the search for inhibitors able to covalently bind the catalytic Lys745 has emerged as a promising approach to identify fourth-generation EGFR inhibitors for the treatment of NSCLC patients.

The application of atomistic investigations aimed at elucidating the catalytic mechanism of lysine sulfonylation is a key and unavoidable process to evaluate novel candidates targeting Lys745. In this work, we applied different hybrid QM/MM approaches using enhanced sampling methods, US and PCVs, to characterize the mechanism of $EGFR^{L858R/T790M/C797S}$ sulfonylation by the reference sulfonyl fluoride agent XO44, and by two recently described inhibitors UPR1444 and UPR1433.

Analysis of reaction energetics and visual inspection of key geometries along the minimum free-energy path led to the identification of the key electrostatic interactions at the basis of an efficient sulfonylation of Lys745. Simulations indicated that the presence of stable protein dipoles in the active site (i.e., −NH backbone groups of Phe723 and Gly724), able to form hydrogen bonds with the sulfonyl fluoride warheads, is critical to efficiently stabilize the main TS of the reaction. This is consistent with recent findings reported by the Sharpless group showing that a "perfect electrostatic matchmaking" between S^{VI}−F warhead and a protein environment is required to observe a sulfur fluoride exchange (SuFEx) process in enzymes.[22,23]

Our simulations pointed to the presence of a flexible linker connecting the S^{VI}−F warhead to the hinge region binding scaffold, as the structural and chemical determinant allowing Lys745 modification. This is the case of XO44 and UPR1444 that, thanks to their methylamino linker, can accommodate their sulfonyl fluoride warhead close to the polar backbone of Phe723 and Gly724. On the contrary, UPR1433 bearing a rigid amide linker fails to accommodate the sulfonyl fluoride group in a position prone to form stabilizing interactions with Phe723 and Gly724 in the main TS (TS1) leading to Lys745 sulfonylation.

Overall, our computational approach was able to provide a sounding mechanism for sulfonylating agents and to discriminate irreversible from reversible inhibitors of $EGFR^{L858R/T790M/C797S}$, thus emerging as a suitable protocol to evaluate novel potential TCIs targeting the catalytic lysine of EGFR.

METHODS

Model Building and Equilibration. The Michaelis complexes of EGFR with **2**−**4** were prepared exploiting the covalent adduct of 5U8L.pdb X-ray structure (**2**) and previous covalent docking models (**3**−**4**).[26] These covalent models were modified by restoring the sulfonyl fluoride warhead of the inhibitors, and the atom type of the terminal nitrogen of Lys745 was conveniently adjusted and modeled in its neutral form accordingly to PROPKA3 prediction (Table S1). The resulting Michaelis complexes were solvated, neutralized, and parametrized with AMBER.[44] AMBER ff03 force field[57,58] and GAFF[59] were applied to model the protein and the inhibitors, respectively. The systems were further minimized with the AMBER ff03 force field and equilibrated under NVT and NPT conditions, increasing the temperature up to 300 K and gradually reducing constraints on both the inhibitor and the protein. The production phase was carried out for 10 ns under NVT conditions. Starting from these equilibrated structures, the reaction mechanisms were investigated using a QM/MM approach (see the Supporting Information, SI, for details).

Application of the QM/MM Potential. We preliminary performed a computational test on semiempirical methods (AM1,[60] PM3,[61] and PM6[43]) available in AMBER aimed at identifying a reliable method to describe the reaction of sulfonylation of a primary amine by an aromatic sulfonyl fluoride compound. These results were compared with calculations performed by modeling the mechanism *m2* at M06-2X[62] and MP2[63−65] levels of theory with 6-31G+(d,p) as basis set[66] employing the Gaussian16 program.[67] Calculations on a cluster model point at PM6 as a fair approach to describe this process in terms of reaction energetics compared to calculations performed at M06-2X[62] and MP2[63−65] levels (see the SI for details).

For this reason, the QM region of EGFR−inhibitor complexes was described by the PM6 Hamiltonian[43] while the rest of the system was described by the AMBER ff03 force field[57,58] (Figure S1). Adiabatic mapping approach was applied for a preliminary evaluation of alternative reaction mechanisms for Lys745 sulfonylation (see the SI for details).

US Simulation for Lys745 Sulfonylation. After the exploration of the different inhibition mechanisms by adiabatic mapping simulations at the PM6/AMBER level of theory, the free-energy surfaces of mechanism *m3* were reconstructed at the same level of theory by means of QM/MM US simulations.[45,68] Potentials of mean force (PMFs) of every single step were obtained using the WHAM approach,[46,47] including in the calculations only the production phase (60 ps) of each simulated window.

For Lys745 sulfonylation by XO44 192 and 51 windows of US simulations for the 2D and 1D FESs were required, respectively. In the case of UPR1444, 156 and 41 US windows were used to reconstruct the 2D and 1D FESs, respectively. For UPR1433, the 2D and 1D FESs were obtained by performing 228 and 49 windows of US simulations, respectively. For all of the inhibitors, each US window was simulated for 60 ps, for a total of 43 ns of simulations performed. Convergence of the computed PMFs is achieved for each US window with an estimated error on the computed energy lower than 0.5 kcal/mol (Figures S9, S15, and S18).

https://doi.org/10.1021/acs.jcim.2c01586
J. Chem. Inf. Model. 2023, 63, 1301−1312

Therefore, differences in E_{act} among the three compounds are statistically significant.

The minimum-energy paths on the surfaces (potential and free energy) were determined using MEPSA software.[69]

Path Collective Variables (PCVs) Simulations. To generate the free-energy landscape of Lys745 sulfonylation by XO44, without dividing it into different chemical steps, we performed path collective variables (PCVs) simulations using the sander module of Amber16[70] coupled with Plumed 2.4.1 (see the SI for details),[71,72] as previously described.[56]

An initial guess-path was generated from representative geometries belonging to the PES of step 1 and step 2 of the proposed mechanism *m3* (taken from Figure 5), following the procedure described by Branduardi et al.[29] The following atoms were used as reference atoms to describe the reaction path: (i) S_{SO2F}, F_{SO2F}, the two O_{SO2F} atoms, and the C atom at position 1 of XO44; (ii) the Cε, N_{Lys745}, and the two H_{Lys745} of Lys745; and (iii) all of the atoms of the crystalized water molecule *wat1*. Forty frames, spaced by an average root-mean-squared deviation (RMSD) of 0.18 Å, represented the initial guess-path connecting the reactants to the products. Two descriptors, S and Z, were used to define the position of the reference atoms (R) with respect to the guess-path. The two variables are defined as

$$S(R) = \frac{\sum_{i=0}^{P} i\ e^{-\lambda(R-R(i))^2}}{\sum_{i=0}^{P} e^{-\lambda(R-R(i))^2}} \quad (1)$$

$$Z(R) = -\frac{1}{\lambda}\ln\left(\sum_{i=1}^{P} e^{-\lambda(R-R(i))^2}\right) \quad (2)$$

where i is a discrete index ranging, in our study, from 0 to 1, $(R - R(i))^2$ is the squared deviation from the i-th frame of the path, while λ is a smoothing parameter. The system was pulled along the variable S from 0 (Michaelis complex) to 1 (products) by applying a force constant of 300 kcal/mol, and the overall simulation was performed at a velocity of 0.3 S steps/ps. The variable Z was constrained by applying a quartic wall to allow the system to explore alternative geometries in proximity of the reference path while preventing it to move exceedingly far from the path. An upper limit was set to 0.005 Å^2 with a force constant of 300 kcal/mol·Å^4. At the end of each steered-MD (SMD) simulation,[53] the frames collected were aligned to the first one and interpolated by applying the Catmull−Rom method,[73] to generate a sufficiently large set of geometries from which to retrieve a smooth guess-path for a subsequent SMD/PCVs simulation. At each iteration, an RMSD-based distance matrix was computed to extract a new set of reference structures, equally separated from each other (∼0.01 Å), to be used as guess-path for the following SMD run. The procedure was applied iteratively for six rounds, when the reaction path did not change anymore moving from an SMD simulation to the following one (Figures S11 and S12). The converged path was finally used to perform five independent SMD/PCVs simulations. The final free-energy profile was obtained by averaging the work of the five simulations, following the application of Jarzynski's equality in SMD simulations.[54,55]

Characterization of the TSs at the DFT/AMBER Level of Theory. Structures of all important states involved in the reaction mechanisms (minima and transition state structures) were optimized at the M06-2X/6-31+G(d,p)/AMBER level,[62,66] starting from representative PM6/AMBER snapshots from the FESs, with Gaussian 09[74] coupled to the fDynamo library,[75,76] as previously described (see the SI for details, Tables S7−S12).

■ ASSOCIATED CONTENT

Data Availability Statement

Molecular structures and input files used to run the simulations are made available as Supporting Information. The Amber16, Amber20, and AmberTools packages (https://ambermd.org) and Gaussian 09 (Revision A.01; https://gaussian.com) are available under license. PLUMED (https://www.plumed.org/) and fDynamo 2.2 (https://www.pdynamo.org/home) are open-source plugins.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.2c01586.

Detailed description of the model building and equilibration procedure, technical details regarding the QM/MM simulations performed at the PM6/AMBER level and cluster model calculations; computational results and details for alternative mechanisms for sulfonylation by AM, US, and PCVs simulations; charge analysis of QM atoms for mechanisms *m1*−*m3*; details of characterization of the TSs at the DFT/AMBER level of theory; and frontier orbital energies for modeled inhibitors (PDF)

Geometries of the transition state (TS1) for EGFR sulfonylation by XO44, UPR1444, and UPR1433 (ZIP)

Input files of Path-CVs simulations (round 6; ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Alessio Lodola − *Dipartimento di Scienze degli Alimenti e del Farmaco, Università degli Studi di Parma, I- 43124 Parma, Italy;* orcid.org/0000-0002-8675-1002; Phone: +39 0521 905062; Email: alessio.lodola@unipr.it; Fax: +39 0521 905006

Authors

Kemel Arafet − *Dipartimento di Scienze degli Alimenti e del Farmaco, Università degli Studi di Parma, I- 43124 Parma, Italy; BioComp Group, Institute of Advanced Materials (INAM), Universitat Jaume I, 12071 Castelló, Spain*

Laura Scalvini − *Dipartimento di Scienze degli Alimenti e del Farmaco, Università degli Studi di Parma, I- 43124 Parma, Italy;* orcid.org/0000-0003-3610-527X

Francesca Galvani − *Dipartimento di Scienze degli Alimenti e del Farmaco, Università degli Studi di Parma, I- 43124 Parma, Italy;* orcid.org/0000-0002-7543-9945

Sergio Martí − *BioComp Group, Institute of Advanced Materials (INAM), Universitat Jaume I, 12071 Castelló, Spain;* orcid.org/0000-0002-1087-7143

Vicent Moliner − *BioComp Group, Institute of Advanced Materials (INAM), Universitat Jaume I, 12071 Castelló, Spain;* orcid.org/0000-0002-3665-3391

Marco Mor − *Dipartimento di Scienze degli Alimenti e del Farmaco, Università degli Studi di Parma, I- 43124 Parma, Italy; Microbiome Research Hub, University of Parma, I-43124 Parma, Italy;* orcid.org/0000-0003-0199-1849

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.2c01586

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

This work was supported by Generalitat Valenciana (APOSTD/2020/015 to K.A.), by the Spanish Ministerio de Ciencia e Innovación (PGC2021-23332OB-C21 to V.M.), and by the Generalitat Valenciana (PROMETEO, CIPROM/2021/079, to V.M.). Regione Emilia-Romagna is kindly acknowledged for support of the Ph.D. program in Drug Science (to F.G.) at the University of Parma through the POR FSE 2014/2020 program.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research benefits from the HPC (High Performance Computing) facility of the University of Parma, Italy (http://www.hpc.unipr.it), and the local computational resources of the Servei d'Informatica of Universitat Jaume I. Associazione Italiana per la Ricerca sul Cancro (AIRC) is also acknowledged.

■ ABBREVIATIONS

EGFR, epidermal growth factor receptor; NSCLC, non-small-cell lung cancer; TCI, targeted covalent inhibitors; SuFEx, sulfur fluoride exchange; QM/MM, quantum mechanics/molecular mechanics; RC, reaction coordinate; AM, adiabatic mapping; US, umprella sampling; PCVs, path collective variables; FES, free-energy surface; PES, potential energy surface; WHAM, weighted histogram analysis method

■ REFERENCES

(1) Hynes, N. E.; Lane, H. A. ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat. Rev. Cancer* **2005**, *5*, 341−354.
(2) Mendelsohn, J.; Baselga, J. The EGF receptor family as targets for cancer therapy. *Oncogene* **2000**, *19*, 6550−6565.
(3) Normanno, N.; De Luca, A.; Bianco, C.; Strizzi, L.; Mancino, M.; Maiello, M. R.; Carotenuto, A.; De Feo, G.; Caponigro, F.; Salomon, D. S. Epidermal growth factor receptor (EGFR) signaling in cancer. *Gene* **2006**, *366*, 2−16.
(4) Finlay, M. R. V.; Anderton, M.; Ashton, S.; Ballard, P.; Bethel, P. A.; Box, M. R.; Bradbury, R. H.; Brown, S. J.; Butterworth, S.; Campbell, A.; Chorley, C.; Colclough, N.; Cross, D. A. E.; Currie, G. S.; Grist, M.; Hassall, L.; Hill, G. B.; James, D.; James, M.; Kemmitt, P.; Klinowska, T.; Lamont, G.; Lamont, S. G.; Martin, N.; McFarland, H. L.; Mellor, M. J.; Orme, J. P.; Perkins, D.; Perkins, P.; Richmond, G.; Smith, P.; Ward, R. A.; Waring, M. J.; Whittaker, D.; Wells, S.; Wrigley, G. L. Discovery of a potent and selective EGFR inhibitor (AZD9291) of both sensitizing and T790M resistance mutations that spares the wild type form of the receptor. *J. Med. Chem.* **2014**, *57*, 8249−8267.
(5) Ramalingam, S. S.; Vansteenkiste, J.; Planchard, D.; Cho, B. C.; Gray, J. E.; Ohe, Y.; Zhou, C.; Reungwetwattana, T.; Cheng, Y.; Chewaskulyong, B.; Shah, R.; Cobo, M..; Lee, K. H.; Cheema, P.; Tiseo, M.; John, T.; Lin, M. C.; Imamura, F.; Kurata, T.; Todd, A.; Hodge, R.; Saggese, M.; Rukazenkov, Y.; Soria, J. C. FLAURA Investigators. Overall Survival with Osimertinib in Untreated, EGFR-Mutated Advanced NSCLC. *N. Engl. J. Med.* **2020**, *382*, 41−50.
(6) Leonetti, A.; Sharma, S.; Minari, R.; Perego, P.; Giovannetti, E.; Tiseo, M. Resistance mechanisms to osimertinib in EGFR-mutated non-small cell lung cancer. *Br. J. Cancer* **2019**, *121*, 725−737.
(7) Grabe, T.; Lategahn, J.; Daniel, R. C797S resistance: the undruggable EGFR mutation in non-small cell lung cancer? *ACS Med. Chem. Lett.* **2018**, *9*, 779−782.
(8) Callegari, D.; Ranaghan, K. E.; Woods, C. J.; Minari, R.; Tiseo, M.; Mor, M.; Mulholland, A. J.; Lodola, A. L718Q mutant EGFR escapes covalent inhibition by stabilizing a non-reactive conformation of the lung cancer drug osimertinib. *Chem. Sci.* **2018**, *9*, 2740−2749.
(9) Thress, K. S.; Paweletz, C. P.; Felip, E.; Cho, B. C.; Stetson, D.; Dougherty, B.; Lai, Z.; Markovets, A.; Vivancos, A.; Kuang, Y.; Ercan, D.; Matthews, S. E.; Cantarini, M.; Barrett, J. C.; Jänne, P. A.; Oxnard, G. R. Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. *Nat. Med.* **2015**, *21*, 560−562.
(10) Scalvini, L.; Castelli, R.; La Monica, S.; Tiseo, M.; Alfieri, R. Fighting tertiary mutations in EGFR-driven lung-cancers. *Biochem. Pharmacol.* **2021**, *190*, No. 114643.
(11) Shi, K.; Wang, G.; Pei, J.; Zhang, J.; Wang, J.; Ouyang, L.; Wang, Y.; Li, W. Emerging strategies to overcome resistance to third-generation EGFR inhibitors. *J. Hematol. Oncol.* **2022**, *15*, No. 94.
(12) Heppner, D. E.; Günter, M.; Wittlinger, F.; Laufer, S. A.; Eck, M. J. Structural basis for EGFR mutant inhibition by trisubstituted imidazole inhibitors. *J. Med. Chem.* **2020**, *63*, 4293−4305.
(13) He, J.; Zhou, Z.; Sun, X.; Yang, Z.; Zheng, P.; Xu, S.; Zhu, W. The new opportunities in medicinal chemistry of fourth-generation EGFR inhibitors to overcome C797S mutation. *Eur. J. Med. Chem.* **2021**, *210*, No. 112995.
(14) Eno, M. S.; Brubaker, J. D.; Campbell, J. E.; De Savi, C.; Guzi, T. J.; Williams, B. D.; Wilson, D.; Wilson, K.; Broojimans, N.; Kim, J.; Özen, A.; Perola, E.; Hsieh, J.; Brown, V.; Fetalvero, K.; Garner, A.; Zhang, Z.; Stevison, F.; Woessner, R.; Singh, J.; Timsit, Y.; Kinkema, C.; Medendorp, C.; Lee, C.; Albayya, F.; Zalutskaya, A.; Schalm, S.; Dineen, T. A. Discovery of BLU-945, a reversible, potent, and wild-type-sparing next-generation EGFR mutant inhibitor for treatment-resistant non-small-cell lung cancer. *J. Med. Chem.* **2022**, *65*, 9662−9677.
(15) Mukherjee, H.; Grimster, N. P. Beyond cysteine: recent developments in the area of targeted covalent inhibition. *Curr. Opin. Chem. Biol.* **2018**, *44*, 30−38.
(16) Cuesta, A.; Taunton, J. Lysine-targeted inhibitors and chemoproteomic probes. *Annu. Rev. Biochem.* **2019**, *88*, 365−381.
(17) Pettinger, J.; Jones, K.; Cheeseman, M. D. Lysine-targeting covalent inhibitors. *Angew. Chem., Int. Ed.* **2017**, *56*, 15200−15209.
(18) Lonsdale, R.; Ward, R. A. Structure-based design of targeted covalent inhibitors. *Chem. Soc. Rev.* **2018**, *47*, 3816−3830.
(19) Dong, J.; Krasnova, L.; Finn, M. G.; Sharpless, K. B. Sulfur(VI) fluoride exchange (SuFEx): another good reaction for click chemistry. *Angew. Chem., Int. Ed.* **2014**, *53*, 9430−9448.
(20) Jones, L. H.; Kelly, J. W. Structure-based design and analysis of SuFEx chemical probes. *RSC Med. Chem.* **2020**, *11*, 10−17.
(21) Mukherjee, H.; Debreczeni, J.; Breed, J.; Tentarelli, S.; Aquila, B.; Dowling, J. E.; Whitty, A.; Grimster, N. P. A study of the reactivity of S(VI)−F containing warheads with nucleophilic amino-acid side chains under physiological conditions. *Org. Biomol. Chem.* **2017**, *15*, 9685−9695.
(22) Zheng, Q.; Woehl, J. L.; Kitamura, S.; Santos-Martins, D.; Smedley, C. J.; Li, G.; Forli, S.; Moses, J. E.; Wolan, D. W.; Sharpless, K. B. SuFEx-enabled, agnostic discovery of covalent inhibitors of human neutrophil elastase. *Proc. Natl. Acad. Sci. U.S. A.* **2019**, *116*, 18808−18814.
(23) Chen, W.; Dong, J.; Plate, L.; Mortenson, D. E.; Brighty, G. J.; Li, S.; Liu, Y.; Galmozzi, A.; Lee, P. S.; Hulce, J. J.; Cravatt, B. F.; Saez, E.; Powers, E. T.; Wilson, I. A.; Sharpless, K. B.; Kelly, J. W. Arylfluorosulfates inactivate intracellular lipid binding protein(s) through chemoselective SuFEx reaction with a binding site Tyr residue. *J. Am. Chem. Soc.* **2016**, *138*, 7353−7364.
(24) Gehringer, M.; Laufer, S. A. Emerging and re-emerging warheads for targeted covalent inhibitors: applications in medicinal chemistry and chemical biology. *J. Med. Chem.* **2019**, *62*, 5673−5724.

https://doi.org/10.1021/acs.jcim.2c01586
J. Chem. Inf. Model. 2023, 63, 1301−1312

(25) Zhao, Q.; Ouyang, X.; Wan, X.; Gajiwala, K. S.; Kath, J. C.; Jones, L. H.; Burlingame, A. L.; Taunton, J. Broad-spectrum kinase profiling in live cells with lysine-targeted sulfonyl fluoride probes. *J. Am. Chem. Soc.* **2017**, *139*, 680−685.
(26) Ferlenghi, F.; Scalvini, L.; Vacondio, F.; Castelli, R.; Bozza, N.; Marseglia, G.; Rivara, S.; Lodola, A.; La Monica, S.; Minari, R.; Petronini, P. G.; Alfieri, R.; Tiseo, M.; Mor, M. A sulfonyl fluoride derivative inhibits $EGFR^{L858R/T790M/C797S}$ by covalent modification of the catalytic lysine. *Eur. J. Med. Chem.* **2021**, *225*, No. 113786.
(27) Field, M. J.; Bash, P. A.; Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* **1990**, *11*, 700−733.
(28) van der Kamp, M. W.; Mulholland, A. J. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry* **2013**, *52*, 2708−2728.
(29) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in free energy space. *J. Chem. Phys.* **2007**, *126*, No. 054103.
(30) Chudyk, E. I.; Beer, M.; Limb, M. A. L.; Jones, C. A.; Spencer, J.; van der Kamp, M. W.; Mulholland, A. J. QM/MM simulations reveal the determinants of carbapenemase activity in class A β-lactamases. *ACS Infect. Dis.* **2022**, *8*, 1521−1532.
(31) Ramanan, R.; Chaturvedi, S. S.; Lehnert, N.; Schofiled, C. J.; Karabancheva-Christova, T. M.; Christov, C. T. Catalysis by the JmjC histone demethylase KDM4A integrates substrate dynamics, correlated motions and molecular orbital control. *Chem. Sci.* **2020**, *11*, 9950−9961.
(32) Arafet, K.; Serrano-Aparicio, N.; Lodola, A.; Mulholland, A. J.; González, F. V.; Świderek, K.; Moliner, V. Mechanism of inhibition of SARS-CoV-2 M^{pro} by N3 peptidyl Michael acceptor explained by QM/MM simulations and design of new derivatives with tunable chemical reactivity. *Chem. Sci.* **2021**, *12*, 1433−1444.
(33) Mihalovits, L. M.; Ferenczy, G. G.; Keserű, G. Affinity and selectivity assessment of covalent inhibitors by free energy calculations. *J. Chem. Inf. Model.* **2020**, *60*, 6579−6594.
(34) Galvani, F.; Scalvini, L.; Rivara, S.; Lodola, A.; Mor, M. Mechanistic modeling of monoglyceride lipase covalent modification elucidates the role of leaving group expulsion and discriminates inhibitors with high and low potency. *J. Chem. Inf. Model.* **2022**, *62*, 2771−2787.
(35) Liu, R.; Yue, Z.; Tsai, C.-C.; Shen, J. Assessing lysine and cysteine reactivities for designing targeted covalent kinase inhibitors. *J. Am. Chem. Soc.* **2019**, *141*, 6553−6560.
(36) Tsai, C.-C.; Yue, Z.; Shen, J. How electrostatic coupling enables conformational plasticity in a tyrosine kinase. *J. Am. Chem. Soc.* **2019**, *141*, 15092−15101.
(37) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical pK_a predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525−537.
(38) Lonsdale, R.; Ranaghan, K. E.; Mulholland, A. J. Computational enzymology. *Chem. Commun.* **2010**, *46*, 2354−2372.
(39) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.
(40) Zheng, Q.; Xu, H.; Wang, H.; Du, W.-G. H.; Wang, N.; Xiong, H.; Gu, Y.; Noodleman, L.; Sharpless, K. B.; Yang, G.; Wu, P. Sulfur [^{18}F]fluoride exchange click chemistry enabled ultrafast late-stage radiosynthesis. *J. Am. Chem. Soc.* **2021**, *143*, 3753−3763.
(41) Page, M. I. Beta-sultams-mechanism of reactions and use as inhibitors of serine proteases. *Acc. Chem. Res.* **2004**, *37*, 297−303.
(42) Sung, D. D.; Kim, T. J.; Lee, I. Theoretical studies of the nucleophilic substitution of halides and amine at a sulfonyl center. *J. Phys. Chem. A* **2009**, *113*, 7073−7079.
(43) Stewart, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13*, 1173−1213.
(44) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I. Y.; Berryman, J. T.; Brozell, S. R.; Cerutti, D. S.; Cheatham, I. T. E.; Cisneros, G. A.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izardi, S.; Izmailov, S. A.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O'Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shajan, A.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R.M.; Wu, X.; Xiong, Y.; Xue, Y.; York, D. M.; Zhao, S.; Kollman, P. A. *AMBER 2020*; University of California: San Francisco, 2020.
(45) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(46) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(47) Grossfield, A. *Grossfield, A. WHAM: the weighted histogram analysis method, ver. 2.0.10*. http://membrane.urmc.rochester.edu/wordpress/?page_id=126.
(48) Gilbert, K. E.; Vuorinen, A.; Aaktar, A.; Pogány, J.; Pettinger, J.; Grant, E. K.; Kirkpatrick, J. M.; Rittinger, K.; House, D.; Burley, G. A.; Bush, J. T. Profiling sulfur(VI) fluorides as reactive functionalities for chemical biology tools and expansion of the ligandable proteome. *ACS Chem. Biol.* **2023**, DOI: 10.1021/acschembio.2c00633.
(49) Ensing, B.; De Vivo, M.; Liu, Z.; Moore, P.; Klein, M. L. Metadynamics as a tool for exploring free energy landscapes of chemical reactions. *Acc. Chem. Res.* **2006**, *39*, 73−81.
(50) Lodola, A.; Branduardi, D.; De Vivo, M.; Capoferri, L.; Mor, M.; Piomelli, D.; Cavalli, A. A catalytic mechanism for cysteine *N*-terminal nucleophile hydrolases, as revealed by free energy simulations. *PLoS One* **2012**, *7*, No. e32397.
(51) Capoferri, L.; Lodola, A.; Rivara, S.; Mor, M. Quantum Mechanics/Molecular Mechanics modeling of covalent addition between EGFR−cysteine 797 and n-(4-anilinoquinazolin-6-yl) acrylamide. *J. Chem. Inf. Model.* **2015**, *55*, 589−599.
(52) Elisi, G. M.; Scalvini, L.; Lodola, A.; Mor, M.; Rivara, S. Free-energy simulations support a lipophilic binding route for melatonin receptors. *J. Chem. Inf. Model.* **2022**, *62*, 210−22.
(53) Grubmüller, H.; Heymann, B.; Tavan, P. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* **1996**, *271*, 997−999.
(54) Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **1997**, *78*, 2690−2693.
(55) Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J. Chem. Phys.* **2003**, *119*, 3559−3566.
(56) Scalvini, L.; Ghidini, A.; Lodola, A.; Callegari, D.; Rivara, S.; Piomelli, D.; Piomelli, D.; Mor, M. *N*-Acylethanolamine acid amidase (NAAA): mechanism of palmitoylethanolamide hydrolysis revealed by mechanistic simulations. *ACS Catal.* **2020**, *10*, 11797−11813.
(57) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(58) Lee, M. C.; Duan, Y. Distinguish protein decoys by Using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized born solvent model. *Proteins* **2004**, *55*, 620−634.
(59) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(60) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(61) Stewart, J. J. P. Optimization of parameters for semiempirical methods II. Applications Method. *J. Comput. Chem.* **1989**, *10*, 209−220.
(62) Zhao, Y.; Thrulhar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(63) Møller, C.; Plesset, M. S. Note on an approximation treatment for many-electron systems. *Phys. Rev.* **1934**, *46*, 618.
(64) Pople, J. A.; Seeger, R.; Krishnan, R. Variational configuration interaction methods and comparison with perturbation theory. *Int. J. Quantum Chem.* **2009**, *12*, 149.
(65) Bartlett, R. J. Many-body perturbation theory and coupled cluster theory for electron correlation in molecules. *Annu. Rev. Phys. Chem.* **1981**, *32*, 359.
(66) Hehre, W. J.; Radom, L.; Schleyer, P. V. R.; Pople, A. J. *Ab Initio Molecular Orbital Theory*; John Wiley: New York, 1986.
(67) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A.; Peralta, J. E., Jr.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 (revision B.01)*, 2016.
(68) Lodola, A.; Callegari, D.; Scalvini, L.; Rivara, S.; Mor, M. Design and SAR Analysis of Covalent Inhibitors Driven by Hybrid QM/MM Simulations. In*Quantum Mechanics in Drug Discovery*; Springer, 2020; pp 307−337.
(69) Marcos-Alcalde, I.; Setoain, J.; Mendieta-Moreno, J. I.; Mendieta, J.; Gómez-Puertas, P. MEPSA: minimum energy pathway analysis for energy landscapes. *Bioinformatics* **2015**, *31*, 3853−3855.
(70) Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. M., III; Duke, T. A.; Giese, T. J.; Golke, H.; Goetze, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kolvalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Marmelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; Kollman, P. A. *AMBER 16*; University of California: San Francisco, 2016.
(71) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.
(72) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.
(73) Catmull, E.; Rom, R. A Class of Local Interpolating Splines. In *Computer Aided Geometric Design*; Barnhill, R. E.; Riesenfeld, R. F., Eds.; Academic Press: New York, 1974; pp 317−326.
(74) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A.; Peralta, J. E., Jr.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 09 (revision A. 1)*, 2009.
(75) Field, M. J.; Albe, M.; Bret, C.; Proust-De Martin, F.; Thomas, A. The dynamo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *J. Comp. Chem.* **2000**, *21*, 1088−1100.
(76) Krzemińska, A.; Paneth, P.; Moliner, V.; Swiderek, K. Binding isotope effects as a tool for distinguishing hydrophobic and hydrophilic binding sites of HIV-1 RT. *J. Phys. Chem. B* **2015**, *119*, 917−927.

https://doi.org/10.1021/acs.jcim.2c01586
J. Chem. Inf. Model. 2023, 63, 1301−1312