Structure-Activity Relationships of Hexahydrocyclopenta[*c*]quinoline Derivatives as Allosteric Inhibitors of CDK2 and EGFR

Luca Carlino[+], Michael S. Christodoulou[+], Valentina Restelli, Fabiana Caporuscio, Francesca Foschi, Marta S. Semrau, Elisa Costanzi, Annachiara Tinivella, Luca Pinzi, Leonardo Lo Presti, Roberto Battistutta, Paola Storici, Massimo Broggini, Daniele Passarella, Giulio Rastelli[*]

[+] These authors contributed equally to this work.

Abstract

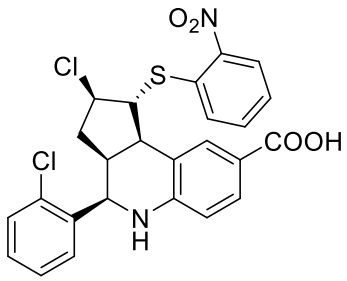
Following the discovery of a type III allosteric modulator of cyclin-dependent kinase 2 (CDK2) characterized by a hexahydrocyclopenta[c]quinolone scaffold, three different series of its derivatives were synthesized and biologically evaluated. Docking of the synthesized compounds into the allosteric pocket of CDK2 allowed the elucidation of structure–activity relationships (SARs). Moreover, the compounds were tested on the wildtype epidermal growth factor receptor (EGFR) kinase domain (KD) and its clinically relevant T790M/L858R mutant form. Herein we describe the first SAR investigation of allosteric ligands that bind to the type III inhibitor pocket of CDK2 and EGFR-KD. Although the activity of the synthesized inhibitors needs to be improved, the obtained results provide clear-cut indications about pharmacophore requirements and selectivity determinants. Remarkably, this study led to the identification of a selective T790M/L858R EGFR allosteric inhibitor that is inactive toward both wild-type EGFR and CDK2. Finally, docking into the T790M/L858R EGFR-KD led us to hypothesize that the compounds bind to the double-mutant EGFR-KD by adopting a binding mode different from that in CDK2, thus rationalizing the observed selectivity profile.

Introduction

Due to their pivotal role in the modulation of cell pathways, protein kinases (PKs) are the targets of choice for a number of human diseases.[1] The standard approach of targeting the ATP binding site has recently come up against selectivity and drug resistance issues, which can be considerably reduced by targeting allosteric pockets.[2] In fact, allosteric modulators of protein kinases bind to less conserved binding sites thus allowing to bypass drug resistance issues clinically observed with ATP-competitive drugs.[3] Moreover, they do not have to compete with the high intracellular levels of ATP, as is the case with type I, I$^{1/2}$, and II inhibitors.

We have recently discovered that the hexahydrocyclopenta[*c*]quinoline derivative **3f** (Figure 1) is an allosteric inhibitor of CDK2.[4] Further investigations by X-ray crystallography, stereoselective synthesis of its pure enantiomers, modelling studies, biological evaluations and mutagenesis experiments provided the first thorough chemical and biological characterization of an allosteric modulator of this protein, and confirmed that **3f** binds CDK2 *via* a type III allosteric mechanism.[5]

In this study, we explored the structure-activity relationships (SARs) of this class of compounds by synthesizing and testing three series of differently substituted derivatives (series **3a-f**, **i**, series **3g**, **h**, **k-q**, **4a**, **b**, **5a**, **b**, **7**, and series **3j**, **r-y**). Moreover, as the CDK2 mechanism of activation was proposed to mimic that of EGFR[6-8] and the two proteins show similar inactive conformations,[9] the compounds were also tested on the kinase domains of *wild-type (wt)* EGFR and the T790M/L858R mutant. The latter double mutant is of particular clinical relevance, because it constitutively activates the *full-length* EGF receptor and is a key driver of drug resistance to EGFR tyrosine kinase inhibitors gefitinib, erlotinib and afatinib.[10,11] Induced-fit docking simulations of the synthesized compounds to the CDK2 and the *wt* and double mutant EGFR allosteric pockets allowed to rationalize and discuss the obtained SARs.



3f

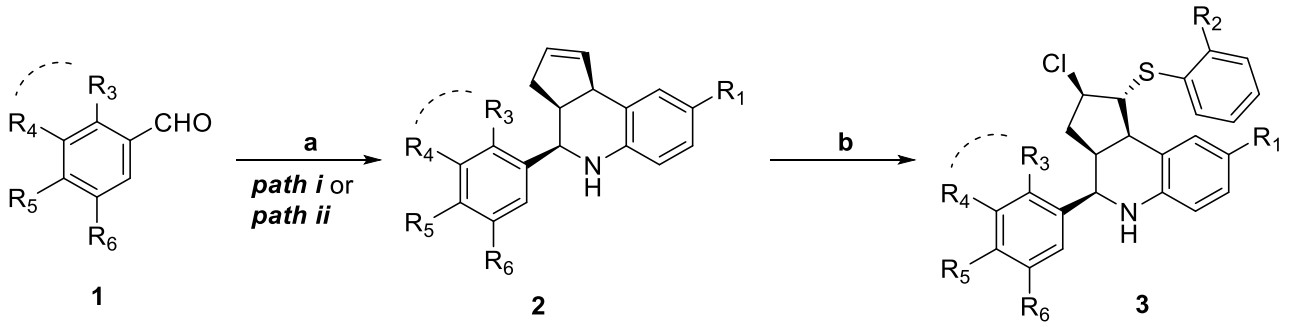
Figure 1: Chemical structure of compound **3f**.

Results and Discussion

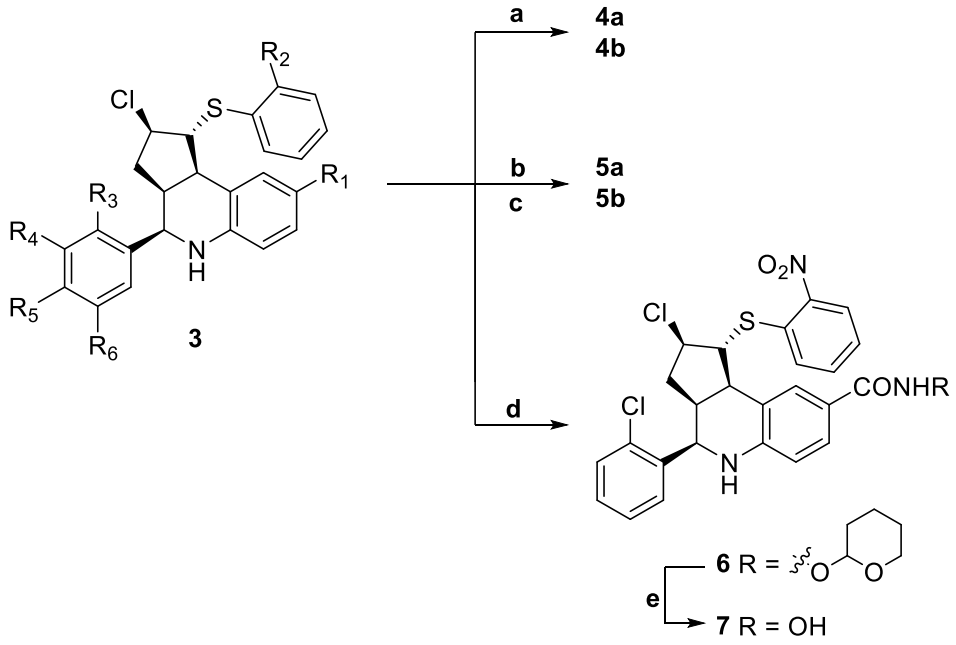
Chemistry

The synthesis of compounds **3a-y** (Scheme 1, Tables 1-3) is based on a multicomponent Povarov reaction to give compounds **2a-y**, followed by an addition reaction with ortho-phenylsulfenyl chloride derivatives. In the Povarov reaction TFA (a. *path i*) was used as catalyst in the presence of either electron-deficient or inactivated aldehydes **1** (synthesis of compounds

2a-u).[5] On the other hand the use of $AlCl_3$ (a. *path ii*) is essential to promote both the imine formation and the cycloaddition in the presence of electron rich aldehydes (synthesis of compounds **2v-y**). For the synthesis of racemic compounds **3a-y**, 2-nitrobenzenesulfenyl chloride was purchased from Alfa Aesar, whereas all the other *ortho*-phenylsulfenyl chlorines were prepared according to the published procedures.[12] Racemic compounds **4a** and **4b** were prepared by a condensation reaction of compounds **3f** and **3i,** respectively, whereas racemic compounds **5a** and **5b** were obtained from compound **3m** by using 1 and 2.5 equiv of MeI, respectively (Scheme 2). Finally, the hydroxamic acid **7** was furnished by deprotection of compound **6** in the presence of HCl (Scheme 2). NMR studies confirmed the relative configuration of the obtained molecules. Further X-ray diffraction experiments of a single crystal of **3j** confirmed the structure (see Experimental Section and Supporting information).



Scheme 1. Reagents and Conditions: a) *path i*-Synthesis of **2a-u**: various amines, cyclopentadiene, TFA, MeCN, 0 °C, 2h, 50 – 90%; *path ii*-Synthesis of **2v-y**: various amines, cyclopentadiene, $AlCl_3$, MeCN, 0 °C, 8h, 58 – 76%; b) various *ortho*-phenylsulfenyl chlorines, MeCN, r.t., 30 min, 50 – 90%.



Scheme 2. Reagents and Conditions: a) $MeNH_2$, HATU, DIPEA, THF, r.t., overnight, 85 – 90%; b) MeI (1.2 equiv), K_2CO_3, DMF, 80 °C, 4h, 70%; c) MeI (2.4 equiv), K_2CO_3, DMF, 80 °C, 4h, 75%; d) NH_2OTHP, HATU, Et_3N, THF, r.t., overnight, 58%; e) HCl (3M), MeOH, r.t., 3 h, 15%.

Structure-activity relationships in CDK2

The progenitor molecule 3 f (IC_{50}=10.2:2.5 mm; Figure S1N) was thoroughly characterized by X-ray crystallography and the structural model of its interaction with CDK2 was obtained by molecular docking to selected CDK2 crystal structures and extensive molecular dynamics simulations of the resulting protein–ligand complexes. The predicted binding mode was validated through mutagenesis experiments and direct binding experiments on the wild-type and mutant CDK2 with microscale thermophoresis.[5] As previously reported, 3 f was proposed to bind to the allosteric pocket of CDK2 with the carboxylate moiety forming a salt bridge with the polar head of the conserved Lys33 residue, and the nitro group making hydrogen bonds with the backbone NHs of the Asp145 and Phe146 residues of the DFG motif (Figure 2).[5]

Therefore, the first series of compounds was designed to investigate the importance of the carboxylic acid and the nitro group (Table 1). Biological evaluations clearly showed that the removal of the carboxylic acid (3c and 3d), the nitro group (3e), or both (3b and 3a), completely abolished binding to CDK2. These results are consistent with the previously reported mutagenesis studies that showed that mutation of Lys33 to an alanine significantly impaired binding of 3 f to CDK2, and provided further validation of the proposed binding mode.[5]

Substitution with an additional chlorine atom in the para position of the 2-phenyl ring attached to the tetrahydroquinoline moiety as shown by compound 3i (IC50=25.8:11.9 mm; Figure S1N) and compound 3d did not affect the CDK2 inhibitory activity. In conclusion, results obtained from this first series of compounds indicated that the carboxylic acid and the nitro group are mandatory for the CDK2 inhibitory activity. Having ascertained that the carboxylic acid and the nitro groups cannot be removed, the second series of compounds was designed to investigate whether the two groups could be replaced by different chemical moieties (Table 2). The acetic acid derivative 3o partially retained CDK2 inhibitory activity. Induced-fit docking of this compound to the CDK2 allosteric site showed that the acetic acid moiety of 3o forms a salt bridge with the Lys33 residue and, in addition, it hydrogen bonds to the backbone NHs of Thr14 and Tyr15 (Figure S1A). On the contrary, the esterification of the carboxylic acid (3k and 3l) or replacement of the carboxylic group with an acetyl group (3p) provided inactive compounds. These modifications strongly affected the mandatory electrostatic interaction with the Lys33 side chain without forming any additional interaction, and an ethyl or a methyl moiety are

accommodated in the proximity of the polar region of the cavity, thus resulting in loss of activity. The modification of the carboxylic acid group with a carboxamide (4a and 4b), an acetamido group (3q), or a hydroxamate (7) provided compounds able to retain CDK2 inhibitory activity. In the CDK2/4a and CDK2/4b complexes, the carbonyl and the amino groups of the amide hydrogen bond to the Lys33 side chain and the Lys34 backbone carbonyl, respectively (Figure S1B). In the acetamido analogue 3q, the amino and the carbonyl groups make hydrogen bonds with the Asp145 side chain and with the Tyr15 and Gly16 backbone carbonyls, respectively (Figure S1C). The hydroxamate moiety of 7 forms hydrogen bonds with the Asp145 side chain (Figure S1D). Moreover, the substitution of the carboxylic acid with sulfonamide moieties provided compounds whose activities are in the order: SO2 NH2>SO2NHMe>SO2NMe2 (compounds 3m, 5a, 5b, respectively). The unsubstituted sulfonamide 3m (IC50=59.7:21.1 mm; Figure S1N) hydrogen bonds to the polar head of Lys33 via both the SO2 and the amino groups (Figure S1E). On the contrary, the methyl-substituted sulfonamides 5a and 5b only interact via the SO2 group, due to the steric repulsion of the methyl or the dimethyl sulfonamide substituents (Figure S1F, S1G).

For the sulfonylmethane derivative 3n a similar behavior is also predicted (Figure S1H). Finally, the modification of the nitro substituent of 3 f with a methoxy (3g) or a trifluorocarbon (3h) group completely abolished the inhibitory activity. This finding, combined with the observed lack of activity of 3a, 3b, and 3e, in which the nitro group was removed, suggested that the interaction of the nitro group with the Asp145 and Phe146 NH groups of the DFG motif (Figure 2) is strictly required for CDK2 inhibitory activity.

The third series of compounds was synthesized to analyze the effect of different substituents on the 2-phenyl ring attached to the tetrahydroquinoline moiety (Table 3). In the CDK2–3 f complex, the ortho-chloro phenyl ring of 3 f is tightly packed against the short helix at the beginning of the activation loop, which is typical of PK inactive conformations, and the hydrophobic residues of the b4-b5 hairpin (Figure 2). In this series, substitution of the ortho-chloro atom with a methyl (3j), a trifluorocarbon (3u), or a bromine (3t) was tolerated, whereas substitution with a less hydrophobic fluorine atom (3s) was detrimental for activity. This result is consistent with the hydrophobic nature of this pocket (Figure 2). The replacement of the phenyl ring with a naphthalene (3r) moiety was also tolerated, whereas the introduction of a polar hydroxyl group in meta or para positions (3v, 3w, 3x, 3y) resulted in less active or inactive compounds, in agreement with the hydrophobic character of this sub-pocket.

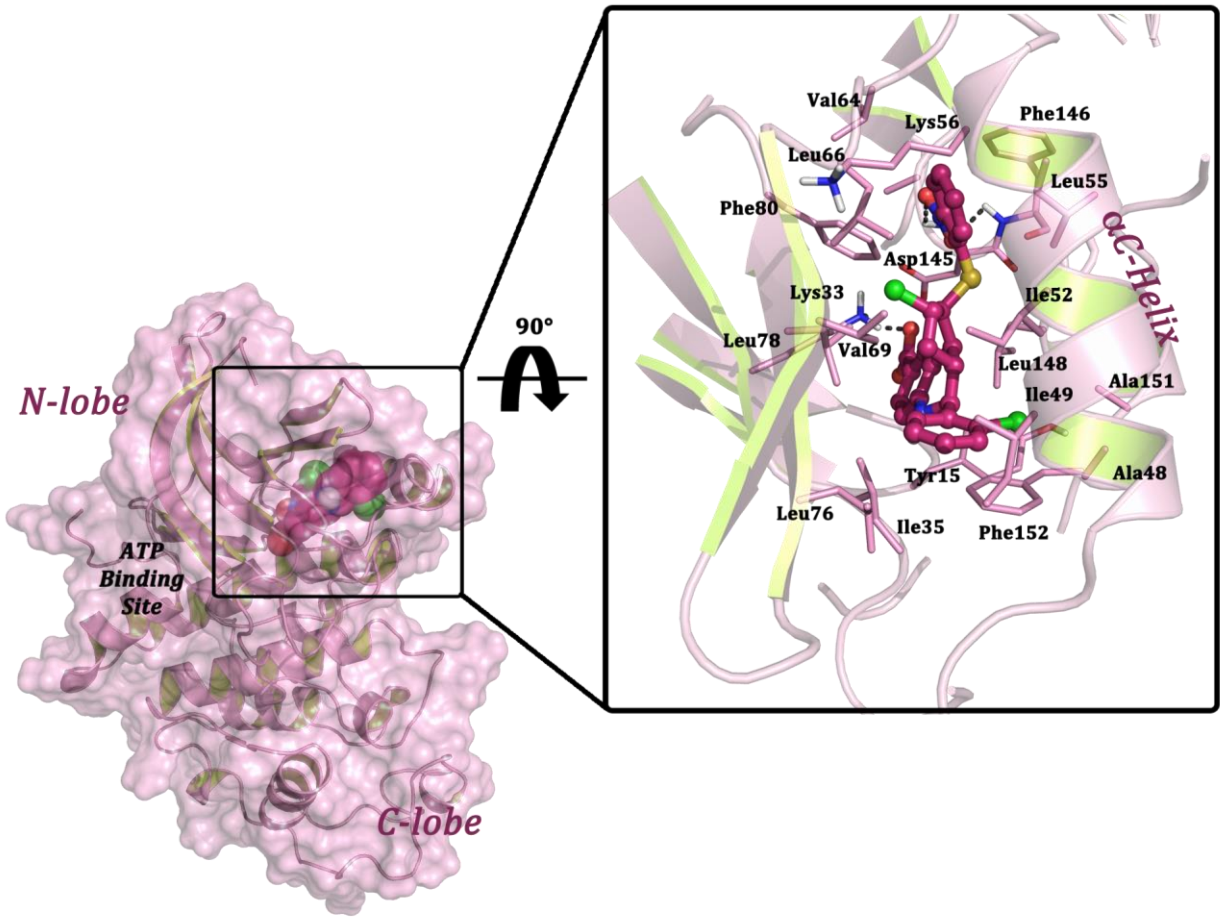
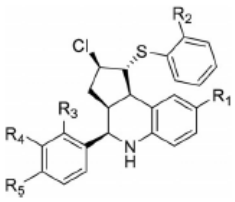


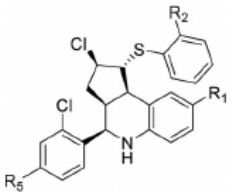
Figure 2: Binding mode of compound 3 f within CDK2 obtained via induced fit docking of the compound to the 2EXM crystal structure (see Christodoulou et al.[5] for details). The protein is represented by a pink surface and pink cartoons, whereas the residues forming the binding cavity are labeled with the three-letter code and shown as pink sticks. Compound 3 f is represented by burgundy spheres and burgundy balls and sticks within the box.

Table 1. Structures and biological activity of the investigated compounds synthesized to explore the importance of the carboxylic acid and nitro groups. Biological activities are reported as percent inhibition at 10 and 50 μM.



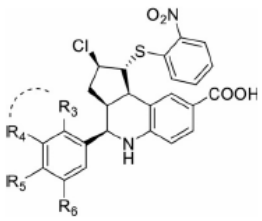
Compd	R_1	R_2	R_3	R_4	R_5	CDK2		wild-type EGFR		L858R/T790M EGFR	
						10 μM	50 μM	10 μM	50 μM	10 μM	50 μM
3a	H	H	H	H	H	0	0	0	12	0	0
3b	H	H	Cl	H	H	0	0	0	29	0	0
3c	H	NO_2	Cl	H	H	0	0	0	6	0	0
3d	H	NO_2	Cl	H	Cl	0	0	0	26	0	0
3e	COOH	H	Cl	H	H	0	0	0	4	0	28
3f	COOH	NO_2	Cl	H	H	52	65	0	38	10	81
3i	COOH	NO_2	Cl	H	Cl	56	66	0	28	18	55

Table 2. Structures and biological activity of the investigated compounds synthesized to explore the effects of replacement of the carboxylic acid and the nitro group with various chemical moieties. Biological activities are reported as percent inhibition at 10 and 50 μM.



Compd	R_1	R_2	R_5	CDK2		wild-type EGFR		L858R/T790M EGFR	
				10 μM	50 μM	10 μM	50 μM	10 μM	50 μM
3g	COOH	OCH_3	H	0	0	0	0	0	24
3h	COOH	CF_3	H	0	0	0	17	23	74
3k	COOEt	NO_2	H	0	0	0	0	0	0
3l	COOEt	NO_2	Cl	0	0	0	0	0	0
3m	SO_2NH_2	NO_2	H	46	63	0	42	0	9
3n	SO_2CH_3	NO_2	H	0	0	0	0	0	0
3o	CH_2COOH	NO_2	H	33	81	0	25	28	74
3p	$COCH_3$	NO_2	Cl	0	0	0	0	0	0
3q	$NHCOCH_3$	NO_2	Cl	16	13	0	17	0	0
4a	$CONHCH_3$	NO_2	H	55	56	0	0	0	0
4b	$CONHCH_3$	NO_2	Cl	57	62	0	8	0	0
5a	SO_2NHCH_3	NO_2	H	23	41	0	0	0	0
5b	$SO_2N(CH_3)_2$	NO_2	H	18	21	0	0	0	0
7	CONHOH	NO_2	H	30	82	0	0	0	0

Table 3. Structures and biological activity of the investigated compounds synthesized to explore the role of phenyl ring substitutions. Biological activities are reported as percent inhibition at 10 and 50 μM.



Compd	R_3	R_4	R_5	R_6	CDK2		wild-type EGFR		L858R/T790M EGFR	
					10 μM	50 μM	10 μM	50 μM	10 μM	50 μM
3j	CH_3	H	H	H	38	40	0	0	17	64
3r	phenyl		H	H	36	36	0	27	41	52
3s	F	H	H	H	0	0	0	23	28	70
3t	Br	H	H	H	55	63	0	0	39	72
3u	CF_3	H	H	H	49	65	0	20	41	79
3v	F	H	H	OH	22	43	0	23	0	17
3w	H	OH	H	F	0	0	19	31	0	30
3x	H	H	OH	F	13	51	0	32	0	30
3y	OH	H	H	F	29	62	19	28	0	33

Structure-activity relationships in EGFR

All compounds were inactive on wild-type EGFR-KD (Tables 1–3). Only two compounds (3w and 3y, Table 3) showed moderate wild-type EGFR-KD inhibitory activity. The superimposition of the structure of the CDK2–3 f complex[5] with the crystal structure of wild-type EGFR-KD in the inactive conformation (PDB ID: 3W32)[13] showed that compound 3 f, if adopting a CDK2-like binding mode, would form severe steric clashes with several residues of EGFR. These residues include Leu858, Val786, Ile759, Met766, Leu861, and Leu747. Indeed, standard

induced-fit docking calculations of 3 f to the allosteric pocket of wild-type EGFR-KD in the inactive conformation[13] failed to produce allosteric binding modes. Induced-fit docking with an extended sampling protocol (see the Experimental Section) generated allosteric binding modes that were quite different from those observed in CDK2. In particular, none of the generated binding modes showed the carboxylate and the nitro moieties, respectively engaged in interactions with the Lys745 (corresponding to Lys33 in CDK2) side chain and with the NH backbone atoms of Asp855 (Asp145) and Phe856 (Phe146), as observed in CDK2. Altogether, these findings provided an explanation for the observed lack of activity of the investigated compounds on wild-type EGFR.

In contrast, the compounds could be readily docked to the crystal structure of the T790M/L858R double mutant EGFR-KD (PDB ID: 3W2R),[14] in which the unfolding of the short helix at the beginning of the activation loop due to the L858R mutation makes the allosteric pocket significantly more accessible to the ligands.[14–16] Induced-fit docking of the carboxylic acid derivatives to the T790M/L858R EGFR-KD crystal structure showed that the compounds bind to the allosteric pocket by accommodating the 2-phenyl-substituted ring into the deeper hydrophobic pocket lined by Leu788, Met766, Met790, and Phe856, with the amino group of the hexahydrocyclopenta[c]quinoline hydrogen bonding with the Phe856 backbone carbonyl of the DFG motif, and the carboxylate group forming a salt bridge with the polar heads of Arg858 and Arg748 (3s, Figure 3). In particular, the interaction of the carboxylate with the mutated Arg858 residue may further explain the observed selectivity of the compounds for the double mutant EGFR-KD with respect to the wild-type EGFR-KD. Indeed, all the compounds in which the carboxylate is removed or modified were inactive when tested on the T790M/L858R protein, except for the acetic acid derivative 3o that was still able to form a salt bridge with Arg858 (Tables 1 and 2). Finally, the different substituents of the 2-phenyl ring, which is accommodated into the deeper hydrophobic pocket lined by the side-chains of Phe856, Leu788, Met766, Met790, Phe723, and Lys745, were found to play a role in modulating the selectivity toward the double mutant EGFR. In particular, the fluorine substituted compound 3s was a selective allosteric inhibitor of the double mutant T790M/L858R EGFR-KD with respect to either wild-typeEGFR-KD and CDK2.

Molecular dynamics simulations of compound 3s within the T790M/L858R EGFR-KD (see the Experimental Section) showed that the proposed binding mode was stable and the key interactions within the allosteric pocket were maintained throughout the simulation (100 ns; two replicas with different starting velocities, Figure S1I).

Dose–response curves (Figure 4) showed that 3s was able to inhibit the T790M/L858R EGFR-KD with an EC50 of 44:4 mm, whereas inhibition was not observed with wild-type EGFR-KD. To demonstrate its allosteric behavior on the T790M/L858R EGFR-KD, compound 3s was also tested in the presence of a higher concentration of ATP (1 mm). The obtained dose–response curves clearly demonstrated that compound 3s maintained its inhibitory profile, thus confirming its allosteric mechanism of action (Figure S1O).

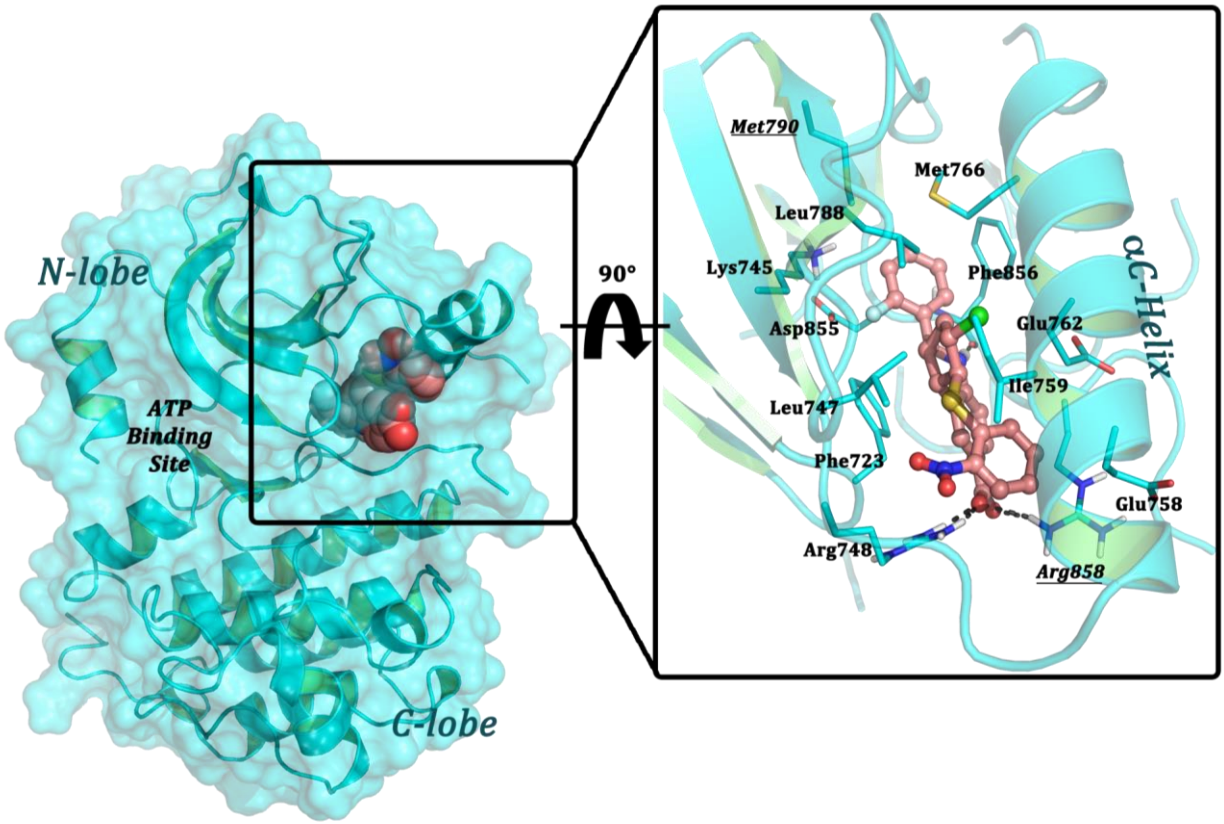


Figure 3: Induced-fit docking binding mode of compound 3s within the T790M/L858R EGFR-KD crystal structure (PDB ID: 3W2R). The protein is represented by a cyan surface and cyan cartoons, whereas the residues forming the binding cavity are labeled with the three-letter code and shown as cyan sticks. Compound 3s is represented by pink spheres and pink balls and sticks within the box.

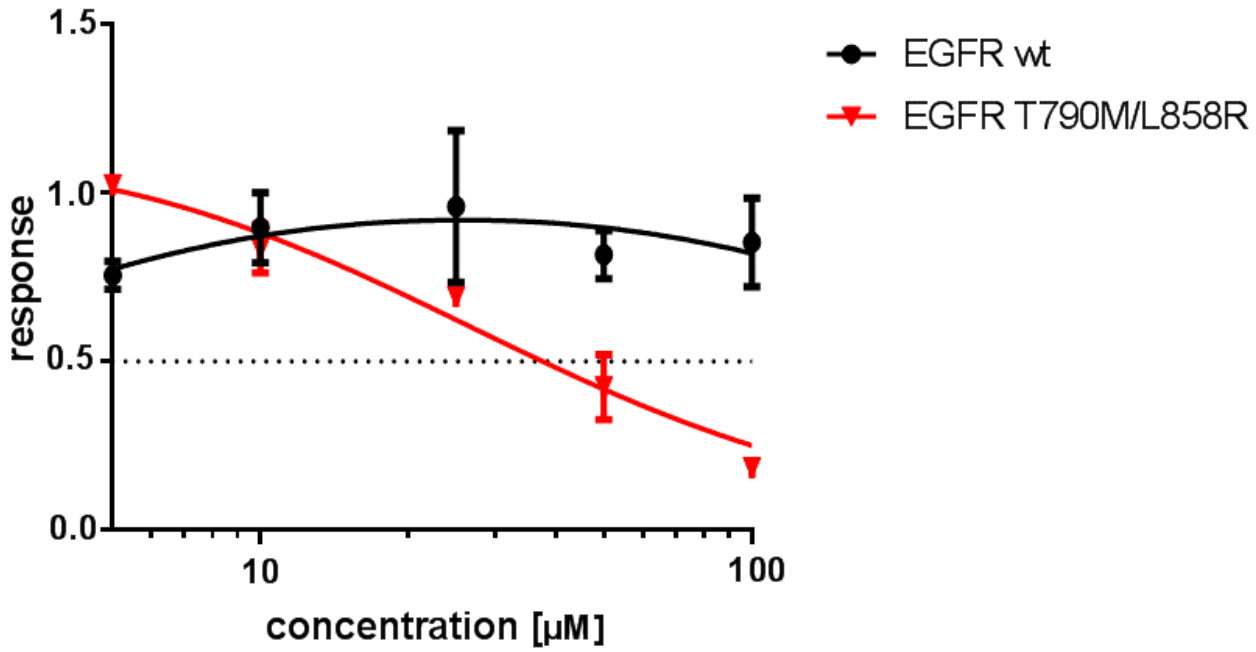


Figure 4: Dose–response curves of compound 3s tested on wild-type and T790M/L858R EGFR.

Chemical diversity with respect to known CDK2 and EGFR inhibitors reported in ChEMBL

The chemical novelty of the presented hexahydrocyclopenta[c]quinolone inhibitors was evaluated by means of 2D similarity analyses with respect to the CDK2 and EGFR ligands collected in the ChEMBL database. Chemical structures of the compounds with activity annotations on human CDK2 and EGFR were extracted from the ChEMBL database and similarity analyses were performed by using the MACCS and circular fingerprints (ECFP6 fingerprint), as described in the experimental section. The 2D fingerprint-based retrospective analyses revealed that the disclosed compounds are, on average, highly dissimilar to the already reported CDK2 and EGFR inhibitors, the averaged similarities being far below the commonly accepted similarity thresholds (Table S1).[17, 18] Therefore, results of the cheminformatic analyses demonstrated that the ligands herein reported are structurally different from those in the analyzed ChEMBL subsets. This result is not surprising considering that the new class of inhibitors bind to an allosteric pocket, whereas the majority of ChEMBL compounds belongs to the type I, $I^{1/2}$, and II classes of PK inhibitors.

Conclusions

This study was devised to explore the SAR of a promising class of type III allosteric inhibitors of CDK2 through target-guided synthetic modifications. Moreover, for the first time, the reported CDK2 allosteric inhibitors were tested on the wild-type GFR kinase domain and its clinically relevant T790M/L858R mutant, leading to the identification of active compounds. Docking studies and molecular dynamics simulations were performed to elucidate the SARs. Through this strategy, the importance of the carboxylic acid and the nitro groups for the CDK2 and T790M/L858R EGFR inhibitory activity were established. The carboxylic acid is strictly required for binding to the T790M/L858R EGFR, whereas it is potentially replaceable in CDK2 by different chemical moieties able to make hydrogen bonds with the side-chain of Lys33. We also demonstrated that 2-phenyl ring substituents that fit into an additional hydrophobic pocket of the allosteric site are important to fine-tune the activity and selectivity of the investigated compounds. Our data, although the synthesized compounds need to be optimized for potency, explored the relatively uncharted field of type III allosteric inhibitors and demonstrated that switching the inhibitory activity from CDK2 to double mutant EGFR is to be considered a feasible strategy. Our results pointed to different binding modes in CDK2 and double mutant EGFR-KD, thus reflecting the peculiarities of the respective allosteric sites andnpotentially enabling the development of selective allosteric ligands. Remarkably, compound 3s was shown to be an allosteric inhibitor of the T790M/L858R EGFR that spares both wildtype EGFR and CDK2, and will be further investigated for the design of more potent and selective T790M/L858R EGFR inhibitors.

Experimental Section

Chemistry: All reactions were carried out in oven-dried glassware and dry solvents under nitrogen atmosphere. Unless otherwise stated, all solvents were purchased from Sigma Aldrich and used without further purification. Substrates and reagents were purchased from Sigma Aldrich and used as received. 2-Nitrobenzenesulfenyl chloride was purchased from Alfa Aesar and used as received. Thin layer chromatography (TLC) was performed on Merck precoated $60F_{254}$ plates. Reactions were monitored by TLC on silica gel, with detection by UV light (254 nm) or by a solution of ninhydrin with heating. Flash chromatography was performed using silica gel (240-400 mesh, Merck). All tested compounds possessed a purity of > 98% confirmed *via* elemental analyses (CHN) with a Perkin Elmer 2400 instrument. ^{1}H-NMR spectra were

recorded on a Bruker DRX-400 instrument and are reported relative to residual $CDCl_3$ and DMSO. ^{13}C-NMR spectra were recorded on the same instrument (100 MHz) and are reported relative to residual $CDCl_3$ and DMSO. Chemical shifts (δ) for proton and carbon resonances are quoted in parts per million (ppm) relative to tetramethylsilane (TMS), which was used as an internal standard. COSY, HSQC, HMBC, and NOESY experiments were used in the structural assignment. MS spectra were recorded using electrospray ionization (ESI) technique on a Waters Micromass Q-Tof micro mass spectrometer.

Detailed methods for the preparation of the compounds reported in Tables 1, 2, and 3 can be found in the Supporting Information.

Single-crystal X-ray analysis of racemic compound 3j: X-ray-quality crystals were obtained by slow crystallization from MeOH. The data collection was performed at room temperature on a three-circle Bruker AXS Smart diffractometer equipped with an APEX-II CCD detector. A 99.3% complete full sphere of data was collected using graphite-monochromated radiation (λ = 0.71073 Å) with a nominal power of 50 kV x 30 mA. Data were corrected for absorption and beam anisotropy effects, and the structural model was obtained up to a maximum resolution of 0.77 Å through the full-matrix least-squares procedure implemented in SHELXL–2014/4.[19] The interested reader can find full information on the crystal structure of **3j** in the Supporting Information. CCDC 1856132 contains the supplementary crystallographic data for this paper. These data can be obtained free of charge from the Cambridge Crystallographic Data Centre.

CDK2 production and purification: *E. coli* BL21(DE3) cells were transformed with a vector enclosing the human CDK2 coding sequence with a GST tag at the N-terminus. Cells were grown for 2-3 h at 37 °C, then the temperature was decreased to 16 °C prior to induction. Expression was induced with 0.1 mM IPTG at OD_{600} = 0.9 for 20-24 h at 16 °C. Cells were harvested by centrifugation for 20 minutes at 4000 *g* at 4 °C. Bacteria were suspended in 50 mM Tris-HCl pH 8.0, 150 mM NaCl, and 1 mM DTT (buffer A) supplemented with the Complete Mini Protease Inhibitor Cocktail (Roche) and were lysed using a French pressure cell press. The lysate was centrifuged for 30 min at 12000 rpm (15500 x *g*, on a Beckman Coulter TA-14-50 rotor) at 4 °C and the supernatant was applied onto a GST-affinity column (GE Healthcare) equilibrated with the buffer A. GST-CDK2 was eluted using the buffer A supplemented with 10

mM reduced glutathione. The final protein was diluted at 1 mg/mL, aliquoted, flash-frozen in liquid nitrogen, and stored at -80 °C.

Wild-type and T790M/L858R EGFR production and purification: DNA encoding for the kinase domain residues 696–1022 of the human EGFR *wt* was inserted in pFB-LIC-Bse and pFB-6HZB transfer vectors (kindly provided by Opher Gileadi, SGC-Oxford). The T790M/L858R mutation was introduced in the *wt* constructs by the Q5 site directed mutagenesis based method (New England Biolabs) and sequence was verified. Recombinant baculovirus was generated by using the Bac-to-Bac® Baculovirus Expression System (Invitrogen by Life Technologies Corporation, CA, USA) by transfecting Sf9 cells seeded in 6-well plates (1.5 x 10^6 cells/well) with FuGENE HT (Promega Italia Srl, Milano, Italy). The high titer viral stock was generated by two rounds of amplifications and used for protein expression in Sf9 cells at 27 °C. The cells were harvested by centrifugation 72 h post-infection. For protein purification, cells were homogenized (Emulsiflex – Avestin) in lysis buffer (50 mM Tris pH 8, 500 mM NaCl, 10% glycerol, 10 mM imidazole, 2 mM TCEP) supplemented with 10 µg/mL DNaseI, 1 mM MgCl2 and protease inhibitor cocktail (cOmplete EDTA-Free, Roche Diagnostics GmbH, Manheim, Germany). The lysate was cleared by centrifugation at 30000 x *g* for 1 h at 4°C and the supernatant was incubated with NiNTA beads (Qiagen, Milano, Italy) for 1 h at 4 °C. The beads were washed with 30 CV of the lysis buffer and the bound proteins were eluted with lysis buffer containing 300 mM imidazole. Eluted protein fractions were dialysed o/n at 4°C against 25 mM Tris pH 8, 0.2 M NaCl, 10% glycerol, 5 mM DTT in the presence of the Tobacco Etch Virus protease to remove tags. The cleaved proteins were purified on a Fractogel® EMD TMAE (M) column (Merck KGaA, Darmstadt, Germany) and eluted with a NaCl gradient from 10 to 1000 mM in 25 mM Hepes pH 8, 10% glycerol, 5 mM DTT in 30 CV. The protein fractions were gel filtered on Superdex 75 (10/300 GL) column (GE Healthcare Life Sciences, Milano, Italy) in 25 mM Hepes pH 8, 250 mM NaCl, 10% glycerol and 2 mM TCEP. Stock protein samples were flash frozen at the final concentration of 0.5 – 1 mg/mL and stored in aliquots at -80 °C till further use. Typical yield of purified EGFR-KD constructs was 0.6 - 1.8 mg/mL from 1 x 10^9 Sf9 cells.

Biological assays: CDK2 inhibitory activities were determined using the previously described ANS displacement assay.[4,20] The assay was performed in 96-well black plates in a total volume of 50 µl of 50 mM Hepes pH 7.5 with DTT 1 mM and ANS (final concentration 62.5 µM). Different concentrations of the compounds were added and the fluorescence (excitation 405 nm, emission 460 nm) measured using an Infinite M200 Microplate Reader (Tecan) to

determine the intrinsic fluorescence of the compounds. Finally, 1 μg of recombinant CDK2 was added and a second fluorescence measurement was taken at the same wavelengths. EGFR inhibitory activities were determined using the ADP-Glo kinase assay. The ADP-Glo kinase assay (Promega) is a luminescent kinase assay that measures the ADP formed from the EGFR kinase reaction. With this method, the luminescent signal generated is proportional to the ADP concentration produced and is correlated with the kinase activity. The kinase reaction was performed in 96-well white plates as already described,[21] in a total volume of 25 μl and incubated for 1 h at room temperature. Then, an equal volume of ADP-Glo reaction was added followed by, after 40 minutes of incubation, the addition of 50μL of kinase detection reagent. The luminescence was read using the GloMax plate reader (Promega). Control experiments with the reference inhibitors are reported in the Supplementary Information (Figure S1M).

Docking calculations: The protein structures were prepared with the Schrödinger Suite 2014-3 Protein Preparation Wizard (PPW) tool. PPW automatically adjusted the ionization and tautomerization state of the proteins at a neutral pH (PROPKA was used to predict the pKa of residues at pH=7), set the orientation of any misoriented groups (Asn, Gln, and His residues), and optimized the hydrogen bond network. Finally, the structures were refined to relieve steric clashes with a restrained minimization with the OPLS2005 force field till a final RMSD of 0.30 Å with respect to the input protein coordinates.

Docking calculations to CDK2 were performed by extracting the protein from the CDK2/**3f** complex obtained in our previous publication.[5] Docking calculations to the *wt* EGFR and the T790M/L850R EGFR kinase domains were performed by using the 3W32[13] and 3W2R crystal structures,[14] respectively, prepared as described above.

All docking calculations were performed with the Induced Fit Docking (IFD) protocol available in the Schrödinger Suite 2014-3, by using default settings. In addition, the *wt* EGFR-KD/**3f** complex was also investigated by using the extended sampling protocol available in the Schrödinger Suite 2014-3, which automatically prepares the receptor by choosing the residues to trim and atom-specific van der Waals scaling factors on the basis of solvent-accessible surface areas, B-factors, the presence of salt bridges, and rotamer searches. In this case, a maximum number of 80 poses were saved for each ligand and submitted to the subsequent Prime side-chain orientation prediction of residues within a shell of 5 Å around the ligand. After the Prime minimization of the selected residues and the ligand for each pose, a Glide SP re-docking of each protein-ligand complex structure was performed. Finally, the binding energy

(IFDScore) for each output pose was calculated.[22] In all cases, grids were centered to the allosteric pocket of the investigated kinases.

Molecular dynamics: The complex between the T790M/L858R EGFR-KD and **3s** obtained by IFD was saved in the PDB format and prepared with the tleap module of the Amber14 suite, using the ff14SB force field. [23,24] The protein-ligand complex was enclosed in a truncated octahedral TIP3P water box with a radius of 10 Å. Na^+ counter ions were added to neutralize the complex.

Long-range electrostatic interactions were treated using the Particle-Mesh Ewald (PME) method,[25] while a short-range cutoff of 8 Å was applied for all non-bonded interactions.

Before data acquisition the system was equilibrated in a stepwise fashion to eliminate bad atomic contacts. The solvent was minimized, while keeping the protein and ligand fixed, by using 800 steps of the steepest descent (SD) method and 800 steps of the conjugate gradient (CG) method. A second minimization was performed by keeping the protein fixed with 800 steps of the SD method and 800 steps of the CG method. A third minimization was performed by keeping the ligand fixed with 800 steps of the SD method and 800 steps of the CG method. Finally, an unrestrained minimization was performed on the whole system by using 800 steps of the SD method and 800 steps of the CG method.

The system was subsequently heated to 300 K in two steps of 150 ps each, in the NVT ensemble. Then, equilibration was performed in six separate runs of 400 ps each, in the NPT ensemble. Constraints on the protein and ligand atoms were gradually decreased throughout the equilibration steps. All constraints were removed in the production run (100 ns). The time step was set to 2 fs, and bonds involving hydrogen atoms were constrained using the SHAKE algorithm.[26] Two replicas with different starting velocities were performed. Trajectories were analysed by calculating the RMSD of protein alpha carbons and ligand heavy atoms.

2D similarity analyses: 2D similarity analyses were carried out on the ChEMBL database (release 23, accessed on May 29th, 2018) to evaluate the chemical novelty of the identified CDK2 and EGFR allosteric inhibitors.[27] In particular, similarity calculations were performed on already known CDK2 and EGFR ChEMBL compounds with activity annotations characterized by an assay confidence score equal or greater than 8. Two different types of molecular fingerprints implemented in the OpenEye OEGraphSim Python Toolkits (release 2017.Jun.1), *i.e.,* MACCS and ECFP6, were used to encode the structural details of the molecules.[28-30] Ligand similarities were evaluated in terms of the Tanimoto similarity score by considering CDK2 or

EGFR active compounds as reference queries. A total of 56 ranks (2 molecular fingerprints x 28 active compounds) were obtained for each target (CDK2 and EGFR) in the 2D retrospective analyses on ChEMBL. Results and statistics of the similarity calculations are summarized in Table S1 of the Supplementary Information.

Acknowledgements

This work was supported by a grant from the Associazione Italiana per la Ricerca sul Cancro (AIRC IG 15993 and IG 16792).

Keywords

Allosteric inhibitors, cyclin-dependent kinase 2, docking, drug design, epidermal growth factor receptor, structure-activity relationships.

References

[1] K.A. Stenberg, P.T. Riikonen, M. Vihinen. *Nucleic Acids Res*. **1999**, 27, 362–364.

[2] Z. Fang, C. Grütter, D. Rauh. *ACS Chem. Biol*. **2013**, 8, 58–70.

[3] P. Wu, M. H. Clausen, T. E. Nielsen. *Pharmacol. Ther*. **2015**, 156, 59–68.

[4] G. Rastelli, A. Anighoro, M. Chripkova, L. Carrassa, M. Broggini. *Cell Cycle* **2014**, *13*, 2296-2305.

[5] M.S. Christodoulou, F. Caporuscio, V. Restelli, L. Carlino, G. Cannazza, E. Costanzi, C. Citti, L. Lo Presti, P. Pisani, R. Battistutta, M. Broggini, D. Passarella, G. Rastelli. *ChemMedChem* **2017**, *12*, 33-41.

[6] X. Zhang, J. Gureasko, K. Shen, P.A. Cole, J. Kuriyan. *Cell* **2006**, *125*, 1137-1149.

[7] N. Jura, X. Zhang, N.F. Endres, M.A. Seelinger, T. Schindler, J. Kuriyan. *Mol. Cell* **2011**, *42*, 9-22.

[8] L. Palmieri, G. Rastelli. *Drug Discov. Today* **2013**, *18*, 407-414.

[9] N. Sturm, A. Tinivella, G. Rastelli. J. *Chem. Info. Model.* **2018**, *58*, 1094-1103.

[10] S.V. Sharma, D.W. Bell, J. Settleman, D.A. Haber. *Nat. Rev. Cancer* **2007**, *7*, 169-181.

[11] C. Ricordel, L. Friboulet, F. Facchinetti, J.C. Soria. *Annals of Oncology* **2018**, *29*, 28-37.

[12] L. Zai-Shun, W. Wei-Min, L. Wei, N. Cong-Wei, L. Yong-Hong, L. Zheng-Ming, W. Jian-Guo. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 3723-3727.

[13] Y. Kawakita, M. Seto, T. Ohashi, T. Tamura, T. Yusa, H. Miki, H. Iwata, H. Kamiguchi, T. Tanaka, S. Sogabe, Y. Ohta, T. Ishikawa. *Bioorg. Med. Chem.* **2013**, *21*, 2250-2261.

[14] S. Sogabe, Y. Kawakita, S. Igaki, H. Iwata, H. Miki, D.R. Cary, T. Takagi, S. Takagi, Y. Ohta, T. Ishikawa. *ACS Med. Chem. Lett.* **2013**, *4*, 201-205.

[15] C.H. Yun, T.J. Boggon, Y. Li, M.S. Woo, H. Greulich, M. Meyerson, M.J. Eck. *Cancer Cell* **2007**, *11*, 217-227.

[16] K.S. Gajiwala, J. Feng, R. Ferre, K. Ryan, O. Brodsky, S. Weinrich, J.C. Kath, A. Stewart. *Structure* **2013**, *21*, 209-219.

[17] S. W. Muchmore, D. A. Debe, J. T. Metz, S. P. Brown, Y. C. Martin, P. J. Hajduk. *J. Chem. Inf. Model.* **2008**, *48*, 941–948.

[18] S. Jasial, Y. Hu, M. Vogt, J. Bajorath. *F1000Research* **2016**, *5*, 591.

[19] Sheldrick, G. M. *ActaCrystallogr. Sect. C*, **2015**, 71, 3-8.

[20] M.P. Martin, R. Alam, S. Betzi, D.J. Ingles, J.-Y. Zhu, E. Schönbrunn. *ChemBioChem* **2012**, *13*, 2128–2136.

[21] F. Caporuscio, A. Tinivella, V. Restelli, M.S. Semrau, L. Pinzi, P. Storici, M. Broggini, G. Rastelli. *Future Med. Chem.* **2018**, 10, 1545-1553.

[22] M.D. Parenti, G. Rastelli. *Biotechnol. Adv.* **2012**, 30, 244-250.

[23] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser and C. Simmerling. *J. Chem. Theory Comput.* **2015**, *11*, 3696-3713.

[24] D.A. Case, V. Babin, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham III, T.A. Darden, R.E. Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, P.A. Kollman, 2014, AMBER14, University of California, San Francisco.

[25] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, *103*, 8577–8593.

[26] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327–341.

[27] A. Gaulton, L.J. Bellis, A.P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, J.P. Overington. *Nucleic Acids Res.* **2012**, *40*, D1100-D1107.

[28] OpenEye Python Toolkits 2017.Jun.1. OpenEye Scientific Software, Santa Fe, NM. Http://www.eyesopen.com.

[29] J. L. Durant, B.A. Leland, D.R. Henry, J.G. Nourse. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273-1280.

[30] D. Rogers, M. Hahn. *J. Chem. Inf. Model.* **2010**, *50*, 742–754.