

A computational workflow for the design of irreversible inhibitors of protein kinases

Alberto Del Rio · Miriam Sgobba · Marco Daniele Parenti ·
Gianluca Degliesposti · Rosetta Forestiero · Claudia Percivalle ·
Pier Franco Conte · Mauro Freccero · Giulio Rastelli

Received: 24 July 2009 / Accepted: 5 March 2010 / Published online: 21 March 2010
© Springer Science+Business Media B.V. 2010

Abstract Design of irreversible inhibitors is an emerging and relatively less explored strategy for the design of protein kinase inhibitors. In this paper, we present a computational workflow that was specifically conceived to assist such design. The workflow takes the form of a multi-step procedure that includes: the creation of a database of already known reversible inhibitors of protein kinases, the selection of the most promising scaffolds that bind one or more desired kinase templates, the modification of the scaffolds by introduction of chemically reactive groups (suitable cysteine traps) and the final evaluation of the reversible and irreversible protein–ligand complexes with molecular dynamics simulations and binding free energy predictions. Most of these steps were automated. In order to prove that this is viable, the workflow was tested on a database of known inhibitors of ERK2, a protein kinase possessing a cysteine in the ATP site. The modeled ERK2–ligand complexes and the values of the estimated binding free energies of the putative ligands provide useful indicators of their aptitude to bind reversibly and irreversibly to the protein kinase. Moreover,

the computational data are used to rank the ligands according to their computed binding free energies and their ability to bind specific protein residues in the reversible and irreversible complexes, thereby providing a useful decision-making tool for each step of the design. In this work we present the overall procedure and the first proof of concept results.

Keywords Protein kinases · ERK2 · Irreversible inhibition · Cysteine trap · Michael acceptors · Binding free energy · Molecular dynamics · Generalized amber force field

Introduction

Kinases and cancer have been intimately linked for 30 years. Over 500 protein kinases have been identified to date, including many involved with cellular signalling pathways whose dysregulation has been implicated in cancer [1, 2]. Many protein kinase (PK) inhibitors have been discovered and developed, some of which were approved and entered the clinic [3, 4]. However, the road to develop kinase inhibitors as marketed drugs proved to be difficult, and many inhibitors failed late in development. One way to extend the efficacy of kinase inhibitors is to hit multiple tumor-promoting signalling pathways, either with drug combinations or with a single multi-targeted compound. Good examples are Lapatinib, which targets EGFR and ErbB2 and shows activity against metastatic breast cancer, and Sunitinib, which targets four kinases and was approved in 2006 for renal cancer and gastrointestinal stromal tumor [5]. While the interest in the last decade has been focused toward small molecules that reversibly inhibit protein kinases, another important approach for the development of anticancer drugs is irreversible inhibition.

Electronic supplementary material The online version of this article (doi:10.1007/s10822-010-9324-x) contains supplementary material, which is available to authorized users.

A. Del Rio · M. Sgobba · M. D. Parenti · G. Degliesposti ·
R. Forestiero · G. Rastelli (✉)
Dipartimento di Scienze Farmaceutiche, Università di Modena
e Reggio Emilia, Via Campi 183, 41100 Modena, Italy
e-mail: giulio.rastelli@unimore.it

C. Percivalle · M. Freccero
Dipartimento di Chimica Organica, Università di Pavia,
Via Taramelli 10, 27100 Pavia, Italy

P. F. Conte
Dipartimento di Oncologia ed Ematologia, Università di Modena
e Reggio Emilia, Via del Pozzo 71, 41100 Modena, Italy

The increasing interest on this topic is testified by the appearance of several recent studies showing how different reactive groups can act towards specific kinases or growth factor receptors with formation of a covalent adduct [6–11]. As a matter of fact, the common prejudice against irreversible inhibitors is belied by the fact that 19 of the 71 enzyme targets of marketed drugs are covalently modified by the drug [11, 12]. Indeed, provided that a selective targeting in a biological environment can be achieved, such inhibitors offer several potential advantages over reversible binding agents. First, the onset of covalent inhibition is a function of time and inhibitor concentration, so selectivity within the target subset might be altered by duration of exposure. Second, the reaction is thermodynamically driven toward the covalent adduct, and high concentrations of inhibitor are not needed to achieve effective inhibition [5, 13]. Third, once covalently bound, such inhibitors do not readily dissociate from the enzyme therefore restoration of enzyme activity requires new synthesis and the inhibition of the enzyme would continue long after the blood levels of the inhibitors had fallen below inhibitory concentrations. Fourth, irreversible inhibition can circumvent drug resistance [14]. Another advantage is the possibility to circumvent problems related to targeting one specific kinase conformation. In fact, in some cases, mutations in response to inhibitor treatment decrease the lifespan of the targeted kinase conformation with consequent loss of binding of the original inhibitor [5]. Imatinib, that targets the inactive conformation of the kinase domain of the Bcr-Abl oncogene, is an example of this [3]. In this context, an irreversible ligand is expected to retain inhibitory efficacy against mutated targets that exhibit such equilibrium shift among the various conformational states of the kinase.

While drug design is still mainly focused on reversible ATP-competitive inhibitors, the strategy of developing irreversible inhibitors remains relatively less explored [7–10, 14–16]. In this paper we describe a computational workflow that was specifically conceived for the design of irreversible inhibitors that target cysteine residues of protein kinases. We start from a description of the design strategy, which has been divided into sequential steps. Then, for each of these steps we present the technical solutions and the workflow implementation. In addition, as a proof of concept, we show the results obtained when the workflow was applied to ERK2, a protein kinase possessing a cysteine residue in the ATP binding site. As shall we see, the workflow can be easily adapted to other biological targets.

Workflow overview

With a bioinformatics analysis Schirmer et al. [16] identified a Cys residue that is conserved in a subset of protein

kinases. This target cysteine, immediately N-terminal to the completely conserved Asp residue responsible of Mg^{2+} -ATP (Fig. 1), is present in a subset of 46 of the 510 protein kinases and forms a conserved motif that is located in the ATP binding pocket. Table 1 shows the kinases that possess the above-mentioned motif. Of the 46 Cys-containing kinases, 38 exist in 8 evolutionarily related clusters, 7 of which lie within 5 major branches of the human kinase tree. Importantly, the 46 kinases condense to major subsets of related targets that encompass multiple components of important pathways. In the same study, Schirmer et al. [16] screened a panel of 124 kinases with the resorcylic acid lactone hypothemycin showing that 18 out of 19 targets containing the conserved Cys were inhibited, and kinetic analysis showed time-dependent inhibition and covalent inactivation of the kinases. Formation of the covalent bond was also demonstrated with X-ray and molecular dynamics of the hypothemycin-ERK2 complexes [17]. Thus, covalent inactivation of the enzyme is possible when a sufficiently electrophile carbon on the ligand allows a nucleophilic attack from the conserved cysteine sulphur. These studies opened new and promising perspectives for the development of novel anticancer drugs.

Here we present a computational workflow that is focused on the design of irreversible inhibitors of protein kinases. The workflow chart is shown in Fig. 2. Four main steps are involved, namely (1) the creation of a database of known reversible inhibitors of protein kinases (2) the computational selection of the most promising scaffolds that bind a protein kinase chosen template (3) the modification of the scaffolds by introduction of suitable *cysteine trap* located close to the thiol group (4) the computational evaluation of reversible and irreversible protein–ligand complexes with molecular dynamics (MD) and binding free energy analysis (ΔG of binding). A description of each step of Fig. 2 is presented and discussed in the next paragraphs. It should be emphasized that this workflow is applicable to any protein kinases shown in Table 1. Moreover, provided that other useful cysteine residues exist in the binding site, the workflow can be potentially extended to the whole kinome. For the aim of the present work we have selected ERK2 as a template for testing the workflow. Our choice is supported by two main reasons. First, several crystal structures of ERK2, alone and complexed with reversible and irreversible inhibitors, have been solved at a good resolution and deposited in the Protein Data Bank; second, since ERK2 is involved in the last step where the signal pathway converges before activation of transcription factors into the nucleus, its inhibition is highly desirable in order to obtain most profound effects downstream of the signal transduction pathway. For sake of simplicity we will refer to ERK2 residues numbering throughout.

Fig. 1 Depiction of protein kinase ATP-binding site regions important for drug design purposes

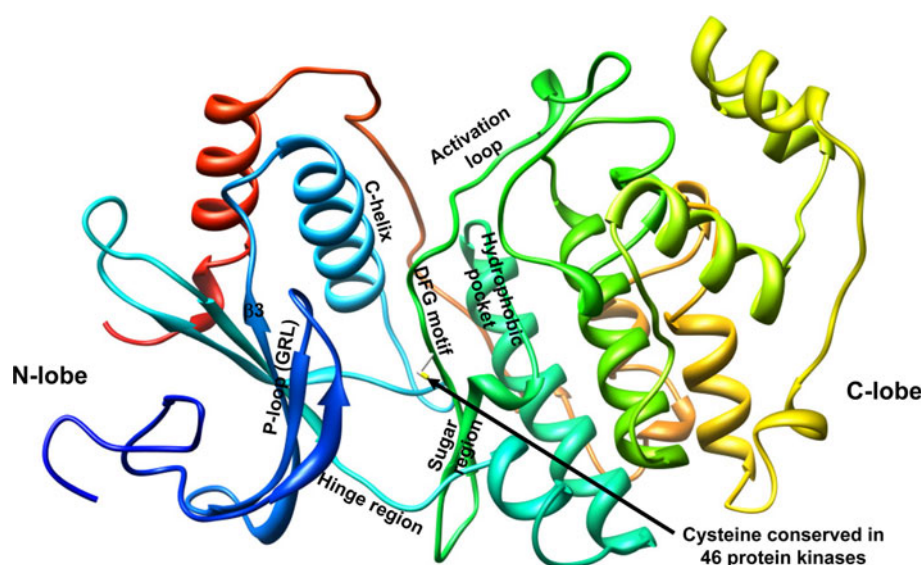


Table 1 Kinases containing the target cysteine

Group names CAMK Calcium/calmodulin-dependent protein kinase; CMGC Containing CDK, MAPK, GSK3, CLK families; STE Homologs of yeast Sterile 7, Sterile 11, Sterile 20 kinases; TK Tyrosine kinase; TKL Tyrosine kinase-like

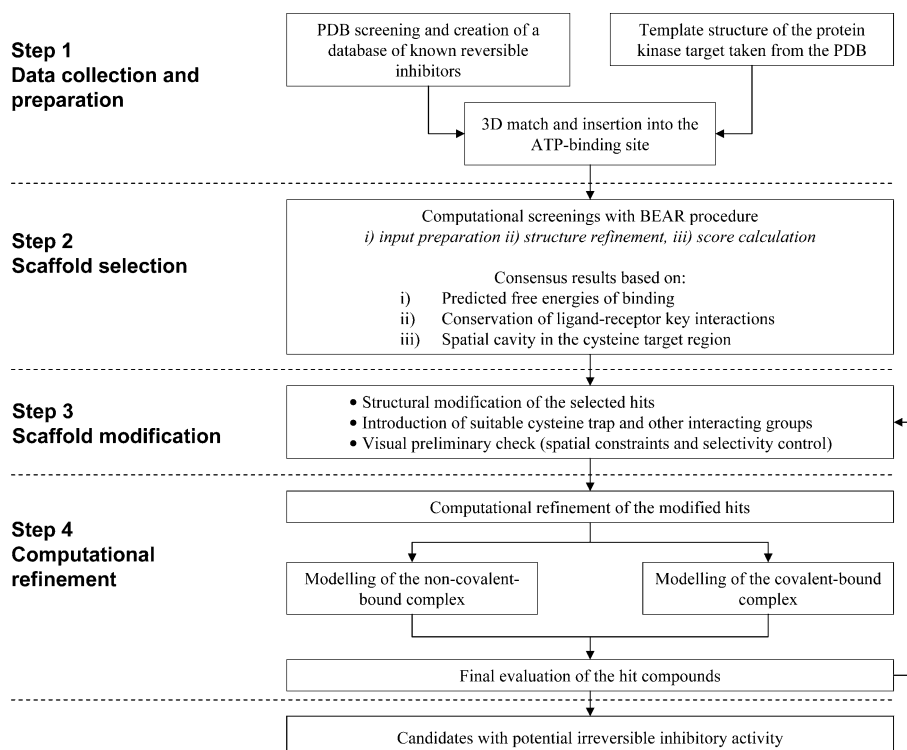
Branches of human kinome	Kinases
TK	VEGFR-1, VEGFR-2, VEGFR-3, FLT3, KIT, PDGFR α , PDGFR β
TKL	TAK1, TGF-BR2, ZAK
CMGC	CDKL1, CDKL2, CDKL3, CDKL4, CDKL5, GSK3 α , GSK3 β , ERK1, ERK2, MAPK15, NLK, PRP4 K
CAMK	MAPKAPK5, MNK1, MNK2, RSK1, RSK2, RSK3, RSK4, PKD1, PKD2, PKD3, SPEG
STE	MEK1, MEK2, MEK3, MEK4, MEK5, MEK6, MEK7, NIK
Other	AAK1, BMP2 K, GAK, STK36, TOPK

Step 1. Data collection and preparation

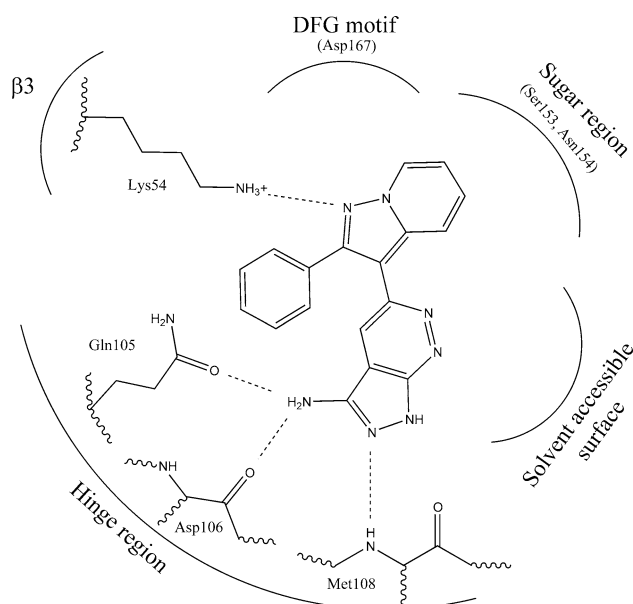
The first step of the workflow consists in the creation of a database of known reversible inhibitors of protein kinases with structures available in the Protein Data Bank. The aim of this step is to gather a restricted number of compounds already possessing structural features prone to protein kinase inhibition, in order to increase the chance of finding active candidates. Figure 1 shows kinase regions that are known to be critical for binding in the ATP pocket. In particular, the so-called *hinge region* connecting the two lobes plays a crucial role since all ATP-competitive inhibitors bind this region via multiple hydrogen bond interactions. In Fig. 3 is reported the scheme of the interactions of a small molecule inhibitor with key residues of ERK2 (PDB code 1TVO). The ligand hydrogen bonds with the backbone nitrogen of Met108 [18] of the hinge. Since this interaction is important and conserved in most ATP-competitive inhibitors, the creation of a database of known inhibitors aims to collect molecular structures able to form such interaction. It is worth noting that since the hydrogen bond involves the backbone nitrogen it will probably be

conserved in most kinases within the subset of 46 shown in Table 1, regardless of the presence of a Met or a different amino acid. It should also be noted from Fig. 3 that Met108 can form additional hydrogen bonds *via* its carbonyl group and stabilize further the binding with the *hinge region*. Another interesting residue is Asp106, which is close to Met108 and provides an additional hydrogen bond anchoring point *via* its backbone carbonyl. Next to Asp106 is the gatekeeper residue, Gln105 in the case of ERK2, which is a key residue for controlling the selectivity among different kinases. In the case of the inhibitor shown in Fig. 3, Gln105 provides an additional hydrogen bond. Other potential hydrogen bonding groups are the backbone carbonyl of Ser153 and the side chain of Asn154 located in the *sugar region*, the side chain of Asp167 belonging to the DFG motif and the side chain of the conserved Lys54 in the β 3 catalytic loop. The remaining residues that line the ATP pocket are mainly hydrophobic.

Interrogation of the Protein Data Bank issued 300 crystal structures of protein kinases complexed with different inhibitors. Allosteric ligands binding outside the Cys-containing pocket were excluded. When structures

Fig. 2 Workflow of the computational procedure

contained multiple copies of the kinase catalytic domain in the asymmetric unit, the chain with fewer missing residues was used for our analyses. However, visual inspection was performed to make sure that the inhibitor orientation was not significantly different along the multiple copies. In a

**Fig. 3** Main interactions established by a reversible inhibitor of ERK2 (PDB code 1TVO)

number of cases, crystallographic structures showed missing residues in the activation loop, the loop between C-helix and the third β -sheet of the kinase domain and the glycine-rich loop (Fig. 1 and Table 1 of the Supplementary Material). In addition, 91% of the kinase structures showed a DFG-in conformation, 6% had the DFG-out while a small number of proteins showed a DFG with non-canonical conformation or had a low resolution in such segment. At the time we performed the analysis, four crystallographic structures of PK with irreversible inhibitors were present, i.e. hypothemycin and a close analog with a double bond in place of the epoxide binding to ERK2 (PDB codes 3C9W and 2E14, respectively) [6, 17] and two quinazoline derivatives that irreversibly bind EGFR (PDB codes 2J5E and 2J5F) [19]. After removal of waters and inorganic salts these structures were superimposed on our template ERK2 structure (PDB code 1TVO) [18] by using the MatchMaker module of Chimera software and the Needleman-Wunsch alignment that includes secondary structure information [20]. The template structure 1TVO was chosen in the PDB by taking a well resolved X-ray structure without missing residues. To refine the match in the ATP-binding region of the different kinases, a number of residues to be aligned were selected among the residues that line the binding site, namely 29–42, 50–56, 64, 67, 68, 71–72, 75, 83–90, 101–114, 147, 149–159, 162–169 (ERK2 numbering). Some exceptions were made when matching p38 α protein family that showed significant differences with respect to ERK2

and for which the above-mentioned procedure generated structural clashes. In these cases we further reduced the list of residues to match and we increased the weight assigned to the secondary structure from 30 to 60%.

The structure-based alignments help us to identify differences in conformation and active site composition among the various kinases. For instance, consistently with previous studies, we noticed that type II inhibitors such as imatinib (pdb code 1IEP) that target the DFG-out state of kinases hardly fit the active form of ERK2 template that, vice versa, is crystallized in a DFG-in conformation. In such cases the three-dimensional matching revealed unfavorable interaction and clashes in proximity of the hydrophobic pocket (Fig. 1) [5]. Importantly, in our protein preparation protocol we visually inspected all structural superimpositions in order to make sure that the binding sites were correctly matched and the inhibitors inserted in ERK2 were positioned without significant steric clashes (e.g. clashes that could not reasonably be relieved with simple energy minimization of the complexes). On the front of the methodology, while it is clear that we could have generated the starting ERK2-ligand complexes by performing molecular docking simulations, in our approach we preferred to rely on protein kinase inhibitors with crystallographic information for the following reasons: (1) we are interested in obtaining valuable scaffolds able to irreversibly bind the whole set of 46 kinases containing our target Cys (Table 1); therefore, taking into account all protein kinase-inhibitor complexes available in the PDB irrespective of the type of kinase make our workflow more general and potentially applicable to anyone of them, even to kinases for which selective ligands are not reported. (2) considering known inhibitors already tested and developed for different PKs instead of a single kinase (e.g. ERK2) may increase the molecular diversity of the scaffolds emerging from our analyses. (3) even if docking methods generally perform well in assigning a correct orientation to the ligands, the variability in inhibitor structures as well as in active site conformation among the PKs could have potentially led to wrong docked structures, which in turn could have resulted in false negative hits. In this context the structural matching performed on the template allowed us to obtain a better set of ligand orientations in the binding pocket of the kinase.

After the superimposition, each inhibitor structure was extracted from the original kinase complex and inserted in the ERK2 template, assuming that the inhibitor maintains a similar binding mode. The compounds were then prepared assigning protonation states at physiological pH by means of ACD/pKa software. Atomic charges were obtained with *antechamber* module of Amber software by using AM1-BCC charges [21, 22]. The database of known reversible inhibitors was saved and further processed to step 2.

Step 2. Scaffold selection

The aim of this step is to select the most promising scaffolds from the database prepared in step 1 (Fig. 2). This can be achieved with an automated procedure developed in our laboratory named *Binding Estimation After Refinement* (BEAR) [23–25], that is based on molecular mechanics minimization and molecular dynamics simulations followed by MM-PBSA and MM-GBSA [25–27] binding free energy estimations as tools for refining and rescoring the structure of protein–ligand complexes. BEAR processes in a fully automated way each ligand-receptor complex through a sequential three-step procedure that includes input files preparation, structural refinement and free energy score calculation. A thoughtful description of the general procedure is available in supplementary material and in the reference articles [24]. The usefulness and accuracy of MM-PBSA and MM-GBSA as scoring functions has been extensively validated on a number of biological targets [23–30]. The performance of these methods has been investigated in terms of their ability to generate computed binding free energies that correlate with experimental values as well as their ability to discriminate known ligands among lists of inactive decoys. Very recently, validation studies were performed also on ERK2 [28]. The study showed that molecular dynamics coupled with MM-PBSA free energy estimation is a suitable tool for investigating the experimental binding activities of ligands to protein kinases. In particular, computed and experimental binding free energies were found to be significantly correlated [28]. These studies further demonstrate the applicability of our procedure to protein kinases.

For scaffold selection, we applied BEAR with different computational settings. In particular, we explored two different temperature ramp schemes and the effects of changing the number of residues around the ligands that were allowed to move during MD. These variations served to probe and compare the effect of different simulation conditions on the resulting ligand-PK complexes and position of the ligands in the ranked lists. The list of parameters used to perform the three rankings is reported in Table 2 of the supplementary material. The analysis of each ranking together with the visual inspection of the ligand-receptor complexes was useful to confirm the usefulness of taking into account different BEAR parameters. For instance, we noticed that a very fast heating to 300 K was useful to relieve possible steric clashes between the ligand and the structural template in the initial conformation (Step 1), especially with the glycine-rich loop which explores significant structural variations among various kinases with different bound ligands (the active site may be more or less open depending on the conformation of this loop). Vice versa, the ranking with a slower heating ramp

produced structures closer to the initial configuration and was more useful to rule out compounds with structural features non complementary with the template. An example of this concerned ligands with lipophilic moieties able to bind a hydrophobic pocket that is accessible in kinases with the DFG-out conformation. Table 3 in supplementary material shows one of these cases in which a type-II inhibitor was ranked in the top position with the fast heating MD scheme while it was penalized with a slower heating ramp, highlighting the need of a consensus scoring. In light of these findings it will be interesting in future studies to apply our workflow to other Cys-containing kinases with the DFG-out conformation, in order to search for type II irreversible inhibitors.

To allow us to keep into account all the information collected from the various rankings we adopted a consensus criteria keeping into account (1) the numerical value of the predicted free energies of binding, (2) the presence of ligand-receptor key interactions in the Asp106/Met108 hinge region (Fig. 1) (3) the presence of a sufficiently large cavity in the ATP-binding site to allow the introduction of an appropriate reactive group on the ligand. Clearly, the consensus rank of the 300 potential ligands was devised not only for providing a unique first filter toward the selection of a subset of inhibitors that fit the protein pocket particularly well but also to track the different behaviour of known kinase inhibitors mentioned above. In all cases we obtained negative values of MM-PBSA and MM-GBSA binding free energies (data not shown), indicating a general affinity of the screened ligands for ERK2. This is not surprising considering that the 300 ligands were selected among protein kinase inhibitors. However, since the energy scores were different and in some cases markedly superior, implying a tighter ligand-receptor binding, the ligand rankings could be used to select the best molecules to be passed to the next steps. In addition to the calculated ΔG of binding, protein–ligand complexes were graphically visualized to check whether the ligand-ERK2 interactions after BEAR were preserved comparing to the ligand matching obtained in step 1. We retained for further steps only ligands showing the key interactions with the *hinge region* (Figs. 1 and 3), leading to a reduced set of 26 compounds. Since our first goal was to make as few modifications as possible to already known protein kinase inhibitors, we extracted scaffolds that left unaltered the ligand-receptor interactions of the *hinge region*. These scaffolds were processed through Step 3.

Finally, it is interesting to note that an alternative way for selecting the most promising scaffolds was also tested. In particular, instead of processing the entire database of known inhibitors, we dissected the molecules in their constituting fragments and for each molecule we retained the fragment that binds the *hinge region*, i.e. the part of the

molecule that establishes the most important and conserved interactions with protein kinases. These fragments were aligned with the protein and then processed with BEAR using the settings already described for the entire ligands. Using the fragments we obtained poorer results for the following reasons: (1) MM-PB(GB)SA free energy scores were significantly lower and their variation along the list spanned a much smaller energy window (few kcal/mol instead of the tens of kcal/mol for the entire ligands), leading to a more difficult selection of top-ranking fragments; (2) deciding where to dissect the molecule and select the fragment was somewhat arbitrary; (3) in many cases we found that, during MD, the fragments loose interaction with the hinge and explore different regions of the binding site. This behavior is likely due to the fact that fragments are not optimized for PK inhibition and are too small to adopt a well defined binding mode. For all these reasons we pursued our calculations with the database of entire inhibitors.

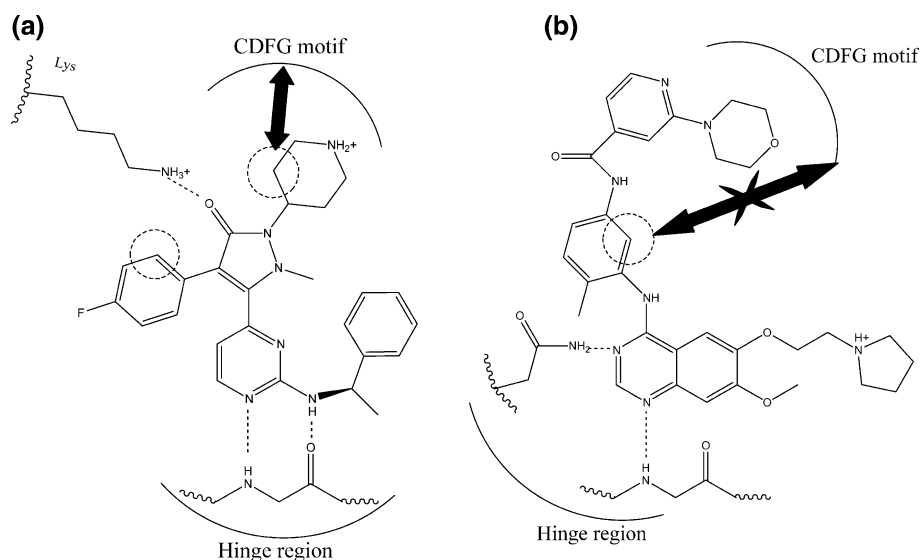
Step 3. Scaffold modification

In this step we started from the 26 scaffolds selected in step 2 and explored modifications of their chemical structure in order to design putative irreversible inhibitors by introduction of suitable reactive groups able to bind the cysteine. Scaffold modifications were guided by the visual inspection of the ligand-receptor reversible complexes and judged on the basis of their scores, the distance between the reactive group and the target cysteine as well as an in-deep analysis of the ligand-receptor interactions of the resulting complexes.

Standing on the information collected in the previous steps, some ligands could approach the target cysteine (Cys166 in ERK2). In these cases, we proposed the introduction of a suitable *cysteine trap* positioned in proximity of the cysteine. It is worth noting that several structures, such as the one shown in Fig. 4a, had enough space to accommodate the reactive group. Vice versa, in some cases, a scaffold modification could not be envisaged without interfering with the basic ligand-receptor interactions. Figure 4b, for instance, shows a ligand for which the introduction of a *cysteine trap* in the ideal location would give structural conflicts. In these cases ligands were either discarded for further analysis or, when possible, simplified so as to obtain pocket-compatible structures (*vide infra*).

The choice of the reactive group was made among a number of different Michael acceptor (MA) functionalities shown in Fig. 5a. An example of introduction of a cysteine trap that keeps into account the required spatial constraints is depicted in Fig. 5b. It is important to note that the different combinations of cysteine traps and scaffolds offer a

Fig. 4 Comparison of a suitable (a) and an unsuitable (b) ligand for their ability to generate a template scaffold. Dashed circles highlight the putative points on the ligand in which the ideal insertion of the cysteine trap would be required



wide range of possibilities, and that the reactivity of the resulting electrophiles may be tuned by introducing suitable substituents. The information collected in Step 1 is especially relevant if one considers that usually commercial libraries contain few compounds that already bear *cysteine traps*. In such situation, the *de-novo* design of small molecules starting from known scaffolds with high affinities for protein kinases appears to be a valid alternative to typical virtual screening performed on commercial compound libraries. Moreover, instead of creating all possible combinations between the selected scaffolds and the MAs of Fig. 5a, we preferred to proceed through the selection of a limited set of modified ligands guided by the visual inspection of the complexes. The reason for doing this is threefold: (1) deciding where to introduce the reactive group was not obvious, e.g. in which position of an aromatic ring, or the length of a linker, if necessary; (2) we could monitor the distance between the chosen reactive group of the ligand and the target cysteine; (3) we limited the far more high- and time-consuming calculations implied in screening all the combinations, considering that each MA can potentially be introduced in many positions of a giving scaffold. Our choice of visually inspecting the ligand-receptor complexes in order to drive the design is supported by the fact that for each scaffold only few specific substitutions appeared to be suitable to obtain reasonable C–S distances and favorable binding features.

An important aspect of the design is how to control selectivity among the various kinases. Searching for irreversible inhibitors that target a given cysteine of the active site is clearly an advantage for selectivity because one can expect that only the kinases possessing that Cys (Table 1) will be irreversibly inhibited, and that reversible binding to kinases without the Cys will be much weaker. Among the kinases in Table 1, the selectivity can be controlled by fine-

tuning differences in amino acid composition of active site residues. To prove that this is viable, we reported in Table 2 the sequence alignments of the segments lining the ATP binding pocket for the 46 kinases with the target Cys [31]. The alignment shows that while some segments and residues are conserved or generally conserved, others clearly differ from kinase to kinase. Among these, the gatekeeper (Gln105 in ERK2) exhibits significant variations. Thus, the introduction of a particular *cysteine trap* that favours interactions with the gatekeeper is a clear advantage for selectivity. Another useful motif in this region is the short linker located downstream of the hinge residues. Table 2 shows that some kinases have a shorter linker that was proven to be important for modulating selectivity [32]. Other putative regions with amino acid variability are the C-helix and the P-loop. It should be noted however, that C-helix is located in the inner part of the ATP-binding pocket and therefore is less prone to interact with ligands. Vice versa the P-loop, whose conformation plays an important role for determining the shape and the opening of the ATP-binding site, may provide other anchoring points important for modulating selectivity [33].

Step 4. Modelling of reversible and irreversible complexes

From step 3 we obtained several compounds that were submitted to extensive computational refinements. In this step, we performed longer MD simulations of the reversible and irreversible ligand-kinase complexes solvated with explicit water molecules. The comparison of their binding modes is useful to investigate possible conformational changes that ligand and receptor may undergo when the

Fig. 5 Examples of cysteine traps moieties that can be considered for the derivatization of the scaffolds of step 2 (a), and a practical application in the case of the scaffold shown in Fig. 4 (b)

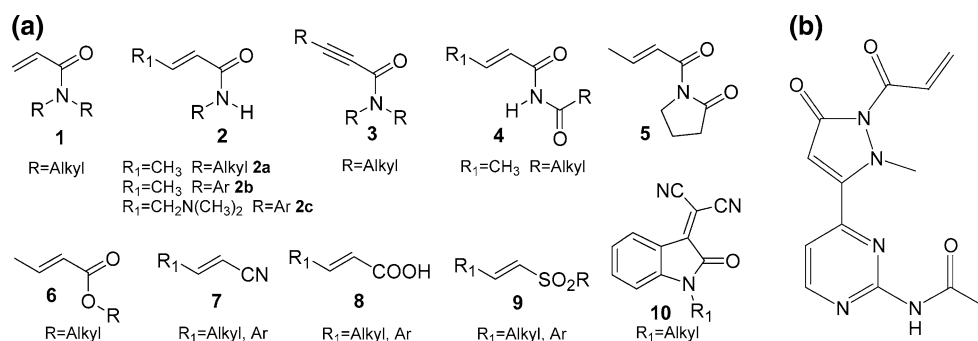


Table 2 Sequence alignment of the protein kinases of Table 1 that have the target cysteine site in the ATP binding pocket

Kinase	P-loop	Catalytic lysine	C-Helix	Hinge region	Hydrophobic pocket	CDFG motif
AAK1/1-304	DEVLAEGFAIVF	MKCALRM	DLQVCKRE	LIIMDFGRGGGV	IIIRDLKVENILL	VLCDFGSA
BMP2K1/1-305	EESLAEGFSTVF	IRCALRM	DLNVCKRE	LIIMEYCRAGGV	IIIRDLKVENILL	VLCDFGSA
CDKL2_Hsap/1-294	IGKIGESYGVVF	QIVAIKF	IKKIALRRLMLKQ	HLVFEFYDHTVL	IIIRDLKVENILL	KLCDFGSA
CDKL2_Hsap/1-293	LGLVCESYGMVM	RIVAIKF	VKKIAMRKLKQ	YLVFEFYDHTVL	IIIRDLKVENILL	KLCDFGSA
CDKL3_Hsap/1-293	LGLVCESYGVVF	QIVAIKF	VNKIAMRKLKQ	HLVFEFYDHTVL	IIIRDLKVENILL	KLCDFGSA
CDKL4_Hsap/1-293	LAKTGESYGVVF	QVVAVKF	VKKIALRRLMLKQ	HLVFEFYDHTVL	IIIRDLKVENILL	KLCDFGSA
CDKL5_Hsap/1-295	LGVVCEAYGVVL	EIVAIKF	VKETTLLRLKMLRT	YLVFEFYVEKNML	IVIRDLKVENILL	KLCDFGSA
ERK1_Hsap/1-269	LQYICERAYGMVS	TRVAIKF	YCQRTLRLKILLR	YIVQDLME-TDL	VLIRDLKVENILL	KLCDFGSA
ERK2_Hsap/1-269	LSYICERAYGMVC	VRVAIKF	YCQRTLRLKILLR	YIVQDLME-TDL	VLIRDLKVENILL	KLCDFGSA
FLT3_Hsap/1-334	GKVLCEAFGKVM	IQVAVML	EREALMSLKMMTQ	LVITEFYCYGDL	CVIRDLKVENILL	KLCDFGSA
GAUK_Hsap/1-314	RRVLAEGFAFVY	REYALRL	KNRAIQVCFMKK	LLLTLELCK-GQL	IIIRDLKVENILL	KLCDFGSA
GSK3A_Hsap/1-285	IKVIGNSFQGVVY	ELVAIKV	---KNRQLQIMRK	NLVLDYVP-ETV	VCIRDLKVENILL	KLCDFGSA
GSK3B_Hsap/1-285	TKVIGNSFQGVVY	ELVAIKV	---KNRQLQIMRK	NLVLDYVP-ETV	VCIRDLKVENILL	KLCDFGSA
KIT_Hsap/1-338	GKTLCEAFGKVM	MTVAVML	EREALMSLKVLSY	LVITEFYCYGDL	CVIRDLKVENILL	KLCDFGSA
MAPK15/1-311	RRQLCEAFGKVM	EVVAIKF	SLQRTFLRLTQGE	YLVFEFYDHTVL	VVIRDLKVENILL	KLCDFGSA
MAPKAPK5_Hsap/1-283	TQKLCEAFGKVM	ERFALIL	---RNEVRLHMM	LIVMEMMEGGE	IAIRDLKVENILL	KLCDFGSA
MAP2K1_Hsap/1-294	ISELCEANGGVVF	LVMAIRL	IRNQIRLEQLVHE	SICMEHMDGGS	IMIRDLKVENILL	KLCDFGSA
MAP2K2_Hsap/1-298	ISELCEANGGVVF	LVMAIRL	IRNQIRLEQLVHE	SICMEHMDGGS	IMIRDLKVENILL	KLCDFGSA
MAP2K3_Hsap/1-292	ISELCEANGGVVF	LVMAIRL	IRNQIRLEQLVHE	SICMEHMDGGS	IMIRDLKVENILL	KLCDFGSA
MAP2K4_Hsap/1-296	LGELCEANGGVVF	QIMAVRI	EQKQLLMDLVVMR	WICMELMS-TS	VIIRDLKVENILL	KLCDFGSA
MAP2K5_Hsap/1-244	RDTLCEANGGVVF	KILAVVI	LQKQIMSELEILYK	SICTEFMDGGS	ILIRDLKVENILL	KLCDFGSA
MAP2K6_Hsap/1-282	IMELCEANGGVVF	QIMAVRI	EQKQLLMDLVVMR	WICMELMS-TS	VIIRDLKVENILL	KLCDFGSA
MAP2K7_Hsap/1-288	LGEMCEANGGVVF	HVIAIRL	ENKRIILMDLVVLK	FIAELMLO-TCA	VIIRDLKVENILL	KLCDFGSA
MINK1_Hsap/1-286	SELLCEANGGVVF	KEYAVVI	SRSRVFVEVETLYQ	YLVFEFYDHTVL	IAIRDLKVENILL	KLCDFGSA
MINK2_Hsap/1-285	EDVLCERAHARVQ	QEVAVVI	IRSRVFVEVETLYQ	YLVFEFYDHTVL	IAIRDLKVENILL	KLCDFGSA
NIK_Hsap/1-287	QLRLCEANGGVVF	FOCAVVI	---AEELMAAG	NIFMELLEGGSL	ILIRDLKVENILL	KLCDFGSA
NLK_Hsap/1-290	DRPICEANGGVVF	KRVAVVI	SKRVFRLKMLCF	YVVTLEMG-SD	ILIRDLKVENILL	KLCDFGSA
PDGFR2_Hsap/1-358	GRVLCEANGGVVF	MKVAVVI	EQKQLLMDLVVMR	YIITEFYDHTVL	CVIRDLKVENILL	KLCDFGSA
PDGFRb_Hsap/1-359	GRVLCEANGGVVF	MKVAVVI	EQKQLLMDLVVMR	YIITEFYDHTVL	CVIRDLKVENILL	KLCDFGSA
PKD1_Hsap/1-287	DEVLCEANGGVVF	RDVAIRL	EQSGLRNEVAILQ	FVVMELKLGDM	IVIRDLKVENILL	KLCDFGSA
PKD2_Hsap/1-287	DEVLCEANGGVVF	RDVAIRL	EQSGLRNEVAILQ	FVVMELKLGDM	IVIRDLKVENILL	KLCDFGSA
PRK4_Hsap/1-313	YGVCEANGGVVF	QEVAVVI	-QKTGLKLEFLLK	CLVFEPLS-MNL	ILIRDLKVENILL	KLCDFGSA
RSK1_Hsap/1-375	KETICEANGGVVF	MEYAVVI	---PSEIEILLR	YLVTELMRGGE	VVIRDLKVENILL	KLCDFGSA
RSK2_Hsap/1-373	KEDICEANGGVVF	MEYAVVI	---PTEIEILLR	YVVTLELMRGGE	VVIRDLKVENILL	KLCDFGSA
RSK3_Hsap/1-376	KEDICEANGGVVF	MEYAVVI	---PSEIEILLR	YLVTELMRGGE	VVIRDLKVENILL	KLCDFGSA
RSK4_Hsap/1-384	KEDICEANGGVVF	MEYAVVI	---PSEIEILLR	YLVTELMRGGE	VVIRDLKVENILL	KLCDFGSA
SPGK_Hsap/1-254	HQELCEANGGVVF	LEFAIRL	-KASARRARLLAR	VIVTELETEEL	VLIRDLKVENILL	KLCDFGSA
Fused_Hsap/1-251	LEMICEANGGVVF	QVVALFI	ELRNLRQVIEIMRG	VVVTDYAE-GE	ILIRDLKVENILL	KLCDFGSA
TAK1_Hsap/1-248	EEVVCERAFGVVF	KDVAIRL	ERKAFIVELQLSR	CLVMEYAEAGGS	LIIRDLKVENILL	KLCDFGSA
TGFR2_Hsap/1-296	DTLVCEANGGVVF	ETVAVVI	TEKDFISQINL---	WLITAFHAKGN	IVIRDLKVENILL	KLCDFGSA
PBK_Hsap/1-287	MOKLCEANGGVVF	SPWAVVI	YQKRLMDAKILKS	CLAMEYGEKS	LLIRDLKVENILL	KLCDFGSA
FLT1_Hsap/1-328	GKSLCEANGGVVF	RTVAVVI	EYKALMTLKLTH	MVIVEFYKYGN	CIIRDLKVENILL	KLCDFGSA
KDR_Hsap/1-327	GKPLCEANGGVVF	RTVAVVI	EHRALMSLKILIH	MVIVEFYKYGN	CIIRDLKVENILL	KLCDFGSA
FLT4_Hsap/1-325	GRVLCEANGGVVF	DTVAVVI	EHRALMSLKILIH	MVIVEFYKYGN	CIIRDLKVENILL	KLCDFGSA
ZAK_Hsap/1-244	FENCCEANGGVVF	KEVAVVI	---IEKAEILSV	GIVTEYASLGSL	VIRDLKVENILL	KLCDFGSA

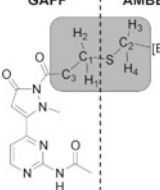
covalent bond is formed and to assess the position of the cysteine trap relative to the cysteine. MD of the complexes was performed by setting up an automated procedure able to assign parameters, prepare input files, and run the MD simulations with the Amber software. In the case of the covalently bound complexes, the covalent linkage between MAs and Cys166 was conceived by mixing generalized amber force-field (GAFF) [34] parameters for the ligand and Amber force field for the cysteine as depicted in Table 3. Such mixing required a re-parametrization of the terms located at the borderline between the two force fields (Table 3). In particular bond distances and angles were calibrated to reproduce the geometry of the ligand-cysteine adducts energy-minimized with the semi-empirical method AM1. Dihedral parameters were assigned on the basis of analogous parameters already available in the Amber force

field for similar atom types. For each molecule, energy-minimization with AMBER was performed to make sure that the optimized conformation was in agreement with the AM1-optimized conformation.

In all cases, the complexes were first energy-minimized and then subjected to extensive MD simulations of at least 6 ns at 300 K in a periodic box of water molecules. The MD trajectories were saved and used for structural analyses and calculation of MM-PBSA and MM-GBSA binding free energy scores. Within the computational workflow of Fig. 2, a molecule is considered a good candidate inhibitor when, according to MD trajectories of reversible and irreversible complexes, it shows:

- Conservation of key interactions in the hinge region and other important residues in the ATP pocket.

Table 3 Custom parameterization of borderline terms between AMBER and GAFF force fields for the modeling of covalently bound complexes

GAFF	AMBER	Atoms involved	Amber and GAFF atom types	Potential term type
		S, C1	S -c3	Bond
		C1, S, C2	c3- S-CT	Angle
		H1, C1, S	h1-c3-S	Angle
		H2, C1, S		
		C3, C1, S	c3-c3-S	Angle
		H1, C1, S, C2	X -c3- S-X	Dihedral
		H2, C1, S, C2		

Lower case and upper case atom types are used in the GAFF or AMBER force field, respectively

- ii) Proximity of the reactive group to the cysteine sulphur
- iii) Similar reversible and irreversible binding modes without important conformational rearrangement of the ligand or the kinase.
- iv) Particularly favorable predicted free energies of binding.

In this respect, a sensible change in ligand conformation, e.g. a rearrangement of the ligand with loss of the key interaction in the hinge, or a sensible variation in the kinase conformation indicate that thermodynamic penalties have to be paid to achieve effective irreversible inhibition. In this context, longer molecular dynamics simulations with explicit solvent are needed to better explore the potential energy surface of the modified scaffolds around the structural template of the biological target. This is particularly true for the scaffolds that in step 3 were more manipulated. At this stage, the importance of this conformational search relates to the monitoring of solvent effects with explicit water molecules and served to highlight time-dependent structural features already reported to be critical for kinases [17, 28] and other biological targets. In these cases, without an extensive conformational sampling the time-dependent structural features important for inhibitor binding could not have been detected. The test cases that will be presented in the next paragraph exemplify the importance of providing detailed information on the stability of both reversible and irreversible complexes over simulation time, including the spatial position of reactive group in reversible complexes.

Test cases

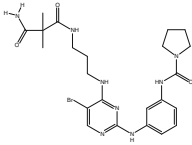
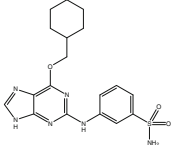
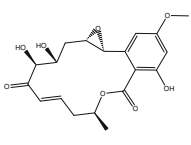
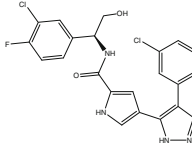
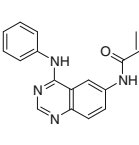
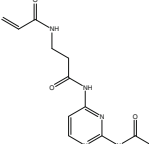
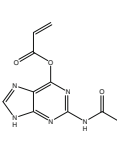
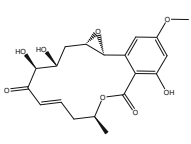
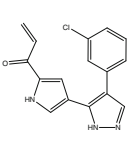
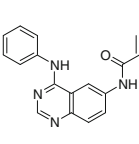
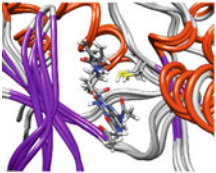
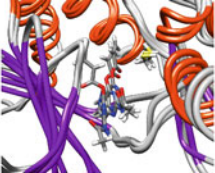
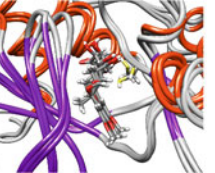
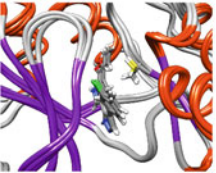
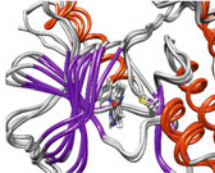
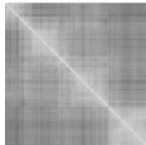
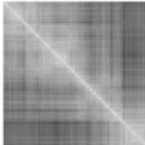
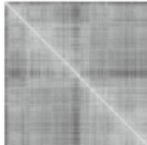
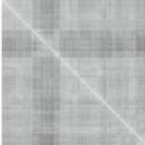
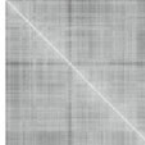
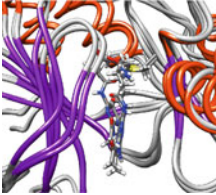
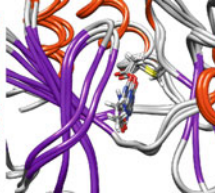
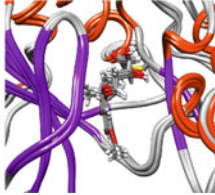
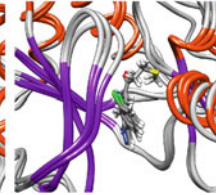
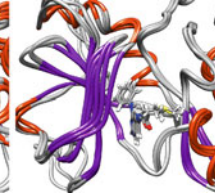
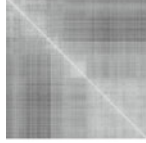
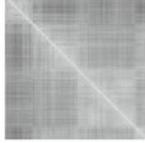
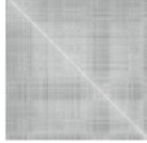
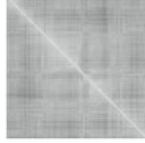
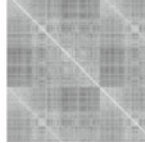
Table 4 summarizes the results obtained in the fourth step for some validation cases. As test cases, we have selected hypothemycin (3), a known irreversible inhibitor of ERK2 [16, 17], and three compounds that passed the first 3 steps of our workflow (1,2 and 4). In addition, as a further validation we have selected a quinazoline irreversible

inhibitor (5) covalently bound to EGFR, a protein kinase with a cysteine in a position different from ERK2 [19].

The extensive MD simulations performed on compound 3 (hypothemycin) showed a remarkable similarity of the structure of its reversible and irreversible complexes with ERK2. This similarity can be depicted from the representative structures collected along the MD trajectories and from the comparison of the RMSd plots shown in Table 4. The latter plots, in which light pixels correspond to low RMSd values among each pair of MD frames, show that the conformation of the hypothemycin complexes is fairly stable over time. In general, for compound 3 we observed: (1) a conservation of the key interactions with the hinge residues as evidenced from the superimposition of the representative structures of the reversible and covalent complexes (Table 4), (2) a correct orientation of the Michael acceptor toward the cysteine and a short distance between the reactive carbon and the sulphur in the reversible complex (C–S distance of 4.3 ± 0.6 Å), (3) a reversible binding free energy score of -22.6 ± 1.7 kcal/mol that reflects the favorable interaction observed experimentally, (4) a small variation of the RMSd values along the whole trajectory length of 10 ns as depicted by the off-diagonal pixels of the RMSd plots in Table 4. Importantly, the computational results obtained for hypothemycin are consistent with the experimental evidences that this compound is a potent irreversible inhibitor of ERK2 [16]. It should be noted that the starting point used to perform the test case was not the X-ray structure of the ERK2-hypothemycin complex recently reported by Rastelli et al. [17] (PDB code 3C9W), but the same structure of ERK2 used to process all the ligands in the workflow (PDB code 1TVO). Therefore, the fact that we obtained positive results with a protein structure not pre-organized around the ligand is an indication that the overall procedure is sensible. In addition, the results are consistent with the crystal structure of the ERK2-hypothemycin complex [17].

Having obtained good agreement with the experimental data of compound 3, we have exemplified in Table 4 the computational results of three putative irreversible inhibitors generated with the workflow of Fig. 2. The free energy estimation of compound 1 is identical to that of hypothemycin (ΔG of -22.6 ± 2.1 kcal/mol), indicating a good affinity of the molecule towards ERK2 in the reversible complex. Equally, the analysis of the representative structures shows that the key interactions in the *hinge region* are conserved along the MD simulation. In spite of this, some conformational changes, e.g. a re-orientation of the scaffold, appears when comparing the reversible and covalent complexes. Moreover, the C–S distance of more than 8 Å is not favorable for nucleophilic attack of the cysteine (Table 4). On the other hand, compound 2 shows a remarkably lower free energy score (ΔG of –

Table 4 Results overview for three selected compounds and hypothemycin

Origin PDB code (protein kinase).	1Z5M (PDK1)	1OIU (CDK2)	3C9W (ERK2)	2OJJ (ERK2)	2J5E (EGFR)
Co-crystallized ligand structure.					
	Compound 1	Compound 2	Compound 3	Compound 4	Compound 5
Modified ligand structure ^a					
Representative structures of the reversible-bound complexes ^{b,f}					
Free energy estimation (reversible) ^c	-22.6 (2.1)	-14.7 (7.1)	-22.6 (1.7)	-29.2 (1.0)	-25.2 (0.6)
C-S distance ^d	8.4 (1.3)	5.8 (1.8)	4.3 (0.6)	3.8 (0.4)	5.4 (0.9)
RMSd plot for reversible-bound complexes ^{e,f}					
Representative structures of the covalent-bound complexes ^{b,f}					
RMSd plot for covalent-bound complexes ^{e,f}					

^a Structure modified at the end of Step 3. Compound 3 (hypothemycin) was already reported in the literature as an irreversible inhibitor of ERK2 therefore no chemical modification was made for this compound

^b MDs (Step 4) are carried out for 10 ns of simulation. Trajectories are submitted to *ptraj* software to cluster the snapshots. The representative structures of the three (out of 10) most populated cluster are obtained and overlaid in the figure

^c Free energy estimation of reversible-bound complexes. Average value and standard deviation in parenthesis obtained from five consecutive MM-PBSA determinations of 2 ns each. Values are in kcal/mol

^d Distance between the electrophile carbon of the ligand and the nucleophilic sulphur of the cysteine reported as the average value over the MD frames with their standard deviation in parenthesis. All values are in Å

^e The plots show all-by-all pair-wise root-mean-square deviations (RMSd) among the frames collected on 10 ns trajectory. RMSd values are calculated within 8 Å around the ligand. White pixels are assigned an RMSd of 0.0 while black pixels an RMSd of 3.0. Intermediate values are mapped to grayscale

^f Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081) [20]

14.7 ± 7.1 kcal/mol) if compared with compounds **1** and **3**. The high standard deviation of the free energy result is due to the relatively high structural fluctuations of **2** as evidenced by the analysis of the most representative structures of the reversible-bound complexes and the corresponding RMSd plot (Table 4). The C–S distance of 5.8 ± 1.8 Å is significantly better than that of compound **1**, though still 1.5 Å higher than hypothemycin. The RMSd plot obtained for the covalent-bound simulation shows lower fluctuations that are consistent with the structural constraints imposed by the presence of a covalent bond. However, the prominent structural fluctuations of compound **2** in the reversible complex imply that important conformational rearrangements are needed to undergo covalent reaction with the target cysteine. Differently to compounds **1** and **2**, compound **4** is predicted to be an effective irreversible inhibitor of ERK2 by the computational procedure. This newly conceived ligand shows even better profile with respect to hypothemycin **3**. In particular, compound **4** passed Steps 1–3 and showed (1) an excellent MM-PBSA score of the reversible complex (ΔG of -29.2 ± 1.0 kcal/mol) (2) a distance between the MA and the target cysteine shorter than 4 Å that also shows minimal fluctuations during MD (3.8 ± 0.4 Å) (3) the binding mode of the reversible and irreversible complexes were similar and conserved along the MD trajectories, as evidenced by the RMSd plots. It will be interesting to see if this compound is an irreversible inhibitor of ERK2 and related kinases with our target Cys. Finally, as a further validation step and in order to assess the suitability of the workflow to other reactive cysteines, we applied the computational workflow to a structural template of EGFR covalently bound to a known quinazoline irreversible inhibitor (PDB code 2J5E) [19]. In this template the reactive cysteine is located a few residues downstream of the *hinge*, differently from the cysteine described in Step 1 which is located right before the DFG-motif. The computational results obtained on the EGFR-quinazoline complex clearly show a computational profile similar to that of ERK2-hypothemycin (Table 4), thus providing an additional successful validation of the workflow even for other protein kinases beyond ERK2 that may undergo covalent inactivation in the ATP-binding site.

Conclusions

An important approach for the development of anticancer drugs is irreversible inhibition. In this context, computer-aided drug design strategies that address this need are much sought. The computational workflow that we have presented is specifically conceived to assist drug design efforts in this direction. The workflow takes the form of a multi-step

procedure that includes: the creation of a database of known reversible inhibitors of protein kinases, the selection of the most promising scaffolds that bind a specific kinase template, the modification of the scaffolds by introduction of suitable cysteine traps and the final computational evaluation of reversible and irreversible complexes with molecular dynamics and free energy analysis. The workflow that was tested on ERK2 and a database of protein kinase inhibitors is useful to detect the aptitude of ligands to form stable reversible complexes and to assess their suitability to form covalent complexes with the target cysteine. Moreover, the structural observations emerging from the various steps of the workflow constitute valuable insights for the design of putative irreversible inhibitors, helping in deciding which structures are more interesting and giving useful hints for compound modification and optimization. Its validation on hypothemycin, a known irreversible inhibitor of ERK2, showed that the procedure is sensible and suitable to suggest novel compounds (e.g. compd **4**). The workflow may be easily and quickly adapted to other biological targets, especially other protein kinases that may undergo covalent inactivation in the ATP-binding site. In addition such procedure may be applicable to Cys-containing crystal structures that adopt the DFG-out conformation. This perspective may be relevant for developing type II irreversible inhibitors.

Clearly, the definitive test of the robustness of our computational workflow will come from its application in the field, which implies synthesis of new derivatives and determination of inhibitory activities. Such evaluations are currently underway on molecular scaffolds not related to hypothemycin.

Acknowledgments This work was supported by grants from AIRC, the Italian Association for Cancer Research (Research grant “Novel irreversible protein kinase inhibitors targeting a conserved active site cysteine”).

References

1. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) Science (New York, N.Y) 298:1912–1934
2. Krause DS, Van Etten RA (2005) New Engl J Med 353:172–187
3. Liao JJ (2007) J Med Chem 50:409–424
4. Margutti S, Laufer SA (2007) Chem Med Chem 2:1116–1140
5. Bikker JA, Brooijmans N, Wissner A, Mansour TS (2009) J Med Chem 52:1493–1509
6. Otori M, Kinoshita T, Yoshimura S, Warizaya M, Nakajima H, Miyake H (2007) Biochem Biophys Res Commun 353:633–637
7. Wissner A, Fraser HL, Ingalls CL, Dushin RG, Floyd MB, Cheung K, Nittoli T, Ravi MR, Tan X, Loganzo F (2007) Bioorg Med Chem 15:3635–3648
8. Engelman JA, Zejnullahu K, Gale CM, Lifshits E, Gonzales AJ, Shimamura T, Zhao F, Vincent PW, Naumov GN, Bradner JE, Althaus IW, Gandhi L, Shapiro GI, Nelson JM, Heymach JV,

- Meyerson M, Wong KK, Janne PA (2007) *Cancer Res* 67:11924–11932
9. Pan Z, Scheerens H, Li SJ, Schultz BE, Sprengeler PA, Burrill LC, Mendonca RV, Sweeney MD, Scott KC, Grothaus PG, Jeffery DA, Spoerke JM, Honigberg LA, Young PR, Dalrymple SA, Palmer JT (2007) *Chem Med Chem* 2:58–61
 10. Cohen MS, Zhang C, Shokat KM, Taunton J (2005) *Science* (New York, N.Y.) 308:1318–1321
 11. Potashman MH, Duggan ME (2009) *J Med Chem* 52:1232–1246
 12. Robertson JG (2005) *Biochemistry* 44:5561–5571
 13. Knight ZA, Shokat KM (2005) *Chem Biol* 12:621–637
 14. Kwak EL, Sordella R, Bell DW, Godin-Heymann N, Okimoto RA, Brannigan BW, Harris PL, Driscoll DR, Fidias P, Lynch TJ, Rabindran SK, McGinnis JP, Wissner A, Sharma SV, Isselbacher KJ, Settleman J, Haber DA (2005) *Proc Natl Acad Sci USA* 102:7665–7670
 15. Zhang JM, Yang PL, Gray NS (2009) *Nat Rev Cancer* 9:28–39
 16. Schirmer A, Kennedy J, Murli S, Reid R, Santi DV (2006) *Proc Natl Acad Sci USA* 103:4234–4239
 17. Rastelli G, Rosenfeld R, Reid R, Santi DV (2008) *J Struct Biol* 164:18–23
 18. Ohori M, Kinoshita T, Okubo M, Sato K, Yamazaki A, Arakawa H, Nishimura S, Inamura N, Nakajima H, Neya M, Miyake H, Fujii T (2005) *Biochem Biophys Res Commun* 336:357–363
 19. Blair JA, Rauh D, Kung C, Yun CH, Fan QW, Rode H, Zhang C, Eck MJ, Weiss WA, Shokat KM (2007) *Nat Chem Biol* 3:229–238
 20. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) *J Comput Chem* 25:1605–1612
 21. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) *J Comput Chem* 26:1668–1688
 22. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Mathews DH, Schafmeister C, Ross WS, Kollman PA (2006) *AMBER*. University of California, San Francisco CA
 23. Ferrari AM, Degliesposti G, Sgobba M, Rastelli G (2007) *Bioorg Med Chem* 15:7865–7877
 24. Rastelli G, Degliesposti G, Del Rio A, Sgobba M (2009) *Chem Biol Drug Des* 73:283–286
 25. Rastelli G, Del Rio A, Degliesposti G, Sgobba M (2010) *J Comput Chem* 31:797–810
 26. Stoica I, Sadiq SK, Coveney PV (2008) *J Am Chem Soc* 130:2639–2648
 27. Kuhn B, Gerber P, Schulz-Gasch T, Stahl M (2005) *J Med Chem* 48:4040–4048
 28. Del Rio A, Baldi BF, Rastelli G (2009) *Chem Biol Drug Des* 74:630–635
 29. Brown SP, Muchmore SW (2007) *J Chem Inf Model* 47:1493–1503
 30. Lyne PD, Lamb ML, Saeh JC (2006) *J Med Chem* 49:4805–4808
 31. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) *Bioinformatics* (Oxford, England) 25:1189–1191
 32. Katayama N, Orita M, Yamaguchi T, Hisamichi H, Kuromitsu S, Kurihara H, Sakashita H, Matsumoto Y, Fujita S, Niimi T (2008) *Proteins* 73:795–801
 33. Kinoshita T, Warizaya M, Ohori M, Sato K, Neya M, Fujii T (2006) *Bioorg Med Chem Lett* 16:55–58
 34. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) *J Comput Chem* 25:1157–1174