Assessing protein kinase target similarity: Comparing sequence, structure, and cheminformatics approaches


*BBA Proteins and Proteomics*, 2015, 1854, 1605-1616
Assessing protein kinase target similarity: Comparing sequence, structure, and cheminformatics approaches☆

Osman A. Gani, Balmukund Thakkar, Dilip Narayanan, Kazi A. Alam, Peter Kyomuhendo, Ulli Rothweiler, Veronica Tello-Franco, Richard A. Engh *

The Norwegian Structural Biology Centre, Department of Chemistry, University of Tromsø, Tromsø, Norway

A R T I C L E   I N F O

Article history:
Received 15 March 2015
Received in revised form 8 May 2015
Accepted 11 May 2015
Available online 19 May 2015

Keywords:
Structure based drug design
ABL
Aurora
Drug repurposing
Crystal structure
Cheminformatics

A B S T R A C T

In just over two decades, structure based protein kinase inhibitor discovery has grown from trial and error approaches, using individual target structures, to structure and data driven approaches that may aim to optimize inhibition properties across several targets. This is increasingly enabled by the growing availability of potent compounds and kinome-wide binding data. Assessing the prospects for adapting known compounds to new therapeutic uses is thus a key priority for current drug discovery efforts. Tools that can successfully link the diverse information regarding target sequence, structure, and ligand binding properties now accompany a transformation of protein kinase inhibitor research, away from single, block-buster drug models, and toward “personalized medicine” with niche applications and highly specialized research groups. Major hurdles for the transformation to data driven drug discovery include mismatches in data types, and disparities of methods and molecules used; at the core remains the problem that ligand binding energies cannot be predicted precisely from individual structures. However, there is a growing body of experimental data for increasingly successful focussing of efforts: focussed chemical libraries, drug repurposing, polypharmacological design, to name a few. Protein kinase target similarity is easily quantified by sequence, and its relevance to ligand design includes broad classification by key binding sites, evaluation of resistance mutations, and the use of surrogate proteins. Although structural evaluation offers more information, the flexibility of protein kinases, and differences between the crystal and physiological environments may make the use of crystal structures misleading when structures are considered individually. Cheminformatics may enable the “calibration” of sequence and crystal structure information, with statistical methods able to identify key correlates to activity but also here, “the devil is in the details.” Examples from specific repurposing and polypharmacology applications illustrate these points. This article is part of a Special Issue entitled: Inhibitors of Protein Kinases.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The first protein kinase structure determinations [1,2] initiated the era of kinase inhibitor structure based drug design. The rho-kinase inhibitor HA1077 [3] (Fasudil) was approved in Japan in 1995, but it was the approval of imatinib (Glivec) [4,5] in 2001 that firmly established protein kinases as high priority drug targets, particularly in oncology. Since then, the structural information available for drug design, has grown massively. Now, structure and data driven approaches even may rationally attempt to optimize target selectivity profiles based on biological data, using information from thousands of known inhibitors. Assessing the prospects for adapting known compounds to new therapeutic uses is thus a key priority for current drug discovery efforts, and repurposing or redesigning known compounds may be most efficient [6–11]. Tools that can successfully link the diverse information regarding target sequence, structure, and ligand binding properties have the potential to transform kinase inhibitor research away from single, block-buster drug models, and into “personalized” and other niche areas where also academic groups may specialize.

An understanding of “where to look” aids these efforts, and the evaluation of protein kinase target similarity is part of this. This is most easily quantified by sequence, and such a bioinformatics approach is familiar to the broadest audience, usually using phylogenetic trees of whole sequences. Broad and useful similarity classification can be made by identifying key binding sites, supporting an evaluation of resistance mutations, and the use of surrogate proteins to aid experiment. However, ligand design requires an understanding of ligand–target interactions, and this is most directly a structural topic. Although structural evaluation uses and offers more information, the flexibility of protein kinases, and differences between the crystal and physiological environments may make the use of crystal structures misleading when structures are considered individually. For use of structures collectively, informatics

☆ This article is part of a Special Issue entitled: Inhibitors of Protein Kinases.
* Corresponding author.
E-mail address: Richard.Engh@uit.no (R.A. Engh).
methods must be used. Such methods may enable the “calibration” of sequence and crystal structure information, with statistical methods able to identify key correlates to activity but also here, it is still true that “the devil is in the details” [12].

2. Results and discussion

2.1. Similarity by sequence

2.1.1. Key residues

A direct way to visualize some key aspects of protein kinase similarity is simply to plot the distribution of key residues on a phylogenetic tree with the same layout as the original kinase analysis of Manning et al. [13], which has become intimately familiar to most protein kinase researchers. Thus, the gatekeeper distribution (Fig. 1A) readily shows the clustering of the most common gatekeeper residues: Met, Thr, Leu, and Phe, and also identifies some potentially surprising connections, such as Flt3 with CMGC kinases. Less well known is the conservation of the “gatekeeper + 2 residue” (Fig. 1B), which is usually an aromatic amino acid (tyrosine or phenylalanine, sometimes tryptophan as in BRAF) or leucine. This site is particularly interesting as one that is often found in drug resistant cancers, and in CML is often the one with the most rapid appearance [14]. Other key residues include the glycine residues of the glycine-rich loop, with the consensus sequence GxGxxG. Although the function of these residues is not entirely clear, besides sterically allowing ATP binding [15], they contribute the high flexibility seen for the glycine-rich loop in response to inhibitor binding, with possibly enhanced flexibility for Ab1, with its GGGxxG sequence, and reduced flexibility for protein kinases lacking the third glycine (Fig. 1C; the first two are most highly conserved).

Although plots of individual residues are informative with respect to specific features, they cannot suggest overall inhibitor binding similarities between kinases. On the other hand, the phylogenetic tree itself does represent overall homology, but not specifically for inhibitor binding. An intermediate similarity measure would be the use of pseudosequences, chosen to represent residues important for inhibitor binding. The choice of these residues is however not unique. The differing binding geometries of different inhibitors involve different side chains, and many residues that play a role in binding may make no contact at all, but may influence other properties, such as flexibility.

Fig. 2 shows pseudosequence similarity plots for Aurora B and Ab1 kinases, calculated using the Needleman–Wunsch algorithm and a Blosum62 similarity matrix as implemented in Mathematica, on a pseudosequence of key residues. For comparison, Fig. 2 also shows the correlations of inhibitor binding energies for the same kinases with the protein kinases in the Ambit kinase profiling set of 2011 [16]. These pseudosequences show the Aurora kinases to be quite unique (Fig. 2A), with some cognates in the CAMK group, rather more specific than the kinase binding data show (Fig. 2B). In contrast, Ab1 pseudosequence similarity clusters within the tyrosine kinase subfamily (Fig. 2C), with better agreement with the inhibitor data.

2.2. Similarity by structure

Sequence determines structure, and structure determines binding energetics, so structure represents a higher level of information content for evaluating target similarity; efforts to contribute to and use the information from worldwide Protein Data Bank [17,18] reflect its central importance. However, even though sequence determines structure, a unique sequence does not guarantee a unique structure, despite a persistent prejudice to the contrary. Throughout the period of protein crystallography, protein structures have been known to be dynamic and dependent on total chemical composition (i.e. posttranslational modifications), environment (pH, temperature, ionic strength), binding partners (proteins, small molecules), and so on. As the PDB grows, more and more of these effects can be recognized and characterized, enabling meaningful PDB wide searches, e.g. repurposing opportunities [19,20]. Because the structural variability is large compared to what determines ligand binding energetics, conformational space remains enormous compared to the size of the PDB.

Many of the key states of protein kinases have been determined; key activity modulation mechanisms involving especially “helix C”, the activation loop (including phosphorylation and the “DFG-in”, “DFG-out”, and intermediate states), and the glycine-rich loop have been identified. However, their observation in association with a particular inhibitor does not guarantee that that is the only, or even the lowest energy state of the complex. Crystalization conditions, the energy of crystal packing contacts, and the state of the protein used for crystalization can be major determinants for the observed state. Compounding this problem is the fact that inhibitors are usually assumed to possess a single binding geometry: 1) Crystals lacking the resolution to identify structural heterogeneity will lead to a single modeled structure, as a rule, 2) inhibitors that bind with structural heterogeneity may cause such a moderate resolution, and 3) optimization of crystalization conditions to maximize resolution may be a search for conditions to eliminate alternate binding geometries that occur in a biological environment. The examples presented in the section illustrate some of the difficulties.

2.2.1. Understanding the binding mode(s) of VX680

The inhibitor VX680 (or MK0457), originally identified as an Aurora kinase inhibitor, has been in several clinical trials for cancer indications including solid tumors [21], leukemias [22–24], and lung cancer, as reflected by its low nanomolar inhibition of Ab1, Aurora, and Flt3 kinases, including the drug resistant Ab1 mutant T315I [16]. The cross-reactivity between Ab1 and Aurora is in apparent contradiction to their overall similarity (unlike the cross reactivity between Aurora and Flt3, which is identified, Fig. 2). One notable feature that Aurora and Ab1 kinases share when binding to VX680 is a reconfiguration of the glycine-rich loop to form a pi–pi stacking arrangement between the inhibitor and the highly conserved aromatic amino acid at the beta-hairpin turn of the loop. If the two kinases shared an anomalous propensity for such a reconfiguration, the cross-reactivity might be explained. And indeed, Ab1 is unusually glycine-rich, with a GGGxYxG motif. But Aurora’s GxGxY is not remarkable in this respect. Further, VX680 is seen bound to Aurora both with and without the pi–pi interaction (Fig. 3). Mutational studies indicate that the pi–pi interaction is important for binding independent of binding to co-factor TPX2 [25]. Ab1 kinase has been observed in DFG-in, DFG-out, and intermediate states, and SRC-like, as recently reviewed [26]. VX680 binds to Ab1 kinase in both active and inactive forms (Fig. 3B). This is consistent with measurements of VX680 binding to phosphorylated Ab1 variants by Ambit Biosciences in 2011 [16]. Here, phosphorylation of the Ab1 kinase domain had only small effects on binding of VX680, and variously tightened or weakened binding, depending on the mutant form of the kinase. In contrast to the weak effect of phosphorylation, mutation of the hinge aromat (gatekeeper + 2, see Fig. 1B) residue from phenylalanine to leucine or especially isoleucine weakened binding by two to three orders of magnitude. One clear consequence of these observations is that crystal structures may not, in isolation, be considered to be proof of the “true” or even minimum energy binding geometry as it occurs in crystallo. Another may be that the anomalous cross reactivity of binding to both Ab1 and Aurora kinases stems from a propensity to bind to multiple target structures.

2.2.2. Understanding the polypharmacology of crizotinib

Crizotinib was designed [33] as a dual inhibitor of Alk and Met kinases (low or subnanomolar for Alk, Met, but also Mertk, Ros1, EphB6, Axl, and Ab1 kinases [16]). Inhibitor correlation analyses show moderate similarities for Alk and Met (see also 2.3 Cheminformatics,
below). The availability of X-ray structures for crizotinib in complex with both Alk (PDB ID: 2YFX [34]) and Met (PDB ID: 2WGJ [33]) might explain the similarity. Examination of the structures shows that, although the hinge binding and overall geometry are the same, crizotinib has no side chain interactions in common between the structures. Thus, the similarity is not likely to be clearly reflected in a sequence based analysis. In this case, binding pocket analyses might be the proper approach. However, comparison of the available structures.
for Met and Alk show that they have been crystallized in very different states, especially with respect to activation loop geometries, with no clear choice for making direct comparisons (Fig. 4).

2.2.3. Docking to similar kinases

The examples above illustrate how the flexibilities of target structures complicate binding pocket analysis. Generally, even high resolution crystal structures will represent at best a subset of the binding site conformations that are energetically available for ligand binding. Docking provides a way to observe specific effects that may arise in automated methods, and indeed, automated docking procedures fail more often than not even with correct target structure, at least for protein kinases [35], and the procedures may be tailored to optimize results [36,37]. For unknown protein kinase target structures, homology

Fig. 2. Pseudosequence similarity plots for Aurora B (2A) and Abl (2C), compared to the inhibition profile correlations of Aurora B (2B) and Abl (2D) with other tested kinases (see text).
modeling introduces additional error, such that it may be better to dock against the template structure than the modeled target. However; the modeled targets may even perform better[38]. Because of the uncertainties, experienced medicinal chemists will view docking results skeptically, but they will view them, because inspection of the range of docking poses and characteristics of the binding site will generate ideas for further optimization and testing. Docking done with diverse methods and diverse targets provides a test for the robustness and variability of the predictions.

Here we show tests of docking ROCK1 inhibitors to ROCK1, PKA, and PKA–ROCK1 hybrid target structures (Fig. 5), looking at the predicted minimum energy poses as functions of target structure, minimization of target structure, protonation state of the inhibitor. Although only a small fraction of docking experiments reproduced the experimental binding pose in its entirety (dark green), somewhat over half reproduced at least the hinge binding interactions (dark and light green), and a clear majority predicted at least some hinge binding (yellow and greens). The failure of staurosporine to dock successfully in any target except its parent structure (and here only with XP precision docking) is notable. Because staurosporine, with its extended planar and aromatic structure, significantly expands the binding site [39], docking requires adequate prediction of flexibility prediction for the target, as has been noted previously [40,41]. The target 1SVG also usually failed to predict hinge binding. Here, due to definition of the pocket grid based on the extent of the native ligand, the more extended inhibitor of 1SVG led to a pocket that included more residues distant from the hinge. Several of these created a hydrophobic/aromatic site that the docking algorithm ranked higher than hinge binding interactions. Although the minimization of the target molecule often significantly changed the outcome, there is no overall correlation with respect to the quality of the prediction. The proton assignment could be decisive: Unfavorable assignment of the protonation state of indazole-18 search prevented correct docking. The homology models performed diversely: the ROCK1 model (based on 1q8w) was as good as the best ROCK1 structures, while the PKA homology model was quite poor.

2.2.4. Water and structural variability

Analogous to the problem of treating protein flexibility, the mobilites and potential bridging functions of water in ligand binding pose a serious challenge to structural methods to predict ligand binding [42–44]; SAR analyses of filling a pocket may be complicated by the displacement of a water molecule and partial cavity formation and unfavorable energetics. Analysis of target similarity is less severely impacted, as targets with similar geometric and electrostatic properties would be likely to share similar effects of solvent, even if these are unpredictable. But side chain and other flexibilities determine solvent positions. So, as with flexibility, multiple structure determinations with varying conditions and ligands enable the identification of conserved water positions [45]. Aurora kinases have shown clusters of water; some of these have evolved into distributions. One cluster may be observed in a deep pocket between ATP, the gatekeeper, and the aspartic acid of the DFG...
While strategies to replace this water molecule with a corresponding polar group might be successful, superposition of inhibitor complexes with Aurora show that this pocket may be eliminated entirely by breaking the salt bridge made by the catalytic lysine residue with the helix C glutamic acid and displacing the residues (Fig. 6b).

2.3. Similarity by cheminformatics: calibrating sequence and structure?

Ultimately, it seems, the complexities and many uncertainties of structural analysis will be dealt with by automated application of empirical data. The growing set of computational algorithms for calculating hundreds or thousands of molecular properties increasingly enables

---

**Fig. 5.** Docking of Rho kinase inhibitors to ROCK1, PKA, and PKA-ROCK1 hybrid target structures. a) Overall results. Each target structure (in rows) was used with (bottom half of row) and without (top half of row) minimization. Inhibitors with multiple possible charge or protonation states have multiple columns. Hatching indicates that the inhibitor was extracted from that target structure. The degree to which the energy minimum docked pose is correct is shown by color. Green indicates the correct pose. Light green indicates that the hinge-binding moiety is correctly docked, but not all of the molecule. Yellow indicates that the inhibitor docked at the hinge, but in the incorrect pose. Red indicates that the minimum energy docked pose was not at the hinge. Black indicates that there was no docked solution found at all. b) The failure of docking in target structure 1SVG arises from the extended definition of the binding site grid from the native inhibitor (yellow). Without constraints to explicitly require binding at the hinge (white sticks), the minimum energy docking poses place the fasudil variants (purple) near the tip of the glycine-rich loop, where the binding pocket is bounded by residues rich in aromatic side chains (cyan).
2.3. Applications of “activity homology”

Posy et al. of Bristol-Myers Squibb published data derived from over 20,000 compounds [61] in Ambit assays. That extent of data enables the definition of “activity homology”, or “the prior probability that a compound will be active for kinase B given that it is active for kinase A”. This is calculated as the percentage of the inhibitor set of kinase A that inhibits kinase B, and provides an interesting method for evaluating target similarity. It is related to inhibitor correlation, but compares binding strength rather than binding patterns.

Fig. 7 shows “activity homology” analyses from the BMS data for three sets of targets. First, activity homologies are plotted for Aurora A and B, and Abl kinase, with kinases ordered along the abscissa descending according to their activity homology with Aurora A. Since the activity homology of a protein kinase with itself is 100%, the yellow green points at 100% mark the position along the abscissa of Aurora B and Abl, respectively. Aurora B has an activity homology of roughly 35% with Aurora A, with Abl at about 15%. Aurora B is more homologous to Aurora A in this measure than Abl, but both show a similar pattern of lower homologies with the kinases on the right (with lowest homology to Aurora A). Also apparent are the many outliers to this pattern for Abl, presumably due to different mechanisms for homology (or inhibitor binding) available to Abl and the outlier kinases that are not available to the Aurora kinases. Another pattern is evident: Aurora B has generally higher activity homology to other kinases than Aurora A, that is, Aurora A is a more idiosyncratic kinase.

Comparisons of Alk, Met, and EGFR, all lung cancer targets, show a different pattern (Fig. 7b). Here the similarity of Alk and Met is evident, as is the dissimilarity of EGFR. For polypharmacological targeting of these three kinases, either exceptional compounds with cross-reactivity for all three must be found, or inhibitors must be designed with “orthogonal” binding mechanisms. Covalent binding to the cysteine found at the gate-keeper + 7 hinge position of EGFR would be such a mechanism [62,63]. Finally, Fig. 7c shows a comparison of targets found to be inhibited by firefly luciferin and derivatives [64]. Many of these fragment-like compounds were found to be selective for Dyrk kinases, but individual derivatives were found to target other seemingly unrelated kinases. A full spectrum of patterns is evident, ranging from the close similarity of Dyrk1A and 1B, to the reversed pattern for Aurora A.

2.4. Applying target similarity to efforts against parasites and infectious disease

The use of target similarity to guide drug repurposing or retargeting is especially significant in key areas of infectious disease, especially tuberculosis, malaria, and other protozoan parasite diseases. Some of the focus here is motivated by a market potential considered to small to support large scale research programs, despite the clear medical need; indeed, widespread tropical diseases may be considered orphan diseases in developed countries [65,66]. With its enormous amount of available information and compounds (many coming off patent), protein kinase inhibition for these applications is promising but distinct in several ways. Primary among these is that non-human target kinases are more likely to have novel features compared to the well established human targets. Therefore, there may be greater need for modification of the known compounds. Closely related to this, selectivity for the disease targets over potential human off-target will be a high priority. Thus, the new target must be similar enough to benefit from a retargeting approach, but different enough to allow for the design of selectivity. This may focus attention on more promiscuous inhibitors or fragments in early discovery processes.
Fig. 7. Activity homology plots for selected sets of kinases: A) Aurora A, Aurora B, and Abl; B) Alk, Met, and EGFR, and C) Dyrk1A, Dyrk1B, ERK7, CK2A1, Aurora B, and Aurora A.
One approach is simply to test known inhibitors for activity. A study performed at GSK [67] screened the approximately 2 million compounds of in-house collection against *P. falciparum*, among which around 13,500 had antimalarial activity. They were clustered into different chemotypes, and possible targets were analyzed. 51 possible *Pfalciparum* targets were found, 30 of which are proposed to be kinases. In some cases, no obvious target was found in the malaria genome, which presents the possibility that "*P. falciparum* could have essential proteins that are structurally and functionally similar to the human targets yet have no significant primary amino acid homology."

Another example concerns trypanosomiasis: Based on the observation that tyrosine kinase inhibitors could inhibit the diacylglycerol-stimulated endocytosis of transferrin in *T. brucei*, Patel et al. optimized EGFR inhibitors based on tests in full cell cultures [68], with toxicity monitored in parallel, to create a compound with promising antitrypanosomal activity confirmed in mouse model studies. In a similar approach, the Aurora inhibitor Hesperadin was tested and modified for antiproliferative activity against several targets [69]. Also, pyrrolopyrazines, originally being developed as herbical agents were found to be effective antimalarials.

Computational studies revealed kinases as potential targets, observing inhibitory activity against human IRK, RAF-1, Src, TrkA and PPKs, the plasmoidal homolog of human CD2 [70].

Target based approaches, either at initiation or as a followup from screening as described above, require sequence or structural information. In the case of *M. tuberculosis*, genome sequence analysis found 11 eukaryotic-like Ser/Thr kinases [71,72], with low sequence identity with human kinases but some similarity to the CK1 group [73]. To date, crystal structures of the kinase domains of 5 Mtb kinases [18] and other surrogate kinase targets [74] are available for drug design.

Kinomes of *P. falciparum* [75] and protozoan parasites more generally [76] have been characterized, identifying parasite kinases with some 35–60% identity with their human homologues. They differ in several significant ways: large C- and/or N-terminal extensions, insertions in the catalytic domain, and differences in regulatory domain and at the activation site. Some 89–99 kinases have been identified, including 65–80 eukaryotic kinases, and a new group (FKIK) [75,77] exclusively within the Apicomplex phylum [78]. Espe- sially noteworthy is a family of seven calcium-dependent kinases (PcCDPKs) belonging to the CAMK group, absent from mammalian hosts but conserved among apicomplexans [78], including several that are necessary for plasmodial survival [79]. In addition, a family of CMGC protein kinases include seven CDK cognates, some shared among protists (Pfmrk) or apicomplexans. Other "orphan" kinases are not assigned to established subfamilies; one of these (PpIK7) is exclusive to Plasmodium spp [78]. Currently, 5 different *P.falciparum* protein kinase crystal structures are available on the PDB [18], along with 7 other structures for different apicomplexa organisms. Reviews on kinase targets of interest for *P. falciparum* [80] and approaches to inhibition have been published [79,81].

The PCDPKs have been of special interest as targets for the development of antimicrobials against malaria, in part because they lack a human homolog. Lead compounds have been identified through screening [82–85].

An alternative way of exploiting the exclusivity of CDPK targets to apicomplexa has been demonstrated by Vidalada et al. [86]. They developed a series of pyrazolopyrimidines that display important selectivity against *Toxoplasma gondii* and *Cryptosporidium parvum* CDPKs based on the observation of a glycine at the gatekeeper position (which does not occur in the human kinome, see Fig. 1) and on binding to an adjacent "ribose pocket" [87].

The kinase profiling data of Davis et al. [16] included kinase domains from three pathogen targets, one from *M. tuberculosis* (PKNB), and two from *P. falciparum* (PcCDPK1 and PpPK5). As described above, high correlations of the apparent inhibitor binding energies for the same series of inhibitors compared for two targets should reflect average (dynamic) structural similarity. Fig. 8 shows the correlations of the

inhibitors of the three pathogen targets with the human kinases in the panel. For PKNB, there is a rather broadly distributed similarity, with PAK4 and PAK5 having the highest values. For PCDPK1, there is low similarity. In contrast to these, PpPK5 shows clearly clustered similarity with CDK2 kinases, and a few other kinases in the CAMK and STE groups.

3. Methods

Pseudosequence similarities between protein kinases were calculated using the Needlemen–Wunsch similarity algorithm as implemented in Mathematica (www.wolframresearch.com), with a prohibitive gap penalty (100) and the Blosum62 matrix for similarity scoring. The pseudosequence used for the plots of this manuscript was chosen from the kinase residues that align with the following PKA residues 49, 54, 57, 72, 91, 95, 104, 120, 122, 123, 127, 170, 171, and 183. These correspond to the initiator of the glycine-rich loop and partial "ceiling" of the adenine site (Leu49), the typically aromatic amino acid at the tip of the glycine-rich loop (Phe54), the terminator of the glycine-rich loop and partial "ceiling" of the adenine site (Val57), the innermost side chain of helix C that extends toward the ATP site (Leu55), the ATP site directed side chain of the alpha4-helix C loop (Val104 of the "molecular brake" assembly), the gatekeeper (Met120), the gatekeeper +2 and +3 residues at the hinge (Y122; the gatekeeper +1 residue is highly conserved as Glu and lacks discriminatory power), the gatekeeper +7 residue (Glu127; often responsible for substrate and inhibitor specificity), Glu170 at the entrance to the ATP site (also a substrate specificity determinant), (Asn171 near the phosphate binding sites), and Leu173 and Thr183 at the base of the adenine binding site. The sequence alignment [13] was downloaded from kinase.com.

Inhibitor correlation plots were generated from the published Ambit data profiling of 72 inhibitors against a panel of over 400 protein kinases [16]. Logarithms of the published Ki values were calculated to obtain values proportional to binding energies, and Pearson’s correlation coefficients for pairs of kinases were calculated with Mathematica using the respective vectors of logarithmic inhibition values.

The Schrodinger software suite (www.schrodinger.com) was used for docking Rock1 inhibitors. The target structures chosen were as follows: ROCK1 (PDB ID: 2ESM [88], PDB ID: 2FUJ [89], PDB ID: 3NCZ [90], PDB ID: 3VBS [91], PDB ID: 4W7P [92]); a five-fold mutant PKA model [93] for ROCK1 (PDB ID: 2GNH, PDB ID: 2GNI, PDB ID: 2GNF), a three-fold mutant PKA model [93] for ROCK1 (PDB ID: 2GNH, PDB ID: 2GNI); and native PKA (PDB ID: 1STC [94], PDB ID: 1SVG [94], PDB ID: 1QBW [95]). The following ligands used for docking were taken from these structures (see Fig. 5a): fasudil, hydroxyfasudil, Y27632, 1Q8W [95]. The following ligands used for docking were taken from these structures (see Fig. 5a): fasudil, hydroxyfasudil, Y27632, 1Q8W [95]. The following ligands used for docking were taken from these structures (see Fig. 5a): fasudil, hydroxyfasudil, Y27632, 1Q8W [95]. The following ligands used for docking were taken from these structures (see Fig. 5a): fasudil, hydroxyfasudil, Y27632, 1Q8W [95]. The following ligands used for docking were taken from these structures (see Fig. 5a): fasudil, hydroxyfasudil, Y27632, 1Q8W [95]. The following ligands used for docking were taken from these structures (see Fig. 5a): fasudil, hydroxyfasudil, Y27632, 1Q8W [95].
targets structures, whereby all rotatable groups were allowed to rotate. The docking was carried out without any constraints, first in "standard precision" (SP) mode and then the poses were subjected to "extra precision" (XP) mode.

**Fig. 8.** Based on correlations of inhibitor binding energies, the similarities of human protein kinases to A) *M. tuberculosis* PknB, B) *P. falciparum* PfCDPK1, and C) *P. falciparum* PfPK5.

**Transparency document**

The Transparency document associated with this article can be found, in the online version.
Acknowledgements
We gratefully acknowledge financial support from the Norwegian Research Council and the Norwegian Cancer Society for this work. V. Tello-Franco gratefully acknowledges support from grant number 232128 from CONACYT (Mexico).

References
O.A. Gani et al. / Biochimica et Biophysica Acta 1854 (2015) 1605–1616