Biochimica et Biophysica Acta 1854 (2015) 1605–1616

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

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

ARTICLE INFO

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

ABSTRACT

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.

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).

http://dx.doi.org/10.1016/j.bbapap.2015.05.004
1570-9639/

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 kinome 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 Abl1, 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 Abl 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, Abl 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. Crystallization conditions, the energy of crystal packing contacts, and the state of the protein used for crystallization 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 crystallization 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 Abl, Aurora, and Flt3 kinases, including the drug resistant Abl mutant T315I [16]. The cross reactivity between Abl 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 Abl 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, Abl is unusually glycine-rich, with a GGGxYG motif. But Aurora's GxGxFG 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]. Abl kinase has been observed in DFG-in, DFG-out, and intermediate states, and SRC-like, as recently reviewed [26]. VX680 binds to Abl kinase in both active and inactive forms (Fig. 3b). This is consistent with measurements of VX680 binding to phosphorylated Abl variants by Ambit Biosciences in 2011 [16]. Here, phosphorylation of the Abl 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 *ex crystallo*. Another may be that the anomalous cross reactivity of binding to both Abl 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 Abl kinases [16]). Inhibitor correlation analyses show moderate similarities for Alk and Met (see also 2.3 Cheminformatics,