

# Tyrosine kinase drug discovery: what can be learned from solved crystal structures?

Loris Moretti, Laura Tchernin, and Leonardo Scapozza\*

*Pharmaceutical Biochemistry Group, School of Pharmaceutical Sciences, University of Geneva,  
University of Lausanne, Quai Ernest-Ansermet 30, CH-1211 Geneva 4, Switzerland  
E-mail: [Leonardo.Scapozza@pharm.unige.ch](mailto:Leonardo.Scapozza@pharm.unige.ch)*

---

## Abstract

Understanding molecular recognition during ligand binding is crucial for rational drug design. Protein kinases are a well known pharmacological target, especially in the therapeutic area of cancer. Solved crystal structures of tyrosine kinase domains in complex with ligands deliver insight into the binding event. These experimental data were collected and analysed by means of molecular modelling techniques. Common molecular recognition patterns were depicted and studied among this class of proteins. The results of this analysis and the consequences for rational design of inhibitors are presented.

**Keywords:** Tyrosine kinase ligands, molecular modelling, ligand core, inhibitor minimal requirements and conserved ligand-kinase interactions

---

## Introduction

Protein kinases are enzymes that transfer the  $\gamma$ -phosphate group of ATP to a peptidic substrate. These proteins are divided in two subfamilies based on the molecular target of the phosphorylation. Serine/threonine and tyrosine kinases transfer phosphates to hydroxyl groups of serine/threonine and tyrosine residues, respectively.

Tyrosine kinases are classified in receptor and non-receptor proteins, according to their function and localization in the cell, and sequence similarity.<sup>1</sup> Both are implicated in cell signalling: they do amplify, translate and integrate signals from inside and outside the cell.<sup>2</sup> Epidermal growth factor receptor (EGFR), platelet-derived growth factor receptor (PDGFR), fibroblast growth factor receptor (FGFR), vascular endothelial growth factor receptor (VEGFR) and insulin receptor (InsR) are examples showing the implications of these proteins in cell growth, differentiation and development. Deregulation of their catalytic activity is correlated to oncogenic and inflammatory processes as well as diabetes and other diseases.<sup>3</sup>

Several types of domains form a functional protein tyrosine kinase. Extra cellular and Trans membrane domains are present only in the case of receptors; while SH2, SH3 and other domains which up- and down-regulate the kinase catalytic activity (under physiological conditions) and interact with several cell components can occur in both classes. The conserved part of members of the protein kinase family is the catalytic domain, so-called kinase domain whose overall three-dimensional structure is depicted in figure 1.

The kinase domain is characterised by structural plasticity resulting in the existence of several conformations as shown by crystallographic studies.<sup>4</sup> The catalytic domain undergoes a conformational change to become enzymatically productive.<sup>5</sup>

According to the catalytic requirements, the kinase domain has two recognition sites: one for the ATP (phosphoryl donor), placed in a cleft between the N- and C-terminal lobes and one for the peptide substrate (phosphoryl acceptor) which is situated within the C-terminal lobe in correspondence of the activation loop C-terminal end (figure 1).

Tyrosine kinase inhibitors interfering with ATP binding are of interest for anticancer treatment. Glivec<sup>®</sup> which is used as first line therapeutic for chronic myelogenous leukemia (CML) represents a successful example of this approach.<sup>6-8</sup> During the last years, a wealth of biochemical and structural data about the binding of these molecules has been produced. In fact, many ligand-kinase domain complexes have been crystallised and their three-dimensional structures have been solved by means of X-ray techniques. The coordinates of these complexes are publicly available at the Protein Data Bank (PDB, <http://www.rcsb.org/pdb>). In this manuscript we report the results of a comparison of these structures combined with multiple sequence alignments of tyrosine kinase domains. Common characteristics of the complexes are depicted and minimal requirements for binding deduced.

## Results and Discussion

### The Kinase domain

The overall three-dimensional structure of the kinase domain, depicted in figure 1, is conserved throughout the protein kinase family. The N-terminal lobe is a twisted  $\beta$ -sheet of five antiparallel  $\beta$ -strands and one  $\alpha$ -helix, and the C-terminal lobe is composed by eight  $\alpha$ -helices and four  $\beta$ -strands. These two lobes are connected by a segment called hinge region and they delimit the cleft representing the ATP-binding site. Other important moieties of this domain are: the nucleotide-binding loop, the  $\alpha$ C helix, the catalytic and the activation loop (see figure 1).

For easier understanding of the presented data, residue numbering and representations are referred to the sequence and crystal structures of the insulin receptor (PDB codes: 1IR3<sup>5</sup> and 1IRK<sup>9</sup>).



**Figure 1.** The three-dimensional structure of the kinase domain. The tertiary structure of the kinase domain of the insulin receptor in complex with adenylyl imidodiphosphate (AMP-PNP), and peptide substrate (ball-and-stick and stick representations, respectively; PDB code: 1IR3) is depicted. Main regions are coloured: nucleotide-binding loop (yellow),  $\alpha$ C helix (red), hinge region (green), activation loop (violet) and catalytic loop (cyan).

### Variety of ligands

The PDB was screened and Cartesian coordinates of 44 structures of the catalytic domain of several tyrosine kinases in complex with ligands were retrieved (PDB entries are found in the supplementary information section). These ligands have  $IC_{50}$  values ranging from nM to  $\mu$ M.<sup>10-21</sup>

The first analysis revealed the presence of 15 structures of kinase domains in complex with ATP-like molecules. Interesting, among these there are two kinase domains in complex with ATP analogue chemically linked with peptide that behaves as bisubstrate ligand.

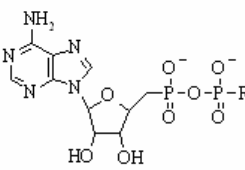
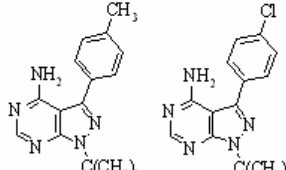
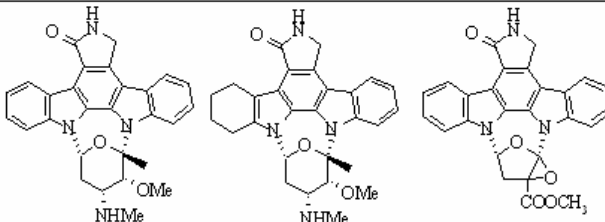
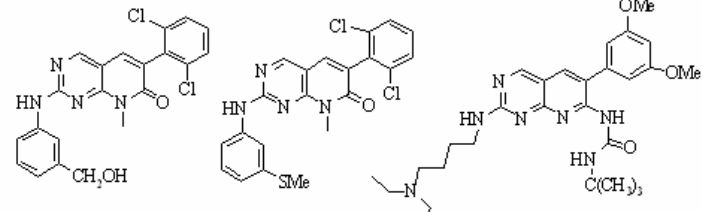
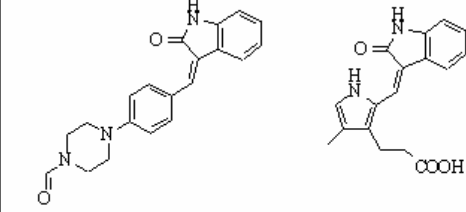
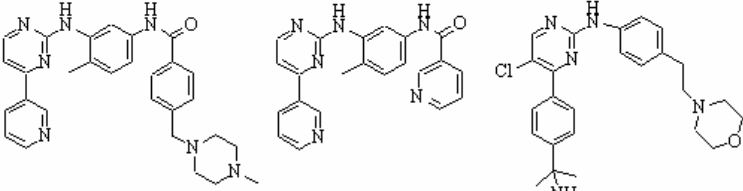
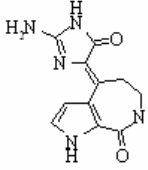
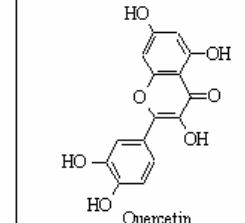
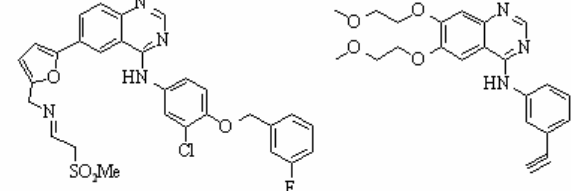
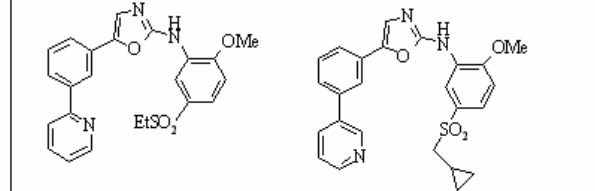
The remaining 29 structures are kinase domains in complex with various types of inhibitors that belong to several chemical families (figure 2). The pyrimidine ring of the adenine which characterised ATP-like ligands is found as basic moiety in several groups of inhibitors: aminopyrimidine, pyrazolopyrimidine, pyridopyrimidine and quinazole derivatives.

Dihydroindolone and an isomer of it characterise other two classes of compounds. Staurosporine, which is a well-known ATP-competitive inhibitor with a wide range of action among protein kinases<sup>17</sup> and its analogues are dihydroisoindolone based molecules.

Recently, tyrosine kinase domains have been crystallised in complex with two aminoxazole derivatives. Finally, the last two inhibitors taken into consideration are a flavonol (quercetin) and an azepinone (debromohymenialdisine).

In order to analyse the protein-ligand interactions computer-aided visualization of the complexes was used. The ligands interact with the kinase domain via hydrogen bonds to the

hinge region and more or less extended hydrophobic interactions with the region expanding from the hinge towards the nucleotide-binding loop and/or the  $\alpha$ C helix.

IMIDAZOLEPYRIMIDINES [A]	PYRAZOLOPYRIMIDINES [A]	INDOLES [E]		
 ATP-like compounds	 PP1                      PP2	 Staurosporine                      AFN941                      K252A		
PYRIDOPYRIMIDINES [A]		1,3-DIHYDRO-INDOL-2-ONES [F]		
 PD166326                      PD173955                      PD173074	 SU4984                      SU5402			
AMINOPYRIMIDINES [A]		AZEPIN-2-ONE [H]	CHROMEN-4-ONE [G]	
 ST1571 [A & C]                      PRC                      AA2	 DBQ	 Quercetin		
QUINAZOLINES [B]		AMINOXAZOLES [D]		
 Lapatinib                      Erlotinib	 AAZ                      AAX			

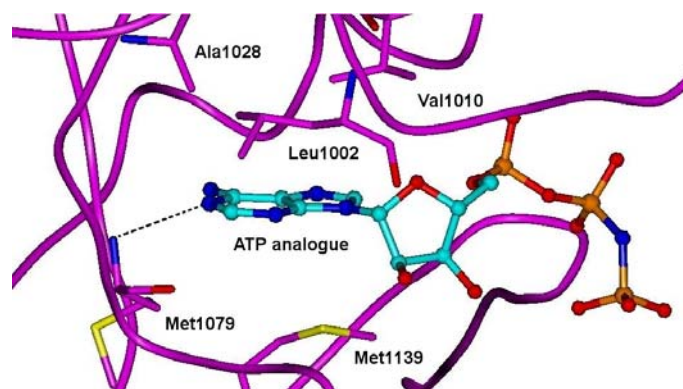
**Figure 2.** Families of compounds. 2D-molecular structures of the ligands found in complex with the kinase domain. Names of compounds and families are reported. Annotations of the ligand cores are given in squared brackets (referring to figure 4).

### Common residues in the complexes

For each complex the residues interacting with the ligand were studied by examination of a protein area of 4.5 Å in the proximity of the ligand. The amino acids found were compared at the sequence level by means of multiple sequence alignment of all structures taken into account.

This procedure allowed the highlighting of common residues involved in conserved interactions between the ligands and the kinase domains. In all complexes the common ligand-protein interaction pattern is constituted by hydrophobic interactions involving 4 residues and a single hydrogen bond.

These residues in the case of insulin receptor correspond to: Ala1028 (from the core of the N-terminal lobe), Met1139 (from the C-terminal lobe), Leu1002 and Val1010 of the nucleotide-binding loop and Met1079 of the hinge region (figure 3).



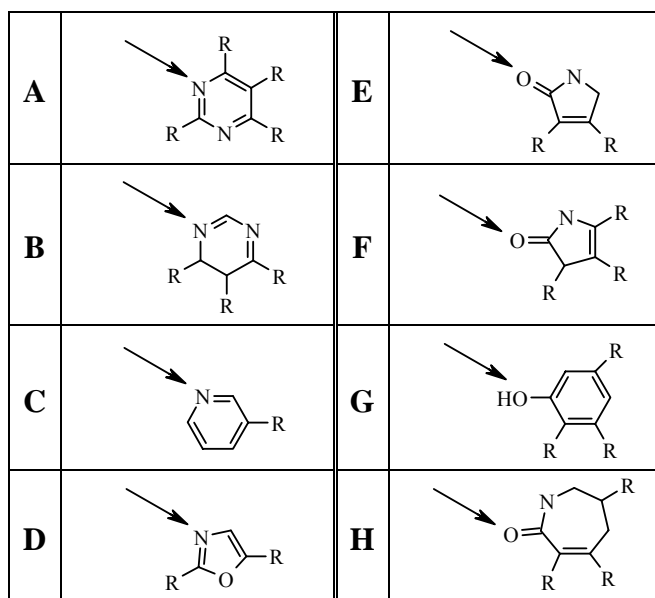
**Figure 3.** Common interaction pattern in the ligand-kinase domain complexes. Secondary structure of the kinase domain of insulin receptor (1IR3) is represented as a tube (magenta), ATP analogue in ball-and-stick representation and the common residues as colour-coded sticks (carbon in cyan for the ligand and magenta for the protein, oxygen in red, nitrogen in blue, sulphur in yellow and phosphorus in orange and the hydrogen bond as black dashed line).

### Defining and determining the ligand core

Superposition and visualisation of the complexes permitted to depict in which part of the protein the conserved residues interacting with the ligand are localised. Common interactions are found at the same binding region in all structures, called adenine region in a previous work.<sup>23</sup>

The moiety of the ligand interacting with this region was examined and defined as “ligand core”. The results of this analysis are summarised in figure 4.

It was generally possible to cluster ligand cores based on the ligand families with each family having a single ligand core. However, for STI-571, the ligand core was determined twice because of the two different modes of binding found in 4 complexes e.g. one involving the pyridine and the other the pyrimidine moiety [namely STI-571 bound to Abl (pdb codes: 1IEP<sup>15</sup> and 1OPJ<sup>13</sup>), Kit (1T46<sup>19</sup>) and Syk (1XBB<sup>20</sup>)].



**Figure 4.** Ligand cores. In the drawn structures, R is in accordance with the rest of the molecules as in figure 2. Arrows indicate, for each ligand core, the atom involved in the hydrogen bond to the hinge region.

#### Amino acid variability around the ligand core

For the 4 residues involved in hydrophobic interactions with the ligand cores, a consensus on the type of amino acid was expected. In fact, the hydrophobic character of these residues was checked by means of multiple sequence alignment (data not shown) of catalytic domains of 90 tyrosine kinases.<sup>1,24</sup> The amino acids variability at the four positions was calculated based on the multiple sequence alignment and is reported in table 1.

The table shows that at each position there is a rather dominant amino acid type. The alternative residues have always hydrophobic side chains, and thus the variability concerns only their bulkiness.

**Table 1.** Amino acid variability of the hydrophobic pattern interacting with the ligand core

Leu1002	69%	Leu
	30%	Ile
	1%	Val
Val1010	96%	Val
	3%	Ile
	1%	Leu
Ala1028	92%	Ala
	4%	Ile & Val
	4%	Cys, Phe & Leu
Met1139	89%	Leu
	9%	Met
	2%	Val & Phe

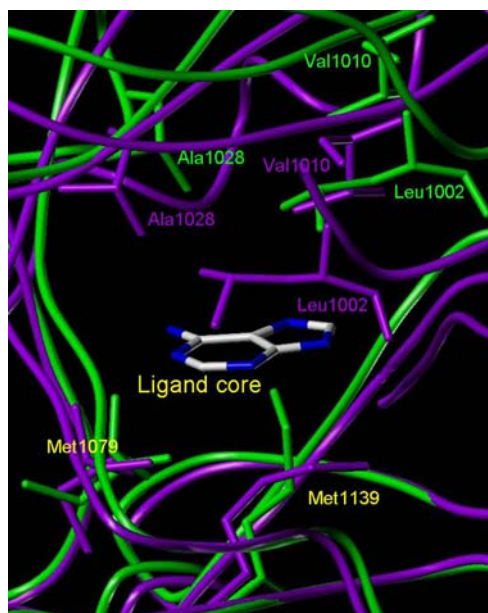
Variance (expressed in percentage) of the amino acid type at the positions (numbering referred to insulin receptor) involved in hydrophobic contacts with the ligand core. The data reported about the amino acid variability of the protein kinases<sup>23</sup> differ slightly in their values from the data here presented. This is most likely due to the used data sets, here only protein tyrosine kinases are considered which are a subfamily of the protein kinases. Nevertheless, the relative preference of amino acids shown here corresponds to the previous report<sup>23</sup>.

#### Active/inactive: residues movements

The superposition of the apo and the ATP-bound crystal structures of the kinase domain of the insulin receptor revealed that a conformational change occurs upon ligand binding (see figure 5). The N-terminal lobe is moved towards the second lobe when the ligand is bound, narrowing the cleft of the ATP-binding site.

To estimate the domain's closure around the ligand core, distances between amino acids in the N-terminal and C-terminal lobe were measured (reported in table 2). The selected residues are the ones involved in the common interactions: from the C $\alpha$  of Met1139 (C-terminal lobe) to the C $\alpha$  of Leu1002, Val1010 and Ala1028 (N-terminal lobe).

Distance variations (range 0.8-2.2Å) shown in table 2 are meaningful of a conformational change between the productive and the unproductive structures. Whereas the variation of distances measured in other regions of these two complexes is significantly smaller (0.1-0.3 Å).



**Figure 5.** Conformational changes around the ligand core. The superposition of active and non-active conformations of insulin receptor is depicted here. The protein main chain (tube representation) and the common residues (sticks representation) are coloured in magenta (1IR3) and in green (1IRK). Adenine of the AMP-PNP is also drawn as sticks (in blue the nitrogen atoms and in light grey the carbon atoms), the pyrimidine moiety represents the ligand core.

**Table 2.** Distances between residues of the N-terminal lobe and Met1139. Distances in Angstrom from C $\alpha$  of the Met1139 to C $\alpha$  of the N-terminal lobe residues are reported. The inactive conformation is labeled “Inact” (resolution: 2.1 Å) and the active conformation “Act” (resolution: 1.9 Å)

	Inact	Act	$\Delta$
Ala1028	12.2	11.4	0.8
Leu1002	13.6	11.8	1.8
Val1010	15	12.8	2.2

## Conclusions

In the present study, by the comparison of the structures of the tyrosine kinase domain in complex with ligands, we could show that binders of this protein family share a common feature that we called “ligand core”. As a consequence, the interactions involving this part of the molecule were understood as minimal requirements for binding: a hydrogen bond to a main chain atom of the hinge region and apolar interactions with the side chain of 4 residues with conserved hydrophobic character.



This represents a reduction of the complexity for the pharmacophore compared to the adenine of the ATP that binds the hinge region via 2 hydrogen bonds.

The domain's closure upon ligand binding, calculated for the insulin receptor structures reveals that the ligand is likely to stabilize the protein conformation with the ligand core playing the major role. Other tyrosine kinase domains have been crystallised in their apo and ligated forms, and their conformational changes were checked by measurements of the distances earlier mentioned. The distance's variations are smaller and less appreciable than the ones reported for insulin receptor. This fact is due to several experimental aspects (size of ligand, presence of other domains interfering with the catalytic one, phosphorylation of the activation loop, conditions of crystallisation, etc.) that might lead to differences in the degree of closure of the two lobes. This is in agreement with the general idea of protein "breath".<sup>25</sup>

In the year 2004 a similar study was published by Vulpetti and Bosotti.<sup>23</sup> The approach used and the considerations and point of view in this study and theirs are slightly different leading towards different and complementary results.

While we focused on defining the ligand core, the authors identified 38 residues of the catalytic domain of protein kinases which built up the ATP pocket. They studied the amino acid variance among the protein family focusing on how to achieve ligand binding selectivity. This work and the list of the ligand cores will extend with the number of the ligand-tyrosine kinase structure solved. The knowledge of the ligand cores might be useful in drug design when hypothetical ligand-binding modes are considered and suggested, for example while validating the results achieved from docking methods.

## Supplementary Information Available

PDB code	kinase	ligand	year	PDB code	kinase	ligand	year
1IR3	InsR	ANP	1997	1OPJ	Abl	STI 571	2003
1GAG	InsR	ATP-pept	2001	1OPK	Abl	PD166326	2003
1I44	InsR	ACP	2001	1OPL	Abl	PD166326	2003
1RQQ	InsR	ATP-pept	2003	1XBB	SYK	STI 571	2004
1JQH	IGFR	ANP	2002	1XBC	SYK	Staurosporine	2004
1K3A	IGFR	ACP	2001	1U59	ZAP-70	Staurosporine	2004
1AGW	FGFR1	SU4984	1998	1BYG	CSK	Staurosporine	1999
1FGI	FGFR1	SU5402	1998	1QPJ	LCK	Staurosporine	2000
2FGI	FGFR1	PD173074	1999	1QPD	LCK	Staurosporine	2000
1OEC	FGFR1	AA2	2004	1QPC	LCK	ANP	2000
1M17	EGFR	Erlotinib	2002	1QPE	LCK	PP2	2000
1XKK	EGFR	Lapatinib	2004	1SNU	ITK	Staurosporine	2004
1Y6A	VEGFR2	AAZ	2005	1SM2	ITK	Staurosporine	2004
1Y6B	VEGFR2	AAX	2005	1AD5	HCK	ANP	1997
1PKG	KIT	ADP	2003	2HCK	HCK	quercetin	1997
1T46	KIT	STI 571	2004	1QCF	HCK	PP1	1999
1R0P	MET	K-252A	2003	1YVJ	JAK3	Stauro-anal	2005
1JPA	EPHR-B2	adenine	2001	1U54	ACK1	ACP	2004
1MQB	EPHR-A2	ANP	2003	1U4D	ACK1	DBQ	2004
1FPU	Abl	PRC	2000	2SRC	Src	ANP	1999
1IEP	Abl	STI 571	2001	1KSW	Src	ATP anal	2002
1M52	Abl	PD173955	2002	1MP8	FAK	ADP	2003

The table collects the PDB codes and other information of all tyrosine kinase complexes taken into consideration in the present study.

## Acknowledgements

This study was supported by a protein kinase research project (ProKinase European Contract, PF6 503467).

## References and Footnotes

1. Robinson, D.R.; Wu, Y.M.; Lin, S.F. *Oncogene* **2000**, *19*, 5548.
2. Hunter, T. *Cell* **2000**, *100*, 113.
3. Blume-Jensen, P.; Hunter, T. *Nature* **2001**, *411*, 355.
4. Huse, M.; Kuriyan, J. *Cell* **2002**, *109*, 275.
5. Hubbard, S. R. *EMBO J.* **1997**, *16*, 5572.
6. Druker, B. J.; Tamura, S.; Buchdunger, E.; Ohno, S.; Segal, G. M.; Fanning, S.; Zimmermann, J.; Lydon, N. B. *Nature Med.* **1996**, *2*, 561.

7. Sawyers, C. L.; Hochhaus, A.; Feldman, E.; Goldman, J. M.; Miller, C. B.; Ottmann, O. G.; Schiffer, C. A.; Talpaz, M.; Guilhot, F.; Deininger, M. W.; Fischer, T.; O'Brien, S. G.; Stone, R. M.; Gambacorti-Passerini, C. B.; Russell, N. H.; Reiffers, J. J.; Shea, T. C.; Chapuis, B.; Coutre, S.; Tura, S.; Morra, E.; Larson, R. A.; Saven, A.; Peschel, C.; Gratwohl, A.; Mandelli, F.; Ben-Am, M.; Gathmann, I.; Capdeville, R.; Paquette, R. L., Druker, B. J. *Blood* **2002**, *99*, 3530.
8. Holtz, M. S.; Slovak, M. L.; Zhang, F.; Sawyers, C. L.; Forman, S. J.; Bhatia, R. *Blood* **2002**, *99*, 3792.
9. Hubbard, S. R.; Wei, L.; Ellis, L.; Hendrickson, W. A.; *Nature* **1994**, *372*, 746.
10. Harris, P. A.; Cheung, M.; Hunter, R. N. 3rd; Brown, M. L.; Veal, J. M.; Nolte, R. T.; Wang, L.; Liu, W.; Crosby, R. M.; Johnson, J. H.; Epperly, A. H.; Kumar, R.; Luttrell, D. K.; Stafford, J. A. *J Med Chem.* **2005**, *48*, 1610.
11. Stamos, J.; Sliwkowski, M. X.; Eigenbrot, C. *J Biol Chem.* **2002**, *277*, 46265.
12. Wood, E. R.; Truesdale, A. T.; McDonald, O. B.; Yuan, D.; Hassell, A.; Dickerson, S. H.; Ellis, B.; Pennisi, C.; Horne, E.; Lackey, K.; Alligood, K. J.; Rusnak, D. W.; Gilmer, T. M.; Shewchuk, L. *Cancer Res.* **2004**, *64*, 6652.
13. Nagar, B.; Hantschel, O.; Young, M. A.; Scheffzek, K.; Veach, D.; Bornmann, W.; Clarkson, B.; Superti-Furga, G.; Kuriyan, J. *Cell* **2003**, *112*, 859.
14. Mohammadi, M.; Froum, S.; Hamby, J. M.; Schroeder, M. C.; Panek, R. L.; Lu, G. H.; Eliseenkova, A. V.; Green, D.; Schlessinger, J.; Hubbard, S. R. *EMBO J.* **1998**, *17*, 5896.
15. Nagar, B.; Bornmann, W. G.; Pellicena, P.; Schindler, T.; Veach, D. R.; Miller, W. T.; Clarkson, B.; Kuriyan, J. *Cancer Res.* **2002**, *62*, 4236.
16. Zhu, X.; Kim, J. L.; Newcomb, J. R.; Rose, P. E.; Stover, D. R.; Toledo, L. M.; Zhao, H.; Morgenstern, K. A. *Structure Fold Des.* **1999**, *7*, 651.
17. Meggio, F.; Donella Deana, A.; Ruzzene, M.; Brunati, A. M.; Cesaro, L.; Guerra, B.; Meyer, T.; Mett, H.; Fabbro, D.; Furet, P.; Dobrowolska, G.; Pinna, L. A. *Eur. J. Biochem.* **1995**, *234*, 317.
18. Schiering, N.; Knapp, S.; Marconi M.; Flocco, M. M.; Cui, J.; Perego, R.; Rusconi, L.; Cristiani, C. *PNAS* **2003**, *100*, 12654.
19. Mol, C. D.; Dougan, D. R.; Schneider, T. R.; Skene, R. J.; Kraus, M. L.; Scheibe, D. N.; Snell, G. P.; Zou, H.; Sang, B. C.; Wilson, K. P. *J. Biol. Chem.* **2004**, *279*, 31655.
20. Atwell, S.; Adams, J. M.; Badger, J.; Buchanan, M. D.; Feil, I. K.; Froning, K. J.; Gao, X.; Hendle, J.; Keegan, K.; Leon, B. C.; Muller-Dieckmann, H. J.; Nienaber, V. L.; Noland, B. W.; Post, K.; Rajashankar, K. R.; Ramos, A.; Russell, M.; Burley, S. K.; Buchanan, S. G. *J Biol Chem.* **2004**, *279*, 55827.
21. Mohammadi, M.; McMahon, G.; Sun, L.; Tang, C.; Hirth, P.; Yeh, B. K.; Hubbard, S. R.; Schlessinger, J. *Science* **1997**, *276*, 955.
22. Sicheri, F.; Moarefi, I.; Kuriyan, J. *Nature* **1997**, *385*, 602.
23. Vulpetti, A.; Bosotti, R. *Farmaco* **2004**, *59*, 759.

24. Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. *Science* **2002**, 298, 1912.
25. Pickover, C. A. *Science* **1984**, 223, 181.