Protein Kinase Inhibitors: Insights into Drug Design from Structure

Martin E. M. Noble *et al.*

Science **303**, 1800 (2004);
DOI: 10.1126/science.1095920

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

Protein Kinase Inhibitors: Insights into Drug Design from Structure

Martin E. M. Noble,* Jane A. Endicott,* Louise N. Johnson*†

Protein kinases are targets for treatment of a number of diseases. This review focuses on kinase inhibitors that are in the clinic or in clinical trials and for which structural information is available. Structures have informed drug design and have illuminated the mechanism of inhibition. We review progress with the receptor tyrosine kinases (growth factor receptors EGFR, VEGFR, and FGFR) and nonreceptor tyrosine kinases (Bcr-Abl), where advances have been made with cancer therapeutic agents such as Herceptin and Gleevec. Among the serine-threonine kinases, p38, Rho-kinase, cyclin-dependent kinases, and Chk1 have been targeted with productive results for inflammation and cancer. Structures have provided insights into targeting the inactive or active form of the kinase, for targeting the global constellation of residues at the ATP site or less conserved additional pockets or single residues, and into targeting noncatalytic domains.

A number of diseases, including cancer, diabetes, and inflammation, are linked to perturbation of protein kinase–mediated cell signaling pathways. The human genome encodes some 518 protein kinases (*1*) that share a catalytic domain conserved in sequence and structure but which are notably different in how their catalysis is regulated. The ATP-binding pocket is between the two lobes of the kinase fold (Fig. 1). This site, together with less conserved surrounding pockets, has been the focus of inhibitor design that has exploited differences in kinase structure and pliability in order to achieve selectivity. Drugs are in clinical trials that target all stages of signal transduction: from the receptor tyrosine kinases that initiate intracellular signaling, through second-messenger generators and kinases involved in signaling cascades, to the kinases that regulate the cell cycle that governs cellular fate (*2*–*5*).

Protein Kinases as Targets for Inhibitor Design

Receptor tyrosine kinases. Dysregulation of growth factor signaling networks has been reported in multiple human cancers. Binding of growth factors to extracellular domains of receptor tyrosine kinases activates the intracellular kinase domain. The epidermal growth factor receptor (EGFR) is normally activated by oligomerization in response to ligand binding, but in cancer cells, family members [EGFR (ErbB1, HER1) and its homologs HER2, HER3, HER4] are frequently overactive. To block the EGFR signal, different therapeutic agents have been developed that target the extracellular ligand-binding and intracellular kinase domains.

The HER2/Neu gene product is upregulated in the tumor cells of about 30% of breast cancer patients (*6*). This finding provided the rationale for the development of Herceptin, a humanized monoclonal antibody that binds the HER2 receptor and induces receptor internalization. In clinical trials, Herceptin alone proved effective in treatment for 15% of patients with HER2-overexpressing metastatic breast cancer and was more effective when used in combination with chemotherapy agents such as paclitaxel (*7*).

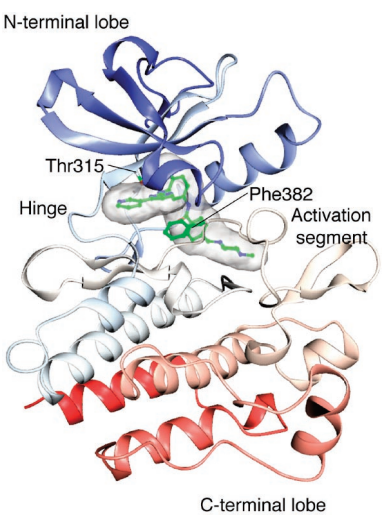

Fig. 1. The structure of the catalytic domain of cAbl in complex with Gleevec (*51*). The N-terminal lobe consists of a β sheet and one conserved α helix (helix C). The C-terminal lobe is largely helical and contains a segment, the activation segment, which includes residue(s) that in many kinases are phosphorylated for activity (*72*). The hinge region connects the two lobes. The protein structure is color ramped so that residues close to the N terminus are blue, and those close to the C terminus are red. Gleevec is shown bound to the ATP-binding site, from which it extends under the C helix. Thr^{315}, the "gatekeeper" residue, and Phe^{382}, the conserved phenylalanine that marks the beginning of the activation segment, are labeled.

Iressa (*8*) and Tarceva (*9*) (Fig. 2) are small-molecule inhibitors that bind to the EGFR tyrosine kinase domain. Iressa has been registered for treatment of metastatic non–small cell lung cancer where other treatments have failed, and Tarceva is currently in phase III clinical trials for several tumor types such as non-small cell lung cancer and pancreatic cancer. Inhibitors that bind irreversibly to the EGFR through covalent bond formation with a cysteine residue in the ATP pocket are even more effective as kinase inhibitors (*10*, *11*). Their clinical efficacy is being evaluated.

A second class of receptor tyrosine kinases rationally targeted in anticancer drug development are those that promote angiogenesis (*12*), particularly the vascular endothelial growth factor receptor (VEGFR). This strategy is based on the rationale that formation of a blood supply is required for progression of solid tumors (*13*). Agents that target VEGFR signaling include the VEGF-specific antibody bevacizumab, the small-molecule tyrosine kinase inhibitors such as SU5416 (Fig. 2) and the anilino-phthalazine PTK787. Some of these agents have met with limited success in monotherapy against solid tumors [e.g., (*14*)], but may be more effective in combination with other agents. The fibroblast growth factor receptor (FGFR) is a further target for rational inhibitor design, particularly for angiogenesis in a spectrum of pathologies that include cancer, rheumatoid arthritis, diabetic retinopathy, and atherosclerosis [see references in (*15*)].

Nonreceptor tyrosine kinases. About one-third of tyrosine kinases are classed as nonreceptor tyrosine kinases. They are found in the cytoplasm, lack a transmembrane section, and generally function downstream of the receptor tyrosine kinases. Chronic myeloid leukemia (CML) is a relatively rare cancer (~5000 cases per year in the United States), often associated with reciprocal translocation of chromosomes 9 and 22. This event fuses

Laboratory of Molecular Biophysics, Department of Biochemistry, Rex Richards Building, University of Oxford, Oxford 3X2 3QU, UK.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: louise@biop.ox.ac.uk

the *Bcr* (breakpoint cluster region) gene to the *Abl* (Abelson leukemia virus) gene at a region corresponding to the N-terminal part of Abl tyrosine kinase (a Src family relative) (*16*). The fusion protein Bcr-Abl has all of the Abl kinase preserved except for a small N-terminal region, upstream of the Src homology 3 domain. Crystal structures have shown that the N-terminal myristoyl modification of Abl is important for autoinhibition, and loss of this regulation may contribute to the unregulated activity of the Bcr-Abl fusion protein (*17*). Gleevec (Fig. 2), an inhibitor of Abl kinase (*18*, *19*), was approved in May 2001 for therapy against CML. Treatment with Gleevec results in remission in nearly 100% of newly diagnosed patients in the early stages of the disease. The success of Gleevec in treating CML was extended quickly to the treatment of gastrointestinal stromal tumors in which c-Kit receptor tyrosine kinase activity is elevated by mutation (*20*–*22*).

Phosphatidylinositol 3′-kinase (PI3K). The PI3 lipid kinase (PI3K) exerts its biological effects through the generation of 3′-phosphorylated phosphoinositides that act as second messengers within the cell [reviewed in (*23*)]. PI3K exhibits a protein kinase fold and hence is included in this review, because its ATP site has much in common with protein kinases. Resistance to radiation treatment in a number of cancers has been linked to activation of the PI3K-AKT pathway, which suggests that inhibition of PI3K to overcome resistance and to improve the efficacy of radiation treatment is an attractive clinical goal (*24*). Wortmannin (Fig. 2) is a PI3K-selective inhibitor, whose structure in complex with PI3K has been determined (*25*). Wortmannin analogs are reported in phase I clinical trials for treatment of osteoporosis. Although wortmannin is also an inhibitor of the DNA protein kinase family, it has been shown to be a potent inhibitor of osteoclast-induced bone resorption in vitro through its ability to inhibit PI3K (*26*).

Signal-transducing serine-threonine kinases. Cells respond to certain stresses by activating members of the mitogen-activated protein kinase (MAPK) family, which in turn activate distinct (but overlapping) effector pathways [reviewed in (*27*)]. Activation of p38α MAPK can lead to increased activities of proinflammatory cytokines, such as tumor necrosis factor–α and interleukin 1β. This observation suggested that p38 selective inhibition could be a therapeutically useful

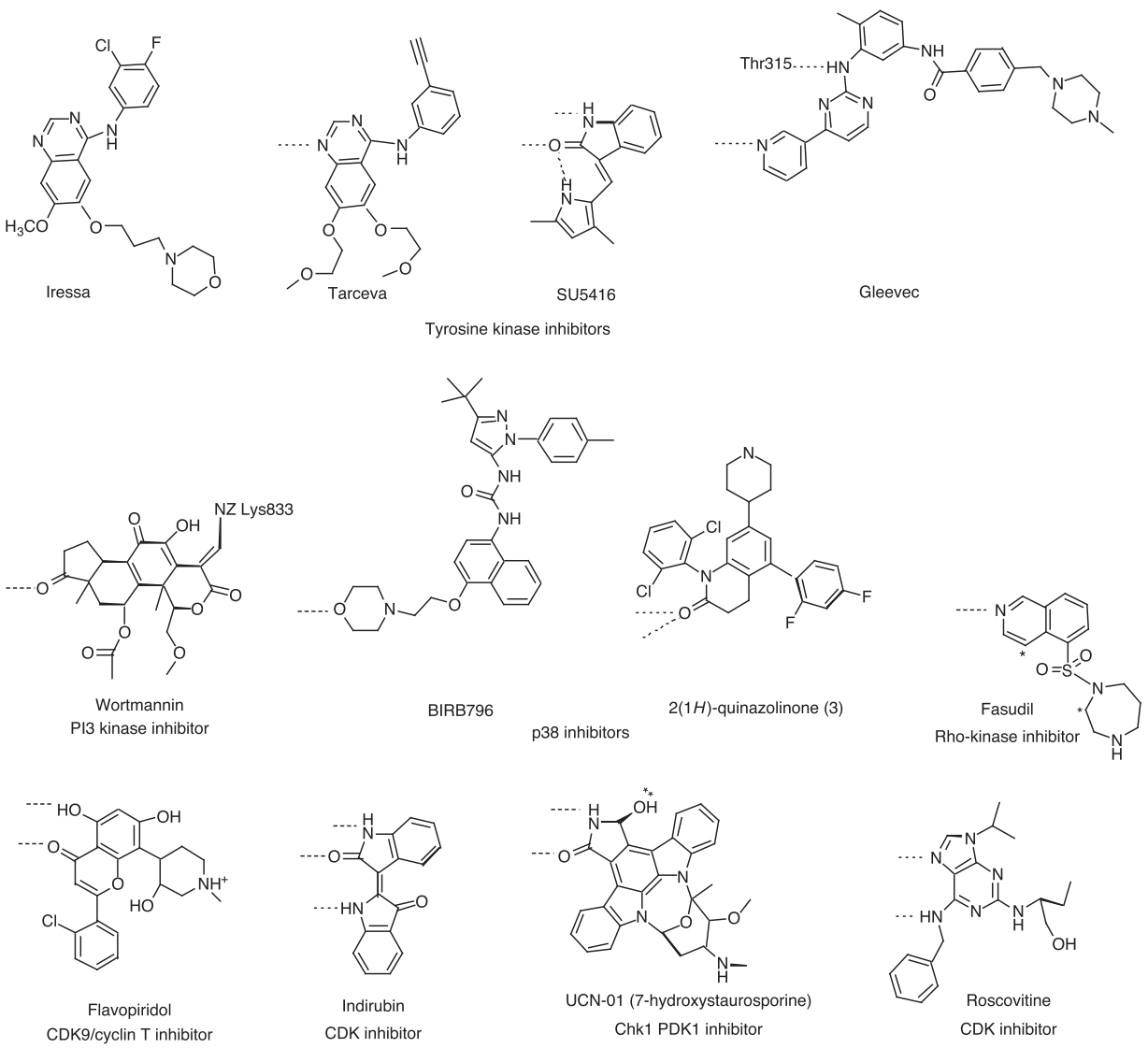

Fig. 2. Chemical structures of selected kinase inhibitors that are in the clinic or in clinical trials. The compounds are oriented so that the groups that hydrogen bond to main-chain atoms of the hinge region are roughly aligned. The hydrogen bonds are indicated by the dotted lines. The covalent bond between wortmannin and Lys^{833} in PI3K is formed by electrophilic attack of epsilon amino group of the lysine NZ, leading to opening of the furan ring. Two asterisks on fasudil indicate the sites where the 200 times as potent Rho-kinase inhibitor H-1152P carries two additional methyl groups. A double asterisk on UCN-01 indicates the 7-hydroxy group that distinguishes UCN-01 from staurosporine.

route to treatment of a number of inflammatory and autoimmune diseases. At present, BIRB 796 {1-(5-*tert*-butyl-2-*p*-tolyl-2*H*-pyrazol-3-yl)-3-[4-(2-morpholin-4-yl-ethoxy) naphthalen-1-yl]urea, a diaryl urea (*28*)} (Fig. 2) is in clinical trials for the treatment of rheumatoid arthritis and Crohn's disease (*29*) and VX-702 for acute coronary syndromes (*30*). Fasudil (Fig. 2) is in the clinic, approved for treatment of cerebral vasospasm (*31*). It has a significant vasodilatory effect attributed to its potent inhibition of Rho-kinase signaling to myosin light-chain kinase.

Cyclin-dependent kinases (CDKs) and other cell cycle–control kinases. In response to mitogenic signals, cells enter the cell cycle, a process that is tightly regulated by members of the CDK family. CDK-mediated phosphorylation of the tumor suppressor protein pRb is required for progression through G_1 phase, and many human cancers have abnormalities in pRb or its regulation (*32–34*). A further hallmark of the transformed state is failure of checkpoint control, resulting in aberrant response to cell damage (*35*). Undoubtedly, the choice of therapeutic agent for treatment of cancers arising from aberrations in cell cycle control demands an understanding of the nature of the molecular defect(s) (*36*).

Nature has provided a number of useful CDK inhibitors that have reached clinical trials. Flavopiridol (Fig. 2) is a synthetic flavonoid analog of a natural alkaloid extracted from the stem bark of the Indian plant, *Dysoxylum binectariferum* (*37*), and as a pan-CDK inhibitor has shown activity in phase I and phase II trials [reviewed in (*38*)]. Its action as a potent CDK9/cyclin T inhibitor may underlie its ability to block human immunodeficiency virus HIV-1 propagation in cultured cells (*39*), an observation that suggests it may also have a role in HIV therapy. It is not yet clear if the antiproliferative properties of flavopiridol are due to its inhibition of cell cycle CDKs or to its ability to inhibit transcription through CDK9/cyclin T or through action on other targets. Flavopiridol also alters protein expression probably through the nuclear factor–κB (NF-κB) pathway (*40*) and is an inhibitor of glycogen phosphorylase (*41*, *42*).

Indirubin (Fig. 2) is a minor constituent of a Chinese prescription Danggui Longhui Wan (a mixture of 11 herbal medicines) traditionally used to treat certain types of leukemias (*43*). It has undergone clinical trials in China to treat CML with promising results, although its mechanism of action remains unknown (*44*). When tested in vitro against a panel of kinases, it proved to be an inhibitor of CDK2 and also a potent inhibitor of CDK5/p25, which suggests a potential role for treatment of neurodegenerative disorders (*45*). UCN-01 (Fig. 2) is currently undergoing clinical trials in the United States and Japan, having shown antiproliferative effects and cytostatic properties in several human tumor cell lines (*46*). The microbial alkaloid staurosporine, the parent compound of UCN-01, is a potent but nonspecific kinase inhibitor and is too toxic to use in therapy.

High-throughput screening of compound libraries has also identified a number of CDK inhibitors. Aminothiazole-based compounds are relatively selective CDK2 inhibitors compared with CDK1 and CDK4 and the determination of structures for selected members of this series, in complex with CDK2, provided a rationalization of their potency and selectivity (*47*). The results of phase I clinical trials with BMS 387032 have recently been reported, and phase II trials are planned (*48*). Roscovitine (Fig. 2) was developed through a series of structure-lead structure-activity relation studies from 6-dimethylaminopurine, a compound that had been widely used in early cell cycle studies as an inhibitor of H1 kinase activity in starfish oocyte extracts (*49*). R-roscovitine (CYC202) is currently undergoing phase II clinical trials in combination with standard chemotherapy regimes for advanced breast cancer and stage IIIB/IV non–small cell lung cancer (*49*).

Principles for Structure-Based Lead Optimization

The results of structural studies have identified the properties of the protein kinase ATP-binding site that are exploited by the relatively potent and selective subset of inhibitors that are showing promise in clinical trials. Although conserved in architecture even in PI3K, the ATP-binding site presents distinct features, some of which are illustrated in Fig. 3. From such structures, the sites of different kinases that are potential drug targets can be classified on the basis of shape and amino acid composition. The following are illustrative examples of inferred routes to kinase selectivity.

Targeting a unique inactive conformation. In paraphrase of the opening sentence of Anna Karenina, all active kinases are alike but an inactive kinase is inactive after its own fashion. Thus, in the search for selectivity, the inactive forms of kinases are attractive targets for drug design. The crystal structure of the Abl kinase domain in complex with Gleevec shows that the pyrimidine and the pyridine rings of the drug overlap with the ATP-binding site and are surrounded by a hydrophobic cage (Figs. 1 and 3). The rest of the molecule is wedged between the activation segment and the C helix, which locks the kinase in an inactive conformation (*50*, *51*). In particular, the aspartate and phenylalanine of the DFG (from the single-letter amino acid code) motif at the start of the activation segment are flipped with reference to the active conformation so that the aspartate can no longer coordinate the magnesium at the catalytic site, and the phenylalanine points toward the ATP site. The pocket that accommodates the 6-methyl group on the Gleevec phenyl ring is guarded by a threonine residue, Thr^{315}, to which Gleevec hydrogen bonds. However, in many kinases, a bulkier nonpolar residue replaces the threonine, and as a result, access to the pocket is blocked. Src kinases do not bind Gleevec with high affinity, although most of the residues that are in contact with Gleevec in c-Abl are conserved in Src kinases. An explanation lies in kinase conformational pliability and stability. Src is unable to adopt the particular inactive conformation required for binding Gleevec with high affinity (*17*).

Despite the logic behind the approach of targeting an inactive conformation of a protein kinase, there are also potential advantages of targeting an active conformation. The active conformation requires conservation of 3D structure and, therefore, is likely to be less tolerant of resistant mutations. It is interesting to note that some successful kinase inhibitors such as Tarceva and UCN-01 (see below) bind to the active form of their target kinase. PD173955, a pyrido-[2,3]pyrimidine–based compound, is a more potent inhibitor of Abl than Gleevec, and the structure (*51*) has suggested that its greater potency may be because it can bind to multiple conformations of Abl (active or inactive), whereas Gleevec requires a specific inactive conformation. However, PD173955 is a less selective inhibitor. Both active and inactive protein kinase structures have been used in inhibitor binding studies. Whereas in the majority of structures there are only subtle conformational changes induced by the inhibitor, in a few examples, such as K-252a binding to c-Met, the inhibitor induces significant conformational changes (*52*). This phenomenon might permit differences in protein kinase plasticity to be exploited in the design of specific inhibitors.

Targeting the global constellation of residues within the ATP site. A recent structure determination (*53*) of PKA in complex with Fasudil (Figs. 2 and 3) and with a more potent Rho-kinase inhibitor H-1152P, which differs by only two methyl groups, demonstrated characteristic binding within the ATP site. Homology-modeling comparisons with Rho-kinase indicated the likely stereochemical basis for the selectivity of the compound and its high affinity for Rho-kinase. The authors address the question of whether selectivity arises from a unique combination of specific amino acids or simply from the sum of a small number of individual interactions that can be considered independently of one another. From sequence alignment of 491 kinases, they show that only six kinases have the same combination of residues at the ATP site as Rho-kinase, although individual resi-

dues show a considerable degree of conservation. It is suggested that it is the combination of residues at the ligand-binding site, which generates a uniquely shaped inhibitor-binding pocket that confers selectivity.

Targeting less conserved additional pockets. The structure of deschloro-flavopiridol in complex with inactive CDK2 (*54*) showed that the compound binds at the ATP site (Fig. 3) with the O5 hydroxyl and the O4 groups hydrogen bonded to hinge-region main-chain groups as donor and acceptor, respectively. These mimic the hydrogen bonds made by the adenine of ATP. The phenyl ring extends into a pocket on the surface of the CDK2 C-terminal lobe below the site occupied by ATP and exploited by a number of CDK selective inhibitors (*49*, *55*).

Pyrimidine imidazoles, quinazolinones, and pyridol-pyrimidines are inhibitors that derive considerable selectivity for p38α from probing a small hydrophobic pocket at the back of the ATP-binding site that ATP does not exploit (*56*, *57*). In the majority of protein kinase structures, entry to this pocket is blocked by bulky amino acid side chains (for example Phe^{80} in CDK2 and Asp or Glu residues in most MAPK family members) but in p38α this so-called "gatekeeper" residue is a much smaller threonine residue as in Abl kinase. The gatekeeper residue can also offer an interaction site for inhibitor binding, as observed in the binding of the isopropyl group of roscovitine to CDK2 (Fig. 3).

The crystal structures of the EGFR kinase domain and its complex with Tarceva (*58*) show that the EGFR kinase structure exhibits a conformation consistent with the active form of protein kinases (Fig. 3). Tarceva binds in a mode similar to that of the anilinoquinazolines bound to CDK2 and p38 (*59*). The interplanar angle of the aromatic rings is 42° and the acetylene moiety is directed into the pocket guarded by the gatekeeper threonine residue (Thr^{766}). This pocket is similar to that occupied by the Gleevec 6-methyl group bound to Abl kinase. Comparison of the structures shows that they have much in common. However, an important difference is that the second phenyl ring of Gleevec bound to inactive Abl kinase occupies the position of the phenylalanine of the DFG motif in the active conformation of EGFR kinase (Fig. 3).

Targeting single residues. The structures of UCN-01 and staurosporine bound to the active conformations of Chk1 (*60*), phospho-CDK2/cyclin A (*61*, *62*) and PDK1 (*63*) have been determined. Both the extensive nonpolar contacts of the inhibitors' three indole rings, which extend beyond those made by the adenine of ATP, and the shape complementarity between enzyme active site and inhibitor contribute to the high affinity of staurosporine and its analogs (*64*). The most important differences between the staurosporine and UCN-01 complexes are the contacts made by the 7-hydroxy group of UCN-01. Komander *et al*. (*63*) have analyzed the relative affinities of staurosporine and UCN-01 for 29 different kinases and have shown that those potently inhibited by UCN-01 tend to have a side chain that can directly hydrogen bond to the 7-hydroxyl. Kinases that are more potently inhibited by staurosporine lack a potential hydrogen bonding partner.

3,4-Dihydropyridol[3,2-*d*]pyrimidines and 2(1H)-quinazolinones (Fig. 2) have been reported as two new classes of p38α inhibitor that

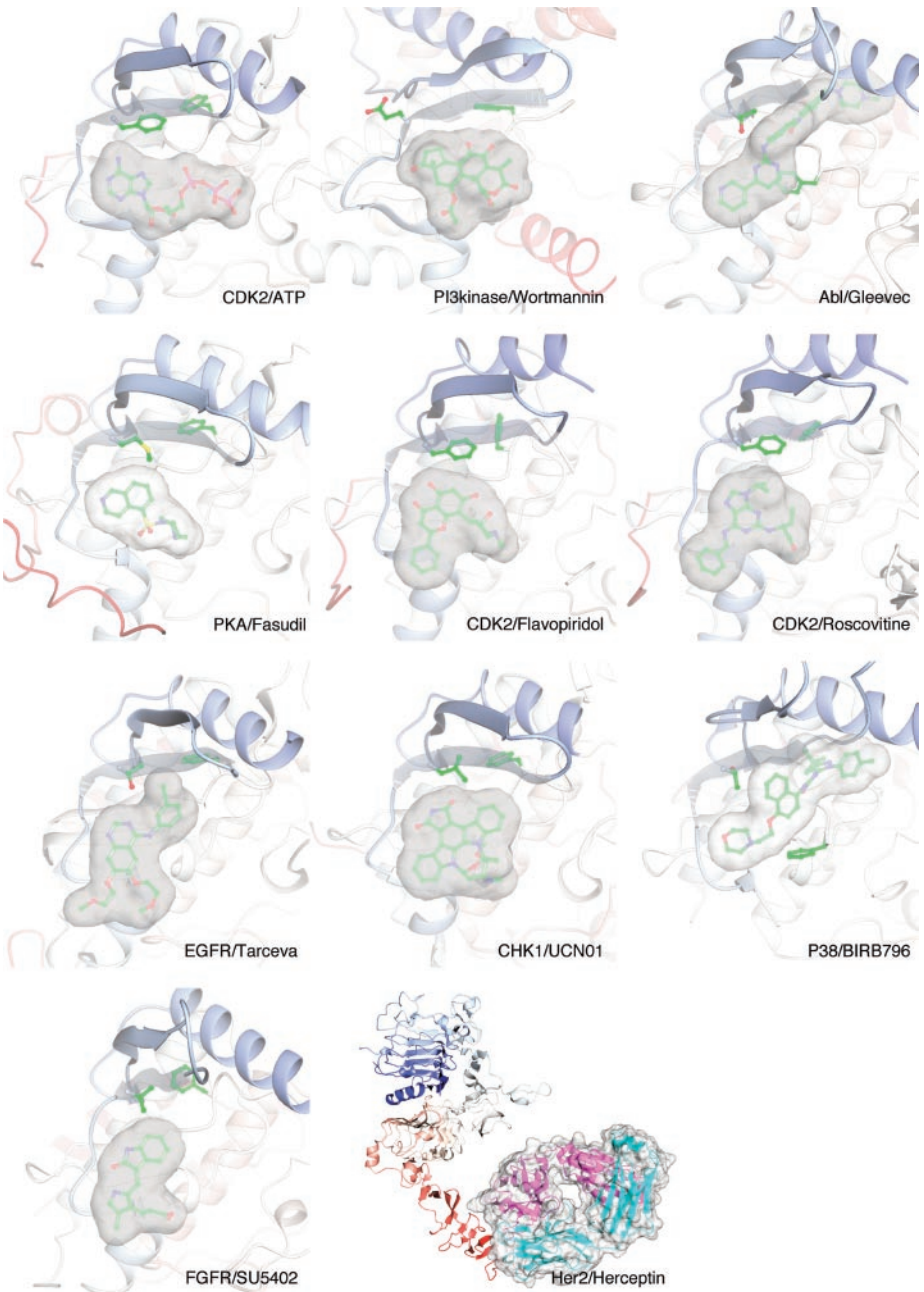

Fig. 3. Representative examples of the action of families of clinically tested inhibitors of protein kinase signaling. The view is rotated about a horizontal axis by 90° relative to that of Fig. 1, with the N-terminal lobe cut away to expose the inhibitor-binding site. The binding of ATP to the fully active state of CDK2 (top left) is shown for reference. The hinge region (left side of each panel) begins with the gatekeeper residue, which is drawn in ball-and-stick representation (Phe in CDK2). The C helix (top of each panel) adopts a range of orientations, often characteristic of the state of activation of the kinase. The phenylalanine residue of the conserved DFG motif that defines the start of the activation segment is also shown and serves as another marker of the state of kinase activation. Inhibitors bind to a range of conformational states, ranging from inactive (e.g., Abl-Gleevec) to active (e.g., EGFR-Tarceva). The last panel illustrates the binding of the Fab fragment of Herceptin to the juxtamembrane domain of the extracellular portion of HER2.

have greater selectivity for p38 than for c-Jun N-terminal kinases (JNKs) and extracellular signal–regulated kinases (ERKs), which could not be rationalized on the basis of predicted p38α-inhibitor interactions (*57*). Structure determination revealed that the enzyme undergoes a rather subtle conformational change in the hinge sequence to accommodate an inhibitor hydrogen bond–accepting group (*57*). To interact more favorably with this carbonyl moiety, the hinge backbone undergoes a peptide flip to present two backbone amide groups to the ligand. This conformational change is calculated to be more energetically favorable in p38α, β, and γ, where the hinge region sequence is Met^{109}-Gly^{110} than in other JNKs and ERKs where Gly^{110} is replaced with a bulkier side chain. The diaryl ureas, exemplified by BIRB 796 (Fig. 3), are p38α selective inhibitors that have recently been in phase IIb/III clinical trials for the treatment of rheumatoid arthritis and for which structural information has rationalized potency (*28*). Notably, BIRB 796 exploits novel binding sites within the ATP-binding pocket, one of which is created by inducing a conformational change in the enzyme that ultimately yields a structure incompatible with ATP binding. Again, inhibitor binding is targeted to the Phe residue in the conserved DFG motif that is buried in a hydrophobic pocket between the two lobes of the kinase.

The structure of the inhibitor SU5402 bound to FGFR1 tyrosine kinase domain (Fig. 3) (*15*) identifies the inhibitory binding mode of the indolin-2-one family of antiangiogenic molecules that include SU5416, SU6668, and SU11248, potent inhibitors of VEGFR tyrosine kinase activity. Members of this family have a functionalized methylpyrrole ring attached to C3, and the orientation of this ring in the complex of SU5402 with FGFR1 is stabilized by an intramolecular hydrogen bond between the pyrrole nitrogen and the O2 atom of the oxindole ring (Fig. 2). Specificity of SU5402 for FGFR1 derives in part from a hydrogen bond between the carboxyl function of SU5402 and the side chain of Asn^{568} at the end of the hinge.

Targeting noncatalytic domains. Structural studies have shown that the HER2 extracellular domain has a constitutively open structure (*65*), similar to that observed for HER1 bound to EGF. Hence, ligand binding is not required to release HER2 from an autoinhibited form, which could account for the transforming potential of HER2 when overexpressed. The structure of HER2 receptor in complex with Herceptin Fab shows that Herceptin binds HER2 on the C-terminal portion of domain IV at a site that encompasses the binding pocket for other domains in the closed inactive forms of HER1 and HER3 (Fig. 3). The binding close to the juxtamembrane region of the receptor may allow engagement with the endocytotic machinery while avoiding kinase activation, which would explain its inhibitory effects on receptor signaling. Other EGFR antibodies are also showing considerable promise such as IMC-C225 (cetuximab/Erbitux) and Thermacin h-R3 (Cima-her) (the latter developed in Cuba) for the treatment of head and neck tumors in combination with radiotherapy.

Prospects

Although structural results are clearly informing the process of antikinase drug development, the most important consideration for successful therapy design remains the choice of a suitable target. The clearest case is Bcr-Abl, where a single gene product drives the disease. The growing body of structural results offers enticing prospects for the future. After characterization of the human kinome (*1*), genomic scale structure determination is a realizable ambition that may contribute to both target selection and the careful design of inhibitor specificity, exploiting the emerging principles outlined above. This in turn will help to mitigate the drug toxicity that is often the cause for kinase inhibitor failure at the stage of clinical trials. It is also likely to broaden the spectrum of diseases that might be addressed by antikinase therapy. Already, efforts are underway to target protein kinases from pathogens including *Mycobacterium tuberculosis* (*66*) and *Plasmodium falciparum* (*67*). Efforts to address such diseases of the poor are less attractive to the pharmaceutical industry but will be substantially helped by the accumulated knowledge about achieving inhibition and avoiding toxicity in antikinase therapies learned from diseases of the rich. In the discovery of lead compounds, experimental and theoretical approaches are being developed to exploit structural information to short cut the process of lead compound identification. Structural studies are also informing the challenging design of chemical entities that bind to the less conserved noncatalytic domains that often distinguish a kinase from its peers, and so offer a novel route to specificity (*68*). Finally, structural studies are starting to provide the basis for optimism that problems with existing antikinase therapies might be overcome. In certain CML patients, for example, resistance to Gleevec has been observed (*69*), especially for those who have progressed from the chronic phase to the blast crises phase of the disease. Some resistance mutations, such as Thr^{315}Ile (*70*), can be understood from the structure. Other resistant mutations lie outside the kinase domain and may exert their effects by modulating regulatory interactions (*71*). In keeping with success stories from the treatment of HIV, it is to be hoped that kinase structure understanding can be exploited to design novel combination and single-agent therapies to enhance the prospects of mitigating resistance.

References

1. G. Manning, D. B. Whyte, R. Martinez, T. Hunter, S. Sudarsanam, *Science* **298**, 1912 (2002).
2. P. Cohen, *Nature Rev. Drug Discov.* **1**, 309 (2002).
3. D. Fabbro *et al.*, *Pharmacol. Ther.* **93**, 79 (2002).
4. A. Levitzki, *Acc. Chem. Res.* **36**, 462 (2003).
5. D. J. Pratt, J. A. Endicott, M. E. Noble, *Curr. Opin. Drug Discov. Dev.*, in press.
6. D. J. Slamon *et al.*, *Science* **235**, 177 (1987).
7. D. J. Slamon *et al.*, *N. Engl. J. Med.* **344**, 783 (2001).
8. A. J. Barker *et al.*, *Bioorg. Med. Chem. Lett.* **11**, 1911 (2001).
9. M. Hidalgo *et al.*, *J. Clin. Oncol.* **19**, 3267 (2001).
10. J. Baselga, L. A. Hammond, *Oncology* **63**, (Suppl. 1), 6 (2002).
11. A. Wissner *et al.*, *Bioorg. Med. Chem. Lett.* **12**, 2893 (2002).
12. J. Folkman, *N. Engl. J. Med.* **285**, 1182 (1971).
13. F. J. Giles, *Oncology (Huntington)* **16**, 23 (2002).
14. S. S. Sridhar, F. A. Shepherd, *Lung Cancer* **42** (suppl. 1), S81 (2003).
15. M. Mohammadi *et al.*, *Science* **276**, 955 (1997).
16. C. L. Sawyers, *N. Engl. J. Med.* **340**, 1330 (1999).
17. B. Nagar *et al.*, *Cell* **112**, 859 (2003).
18. B. J. Druker *et al.*, *Nature Med.* **2**, 561 (1996).
19. B. J. Druker, N. B. Lydon, *J. Clin. Invest.* **105**, 3 (2000).
20. M. C. Heinrich *et al.*, *Blood* **96**, 925 (2000).
21. C. D. Blanke, B. L. Eisenberg, M. C. Heinrich, *Curr. Treat. Options Oncol.* **2**, 485 (2001).
22. A. T. Van Oosterom *et al.*, *Lancet* **358**, 1421 (2001).
23. R. Katso *et al.*, *Annu. Rev. Cell Dev. Biol.* **17**, 615 (2001).
24. W. G. McKenna, R. J. Muschel, *Genes Chromosomes Cancer* **38**, 330 (2003).
25. E. H. Walker *et al.*, *Mol. Cell* **6**, 909 (2000).
26. P. T. Lakkakorpi, G. Wesolowski, Z. Zimolo, G. A. Rodan, S. B. Rodan, *Exp. Cell Res.* **237**, 296 (1997).
27. G. Pearson *et al.*, *Endocr. Rev.* **22**, 153 (2001).
28. C. Pargellis *et al.*, *Nature Struct. Biol.* **9**, 268 (2002).
29. P. W. Manley *et al.*, *Biochem. Biophys. Acta* (2003).
30. L. W. Hardy, A. Malikayil, www.currentdrugdiscovery.com (December 2003), 15 (2003).
31. N. Ono-Saito, I. Niki, H. Hidaka, *Pharmacol. Ther.* **82**, 123 (1999).
32. P. M. Fischer, J. A. Endicott, L. Meijer, in *Progress in Cell Cycle Research*, L. Meijer, A. Jesequel, M. Rowberger, Eds. (Life in Progress Editions, Roscoff, 2002), vol. 5, pp. 235–248.
33. J. W. Harper, P. D. Adams, *Chem. Rev.* **101**, 2511 (2001).
34. E. A. Sausville, Y. Elsayed, M. Monga, G. Kim, *Annu. Rev. Pharmacol. Toxicol.* **43**, 199 (2003).
35. G. I. Shapiro, J. W. Harper, *J. Clin. Invest.* **104**, 1645 (1999).
36. J. Dancey, E. A. Sausville, *Nature Rev. Drug Discov.* **2**, 296 (2003).
37. G. Kaur *et al.*, *J. Natl. Cancer Inst.* **84**, 1736 (1992).
38. S. Zhai, A. M. Senderowicz, E. A. Sausville, W. D. Figg, *Ann. Pharmacother.* **36**, 905 (2002).
39. S. H. Chao, D. H. Price, *J. Biol. Chem.* **276**, 31793 (2001).
40. Y. Takada, B. B. Aggarwal, *J. Biol. Chem.* **279**, 4750 (2004).
41. A. Kaiser *et al.*, *Arch. Biochem. Biophys.* **386**, 179 (2001).
42. N. G. Oikonomakos *et al.*, *J. Biol. Chem.* **275**, 34566 (2000).
43. W. Tang, G. Eisenbrand, *Chinese Drugs of Plant Origin: Pharmacology and Use in Traditional and Modern Medicine* (Springer, Heidelberg, 1992).
44. Z. Xiao, Y. Hao, B. Liu, L. Qian, *Leuk. Lymphoma* **43**, 1763 (2002).
45. S. Leclerc *et al.*, *J. Biol. Chem.* **276**, 251 (2001).
46. A. M. Senderowicz, E. A. Sausville, *J. Natl. Cancer Inst.* **92**, 376 (2000).
47. K. S. Kim *et al.*, *J. Med. Chem.* **45**, 3905 (2002).

48. G. Shapiro *et al*., *Proc. Am. Soc. Clin. Oncol.* **22**, 199 (2003).
49. L. Meijer, E. Raymond, *Acc. Chem. Res.* **36**, 417 (2003).
50. T. Schindler *et al*., *Science* **289**, 1938 (2000).
51. B. Nagar *et al*., *Cancer Res.* **62**, 4236 (2002).
52. N. Schiering *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12654 (2003).
53. C. Breitenlechner *et al*., *Structure (Cambridge)* **11**, 1595 (2003).
54. W. F. De Azevedo, Jr. *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **93**, 2735 (1996).
55. T. G. Davies *et al*., *Nature Struct. Biol.* **9**, 745 (2002).
56. L. Tong *et al*., *Nature Struct. Biol.* **4**, 311 (1997).
57. C. E. Fitzgerald *et al*., *Nature Struct. Biol.* **10**, 764 (2003).
58. J. Stamos, M. X. Sliwkowski, C. Eigenbrot, *J. Biol. Chem.* **277**, 46265 (2002).
59. L. Shewchuk *et al*., *J. Med. Chem.* **43**, 133 (2000).
60. B. Zhao *et al*., *J. Biol. Chem.* **277**, 46609 (2002).
61. L. N. Johnson *et al*., *Pharmacol. Ther.* **93**, 113 (2002).
62. E. De Moliner, E. D. Lowe, K.-Y. Cheng, L. N. Johnson, unpublished observations; Protein Data Bank ID 1PKD.
63. D. Komander *et al*., *Biochem. J.* **375**, 255 (2003).
64. R. A. Engh, D. Bossemeyer, *Pharmacol. Ther.* **93**, 99 (2002).
65. H. S. Cho *et al*., *Nature* **421**, 756 (2003).
66. T. A. Young, B. Delagoutte, J. A. Endrizzi, A. M. Falick, T. Alber, *Nature Struct. Biol.* **10**, 168 (2003).
67. S. Holton *et al*., *Structure (Cambridge)* **11**, 1329 (2003).
68. G. Kontopidis *et al*., *Structure (Cambridge)* **11**, 1537 (2003).
69. A. S. Corbin, P. La Rosee, E. P. Stoffregen, B. J. Druker, M. W. Deininger, *Blood* **101**, 4611 (2003).
70. M. E. Gorre *et al*., *Science* **293**, 876 (2001).
71. M. Azam, R. R. Latek, G. Q. Daley, *Cell* **112**, 831 (2003).
72. L. N. Johnson, M. E. Noble, D. J. Owen, *Cell* **85**, 149 (1996).

REVIEW

Polyketide and Nonribosomal Peptide Antibiotics: Modularity and Versatility

Christopher T. Walsh

Polyketide (PK) and nonribosomal peptides (NRP), constructed on multimodular enzymatic assembly lines, often attain the conformations that establish biological activity by cyclization constraints introduced by tailoring enzymes. The dedicated tailoring enzymes are encoded by genes clustered with the assembly line genes for coordinated regulation. NRP heterocyclizations to thiazoles and oxazoles can occur on the elongating framework of acyl-*S* enzyme intermediates, whereas tandem cyclic PK polyether formation of furans and pyrans can be initiated by post–assembly line epoxidases. Macrocyclizations of NRP, PK, and hybrid NRP-PK scaffolds occur in assembly line chain termination steps. Post–assembly line cascades of enzymatic oxidations also create cross-linked and cyclized architectures that generate the mature scaffolds of natural product antibiotics. The modularity of the natural product assembly lines and permissivity of tailoring enzymes offer prospects for reprogramming to create novel antibiotics with optimized properties.

Polyketides and nonribosomal peptides comprise two families of natural products biosynthesized with comparable logic by multimodular enzymes acting in assembly line arrays. The monomeric building blocks are organic acids or amino acids, respectively. Copolymerization via mixed modules of enzymatic machinery can be used to assemble hybrid polyketide-nonribosomal peptide molecules of useful structural complexity and therapeutic activity.

Since the topic of harnessing the biosynthetic code for natural products was reviewed in this journal a half-dozen years ago (*1*), knowledge about gene clusters that encode enzymes for the biosynthesis of many therapeutically useful PK, NRP, and hybrid PK-NRP natural products has increased dramatically, with hundreds of identified or predicted clusters available in public databases. In just one example, the genome of the avermectin producer, *Streptomyces avermitilis* (*2*), revealed 24 additional polyketide synthase (PKS) or nonribosomal peptide synthase (NRPS) clusters for unidentified secondary metabolites, suggesting natural product biosynthetic capacity is tremendously underestimated by the products detected in fermentations. The systematic variation of culture conditions to up-regulate transcriptomes selectively, and thereby the encoded proteomes, may begin to coax microbes to reveal their full secondary metabolome capacity (*3*). The need for new antibiotics to combat the widening circle of resistance that ensues each time such a drug is launched into widespread human use puts a premium on the discovery of therapeutically useful molecules. Deciphering the structure of heretofore cryptic microbial secondary metabolites may well hasten the discovery of new antibiotic scaffolds useful as novel therapeutics or as reprogrammable scaffolding elements for semisynthetic modifications.

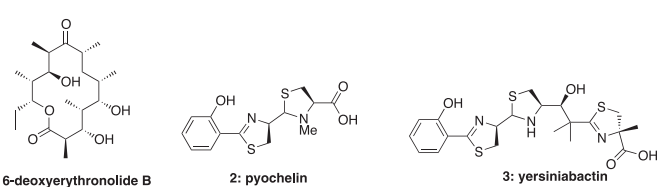

Fig. 1. Polyketide and nonribosomal peptide natural products from reconstituted assembly lines.

The molecular logic used by PKS and NRPS multimodular assembly lines in acyl and peptidyl chain initiation, elongation, and termination reactions has largely been deconvoluted and put to use for mechanistic and synthetic efforts, including making new analogs of the estrogen receptor antagonist R1128 (*4*–*6*). The role of domains and subunits that tailor the growing chains both on the assembly lines and after release have been inventoried (*7*–*9*). Reconstitution of assembly lines from purified assembly line proteins in vitro has been reported for deoxyerythronolide B **1** (DEB) (*10*), the NRP siderophore pyochelin **2** (*11*), and the hybrid PK-NRP iron chelator yersiniabactin **3** (Ybt) (*12*) (Fig. 1). PKS and PKS-NRPS gene clusters encoding DEB (*13*) and Ybt (*14*), respectively, have been moved from producing streptomycetes to *E. coli* and reconstituted in vivo, opening up the genetic toolbox that has been developed over decades for gene and protein studies in *E. coli*. The detection of intersubunit linkers (*15*, *16*) and their portability between PKS and NRPS subunits (*17*) has enabled initiation of mix-and-match module strategies to create new molecular skeletons.

Heterocyclization and Macrocyclization in Assembly Line Enzymology

Among the features that endow polyketide and nonribosomal peptide natural products with high affinity for biological targets are

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA. E-mail: christopher_walsh@hms.harvard.edu