Orally effective FDA-approved protein kinase targeted covalent inhibitors (TCIs)

Robert Roskoski Jr.

Blue Ridge Institute for Medical Research, 3754 Brevard Road, Suite 116, Box 19, Horse Shoe, NC, 28742-8814, United States

ARTICLE INFO

Keywords:
Catalytic spine
Hydrophobic interaction
Protein kinase inhibitor classification
Protein kinase structure
Regulatory spine
Shell residues

Chemical compounds studied in this article:
Acalabrutinib (PubChem CID: 71226662)
Afatinib (PubChem CID: 10184653)
Dacomitinib (PubChem CID: 11511120)
Erlotinib (PubChem CID: 176870)
Gefitinib (PubChem CID: 123631)
Ibrutinib (PubChem CID: 24821094)
Imatinib (PubChem CID: 5291)
Neratinib (PubChem CID: 9915743)
Osimertinib (PubChem CID: 71496458)
Zanubrutinib (PubChem CID: 135565884)

ABSTRACT

Because dysregulation of protein kinases owing to mutations or overexpression plays causal roles in human diseases, this family of enzymes has become one of the most important drug targets of the 21st century. Of the 62 protein kinases inhibitors that are approved by the FDA, seven of them form irreversible covalent adducts with their target enzymes. The clinical success of ibrutinib, an inhibitor of Bruton tyrosine kinase, in the treatment of mantle cell lymphomas following its approval in 2013 helped to overcome a general bias against the development of irreversible drug inhibitors. The other approved covalent drugs include acalabrutinib and zanubrutinib, which also inhibit Bruton tyrosine kinase. Furthermore afatinib, dacomitinib, and osimertinib, inhibitors of members of the epidermal growth factor receptor family (ErbB1/2/3/4), are used in the treatment of non-small cell lung cancers. Neratinib is an inhibitor of ErbB2 and is used in the treatment of ErbB2/HER2-positive breast cancer. The seven drugs considered in this review have a common mechanism of action; this process involves the addition of a protein cysteine thiolate anion (protein−S−) to an acrylamide derivative (CH2=CHC(=O)N(H)R) where R represents the pharmacophore. Such reactions are commonly referred to as Michael additions and each reaction results in the formation of a covalent bond between carbon and sulfur; the final product is a thioether. This process consists of two discrete steps; the first step involves the reversible association of the drug with its target enzyme so that a weakly electrophilic functionality, a warhead, is bound near an appropriately positioned nucleophilic cysteine. In the second step, a reaction occurs between the warhead and the target enzyme cysteine to form a covalently modified and inactive protein. For this process to work, the warhead must be appropriately juxtaposed in relationship to the cysteinyl thiolate so that the covalent addition can occur. Covalent inhibitors have emerged from the ranks of drugs to be avoided to become an emerging paradigm. Much of this recent success can be attributed to the clinical efficacy of ibrutinib as well as the other antagonists covered in this review. Moreover, the covalent inhibitor methodology is swiftly gaining acceptance as a valuable component of the medicinal chemist’s toolbox and is primed to make a significant impact on the development of enzyme antagonists and receptor modulators.

1. Drugs with a covalent mechanism of action including protein kinase antagonists

Because of overexpression and genetic alterations such as mutations and translocations, the dysregulation of protein kinase activity is involved in the pathogenesis of many diseases including inflammatory, cardiovascular, nervous, and autoimmune diseases as well as cancer. Accordingly, this enzyme superfamily has become one of the most important drug targets in the 21st century [1,2]. About one-fourth of drug discovery efforts in the United States and worldwide target protein kinases. The successful use of imatinib for treating Philadelphia chromosome-positive chronic myelogenous leukemias along with its FDA approval in 2001 motivated the pursuit of orally effective therapeutic protein kinase inhibitors [3]. This initial success was the result of the inhibition of the activated chimeric BCR-Ab1 protein-tyrosine kinase, the chief biochemical defect that causes these leukemias, by imatinib.

There are currently about 200 protein kinase inhibitors in clinical trials worldwide [4,5]. Moreover, there are 62 FDA-approved protein

Abbreviations: AS, activation segment; BP, back pocket; BTK, bruton protein-tyrosine kinase; C-spine, catalytic spine; CS1, catalytic spine residue 1; CL, catalytic loop; EGFR, epidermal growth factor receptor; ErbB, erythroblastosis related proto-oncogene of the EGFR family; F, front pocket; GK, gatekeeper; GRL, glycine-rich loop; KLIFS-3, kinase-ligand interaction fingerprint and structure residue-3; HER, human epidermal growth factor receptor; NSCLC, non-small cell lung cancer; PKA, protein kinase A; R-spine, regulatory spine; RS1, regulatory spine residue 1; Sh2, shell residue 2; TCIs, targeted covalent inhibitors.

E-mail address: rrj@brimr.org.

https://doi.org/10.1016/j.phrs.2021.105422
Received 31 December 2020; Accepted 31 December 2020
Available online 9 January 2021
1043-6618/© 2021 Elsevier Ltd. All rights reserved.
kinase antagonists that are directed against about 20 different enzymes (www.brimr.org/PKI/PKIs.htm). However, this represents only a small fraction of the protein kinase superfamily and many other protein kinases are potential therapeutic targets. Although the medicinal community was somewhat wary of developing irreversible covalent drugs for the treatment of diseases, the success of ibrutinib in the treatment of mantle cell lymphomas stimulated interest in the development of covalent protein kinase inhibitors [6].

Manning et al. reported that the human protein kinome consists of 518 members including 478 typical and 40 atypical enzymes [7]. These enzymes catalyze the following reaction;

\[ \text{MgATP}^2+ + \text{protein} - \text{O} - \text{H} \rightarrow \text{protein} - \text{O} - \text{PO}_4^{3-} + \text{MgADP} + \text{H}^+ \]

Note that the phosphoryl ion (PO\(_4^3\)) and not the phosphate (OPO\(_4^2\)) group is transferred from ATP to the protein substrate. Based upon the identity of the phosphorylated —OH groups, these catalysts are classified as protein-tyrosine kinases (90 members), protein-tyrosine-kinase-like enzymes (43) and protein-serine/threonine kinases (385). The protein-tyrosine kinases are composed of both receptor (58) and non-receptor (32) proteins. A small group of enzymes, such as MEK1/2, that catalyze the phosphorylation of tyrosine and then threonine residues within the activation loop of target protein kinases are classified as dual specificity kinases. Assuming a human genome contain of 20,000 genes and a human kinome of about 500 genes, then protein kinases constitute about 2.5 % of all genes. Manning et al. found that 244 protein kinases map to cancer amlicons or to disease loci [7] so that we can expect a significant increase in the number of therapeutic protein kinase targets that will be employed for the treatment of additional diseases including those outside of oncology [8].

Drugs that can form irreversible covalent adducts were disfavored as a drug class owing to toxicity and safety concerns [9]. Aspirin, however, is a covalent inhibitor that has been in the therapeutic armamentarium since 1899. Roth et al. found that aspirin exerts its medicinal effect by acetylating serine 530 of cyclooxygenase 1 [10,11]. Furthermore, irreversible protein pump inhibitors that reduce stomach acid such as omeprazole, esomeprazole, and lansoprazole are effective, safe, and widely used in the treatment of gastroesophageal reflux, dyspepsia, peptic ulcers, and other disorders [9]. That these three medicinal are available to the public without a prescription indicates their measure of safety. These drugs react with an essential gastric proton pump (H\(^+\)/K\(^+\) ATPase) cysteine to form an inactive disulfide adduct [12]. Moreover, seleagine and rasagiline, which are FDA approved for the treatment of Parkinson disease, represent additional examples of irreversible enzyme inhibitors [13]. These acetylenic medicinals, which inhibit type B monoamine oxidase by forming a covalent adduct with the N5 of the monoamine oxidase FAD cofactor, are used as a monotherapy in early Parkinson disease or as an adjunct therapy in more advanced cases [14].

Additionally, bortezomib is a boron-containing drug that selectively inhibits the ubiquitin proteasome pathway that participates in the degradation of many intracellular proteins. It is a selective and irreversible inhibitor of the 26S proteasome and possesses antiproliferative and antitumor activity [15]. The drug, which is given intravenously or subcutaneously, exerts its antineoplastic action by the inhibition of the nuclear factor-xB pathway associated with apoptosis, cell proliferation, and angiogenesis. It was approved for the first-line treatment of patients with multiple myeloma in 2003 and for the first-line treatment (2014) and second-line treatment (2008) of patients with mantle cell lymphoma. Moreover, the drug is used off-label for the treatment of Waldenstrom macroglobulinemia and peripheral T-cell lymphomas. The boron atom binds covalently to the catalytic site of the 26S proteasome with high affinity and specificity. Carfilzomib is the second FDA-approved proteosome inhibitor that is approved for the treatment of multiple myeloma. The drug contains a terminal epoxyketone group that forms an irreversible covalent bond with the proteasome catalytic threonine [16]. The US FDA has approved 32 drugs that form covalent bonds with their target protein since 1990 and thirteen drugs with this mechanism of action were approved in the past decade [6].

Of the 62 small molecule protein kinase inhibitors approved by the US FDA as of 31 December 2020, seven of these drugs form covalent bonds with their target enzymes. The clinical success of ibrutinib following its approval in 2013 has helped to overcome the general bias against the development of irreversible drug inhibitors [6]. The seven targeted covalent inhibitors (TCIs) of protein kinases include acalabrutinib (targeting BTK in mantle cell lymphomas, chronic lymphoblastic leukemias, and small cell lymphomas), afatinib (directed toward EGFR in NSCLC), dacomitinib (targeting mutant EGFR in non-small cell lung cancers), ibrutinib (directed toward BTK in mantle cell lymphomas, marginal zone lymphomas, chronic graft vs. host disease, chronic lymphocytic leukemias, and Waldenstrom macroglobulinemia), neratinib (targeting ErbB2 in HER2-positive breast cancers), osimertinib (directed toward EGFR T970M mutants in NSCLC), and zanubrutinib (targeting BTK in mantle cell lymphomas). Neratinib also inhibits EGFR and ErbB4. Note that EGFR and ErbB2 are receptor protein-tyrosine kinases and BTK is a non-receptor protein-tyrosine kinase. According to Singh et al. [9], a TCI is “an inhibitor bearing a bond-forming functional group of low reactivity that, following binding to the target protein, is positioned to react rapidly with a specific noncatalytic residue at the target site.”

2. Biochemistry of the epidermal growth factor receptor (EGFR) protein-tyrosine kinase family and Bruton protein-tyrosine kinase (BTK)

2.1. An overview of the EGFR family and the family ligands

The EGFR receptor protein-tyrosine kinases and their downstream effectors are among the most studied signal transduction modules in biology [17]. This line of investigation began when Stanley Cohen described epidermal growth factor (EGF), its receptor (EGFR), and its many biochemical activities [18]. Cohen discovered that EGFR possesses protein-tyrosine kinase activity, which was unknown at the time, and not protein-serine/threonine kinase activity (see Ref [19], for a historical review). Cohen et al. found that a single 170-kDa poly-peptide chain bound EGF and also possessed protein kinase activity [20]. Moreover, EGFR was the first receptor that established a relationship between gene mutations, protein overexpression, and cancer [21]. This receptor protein-tyrosine kinase family and its downstream effectors are among the most studied signal transduction modules because of its role in oncogenesis [22–24].

The human epidermal growth factor receptor (HER) family contains four members that belong to the ErbB lineage of proteins (ErbB1/2/3/4) [22]. The ERBB gene name originated from the related avian viral erythralosis oncogene. The four members of this receptor gene family include: (i) EGFR/ERBB1/HER1, (ii) ERBB2/HER2/NEU, (iii) ERBB3/HER3, and (iv) ERBB4/HER4. Although there is considerable variation, the ERBB nomenclature is associated with the biological sciences while the HER nomenclature is used more commonly in clinical papers and reports. Schechter et al. found that several rat neuro-/glioblastomas contain the Neu oncogene, which is related to the rat erbB2 gene of the epidermal growth factor receptor family [25]. This finding argued for the potential role of the ErbB family of receptors in the pathogenesis of various malignancies and led to the use of NEU in human gene nomenclature. The ErbB family of receptors is ubiquitously expressed in mesenchymal, epithelial, and neuronal cells as well as their undifferentiated precursors.

The role of the ErbB family in the pathogenesis of non-small cell lung cancers and breast cancer led to the development of several orally effective small molecule inhibitors of this enzyme family (www.brimr.org/PKI/PKIs.htm). The TCIs that target these and other cancers are listed in Table 1. The number of newly diagnosed lung cancers was estimated to be about 229,000 in the United States in 2020 and the number of deaths was estimated to be 136,000 [26]. The number of
newly diagnosed lung cancers was about 2.09 million worldwide in 2018 and the number of deaths was estimated to be about 1.76 million [27]. An estimated 85% of all lung cancers are classified as non-small cell lung cancers. The total number of newly diagnosed breast cancers in women was about 276,000 in the United States in 2020 and the number of deaths was estimated to be about 42,000 [26]. The total number of newly diagnosed breast cancers in women was about 2.08 million worldwide in 2018 and the number of deaths was estimated to be 627,000 [27]. An estimated 20% of breast cancers overexpress ErbB2. The occurrence of lung and breast cancers rank first and second among all types of malignancies worldwide [27] indicating the significance of developing effective therapies for these illnesses.

Based upon the amino acid sequence of EGFR as determined by cDNA analysis, Ulrich et al. suggested that this receptor consists of an extracellular ligand binding domain and an intracellular protein kinase domain connected by a hydrophobic transmembrane segment [28]. This initial hypothesis has subsequently been found to apply to nearly all receptor protein kinases. The EGFR protein kinase family members possess an extracellular domain consisting of four parts: domains I and III are related leucine-rich ligand binding segments and domains II and IV are cysteine-rich segments that form about a dozen disulfide bonds [22–24]. Furthermore, the second domain participates in both homo and heterodimer formation with EGFR family members, which is required for the activation of their protein kinase activity [22, 23]. The extracellular segment is followed by a single transmembrane portion containing about 25 amino acid residues and this is followed by an

### Table 1

<table>
<thead>
<tr>
<th>Drug (Code) Trade name</th>
<th>Year approved</th>
<th>Primary targets</th>
<th>Therapeutic indications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afatinib (BIBW 2992) Tovok</td>
<td>2013</td>
<td>ErbB1/2/4</td>
<td>Non-small cell lung cancers</td>
</tr>
<tr>
<td>Dacomitinib (PF00298045) Visimpro</td>
<td>2018</td>
<td>ErbB1/2/4</td>
<td>EGFR mutant non-small cell lung cancers</td>
</tr>
<tr>
<td>Neratinib (HKI-272) Herlynx</td>
<td>2017</td>
<td>HER2</td>
<td>HER2-positive breast cancers</td>
</tr>
<tr>
<td>Osimertinib (AZD-9293) Tagrisso</td>
<td>2015</td>
<td>ErbB3</td>
<td>Non-small cell lung cancers</td>
</tr>
<tr>
<td>Acalabrutinib (ACP-196) Calquence</td>
<td>2017</td>
<td>BTK</td>
<td>Mantle cell lymphomas; CLL; SLL</td>
</tr>
<tr>
<td>Ibrutinib (PCI-32,765) Imbruvica</td>
<td>2013</td>
<td>BTK</td>
<td>CLL; graft vs. host disease; marginal zone lymphomas; SLL; Waldenström macroglobulinemias</td>
</tr>
<tr>
<td>Zanubrutinib (BGB3111) Brukinsa</td>
<td>2019</td>
<td>BTK</td>
<td>Mantle cell lymphomas</td>
</tr>
</tbody>
</table>

* Although some of these drugs are multikinase inhibitors, only the primary therapeutic targets are given here.  

* CLL, chronic lymphocytic leukemias; SLL, small lymphocytic leukemias.

Fig. 1. (A) Overall organization of the human epidermal growth factor receptor family members consisting of EGFR/ErbB1/HER1, ErbB2/HER2/NEU, ErbB3/HER3 and ErbB4/HER4. The extracellular component of each receptor consists of four domains (I-IV). Domains I and III participate in ligand binding (except for those of ErbB2/HER2, which are marked with the stop symbol ⊘), and domain II participates in homo or heterodimer formation. The ErbB3/HER3 protein kinase domain, which is marked with the stop symbol ⊘, is catalytically impaired. The numbers represent amino acid residues of the nascent protein including the signal peptide (which is not depicted); each number corresponds to the initial residue of the adjacent segment except for (i) the last residues of the extracellular domains and (ii) the end of the proteins. The growth factor groups (1–4) that bind to the receptors are indicated. EGF, epidermal growth factor; AR, amphiregulin; EPG, epigen; TGFα, transforming growth factor-α; BTC, betacellulin; EPR, epiregulin; HB-EGF, heparin-binding epidermal growth-like factor; Nrg-1/2/3/4, neuregulin-1/2/3/4. PKD, phosphorylated domain. (B) Overall organization of BTK (Bruton protein-tyrosine kinase). The relative size of each segment is indicated by the amino acid residue number given below the diagram. Acalabrutinib binds covalently with C481. AS, activation segment; CL, catalytic loop; PH, pleckstrin homology domain; pY, phosphotyrosine; TH, TEC homology domain.
intracellular portion containing about 550 amino acid residues that consists of (i) a short juxtamembrane section, (ii) a protein kinase domain, and (iii) a long carboxyterminal tail (Fig. 1).

There are two commonly used amino acid numbering schemes for the EGFR family. The initial format employed by Ullrich et al. [28] for EGFR/ErbB1 corresponds to the mature protein and excludes the 24-residue signal peptide. The current format given in the UniProtKB knowledge base corresponds to the protein encoded by their respective mRNAs and includes the signal peptide. Although the use of the mature protein numbering system proposed by Ullrich et al. is ingrained in the EGFR literature, it is easier to use the UniProtKB nascent protein numbers when going from DNA to RNA and then to protein. Accordingly, the UniProtKB numbering scheme including the signal peptide is used in this review.

The growth factors that bind to each of the ErbB receptors are listed in Fig. 1. The term neuregulin (Nrg) corresponds to the NEU gene product and is also called heregulin. Fig. 1 shows that seven growth factors bind to EGFR, none bind to ErbB2, two factors bind to ErbB3, and seven factors bind to ErbB4. Furthermore, the ErbB3 receptor is kinase impaired. The EGFR family, like that of all other receptor protein-tyrosine kinases, consists of functional dimers or higher oligomers [17]. There is one chief isoform of ErbB1, two ErbB2 isoforms that differ slightly as a result of alternative mRNA splicing, and two full-length ErbB3 isoforms, one of which is missing residues 1–59. Adding to the complexity of the EGFR family, there are two different extracellular juxtamembrane versions (JMa and JMb) of ErbB4 and two different versions of the carboxyterminal tail (CTa and CTb) of this receptor.

Accordingly, there are four full-length ErbB4 isoforms that are the result of alternative pre-mRNA splicing: JMaCTa, JMaCTb, JMbCTa, and JMbCTb. Unfortunately, the functional significance of the four ErbB4 isoforms is unknown. Experimental data suggests that ErbB2 is the preferred dimerization partner for all other ErbB family members [29, 30]. Furthermore, Pinkas-Kramarski et al. found that ErbB2 heterodimers with ErbB1 or with ErbB3 exhibit robust signaling activity [31]. That the heterodimer of ErbB2 (which does not bind a growth factor) and ErbB3 (which lacks protein kinase activity) possesses vigorous signaling activity is paradoxical.

2.2. Structures of the small and large lobes and the protein kinase fold

The catalytic domains of the protein kinase family consist of ≈250–300 amino acid residues. Like all other protein kinases, the EGFR protein kinase domains have a small amino-terminal lobe and large carboxyterminal lobe (Fig. 2A) as first described for PKA by Knighton et al. (PDB ID: 2CPK) [32]. The two lobes form a cleft that binds ATP. The small lobe contains a flexible conserved glycine-rich ATP–phosphate–binding loop (GRL) or P-loop, which is near the phosphates of the ATP substrate. The G-rich loop contains a GxGxΦ signature (719GSGAFG724) where Φ refers to a hydrophobic residue and is phenylalanine in EGFR. The glycine-rich loop overlays the ATP-binding site and connects the β1- and β2-strands. The β1- and β2-strands of the small lobe dock with the adenine portion of ATP and they interact with small molecule inhibitors including those listed in Table 1. The β3-strand normally contains a conserved Ala-Xxx-Lys sequence, the

Fig. 2. (A) Active EGFR and its spine residues (B). (C) Inactive BTK and its spine residues (D). (E) Model of the ATP-binding site of EGFR. (F) Catalytic mechanism of action of EGFR. The chemistry occurs within the circle. AS, activation segment; CL, catalytic loop. P refers to P877, which serves as a platform for the protein-tyrosine substrate (Tyr sub). Figs. 2, 3b and 7 were prepared using the PyMOL Molecular Graphics System Version 1.5.0.4 Schrödinger, LLC.
lysin of which in human EGFR (K745) forms an electrostatic bond with a conserved glutamate near the center of the α-helix (E762) (Fig. 2A). The formation of the salt bridge between the β3-strand lysine and the α-helix glutamate is required for the creation of the active enzyme conformation and such a structure corresponds to the “αCin” conformation. In contrast, this salt bridge is absent in many dormant forms of EGFR and BTK and such structures correspond to an inactive “αCout” conformation as depicted for BTK (Fig. 2C). The αCin structure is necessary, but not sufficient, for the expression of maximum protein kinase enzymatic activity.

The large lobe contains a moveable activation segment with an open or extended conformation in operational enzymes (Fig. 2A) and closed conformation in dormant enzymes (Fig. 2C). The activation segment begins with a canonical DFG (Asp-Phe-Gly) sequence. The DF dyad can exhibit two different conformations in the same enzyme. In the functional state, the aspartate side chain carboxylate group extends toward the ATP-binding pocket and coordinates Mg$^{2+}$. This structure is designated as the “DFG-Din” conformation. In the latent activation segment conformation, the aspartate carboxylate of the DFG sequence extends away from the active site in the “DFG-Dout” conformation. The ability (DFG-Din) or inability (DFG-Dout) of the DFG-aspartate to bind or not bind to Mg$^{2+}$ in the active site is the critical consideration. See Ref. [1] for details concerning the two activation segment conformations. However, the dormant conformations of the EGFR family kinases including kinase-impaired ErbB3 generally exist in the DGF-Dout conformation with an inactive closed activation segment or with an inactive αCout conformation. Although the activation segment of protein kinases usually ends with APE (Ala-Pro-Glu), this segment ends with ALH (Alpha-Leu-Glu) in the ErbB family. The last eight residues of the activation loop of the four ErbB family consists of PIKWMALLE and this octet constitutes the protein-substrate positioning loop. The R-group side chain of proline in this sequence serves as a platform that supports the aromatic tyrosyl residue of the protein substrate that is phosphorylated (Fig. 2E) [33]. Although the activation loop of the EGFR family contains a tyrosine residue that can be phosphorylated, unlike many other protein-tyrosine kinases this phosphorylation is not required for enzyme activation [34]. ErbB1/2/4 are operational protein-tyrosine kinases that exist in similar operational and dormant conformations. In contrast, ErbB3 lacks critical catalytic residues and is thus kinase inactive. Moreover, it assumes the structure of a dormant protein kinase. The large lobe of the EGFR family of protein kinases is primarily α-helical with eight conserved segments (αD-αI and αEF1/2) that are found in all protein kinases. The first protein kinase X-ray crystallographic structure (PKA) depicted two short helices proximal to the large lobe αC-helix, which were unannealed at the time (PDB ID: 2CPK). However, these αEF1/2 helices are found in all active protein kinases and they represent the seventh- and eighth-conserved helices in the C-terminal lobe (Fig. 2A). The dormant forms of ErbB1/2/3/4 and BTK possess an αC-helix in the proximal portion of the Activation Loop. The activation loop of active EGFR is open and extends outward while that of the dormant or less active BTK is closed and more compact (Fig. 2A/C). The carboxy-terminal lobe of active protein kinases contains four short β-strands (β6-β9) (Fig. 2A). The β6-strand, the primary structure of which occurs before the catalytic loop, interacts with the activation segment β9-strand. The primary structure of the β7-strand is located between the catalytic loop and the activation segment and it interacts with the adjacent downstream β8-strand. The dormant forms of all four ErbB family members contain the β7- and β8-strands, but they all lack the β6- and β9-strands.

There are two common structural movements associated with the activation and inactivation of all protein kinases including those of the EGFR family. These movements typically involve adjustments in the location of the αC-helix in the small lobe and the activation segment in the large lobe. The interconversion of the dormant and functional forms of the ErbB kinases also involves an electrostatic switch. In dormant enzyme forms, the EGFR β3-lysine (K745) forms an electrostatic bond with the DFG-D (D855) residue. The conversion to the functional enzyme entails an electrostatic switch as the β3-strand lysine forms a polar bond with the αC-helix glutamate with the simultaneous formation of the αCin conformation. The catalytically functional forms of the ErbB1/4 possess the β3-strand lysine–αC-helix glutamate salt bridge (e.g., PDB ID 1M14 for EGFR and 3BCE for ErbB4) and all of the inactive ErbB1-4 enzymes are able to form the β3-strand lysine–DFG-D salt bridge (e.g., PDB ID 4H10 for EGFR, 3RDC for ErbB2, 3KEX for ErbB3, and 3BBW for ErbB4). The important residues of the EGFR family and BTK are listed in Table 2.

2.3. Structures of the hydrophobic spines in the active and in the dormant ErbB/HER protein kinase domains

2.3.1. The regulatory spine

Kornev et al. [35,36] examined the tertiary structures of dormant and functional structures of around two dozen protein kinases and they determined a group of functionally significant residues by a local spatial pattern (LSP) alignment protocol. The residues that make up the catalytic and regulatory spines were determined by their three-dimensional location based upon X-ray crystallographic structures and not by an amino acid signature sequence such as HRD or DFG. Their structural analysis revealed a supporting frame of eight hydrophobic residues that form a catalytic or C-spine and four hydrophobic residues that form a regulatory or R-spine (Fig. 2B/D). These spines consist of residues that occur in both the small and large lobes. The R-spine contains one residue from the αC-helix and another from the activation segment, whose configurations are important in establishing and identifying dormant and functional states. The C-spine promotes ATP binding and thereby mediates catalysis. The correct alignment of both spines is required for the fabrication of an active enzyme. ErbB1/EGFR, ErbB2, and ErbB4 have been observed in both dormant and functional conformations as demonstrated by X-ray crystallography.

The ErbB1 regulatory spine consists of a residue from the beginning of the β4-strand (L777), from the carboxyterminal end of the αC-helix (M766), DFG-F856, along with HRD-H835 of the catalytic loop. M766 and comparable residues from other protein kinases are four residues carboxyterminal to the conserved αC-glutamate. The backbone of H835 is anchored to the α-helix by a hydrogen bond to a conserved aspartate residue (D896). The activation segment, the protein-substrate positioning loop, and the αH loop of protein kinase domains, including the ErbB/HER family, interact hydrophobically with the αC-helix [35,36].

2.3.2. The catalytic spine

The αC-spine of protein kinases consists of residues from the small and large lobes and is completed by the adenosine moiety of ATP [36]. The two residues from the small lobe of ErbB1 that interact with the adenosine group of ATP include V726 near the origin of the β2-strand and A743 from the canonical β3-strand Alac-xxx-Lys. Moreover, L844 from the midpoint of the β7-strand of the large lobe binds to the adenosine moiety of the functional enzyme. V726, A743, and L844 typically make hydrophobic contact with the pharmacophores of ATP-competitive as well as irreversible small molecule inhibitors. V843 and V845, which are next to L844, bind to L798 at the beginning of the αD-helix. L798 interacts with I907 and T903 in the αD-helix. Note that both spines are anchored to the α-helix, a very hydrophobic component that is entirely within the protein. The core α-helix buttresses both spines and they, in turn, secure the protein kinase catalytic residues. See Table 3 for a list of the residues of the spines of human ErbB1/2/3/4 and BTK.

The protein kinase spines play a crucial role in the structure and function of protein kinases and it is not possible to overemphasize their importance in the functioning of the protein kinase superfamily as well as their interactions with small molecule kinase inhibitors. See Refs. [37, 38] for a review of the characteristics of the ALK pleiotrophin and midkine receptor protein-tyrosine kinase spine residues, Refs. [22–24] for

2.3.3. The gatekeeper and other shell residues

Based upon site-directed mutagenesis methodologies, Meharena et al. discovered three residues in protein kinase A (PKA) that stabilize the R-spine and are designated as shell residues [56]. These investigators et al. found that only three of 14 amino acid residues contiguous with the gatekeeper refers to the function of such residues in regulating ligand entry to a hydrophobic back pocket adjacent to the adenine binding site [57,58] that is occupied by fragments of many low molecular weight protein kinase antagonists. Using their local spatial alignment data, Meharena et al. found that only three of 14 amino acid residues contiguous with RS3 and RS4 of PKA are conserved and they found that shell residues help to stabilize the R-spine [56]. A comparison of the R-spines of functional EGFR and those of dormant BTK shows that RS3 of the dormant enzyme is displaced outward when compared with active EGFR, a result that is consistent with the notion that the αC-helix is displaced in inactive BTK (Fig. 2).

2.3.4. The binding pocket for ATP and small molecule inhibitors

The exocyclic amino group of ATP characteristically interacts with a carbonyl group of the first hinge residue. The hinge-linker residues connect the small and large lobes and occur after the β3-strand. Thus, the 6-amino group of the adenine portion of ATP forms a hydrogen bond with the carbonyl oxygen of Q791 (PDB ID: 2G56), the first hinge residue of ErbB1. The adenine N1 nitrogen forms a hydrogen bond with the αC-helix at the -NH group of M793, the third hinge residue. The ATP α-phosphate group binds to K745 of the β3-strand, which in turn makes a salt bridge with E762 of the αC-helix (Fig. 2E). Additionally, the ATP γ-phosphoryl

Table 2

Important amino acid residues in the human EGFR/ErbB family and Bruton tyrosine kinase.

<table>
<thead>
<tr>
<th>No.</th>
<th>EGFR</th>
<th>ErbB2</th>
<th>ErbB3</th>
<th>ErbB4</th>
<th>BTK</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1210</td>
<td>1255</td>
<td>1342</td>
<td>1308</td>
<td>659</td>
</tr>
<tr>
<td>2</td>
<td>1–24</td>
<td>1–22</td>
<td>1–19</td>
<td>1–25</td>
<td>None</td>
</tr>
<tr>
<td>3</td>
<td>25–645</td>
<td>23–652</td>
<td>20–643</td>
<td>26–651</td>
<td>None</td>
</tr>
<tr>
<td>4</td>
<td>646–668</td>
<td>653–675</td>
<td>644–664</td>
<td>652–675</td>
<td>None</td>
</tr>
<tr>
<td>5</td>
<td>669–1210</td>
<td>612–1255</td>
<td>665–1342</td>
<td>676–1308</td>
<td>1–659</td>
</tr>
</tbody>
</table>

Table 3

Human epidermal growth factor receptor family (ErbB1/2/3/4) and Bruton tyrosine kinase (BTK) residues that form the R-spine, C-spine and Shell residues.

<table>
<thead>
<tr>
<th>KLIFS No.</th>
<th>ErbB1</th>
<th>ErbB2</th>
<th>ErbB3</th>
<th>ErbB4</th>
<th>BTK</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>hklfs.net.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

α AS, activation segment.
group forms a salt bridge with Mg^{2+} (1) which in turn binds to DFG-D855 (not shown). Structural studies indicate that the adenine moiety extends only to the β2-strand, but not to the β3-strand. In comparison, most low molecular weight ATP-competitive inhibitors extend to the β3-strand and many extend even further toward the αC-helix into the back pocket as described later.

2.3.5. The catalytic loop and activation segment

Functional protein kinases contain an amino acid signature composed of K/E/D/D (Lys/Glu/Asp/Asp) and these residues play essential mechanistic and structural roles in the ErbB family as well as the Bruton protein-tyrosine kinase (Table 2). The first residue of this signature in EGFR is the β3-strand K745, which forms a salt bridge with the αC-helix E762, the second residue of this signature. The catalytic loop consists of eight residues with a canonical HRDxGxN signature; the HRD-D residue is the first D of K/E/D/D. This loop is made up of an HRDLAARN sequence in most receptor protein-tyrosine kinases including ErbB1/2/4. In contrast, catalytically compromised ErbB3 contains HRNLAARN with an asparagine (N) in place of aspartate (D). The catalytic aspartate (D837) of EGFR serves as a base that abstracts a proton from the protein substrate tyrosyl O—H group (Fig. 2F). Zhou and Adams proposed that the catalytic loop aspartate of protein kinases localizes the substrate O—H group for an in-line nucleophilic attack [59]. Additional studies indicate that the DFG-D855 of protein kinases such as EGFR at the beginning of the activation segment binds Mg^{2+} (1) while the asparagine residue at the end of the catalytic loop (N842) coordinates a second Mg^{2+} (2) as depicted in Fig. 2E. The activation loop DFG-D represents the second D of the K/E/D/D signature. The activation loop of EGFR contains a tyrosine residue that can undergo phosphorylation, but in contrast to most other protein-tyrosine kinases, activation loop phosphorylation is not required for ErbB protein kinase activation [34]. The C-terminal residues of the ErbB/HER activation segments (PIKWMAL) comprise the protein-substrate positioning segment and P877 serves as a platform for the tyrosyl substrate residue (Fig. 2F).

2.4. Bruton protein-tyrosine kinase (BTK)

BTK was originally identified as a non-receptor protein-tyrosine kinase in 1993; a deficiency of this enzyme occurs in X-linked agammaglobulinemia [60–63]. Individuals with this disease possess alterations exclusively in B cells, and this finding is in accordance with the limitation of clinical features to B cell immunity. The Syk and Lyn protein kinases are upstream of BTK while phospholipase γ2 is downstream in the BTK signaling pathway leading to the generation of inositol triphosphate and diacylglycerol [46]. BTK belongs to the Tec family of protein-tyrosine kinases including BTK, ITK (inducible T cell kinase), BMX (bone marrow-expressed kinase), TEC, and TXK (also known as RLK, resting lymphocyte kinase) [7]. BTK contains a short N-terminal pleckstrin homology (PH) domain, followed by a Tec homology (TH) domain, an SH3, SH2, and finally a C-terminal protein kinase domain (Fig. 1B); this architecture occurs in the other members of the Tec family of protein kinases.

Like the ErbB family, protein kinases including BTK contain a glycine-rich loop (GRL) with a GxGxG signature (40%GTPG41%). Recall that Φ refers to a hydrophobic residue and is phenyalaniline in BTK (and EGFR). Because of their role in ATP binding and ADP release, the glycine-rich loops are flexible. Like the ErbB family and other protein kinases, BTK contains a conserved Ala-XXX-Lys (428ALK430) sequence in the β3-strand and the αC-helix a conserved glutamate (E445) that forms a salt bridge with the conserved β3-strand lysine in the active protein kinase conformation. BTK contains the conserved HRD loop, which is located at the beginning of the catalytic loop (319HRDLAARN526). BTK also contains the conserved 359DFG361 sequence and a 565PPE5677 sequence; these represent the beginning and end of the protein kinase activation segment. Although the end of the activation segment of most protein kinases consists of APE, the activation segments of BTK and the EGFR family differs from the most common terminal activation segment sequence.

All protein kinases including the Bruton protein-tyrosine kinase have a small N-terminal lobe and large C-terminal lobe [1] (Fig. 2F) as described above for the ErbB family. BTK also possesses the K/E/D/D (Lys/Glu/Asp/Asp) amino acid residue signature. Although both lobes contribute to ATP binding, the small lobe plays a major role in this process. The BTK β3-strand K430 (the K of K/E/D/D) holds the α- and β-phosphates in position (not shown). The K430 ε-amino group forms a salt bridge with the E445 carboxylate group (the E of K/E/D/D) of the αC-helix that serves to stabilize their interactions with these phosphates. The salt bridge between the αC-glutamate and the β3-lysine is required for the formation of an active protein kinase conformation, which corresponds to an “αC-in” structure. In contrast E445 and K430 of the inactive enzyme fail to make contact and this corresponds to an “αC-out” configuration (Fig. 2C). The αC-in structure is necessary, but not sufficient, for the full expression of catalytic activity. As noted above for the ErbB family of protein kinases, the large lobe of BTK binds to the protein-peptide substrate and participates in the catalytic cycle. The essential residues of BTK are listed in Table 2 and the spine and shell residues are listed in Table 3.

3. Orally effective protein kinase targeted covalent inhibitors (TCIs)

3.1. Selected diseases related to the EGFR family of protein kinases

3.1.1. Classification and treatment of lung cancers

EGFR/ErbB1 plays a major role in the development of many lung cancers. For example, Herbst et al. reported that ErbB1 protein kinase domain mutations were found in 10–40% of lung cancer samples [64]. These investigators found that the most common mutations included (i) a deletion of five exons-19 residues (746ELREA750) that occur proximal to the αC-helix and (ii) an exon-21 arginine for leucine substitution (L858R) within the activation loop. Although dozens of ERBB1 mutations have been reported in NSCLC, these two mutations are most important and account for more than 90% of the activating EGFR mutations found in this malignancy.

The gain-of-function mutations of oncokinases are commonly found in or near important regulatory regions such as the ATP-phosphate binding loop, the αC-helix, or the activation segment. A frequent mechanism for the oncogenic activation of the EGFR involves the destabilization of the dormant enzyme state thereby facilitating the generation of the more active enzyme ensemble. For example, Yun et al. reported enzyme destabilization as the mechanism responsible for the activation of the L858R EGFR mutant [65]. This amino acid residue immediately follows the 855DFG857 sequence at the beginning of the activation segment. The replacement of the hydrophobic leucine R-group with the larger positively charged arginine R-group excludes its presence in the proximal activation segment inhibitory αAL loop while an arginine is readily accommodated in the open configuration of the active EGFR protein kinase domain [65]. Consequently, destabilization of the dormant form of EGFR leads to protein kinase activation and subsequent neoplastic transformation.

The majority of newly diagnosed patients with advanced or metastatic NSCLC are unsuitable candidates for surgery or radiotherapy [66]. First-line platinum-based cytotoxic therapy was formerly the standard of care for these patients according to the European Society for Medical Oncology and the American Society of Clinical Oncology clinical practice guidelines. Alanazi et al. ranked treatments for safety and efficacy in the use of protein-tyrosine kinase inhibitors in the treatment of NSCLCs bearing EGFR mutations [66]. They reported that these EGFR inhibitors improved survival with fewer grade 3 or higher adverse events compared with cytotoxic chemotherapy. These investigators reported that overall-survival outcomes suggest that osimertinib has the highest probability of being the most efficacious, followed by dacomitinib.
Adverse-event results indicate that osimertinib and gefitinib (a non-covalent inhibitor) have the highest probability of being the safest [66].

Results from a clinical trial conducted by Ramalingam et al. on patients with untreated EGF-R-mutant NSCLC indicated that osimertinib-treated patients had a longer progression-free survival than patients treated with gefitinib or erlotinib (18.9 months vs. 10.2 months) [67]. Although the data were incomplete, the preliminary results suggested that overall survival was greater with osimertinib when compared with the gefitinib/erlotinib group. Moreover, the incidence of adverse events was lower in the osimertinib group. Wu et al. reported on the findings of a double-blind clinical trial on the efficacy of osimertinib as an adjuvant therapy in patients who underwent the surgical resection of stage IB to IIIA EGF-R-mutated NSCLC [68]. They reported that 90 % of those patients with stage II to IIIA treated with osimertinib were alive and disease-free after 24 months compared with 44 % of the placebo group. They reported that disease-free survival in the overall population of patients with stage IB to IIIA disease was 89 % in the osimertinib cohort and 52 % in the placebo cohort. This corresponds to an 80 % reduction in the risk of disease recurrence or death in the drug cohort. They found that the benefits were greater at more advanced stages of disease. Osimertinib treatment also decreased the recurrence of brain metastasis (2 %) when compared with the placebo group (11 %). This result parallels the greater effectiveness of osimertinib in decreasing the risk of brain metastasis as the first-line treatment of advanced EGF-R-mutant NSCLC.

3.1.2. Classification and treatment of breast cancers

Breast cancer is the leading cause of death from malignancies predominantly (breast) or exclusively (uterine cervix, uterine corpus, and ovary) occurring in women in the United States and worldwide [26,27]. Breast cancers are grouped into three categories, which are not mutually exclusive, for purposes of treatment: these include (i) overexpression of the ERBB2/HER2/NEU gene, (ii) hormone receptor-positive cancers, and (iii) triple-negative breast cancers. Triple-negative breast cancers refer to those (i) without ERBB2/HER2 overexpression or amplification and those lacking (ii) progesterone and (iii) estrogen receptors. Wittliff found that ERBB2/HER2 overexpression occurs in about 20 % of breast carcinomas while 10–20 % of breast carcinomas are triple-negative [69]. ERBB2/HER2 overexpression was correlated with a poor prognosis prior to the advent of ErbB2/HER2 targeted therapies. He also found that receptors for progesterone, estrogen, or both are detected in about four-fifths of all breast cancers. Moreover, he found that about 56 % of breast cancers contain both the progesterone and estrogen receptors while 9 % contain only the progesterone receptor and 14 % contain only the estrogen receptor while 21 % breast cancers lack both receptors [69].

Surgery is the primary treatment modality for localized breast cancer and is also used for the treatment of some advanced breast cancers. Other treatments include targeted ErbB2/HER2 therapy, radiotherapy, cytotoxic chemotherapy, immunotherapy, and adjuvant hormonal therapy (with tamoxifen or an aromatase inhibitor) for hormone receptor-positive tumors [70,71]. Unfortunately, endocrine-based therapy and immunotherapy are not yet as effective in the treatment of HER2-positive disease as clinicians and patients would desire and chemotherapy is accompanied by substantial toxicity. Numerous cytotoxic drugs are used in the treatment of metastatic breast cancers, especially those malignancies that are triple-negative or hormone receptor-negative [72]. These cytotoxic agents include cyclophosphamide, taxanes, and doxorubicin. Cyclophosphamide is an alkylating agent that forms both intrastrand and interstrand DNA cross links that inhibit the progression of topoisomerase II, and produces oxygen-dependent single and double stranded DNA breaks with subsequent inhibition of DNA function. One of the preferred chemotherapeutic regimens recommended by the National Comprehensive Cancer Network includes cyclophosphamide and doxorubicin followed by paclitaxel. Gemcitabine, capecitabine, vinorelbine, and pemetrexed are additional cytotoxic drugs used in the treatment of advanced breast cancers [22–24]. Gemcitabine is a cytidine analogue that inhibits (i) DNA synthesis, repair, and function, (ii) ribonucleotide reductase, and (iii) RNA function. Capecitabine is prodrug that is metabolized to 5-fluorouracil, which inhibits thymidylate synthase, DNA synthesis and function, and RNA function. Vinorelbine is an antimitotic that binds to tubulin to inhibit microtubule function and arrest mitosis. Pemetrexed is an antifolate that inhibits dihydrofolate reductase, thymidylate synthase, and purine synthesis de novo. Besides the overexpression of wild type ERBB2 in breast cancers, Bose et al. found that about 1.5 % of breast cancer patients possess an ERBB2 mutation [73], which corresponds to about 4000 new cases per year in the United States.

About 20 % of advanced breast cancer cases are HER2-positive [69]. The standard first-line treatment for this disorder includes trastuzumab and pertuzumab in combination with a taxane [71]. Trastuzumab is a monoclonal antibody directed against the extracellular domain of ErbB2/HER2 that results in HER2 internalization and down-regulation and stimulates immune cells to kill the HER2-expressing cell. Pertuzumab is a monoclonal antibody that targets ErbB2/HER2 and precludes its dimerization with other ErB family members. The standard second-line treatment consists of ado-trastuzumab emtansine, an antibody-drug conjugate that delivers the emtansine microtubule inhibitor to ErbB2/HER2-positive cells. One follow-up treatment includes the trastuzumab deruxtecan antibody-drug conjugate as a monotherapy following two previous treatments. Deruxtecan damages DNA and causes apoptosis. Another third-line treatment consists of the non-covalent HER2 antagonist tucatinib in combination with capecitabine and trastuzumab. Additionally, the HER2-TCI neratinib in combination with capecitabine is another third-line treatment option [70,71].

3.2. Diseases related to Bruton tyrosine kinase over activity

Mantle cell lymphomas are B cell disorders that make up about 6 % of non-Hodgkin lymphomas and mantle cell lymphomas usually present with palpable lymphadenopathy at a median age of about 65 years [74]. The male/female ratio is 4.1. Nearly 70 % of patients are at stage IV at the time of diagnosis with peripheral blood, bone marrow, spleen, and gastrointestinal involvement. The historical median overall survival in people with newly diagnosed mantle cell lymphomas is three to four years. Wang et al. found that ibrutinib led to a substantial increase in mantle cell lymphoma cells in the blood after 10 days of treatment followed by a decrease to baseline after 28 days [75]. This transient increase was due to the liberation of these cells from the affected lymphatic tissues. The authors concluded that ibrutinib monotherapy was less stressful and more effective than previous standard therapies. Based upon these studies, the FDA approved ibrutinib for the treatment of mantle cell lymphoma in 2013 [76].

Chronic lymphocytic leukemia, which is a clonal B cell disorder, is the most common type of leukemia in the Western hemisphere accounting for about 40 % of all adult leukemias [77]. The age-adjusted incidence rate is 4.5 per 100,000 inhabitants in the United States, which amount to about 15,000 new cases per year. The median age at diagnosis is about 70 years of age. Its diagnosis is often incidental and based upon routine blood counts. In symptomatic patients, fatigue, fever, and infections may be the presenting features. Clinical trials demonstrated the efficacy of ibrutinib in the treatment of chronic lymphocytic leukemias and small cell lymphomas and these positive outcomes led to its FDA approval for the treatment of these disorders [78].

Waldenström macroglobulinemia is a B cell neoplasm associated with the accumulation of clonal immunoglobulin M secreting lymphoplasmacytic cells [79]. The age-adjusted incidence rate is about 0.38 per 100,000 inhabitants in the United States, which amounts to about 1300
cases per year. The median age at diagnosis is about 70 years of age and the male/female ratio is about 1.5/1. The disorder is quite heterogeneous and can present with fever, chills, fatigue, headaches, recurrent bronchial and sinus infections, gastrointestinal cramping, and diarrhea. Lymphadenopathy and hepatosplenomegaly may also be present. Elevated IgM levels (>7 g/dL) that are more than 25 times greater than the normal value may be present. Such high levels lead to an increase in serum viscosity, which may manifest itself as episodes of mental confusion, blurry vision, and headaches. Studies indicate that BTK is constitutively activated in Waldenström macroglobulinemia cells and clinical trials led to the FDA approval of ibrutinib for the treatment of this illness. See Ref. [78] for a summary of the clinical efficacy of ibrutinib, acalabrutinib, and zanubrutinib in the treatment of B-cell lymphomas and see Ref [80] for a synopsis of the treatment of Waldenström macroglobulinemias with ibrutinib.

4. Protein kinase-inhibitor classification and their binding pockets

Based upon the work of other groups [81–84], we divided the small molecule protein kinase antagonists into seven main groups including reversible (Groups I, I ½, II, III, IV, V) and irreversible inhibitors (Table 4). We followed the lead of Liao [85], van Linden et al. [86], and Kanen et al. [87] in formulating and describing drug-binding pockets. A general overview illustrating the locations of the pockets and subpockets is provided in Figs. 3 and 5. These investigators divided the region between the protein kinase small and large lobes into a front cleft or front pocket, a gate area, and a back cleft. Hydrophobic pocket II (HIPII) or the back pocket includes the gate area and back cleft. The front cleft includes the glycine-rich P-loop, the adenine-binding pocket, the hinge residues, the linker segment connecting the hinge residues to the large lobe αD-helix, and the catalytic loop (HRD(x)N). The gate area includes DFG of the activation segment and the β3-strand of the small lobe. The back-cleft extends to the αC-helix, the αC–β4 back loop, to portions of the β4- and β5-strands of the N-terminal lobe, and to a section of the αE-helix within the C-terminal lobe. van Linden et al. described several components that occur in the drug-binding pocket and these are illustrated in Fig. 3 [86].

van Linden et al. and Kanen et al. created a comprehensive catalogue of ligand and drug binding to more than 5200 human and mouse protein kinase domains [86,87]. Their KLIFS (kinase–ligand interaction fingerprint and structure) directory includes an alignment of 85 ligand binding-site residues occurring in both the N-terminal and C-terminal lobes; this directory facilitates the classification of ligands and drugs depending upon their binding properties that assist in the detection of common interactions. Furthermore, these authors formulated a standard amino acid residue numbering system that facilitates the comparison of different protein kinases. Table 3 indicates the relationship of the KLIFS database nomenclature and the regulatory spine, shell, and catalytic spine amino acid residue numbering system and Fig. 4 depicts the location of the KLIFS residues within the protein kinase domain. Moreover, these investigators established a valuable free and searchable web site that is regularly updated thereby providing complete data on the interaction of protein kinases with ligands and drugs (klifs.net).

Additionally, Carles et al. developed a comprehensive directory of protein kinase antagonists that have been approved or are in clinical trials [4]. They have established a free and searchable web site that is regularly updated and depicts the inhibitor structures, their protein kinase targets, physical properties, therapeutic indications, year of first approval (if applicable), and trade name (http://www.icoa.fr/ pkidb/). The Blue Ridge Institute for Medical Research maintains a web site that lists the FDA-approved protein kinase inhibitors and provides their (i) structures, (ii) number of hydrogen bond donors/acceptors, (iii) calculated logD of the distribution coefficient, (iv) number of rings and rotatable bonds, (v) year of initial approval, (vi) presumed protein kinase targets, (vii) clinical indications, and (viii) FDA labels. This web site, which is regularly updated, is found at www.brim.org/ PKIS. htm.

The seven drugs considered in this review have a common mechanism of action; this process involves the addition of a protein cysteine thiolate anion (protein–S-) to an acrylamide derivative (CH2=CHC=O)N(H)R) where R represents the pharmacophore. Such reactions are commonly referred to as Michael additions and each reaction results in the formation of a covalent bond between carbon and sulfur and the final product is a thioether (Fig. 5A). Most of these Michael addition reactions are irreversible in nature. This process consists of two discrete steps; the first step involves the reversible association of the drug with its target enzyme so that a weakly electrophilic functionality, a warhead, is bound near an appropriately positioned nucleophilic cysteine. The dissociation constant for this part of the reaction is given by the K1 (Fig. 5B). In the second step, a reaction occurs between the warhead and the target enzyme cysteine to form a covalently modified and inactive protein and the rate constant for this process is given by k_{inact}. For this process to work, the warhead must be appropriately juxtaposed in relationship to the cysteiny1 thiolate so that the covalent addition can occur. The overall velocity of the irreversible covalent target modification is described by the second-order rate constant k_{inact}/K1 with units of moles/L per second (M/s). Covalent inhibition is time-dependent and increasing the time of incubation leads to an increase in target inhibition. Consequently, IC50, EC50, or K1 values are not the suitable parameters for comparing the efficacy of covalent ligands because these values vary with time and do not reflect the relative contributions of the k_{inact} and K1 [9]. Although IC50 or EC50 values are reported in much of the older literature owing to the extra work required to determine second-order rate constants, some more recent studies are reporting k_{inact}/K1 values.

5. Structures of EGFR- and BTK-drug complexes

5.1. EGFR-drug complexes

Afatinib is a 3-chloro-4-fluoroanilino-quinoxaline derivative (Fig. 6A) that is FDA-approved for the initial or first-line treatment of patients with NSCLC harboring EGFR-mutations or as a second-line treatment for patients with squamous cell NSCLC progressing after platinum-based chemotherapy [88–91]. Its inhibitory power against other protein kinases has not been reported. Its X-ray crystal structure shows that the quinoxaline N1 forms a hydrogen bond with the EGFR M793 (the third hinge residue) NH group (Fig. 7A). The drug makes hydrophobic contact with four spine residues (R53, G56,7,8) and two shell residues (Sh2/3). Afatinib makes hydrophobic contact with the KLIFS-3 j1-strand L718, K728 of the j2-strand, AXK745 of the β3-strand, E762 and M766 of the αC-helix, and L788 of the β5-strand. Moreover, the drug makes hydrophobic contact with 792LMFGC797 of

Table 4

<table>
<thead>
<tr>
<th>Inhibitor type</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Binds in and around the ATP-binding pocket of an active enzyme</td>
</tr>
<tr>
<td>I ½/A/B</td>
<td>Binds in and around the ATP-binding pocket of an inactive DFG-D_{inact} enzyme</td>
</tr>
<tr>
<td>I/3</td>
<td>Extends into the back cleft</td>
</tr>
<tr>
<td>I ½</td>
<td>Does not extend into the back cleft</td>
</tr>
<tr>
<td>II/A/B</td>
<td>Binds in and around the ATP-binding site of an inactive DFG-D_{inact} enzyme</td>
</tr>
<tr>
<td>II</td>
<td>Extends into the back cleft</td>
</tr>
<tr>
<td>II B</td>
<td>Does not extend into the back cleft</td>
</tr>
<tr>
<td>III</td>
<td>Allosteric inhibitor bound near to the ATP-binding site</td>
</tr>
<tr>
<td>IV</td>
<td>Allosteric inhibitor bound away from the ATP-binding site</td>
</tr>
<tr>
<td>V</td>
<td>Bivalent inhibitor spanning two kinase domain regions</td>
</tr>
<tr>
<td>VI</td>
<td>Covalent inhibitor</td>
</tr>
</tbody>
</table>

* Adapted from Ref. [92].
the hinge-linker segment. The drug makes additional hydrophobic
contract with D800 of the αD-helix, R841 of the catalytic loop, and T854 (the x of xDFG) within the carboxyterminal lobe. The drug forms a co-
valent Michael adduct with C797 that is found in the linker segment immediately before the large lobe αD-helix. The quinazoline group oc-
curs in the adenine pocket and the 3-chloro-4-fluoroanilino group occurs in the gate area (BP-I-A and BP-I-B). Although the drug is bound to an active conformation of EGFR, it is classified as a type VI inhibitor [92].

Like afatinib, dacomitinib is a 3-chloro-4-fluoroanilino-quinazoline
derivative and it forms a covalent bond with EGFR/ErbB2/ErbB4 (Fig. 6B) [24]. It is approved for the initial or first-line treatment of patients with advanced or metastatic NSCLC with EGFR exon 19 deletions or the exon 21 L858R substitution mutation [93–97]. The X-ray crystal structure shows that the N1 nitrogen of the quinazoline group forms a hydrogen bond with the EGFR T790 M mutant M793 NH group, which is the third hinge residue (Fig. 7B). The drug also makes hydro-
phobic contact with three spine residues (CS6/7/8), two shell residues (Sh2/3), and the KLIF-3 residue in the β1-strand. It also makes hydrophobic contact with the β3-strand I744 and K745, 788LITQLMPFGC797 of the hinge-linker segment, the αD-helix D800, T854 (the x of xDFG), and DFG-D855. The dacomitinib acrylamide group forms a covalent Michael addition product with C797 within the hinge-linker of the enzyme. The drug occupies the front pocket, gate area, and sub-pocket BP-I-B (klifs.net). The antagonist is bound to an inactive enzyme with DFG-D in, αC out, and a closed activation segment containing the αAl-helix. Because it is bound covalently to its target, dacomitinib is classified as a type VI inhibitor [92]. As in the case of osimertinib, one of the mechanisms of
resistance to dacomitinib results from an *EGFR* C797S mutation [97,98], the serine of which is unable to form a covalent adduct with the drug.

Osimertinib is an irreversible indole-pyrimidine derivative (Fig. 6C) [89] that is FDA-approved for the initial or first-line treatment of patients with advanced or metastatic NSCLC bearing *EGFR* exon-19 deletions or the exon 21 L858R mutation. It is also approved for the second-line treatment of patients with metastatic *EGFR* T790M-positive NSCLC that has progressed on or after *EGFR* protein-tyrosine kinase inhibitor therapy. Moreover, this was the first targeted medicinal that was approved for patients with the gatekeeper *EGFR* T790M drug-resistant variant [90]. Additionally, osimertinib is more effective than either erlotinib or gefitinib in treating patients with *EGFR*-positive NSCLC brain metastases [91]. The X-ray crystal structure reveals that a pyrimidine nitrogen forms a hydrogen bond with the NH group of M793 (the third hinge residue) and a drug amino NH group forms a hydrogen bond with the carbonyl group of this residue (Fig. 7C) [99]. The drug makes hydrophobic contact with three spine residues (CS6/7/8) and with the KLIFS-3 residue. It also makes hydrophobic contact with the β3-strand K745, L792 (the second hinge residue), and β7-strand L844 (CS6) on the floor of the adenine pocket (Table 6). The four

Fig. 5. (A) Michael addition reaction. (B) Two steps in the Michael addition.

![Fig. 6. Structures of EGFR (A–D) and BTK (E–G) covalent antagonists. The asterisks indicate the site where the covalent modification with the protein-cysteine thiolate anion occurs.](image)

![Fig. 6. Structures of EGFR (A–D) and BTK (E–G) covalent antagonists. The asterisks indicate the site where the covalent modification with the protein-cysteine thiolate anion occurs.](image)
drugs occupy the front pocket and only neratinib extends into the back pocket. Except for osimertinib, the drugs interact hydrophobically with RS3. Afatinib and neratinib occupy the BP-I-A and BP-I-B subpockets.

5.2. Bruton tyrosine kinase-drug complexes

Ibrutinib is an irreversible pyrazolo[3,4-d]pyrimidine derivative (Fig. 6 E) that inhibits BTK and is FDA-approved for the treatment of six diseases as listed in Table 1 [104]. The X-ray crystal structure shows that the N3 pyrimidine forms a hydrogen bond with the backbone NH group of M477 (the third hinge residue) and the 4-amino group forms hydrogen bonds with the OH group of the threonine gatekeeper residue (T474) and the carbonyl group of E475 (Fig. 7 E). The drug carbonyl group also forms a hydrogen bond with the C481 residue within the hinge-linker immediate before the αD-helix. The acrylamide group forms a covalent Michael addition product with the same C481. The medicinal makes hydrophobic contact with five spine residues (RS2/3 and CS6/7/8), three shell residues (Sh1/2/3), and the KLIFS-3 residue preceding the glycine-rich loop. The ligand also makes hydrophobic contact with G409 (first residue of the glycine-rich loop), the β3-strand A1K-K430, the hinge-linker residues Y476, M477, and C481, the αD-helix N484, S538 (the x of xDFG), DFG-D539, F540, and L542 of the activation segment. The drug occupies the front and back pockets, the intervening gate area, and BP-I-B. Ibrutinib binds to an inactive enzyme with αC out, DFG-D in, and a closed activation segment. Ibrutinib is classified as a type VI inhibitor owing to the formation of a covalent bond with its drug target [92].

Zanubrutinib is an irreversible 4,5,6,7-tetrahydropyrazolo[1,5-a]pyrimidine derivative (Fig. 6 F) that inhibits BTK and is FDA-approved for the treatment of mantle cell lymphoma [105, 106]. The X-ray crystal structure shows that the NH of the pyrimidine forms a hydrogen bond with the carbonyl group of E475 and the carboxamide carbonyl group forms a hydrogen bond with the NH group with this same third hinge residue. The carboxamide NH moiety forms hydrogen bonds with the carbonyl group of E475 and the OH group of the gatekeeper T474 (Fig. 7 F). The drug makes hydrophobic contact with five spine residues (RS2/3 and CS6/7/8), three shell residues (Sh1/2/3), and the KLIFS-3 residue preceding the glycine-rich loop. The antagonist also makes

![Fig. 7. (A-D) Structures of EGFR-drug complexes. (E-F) Structures of BTK-drug complexes. The carbon atoms of the drugs are colored yellow. The dashes represent polar bonds. AS, activation segment; CL, catalytic loop.](Image)

**Table 6**

<table>
<thead>
<tr>
<th>Drug-enzyme</th>
<th>PDB ID</th>
<th>RS1</th>
<th>RS2</th>
<th>RS3</th>
<th>RS4</th>
<th>Sh1</th>
<th>Sh2</th>
<th>Sh3</th>
<th>CS5</th>
<th>CS6</th>
<th>CS7</th>
<th>CS8</th>
<th>KLIFS-3</th>
<th>KLIFS pockets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afatinib-EGFR</td>
<td>4g5j</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>F, G, BP-I-A/B</td>
<td></td>
</tr>
<tr>
<td>Dacomitinib-EGFR</td>
<td>4i24</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>F, G, BP-I-B</td>
<td></td>
</tr>
<tr>
<td>Osimertinib-EGFR</td>
<td>6jxt</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>Neratinib-EGFR</td>
<td>3w2q</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>F, G, B, BP-I-A/B</td>
<td></td>
</tr>
<tr>
<td>Ibrutinib-BTK</td>
<td>5p9j</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>F, G, BP-I-B</td>
<td></td>
</tr>
<tr>
<td>Zanubrutinib-BTK</td>
<td>6j6m</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>Φ</td>
<td>F, G, B, BP-I-B</td>
<td></td>
</tr>
</tbody>
</table>

a Klifs.net.
b KLIFS-3, kinase-ligand interaction fingerprint and structure residue-3.

5.2. Bruton tyrosine kinase-drug complexes

Ibrutinib is an irreversible pyrazolo[3,4-d]pyrimidine derivative (Fig. 6E) that inhibits BTK and is FDA-approved for the treatment of six diseases as listed in Table 1 [104]. The X-ray crystal structure shows that the N3 pyrimidine forms a hydrogen bond with the backbone NH group of M477 (the third hinge residue) and the 4-amino group forms hydrogen bonds with the OH group of the threonine gatekeeper residue (T474) and the carbonyl group of E475 (Fig. 7E). The drug carbonyl group also forms a hydrogen bond with the C481 residue within the hinge-linker immediate before the αD-helix. The acrylamide group forms a covalent Michael addition product with the same C481. The medicinal makes hydrophobic contact with five spine residues (RS2/3 and CS6/7/8), three shell residues (Sh1/2/3), and the KLIFS-3 residue preceding the glycine-rich loop. The ligand also makes hydrophobic contact with G409 (first residue of the glycine-rich loop), the β3-strand AIK-K430, the hinge-linker residues Y476, M477, and C481, the αD-helix N484, S538 (the x of xDFG), DFG-D539, F540, and L542 of the activation segment. The drug occupies the front and back pockets, the intervening gate area, and BP-I-B. Ibrutinib binds to an inactive enzyme with αC out, DFG-D in, and a closed activation segment. Ibrutinib is classified as a type VI inhibitor owing to the formation of a covalent bond with its drug target [92].

Zanubrutinib is an irreversible 4,5,6,7-tetrahydropyrazolo[1,5-a]pyrimidine derivative (Fig. 6F) that inhibits BTK and is FDA-approved for the treatment of mantle cell lymphoma [105, 106]. The X-ray crystal structure shows that the NH of the pyrimidine forms a hydrogen bond with the carbonyl group of E475 and the carboxamide carbonyl group forms a hydrogen bond with the NH group with this same third hinge residue. The carboxamide NH moiety forms hydrogen bonds with the carbonyl group of E475 and the OH group of the gatekeeper T474 (Fig. 7F). The drug makes hydrophobic contact with five spine residues (RS2/3 and CS6/7/8), three shell residues (Sh1/2/3), and the KLIFS-3 residue preceding the glycine-rich loop. The antagonist also makes...
hydrophobic contact with the β3-strand AIK-K430, the gatekeeper T474, hinge-linker residues Y476, M477, G480, C481, and α-helix residues L483 and N484, catalytic loop residue R525, S538 (the x of xDFG), DFG-D539, F540, and L542 of the activation segment. The drug forms a covalent linkage with C481 at the end of the hinge-linker segment. Like ibrutinib, zanubrutinib occupies the front and back pockets, the intervening gate area, and BP-I-B. The drug binds to an inactive enzyme with αout, DFG-Dua and a closed activation segment. Zanubrutinib is classified as a type VI inhibitor owing to the formation of a covalent bond with its medicinal target [92]. Unfortunately, we lack an X-ray crystal structure of acalabrutinib with BTK. See Refs. [107,108,109] for synonyms on the use and efficacy of acalabrutinib.

6.0. Epilogue

Covalent drug inhibitors have several potential advantages among them being high potency, effectiveness at low doses (100 ng or less daily), and prolonged inhibition that requires de novo protein biosynthesis to restore target enzyme action [110]. Additionally, targets for such drugs include enzymes, receptors, and proteins with shallow binding sites not responsive to conventional approaches. This potential to inhibit targets previously thought to be undruggable owing to the lack of an appropriate small-molecule binding pocket has been applied to the development of inhibitors of the KRAS protein (Kirsten rat sarcoma viral oncogene homolog).

Mutations of RAS are the most prevalent oncogenic alteration in human cancers [111,112]. RAS proteins are small 21 kDa GTP-binding proteins that participate in the intracellular signaling modules following receptor protein-tyrosine kinase activation. KRAS mutations are the most prevalent followed by NRAS and those of HRAS are comparatively uncommon [113]. More than 80% of pancreatic cancers and more than 30% of colorectal, cholangial, and lung carcinomas harbor activating mutations of the KRAS gene. Exon 2 mutations are the most common KRAS alteration and involve the conversion of glycine-12 to aspartate, valine, alanine, serine, arginine, aspartate, or cysteine. The KRAS G12C mutation occurs in about 35% of lung and ovarian carcinomas, 20% of endometrial and cholangial cancers, and 5% of colorectal carcinomas. KRAS mutations favor the GTP-bound active state that promotes downstream effects including cell proliferation and survival.

The lack of an ideal small molecule binding pocket in the KRAS protein and its high affinity towards GTP makes the design of specific competitive small molecule drugs challenging. The KRAS G12C mutation is potentially druggable by its binding to a pocket contiguous with the GTP-binding site and locking it into an inactive GDP-bound state. One strategy to inhibit the KRAS mutant is to use a GDP-derived inhibitor to covalently lock the protein in the inactive GDP-bound state by targeting the G12 thiol group as a covalent bond. For generating covalent linkages to target proteins. Functional groups that undergo such reactions include acrylamides such as osimertinib, ibrutinib, and zanubrutinib, β-substituted acrylamides including afatinib, dacomitinib, and neratinib, and alkynyl amides such as acalabrutinib (Fig. 6). The addition reactions of such warheads with the cysteine thiol group yield a new covalent carbon-sulfur bond with the concomitant generation of a thioether. Jackson et al. found that substituents at the α- or β-positions of the Michael acceptor warhead can influence the rate of thiol addition [116]. For example, electron withdrawing groups at the α-position increase the rate of addition. In contrast, electron donating groups such as electron-rich aryl groups or alkyl groups at the α- or β-position reduce the rate of thiol addition relative to unsubstituted Michael acceptors.

In the addition to the Michael acceptors, aziridines, epoxides, vinyl sulfones, activated acetylenes, and α-haloacetones are warheads that serve as irreversible inhibitors that react with cysteiny1 residues [117,118]. In contrast, activated ketones, α-ketoamides, cyanamides, α-cyanoacrylamides, carbonitriles, aldehydes, α-ketoheterocycles, and boronic acid derivatives are warheads that function as reversible covalent protein-cysteine inhibitors. Warheads have been designed to react with residues other than cysteine. For example, vinyl sulfones and vinyl sulfonamides form adducts with protein-lysines and cysteines while sulfonoryl fluorides, sulfonimidoyl fluorides, and aryl fluorosulfates form adducts with protein lysines and tyrosines. Additionally, N-Acyl-N-alkyl sulfonamides react with surface-exposed protein-lysines and 2-carbonylarylboronic acids are reversible inhibitors that attack protein-lysines. N-methyl isoazoxolium derivatives react with protein-glutamates, aspartates, and cysteine while oxaziridine compounds react with protein-methionines. See Ref. [118] for a comprehensive list and review of warhead classes, structures, and properties.

One of the possible adverse effects of covalent inhibitors is the potential of such drugs to produce immune-mediated adverse reactions triggered by modification of target or off-target proteins. Although the drug and its protein target separately may not be immunogenic, the combination becomes immunogenic by a process called haptenization where the covalently attached drug is the hapten. In rare instances, drug–protein adducts are immunogenic and trigger idiosyncratic reactions, often weeks or months after initiation of therapy [117]. The mechanisms responsible for adverse immune responses remain ill-defined and such toxicities are unpredictable. Despite issues related to the formation of potentially adverse immunogenic drug-protein covalent complexes, the development of successful covalent-based therapeutics has led to the identification of certain medicinal properties that minimize the risk to covalent drugs. These properties include (i) high selectivity in target engagement allowing for high potency at low doses and (ii) minimizing reactive metabolite formation from cytochrome P450-mediated reactions [119].

In an early assessment of the protein kinase cysteinome, Liu et al. plotted targetable cysteine residues in and around the ATP-binding site [110]. These investigators described six cysteinyl targets including a residue (i) at the hinge-linker found in the EGFR family members ErbB1/2/4 (residues that correspond to EGFR C797), Blk (a Src family member), JAK3, MKK7, and the TEC family including BTK, BMX, TEC, TXK, and ITK, (ii) within the glycine-rich loop of FGFR, (iii) in the roof of the pocket found in the NEK2 and RSK enzymes, (iv) in the solvent area found in JNKs, (v) immediately preceding the DFG of VEGFR, ERK2, and GSK3β, and (vi) in the catalytic loop of PDGFR and BCR. Cysteines are good targets because they compose only about 2% of protein amino acid residues [120] (decreasing the possibility of adventitious reactions) and because they are not conserved in the protein kinase family as are the HRD, DFG, or the APE signatures.

More recently, Zhao and Bourne studied potential cysteinyl targets in the protein kinase kinome [121]. They identified targets in 12 regions in different protein kinases including the (i) front pocket, (ii) the glycine-rich loop, (iii) the x residue of xDFG, (iv) DFG + 2, (v) the gatekeeper, (vi) the gatekeeper + 1, (vii) two residues prior to HRD of
the catalytic loop, (viii) the roof of the ATP-binding site, (ix) a lysine within the roof, (x) the activation loop, (xi) remote cysteines, and (xii) the extended front pocket. Additionally, Leopoldt et al. have reported on the strategy and design of preparing selective covalent protein kinase inhibitors [122]. Generally, the target residues in protein kinases are not evolutionarily conserved and this property adds to the selectivity of the antagonist, a property that can potentially limit drug toxicities. Covalent inhibitors have emerged from the ranks of drugs to be avoided to become an emerging paradigm. Much of this recent success can be attributed to the clinical efficacy of ibritinib [6] as well as the other antagonists covered in this review. Moreover, the covalent inhibitor methodology is swiftly gaining acceptance as a valuable component of the medicinal chemist’s toolbox and is primed to make a significant impact on the generation of enzyme antagonists and receptor modulators [117].

Declaration of Competing Interest

The author is unaware of any affiliations, memberships, or financial holdings that might be perceived as affecting the objectivity of this review.

Acknowledgments

The colored figures in this paper were evaluated to ensure that their perception was accurately conveyed to colorblind readers [123]. The author thanks Laura M. Roskoski for providing editorial and bibliographic assistance. I also thank Josie Rudnicki and Jasper Martinsek for their help in preparing the figures and Pasha Brezina and W.S. Sheppard for their help in structural analyses.

References


J.A. Engelman, K. Zeijnulfah, C.M. Gale, E. Lifshits, A.J. Gonzales, T. Shimamura, et al., PFS029904, an irreversible pan-ERBB inhibitor, is effective in lung cancer models with ERBB2 and EGFR mutations that are resistant to gefitinib, Cancer Res. 67 (2007) 11194–11192.


