A Common Protein Fold Topology Shared by Flavonoid Biosynthetic Enzymes and Therapeutic Targets

Bernadette M. McArdle, Marc R. Campitelli, and Ronald J. Quinn*

Natural Product Discovery, Eskitis Institute, Griffith University, Brisbane, Queensland, 4111, Australia

Received June 27, 2005

The relationship between a natural product's biosynthetic enzyme and its therapeutic target is unknown. The concept of protein fold topologies, as a determining factor in recognition, has been developed through molecular modeling techniques. We have shown that biosynthetic enzymes and the therapeutic targets of three classes of natural products that inhibit protein kinases share a common protein fold topology (PFT) and cavity recognition points despite having different fold type classifications. The clinical agent flavopiridol would have been identified by this new approach.

Natural products are a source of therapeutically useful compounds, with about 40% of the chemical scaffolds displayed in natural products being absent in synthetic compounds.[1] The accepted argument for the wide ranging biological activity of these compounds is that in the course of their interaction with enzymes during their biosynthesis they have incorporated an imprint or "understanding" of biological space within their architecture.[2] This imprint has previously been associated with the overall fold of a protein following a number of cases demonstrating a given natural product's preference to bind to proteins within the same fold classification.[2] Classification at the fold type level is based upon the arrangement of major secondary structures and topological connections and interactions. More recently the association between folds and natural products has been refined with the utilization of overall structural similarity of the ligand-sensing core, guiding the development of compound libraries based on natural product starting points.[3,4] While natural products are shown to be exceptionally biologically relevant and bind with a degree of individual preference to proteins sharing a particular fold type or ligand-sensing core, it is not known if the interactions made during their synthesis with their associated biosynthetic enzymes determine this preference. If a relationship between natural product recognition of biosynthetic enzyme and therapeutic targets could be established, then this would open up a new approach to drug design. Toward this end, we examined the structural relationship between a demonstrated therapeutic target (represented by protein kinases) and the enzymes utilized in the biosynthesis of several natural products that are inhibitors of kinases.

Protein kinases are a large family of proteins that modulate the levels of cellular protein phosphorylation. Because of the crucial role they play in signal transduction pathways, they represent a valid medicinal target for treatment of diseases including cancer, diabetes, Alzheimer's, and inflammation.[5] Understanding the requirements of a protein kinase inhibitor has been greatly aided by the number of solved three-dimensional structures. Presently in the Protein Data Bank (PDB)[6] there is an excess of 200 X-ray crystal structures of protein kinases that have been classified according to their fold as described in the Structural Classification of Proteins (SCOP) database.[7] It is no surprise that as a consequence of their high structural homology, all protein kinase structures are classified as exhibiting the protein kinase-like fold.

The majority of compounds that inhibit protein kinases, including members of three of the classes of natural products, flavonoids, chalcones, and stilbenes (Supporting Information), compete with the natural substrate ATP for its binding site. These classes of natural products are structurally related, and the enzymes, which are responsible for their biosynthesis, reflect this similarity through a number of structural and functional relationships. The primary relationship is their proximity to each other within the overall biosynthetic pathway of flavonoids (Figure 1).[8]

The availability of crystal structures of the biosynthetic enzymes is limited. Only chalcone synthase (CHS), chalcone isomerase (CHI), and anthocyanidin synthase (ANS) have resolved crystal structures. However, each of the remaining biosynthetic enzymes is structurally related to one of these. As a consequence of substrate, mechanistic, sequence, and structural similarities,[9] CHS is also representative of stilbene synthase (STS) and has been classified by SCOP as exhibiting the thiolase-like protein fold. The structure of ANS is classified by SCOP as the double-stranded beta-helix fold and is representative of the other 2-oxoglutarate-dependent oxygenases flavonol synthase (FLS), flavone synthase I (FNS I), and flavanone 3-hydroxylase (FHT) belonging to the same family and sharing other varying degrees of similarity relating to sequence, mechanism, and substrate selectivity.[8,10−14] CHI is not representative of any other biosynthetic enzymes and has the chalcone isomerase fold according to SCOP. Using the available structural information, comparisons can be made between binding of chalcone-, stilbene-, and flavonoid-based natural products in their biosynthetic enzymes and also in their therapeutic target protein kinases.

Results and Discussion

Each of the crystallized biosynthetic enzymes has been cocrystallized with at least one of their natural substrates or products. Docking using Glide was performed to provide a complete dataset of natural product and natural substrate binding to biosynthetic enzymes for analysis. For comparison with the therapeutic target protein kinases, relevant protein kinase/natural product inhibitor complexes, attained either through the PDB[15−18] or by computational docking techniques, were used.

A comparison between the orientations of the natural substrates located within the enzyme cavities proved informative. Within the individual biosynthetic enzymes the same relative orientation is exhibited by the different natural products; however, this is not the case for orientations of the compounds within the therapeutic target. Within the protein kinase-like fold of protein kinases the crystal structure orientations of myricetin and two conformations of quercetin (Figure 2) are differently orientated. This result was also apparent through docking of resveratrol and butein, in which two groups of orientations were observed (not shown in figure). In both cases a particular interaction with the nitrogen of the backbone of the β-sheet bridging the top of the cavity is conserved (Val882 in phosphoinositide 3-kinase (PI3K)). Indeed the equivalent of this interaction is virtually universal among protein kinase-inhibitor complexes and exists in all but two such crystal structures reported in the protein data bank.[19] This conserved interaction is made by

* To whom correspondence should be addressed. Tel: (07)-3735-6006. Fax: (07)-3735-6001. E-mail: R.Quinn@griffith.edu.au.

10.1021/np050229y CCC: $33.50
Published on Web 12/30/2005

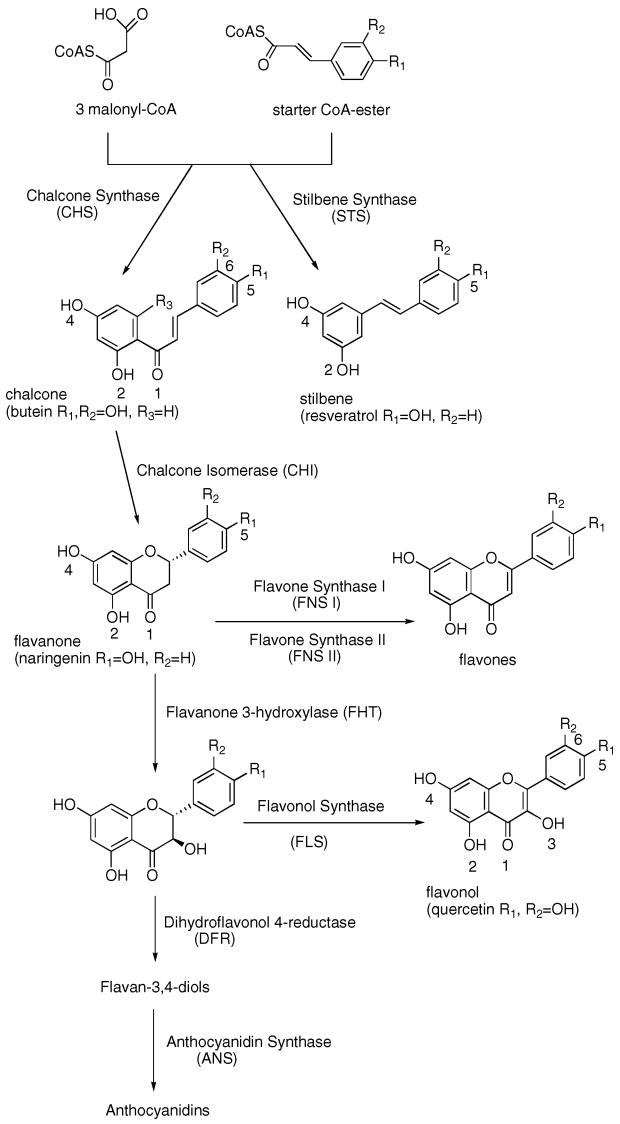

Figure 1. Biosynthetic pathway of the natural products chalcones, stilbenes, and flavonoids. The general structure of chalcones and stilbenes and structures of different types of flavonoids of interest are shown. For the structures of two representative flavonoids (quercetin and naringenin), and a representative stilbene (resveratrol) and chalcone (butein), corresponding recognition sites are numbered 1 through 6.

either position 1/2 or 5/6 on the natural product regardless of class (Figure 1). Upon examination of conserved interactions within the individual biosynthetic enzymes (yellow ball in Figure 2), an interesting observation was made: in CHS the conserved interaction is with the oxygen of the backbone of Gly216, in CHI a hydrogen bond is consistently observed to the hydroxyl of the side chain of Thr48, and in ANS the natural products are always hydrogen bonded with water-601, which is itself coordinated by iron. The position of the natural products making these conserved interactions within the individual biosynthetic enzymes is also either always 1/2 or always 5/6. This seems to indicate that while in the biosynthetic enzyme only one of the two types of orientations is commonly accommodated, the cavity of protein kinases however can accommodate both types. This observation may also explain the preference of individual biosynthetic enzymes for chalcones, stilbenes, flavanones, or flavonols and the absence of this discrimination by protein kinases.

As well as the conserved interaction described above being exhibited in all biosynthetic enzymes and protein kinases, several other interactions exhibited by the natural products were analyzed.

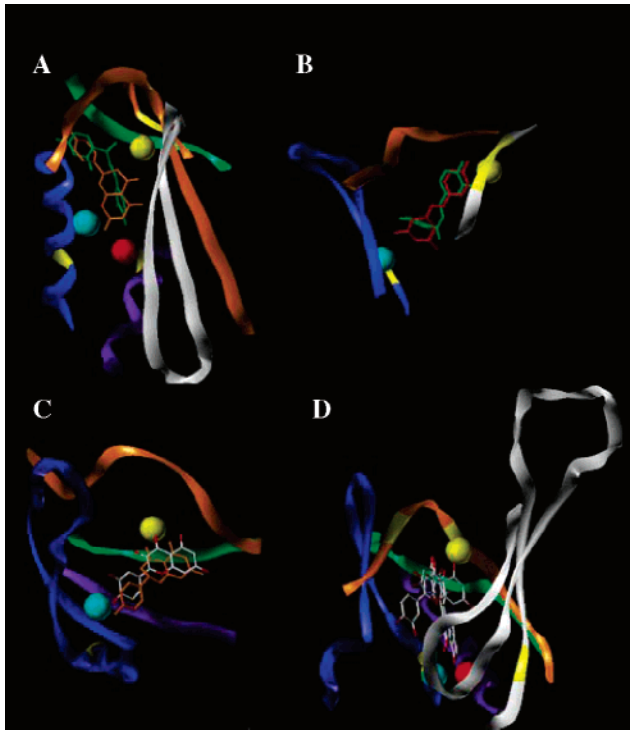

Figure 2. Flavonoid/kinase protein fold topology (PFT). The equivalent points of recognition and PFT shared between protein kinases and the biosynthetic enzymes of chalcones, stilbenes, and flavonoids are shown. For the biosynthetic enzymes CHI (A), CHS (B), and ANS (C), docked and co-crystallized natural product substrates are colored by class. The stilbene resveratrol is shown in red, the chalcone butein in green, the flavanone naringenin in orange, and the flavonol quercetin colored by atom type. In the protein kinase (PI3K) (D), the conformations of crystallized flavonoids quercetin and myricetin are shown colored by atom type. Docked conformations in the protein kinase are omitted for clarity. The yellow ball represents an interaction conserved in all cases. The fold around the binding site of the biosynthetic enzymes and the representative fold around the binding site of the protein kinase PI3K are shown colored by spatial equivalence. Portions colored equivalently represent common features of the shared PFT. The yellow portions of the fold represent the backbone of the residue repeatedly observed to interact with the bound natural products. Spatially equivalent points of interaction are illustrated by the balls and colored by equivalence.

A three-dimensional alignment of these observed interactions by initial alignment of the previously described conserved interaction highlighted other spatially equivalent recognition centers. Comparison of all three representative biosynthetic enzymes and the protein kinases highlighted another relatively conserved recognition point, and this is represented by the cyan ball in Figure 2. For CHS this was the oxygen of the backbone of Asp255, which interacts with position 2/4 of resveratrol (these positions are equivalent) in the crystal structure 1cgz. In CHI the amine of the side chain of Asn113 hydrogen bonds with position 4 of naringenin in the crystal structure 1eyq. In ANS, quercetin in the PDB entry 1gp6, and docked naringenin, hydrogen bond with the amine of the side chain of Gln117 both at position 5. The corresponding position in the protein kinase PI3K is the oxygen of the side chain of Asp964, which interacts with position 5 of quercetin in 1e8w and position 2 of myricetin in 1e90. The equivalent position in Src kinase and EGFR also interacts with docked butein at position 6 and resveratrol at positions 2/4, respectively. An additional recognition point is shared between CHI and protein kinases, shown by the red ball in Figure 2. For CHI the hydroxyl of the side chain of Thr190 interacts with position 5 of butein. The corresponding nitrogen of the side

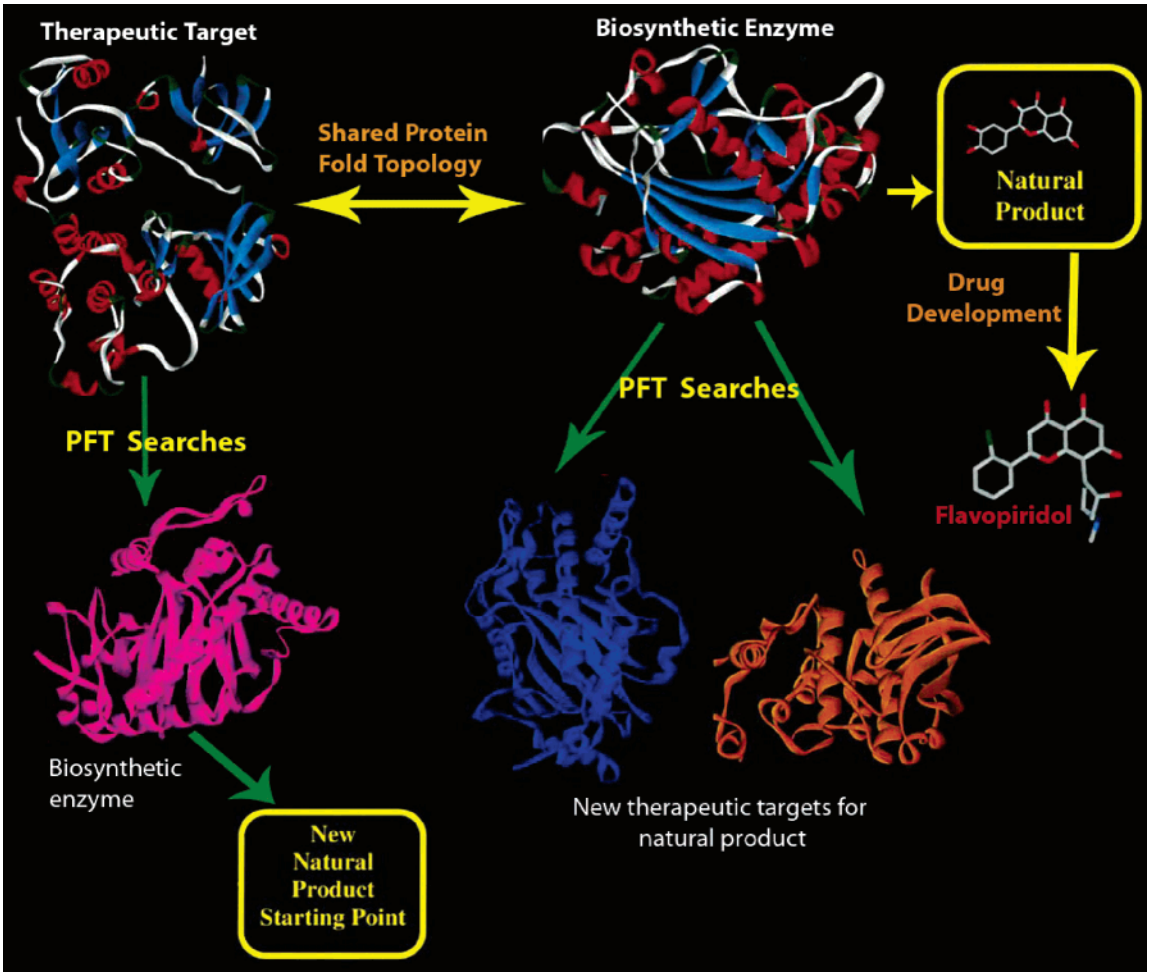

Figure 3. Potential application and validation of protein fold topologies (PFTs) to discovering new therapeutic targets and natural product starting points. The pathway shown in yellow demonstrates how the PFT shared by protein kinases and the biosynthetic enzymes of flavonoids would have resulted in the development of flavopiridol without any previous knowledge of inhibition. The pathways shown in green demonstrate the potential for using shared PFTs as an approach for identifying both new therapeutic targets from natural product/biosynthetic enzyme information and biosynthetic enzymes whose associated natural products could then be used as starting points for the development of drugs aimed toward a target.

chain of Lys833 in PI3K interacts with position 5/6 of quercetin, position 1/2 of myricetin, and position 6 of butein. These cavity recognition points are spatially equivalent and are recognized by only two different types of natural product substrate orientations.

Following the three-dimensional alignment of cavity recognition points, a visual examination of the corresponding fold surrounding the binding site revealed startling similarities. While according to SCOP the fold classifications of CHS, CHI, ANS, and protein kinases are quite different, inspection of the resulting fold alignment enabled the observation of a similar topological arrangement of different secondary structures, or protein fold topology (PFT) (Figure 2). Across all fold types an equivalent portion of β-sheet bridges the top of the cavity with the exception of CHS, for which this portion is less uniform. On the left-hand side of the cavity for CHS and ANS the β-sheet and loop are evident, which can be seen in the protein kinase-like fold of PI3K. In CHI this is substituted by a spatially equivalent α-helix. CHS and CHI also share portions of β-sheet and loop arrangements to the right of the cavity with protein kinase PI3K. Underneath the cavity an α-helix is evident in the protein kinase-like fold, which is replicated in CHI, while being substituted by a spatially equivalent β-sheet in ANS. Finally, CHI and ANS both share the presence of a β-sheet spanning the back of the cavity with the protein kinase-like fold exhibited by PI3K. The reduced number of similarities in PFT shared between CHS and protein kinases may be indicative of the generic nature of CHS and STS, being near the root of the biosynthetic pathway. The described flavonoid/kinase PFT shared by individual biosynthetic enzymes and protein kinases defines a cavity equally recognizable to the natural products.

This correlation is likely to be of value in two different aspects of drug design. As depicted in Figure 3, the defining of PFTs would allow for the identification of targets by analyzing all the potential targets for similarity to the PFT associated with a natural product's biosynthetic enzyme, or identify new natural product starting points through identifying biosynthetic enzymes that have similarity to the PFT of the target. The validity of this approach is demonstrated by flavopiridol, the first potent inhibitor of cyclin-dependent kinases. Flavopiridol (Figure 3) was synthetically derived from the plant alkaloid rohitukine isolated from the Indian plant *Dysoxylum binectariferum*[20] and is presently progressing through clinical trials.[21] Knowledge of binding ability of flavonoids to protein kinases has played an important role in considering equivalency between structural features (α-helix, β-sheet, and loops) that otherwise would not be regarded as equivalent. Once this equivalence is established, the correlation between common PFTs shared between flavonoid biosynthetic enzymes and kinases and subsequent docking studies would have led to the development of flavopiridol de novo.

The flavonoid/kinase protein fold topology defined in this study will lead to the identification and definition of protein fold topologies for other biosynthetic enzyme/therapeutic target pairs. In addition to demonstrating the "knowledge" stored within the

architecture of natural products and their ability to recognize biological space, the implications toward drug design, protein structure and folding, and protein fold classification are broad reaching. This establishes that the recognition of enzymes by natural products during their biosynthesis is translated to recognition of drug targets with corresponding PFT.

Experimental Section

Computational Docking Techniques. To provide a complete dataset of natural product and natural substrate binding to both biosynthetic enzymes and the therapeutic target protein kinases, docking was undertaken of the flavonol quercetin, flavanone naringenin, chalcone butein, and stilbene resveratrol structures (Figure 1) into the available crystal structures[17,18,22−24] of their biosynthetic enzymes and therapeutic target using the program Glide.[25] Specifically, in addition to the re-docking of the already available complexes, the chalcone butein was docked into CHS (1cgk),[22] CHI (1eyq),[23] and protein kinase EGFR (1m17).[17] The stilbene resveratrol was docked into the Src protein kinase (1byg),[18] and the flavanone naringenin was docked into ANS (1gp6).[24]

The preparation for docking using Glide involved using all crystal structure files in PDB format in Maestro, the graphical user interface for all Schrödinger's products including MacroModel[26] and Glide. All the original PDB entries had waters and dummy atoms removed, no hydrogens added, and the atom and bond types of the natural product corrected. The protein preparation in the Maestro interface was performed on the combined protein−ligand complexes and was set to output separate protein and ligand entries. The ligand was prepared further through minimization using the application MacroModel, where all default settings were used. Grid files were prepared using the original natural products to define the grid box centers for each of the proteins and were used in the Glide application for docking. During docking amide bond rotations were allowed and flexible docking was performed to result in a single solution for each docking experiment.

The docked conformations of the compounds in their respective enzymes were validated by comparison with the available experimentally derived crystal structure complexes.

Conserved interactions were identified through a hydrogen-bonding analysis using a designed script. The script finds all heavy atom pairs (nitrogen and oxygen) within 3.9 Å and reports the PP−P−L and LL−L−P angles (PP is protein heavy atom neighboring atom, P is protein heavy atom, L is ligand heavy atom, etc). Donor/acceptor pairs were manually checked such that atom types defined as N.am (nitrogen amide), N.pl3 (nitrogen trigonal planar), and N.4 (nitrogen sp^3 positively charged) could act as donors, O2 (oxygen sp^2), N1 (nitrogen sp), Oco2 (oxygen in carboxylates and phosphates), and N.ar (nitrogen aromatic) could act as acceptors, and O3 (oxygen sp^3), N2 (nitrogen sp^2), and N.3 (nitrogen sp^3) could be both donors and acceptors. Atom definitions are those used in SYBYL, which are required for GOLD. For backbone amides and planar amides (N.am and N.pl3) the donor−hydrogen−acceptor angle was manually checked and only recorded if greater than 90°, and the hydrogen/acceptor distance was checked and discarded if much greater than 3 Å. The final list of hydrogen bonds for a given protein−ligand complex satisfies the criteria of donor/acceptor distance less than 3.9 Å, acceptor atom neighbor−acceptor−donor angle greater than 90°, appropriate donor/acceptor pair, and appropriate directionality and distance for backbone and planar amides. Using this script all potential hydrogen bonds were identified and all tautomers of a ligand are considered.

All protein alignments were performed manually aided by visualization in 3D.

Acknowledgment. B.M.M. acknowledges the award of an Australian Postgraduate Award.

Supporting Information Available: A referenced table of protein kinase inhibition by chalcones, stilbenes, and flavonoids. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Henkel, T.; Brunne, R. M.; Müller, H.; Reichel, F. *Angew. Chem., Int. Ed.* **1999**, *38*, 643−647.
(2) Breinbauer, R.; Vetter, I. R.; Waldmann, H. *Angew. Chem., Int. Ed.* **2002**, *41*, 2879−2890.
(3) Koch, M. A.; Wittenberg, L. O.; Basu, S.; Jeyaraj, D. A.; Gourzoulidou, E.; Reinecke, K.; Odermatt, A.; Waldmann, H. *Proc. Natl. Acad. Sci.* **2004**, *101*, 16721−16726.
(4) Koch, M. A.; Waldmann, H. *Drug Discovery Today* **2005**, *10*, 471−483.
(5) Hopkins, A. L.; Groom, C. R. *Nat. Rev. Drug Discovery* **2002**, *1*, 727−730.
(6) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(7) Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C. *J. Mol. Biol.* **1995**, *247*, 536−540.
(8) Forkmann, G.; Heller, W. In *Comprehensive Natural Products Chemistry*; Barton, D., Nakanishi, K., Ed.; Elsevier: Oxford, 1999; Vol. 1, Chapter 26, pp 713−748.
(9) Schröder, J. In *Comprehensive Natural Products Chemistry*; Barton, D., Nakanishi, K., Eds.; Elsevier: Oxford, 1999; Vol. 1, Chapter 27, pp 749−771.
(10) Turnbull, J. J.; Sobey, W. J.; Aplin, R. T.; Hassan, A.; Schofield, C. J.; Firmin, J. L.; Prescott, A. G. *Chem. Commun.* **2000**, 2473−2474.
(11) Martens, S.; Forkmann, G.; Britsch, L.; Wellmann, F.; Matern, U.; Lukacen, R. *FEBS Lett.* **2003**, *544*, 93−98.
(12) Lukacin, R.; Wellmann, F.; Britsch, L.; Martens, S.; Matern, U. *Phytochemistry* **2003**, *62*, 287−292.
(13) Prescott, A. G.; Stamford, N. P. J.; Wheeler, G.; Firmin, J. L. *Phytochemistry* **2002**, *60*, 589−593.
(14) Welford, R. W. D.; Turnbull, J. J.; Claridge, T. D.; Prescott, A. G.; Schofield, C. J. *Chem. Commun.* **2001**, 1828−1829.
(15) Sicheri, F.; Moarefi, I.; Kuriyan, J. *Nature* **1997**, *385*, 602−609.
(16) Walker, E. H.; Pacold, M. E.; Perisic, O.; Stephens, L.; Hawkins, P. T.; Wymann, M. P.; Williams, R. L. *Mol. Cell* **2000**, *6*, 909−919.
(17) Stamos, J.; Sliwkowski, M. X.; Eigenbrot, C. *J. Biol. Chem.* **2002**, *277*, 46265−46272.
(18) Lamers, M. B.; Antson, A. A.; Hubbard, R. E.; Scott, R. K.; Williams, D. H. *J. Mol. Biol.* **1999**, *285*, 713−725.
(19) Breitenlechner, C.; Gassel, M.; Hidaka, H.; Kinzel, V.; Huber, R.; Engh, R. A.; Bossemeyer, D. *Structure* **2003**, *11*, 1595−1607.
(20) Naik, R. G.; Kattige, S. L.; Bhat, S. V.; Alreja, B.; de Souza, N. J.; Rupp, R. H. *Tetrahedron* **1988**, *44*, 2081−2086.
(21) Shapiro, G. I. *Clin. Cancer Res.* **2004**, *10*, 4270s-4275s.
(22) Ferrer, J. L.; Jez, J. M.; Bowman, M. E.; Dixon, R. A.; Noel, J. P. *Nature Struct. Biol.* **1999**, *6*, 775−784.
(23) Jez, J. M.; Bowman, M. E.; Dixon, R. A.; Noel, J. P. *Nature Struct. Biol.* **2000**, *7*, 786−791.
(24) Wilmouth, R. C.; Turnbull, J. J.; Welford, R. W. D.; Clifton, I. J.; Prescott, A. G.; Schofield, C. J. *Structure* **2002**, *10*, 93−103.
(25) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, 47, 1739−1749.
(26) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caulfield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440−467.
(27) Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97−179.
(28) McDonald, I. K.; Thornton, J. M. *J. Mol. Biol.* **1994**, *238*, 777−793.
(29) Thornton, J. M.; MacArthur, M. W.; McDonald, I. K.; Jones, D. T.; Mitchell, J. B. O.; Nandi, L.; Price, S. L.; Zvelebil. M. J. J. M. *Philos. Trans. R. Soc. London A* **1993**, *345*, 113−129.

NP050229Y