CHEMMEDCHEM

DOI: 10.1002/cmdc.200700075

Functional Classification of Protein Kinase Binding Sites Using Cavbase

Daniel Kuhn,[a] Nils Weskamp,[a, b] Eyke Hüllermeier,[b] and Gerhard Klebe*[a]

Increasingly, drug-discovery processes focus on complete gene families. Tools for analyzing similarities and differences across protein families are important for the understanding of key functional features of proteins. Herein we present a method for classifying protein families on the basis of the properties of their active sites. We have developed Cavbase, a method for describing and comparing protein binding pockets, and show its application to the functional classification of the binding pockets of the protein family of protein kinases. A diverse set of kinase cavities is mutually compared and analyzed in terms of recurring functional recognition patterns in the active sites. We are able to propose a relevant classification based on the binding motifs in the active sites. The obtained classification provides a novel perspective on functional properties across protein space. The classification of the MAP and the c-Abl kinases is analyzed in detail, showing a clear separation of the respective kinase subfamilies. Remarkable cross-relations among protein kinases are detected, in contrast to sequence-based classifications, which are not able to detect these relations. Furthermore, our classification is able to highlight features important in the optimization of protein kinase inhibitors. Using small-molecule inhibition data we could rationalize cross-reactivities between unrelated kinases which become apparent in the structural comparison of their binding sites. This procedure helps in the identification of other possible kinase targets that behave similarly in "binding pocket space" to the kinase under consideration.

Introduction

The reversible phosphorylation of proteins is a central control and regulation mechanism in cell growth, signaling, and regulation. Kinases catalyze the transfer of a terminal phosphate group from adenosine triphosphate (ATP) to serine, threonine, or tyrosine residues of other proteins. They account for 1.7% of all genes found in the human genome[1] and constitute a densely populated protein family. Protein kinases form a large superfamily of proteins. Their catalytic domains are related by sequence and folding similarities (protein kinase fold); nevertheless, they exhibit a rich diversity of regulation modes and substrate specificities.[2–4] The ATP binding site is located at the interface between two subdomains (lobes) forming the specific kinase fold. Interference with protein-kinase-mediated cell-signaling pathways is involved in a variety of diseases including cancer, neurological diseases, and inflammation.[5–7] Consequently, protein kinases have emerged as attractive drug targets for pharmaceutical research. In 2001, the first orally administered kinase inhibitor (imatinib, Gleevec) was approved as a potent agent against chronic myelogenous leukemia (CML), proving that the structurally conserved ATP pocket can be selectively addressed by small molecules. Currently, a vast number of medicinal chemistry projects are being followed in pharmaceutical research to target the ATP binding site of different kinases with ATP-competitive inhibitors.[7–12] Accordingly, the elucidation of selectivity-discriminating features exposed at the catalytic sites of various protein kinases is a fundamental prerequisite for the development of selective inhibitors,[13] and provides a novel taxonomy for the classification of kinases.

Similarity between proteins can be described using different types of information. The most straightforward approach is the classification of protein kinases based on sequence similarity. Manning et al. presented a classification of 518 non-redundant kinase genes in the human genome (human kinome),[1] providing an extensive sequence-based annotation of protein kinases.

A classification focusing on both sequence and structural information was performed by Cheek and co-workers[14,15] for all available sequences of phosphate-transferring proteins. The authors clustered over 59400 kinase sequences (also considering kinases acting on non-protein substrates) into 25 distinct families based on sequence similarities. Analysis of the 22 sequence families (covering 98.8% of all sequences) for which structural information exists reveals that they fall into ten general fold groups, which include the most widespread protein folds (e.g., Rossmann, ribonuclease H, or ferredoxin fold). The protein kinase fold represents one fold subgroup.

A similarity analysis of protein kinases based on folding similarities was presented by Shindyalov and Bourne.[16] The authors used a fold comparison algorithm (combinatorial exten-

[a] *Dr. D. Kuhn, Dr. N. Weskamp, Prof. Dr. G. Klebe*
Department of Pharmaceutical Chemistry
University of Marburg
Marbacher Weg 6, 35032 Marburg (Germany)
Fax: (+49) 6421-28-28994

[b] *Dr. N. Weskamp, Prof. Dr. E. Hüllermeier*
Department of Mathematics and Computer Science
University of Marburg
Hans-Meerwein-Straße, 35032 Marburg (Germany)

Supporting information for this article is available on the WWW under http://www.chemmedchem.org or from the author.

sion (CE)) to align structurally similar regions of different kinase structures.

Several methods have been presented that recognize similarities in the folding pattern of protein domains and use these to build up a hierarchical classification. Two prominent examples of this strategy are SCOP (structural classification of proteins)[17] and CATH.[18] Both methods apply automatic sequence and fold comparison techniques; however, they involve manual annotation. In the SCOP hierarchy, proteins are classified at three major levels, corresponding to an increasing degree of similarity: fold (major structural similarity), superfamily (low sequence similarity, but proteins exhibiting functional and structural similarity), and family (sequence similarity greater than 30%). By comparison, CATH clusters proteins at four levels (class (C), architecture (A), topology (T) and homologous superfamily (H)). These levels reflect the secondary-structure composition and topology.

Besides the classification of proteins based on sequences or secondary structure elements, several approaches focus on the 3D structure or the physicochemical properties of the binding sites.[19] Whereas some of the approaches use predefined three-dimensional templates composed of several amino acids (motifs) to query protein databases,[20–26] other methods operate independently of any reference template mutually comparing entire protein structures or predefined regions of interest.[27–35]

Naumann and Matter performed a structural classification of protein kinases into subfamilies exhibiting similar protein–ligand interactions, based on their crystal structures.[36] In an iterative superposition procedure, structurally conserved residues were used to obtain a spatial alignment of the kinase structures. Putative binding properties were mapped using selected probes of the GRID force field (N1, O, and DRY, representing hydrogen bond donor, hydrogen bond acceptor, and hydrophobic interaction properties, respectively).[37,38] The 3D binding site information is encoded in the obtained GRID molecular interaction fields. A principal components analysis (PCA) and consensus PCA[39] were used for the analysis of molecular interaction fields to extract features common to the different kinase binding sites, with particular focus on the identification of those areas that differ among the subfamilies. By this method (also called "target family landscapes"), the authors were able to classify the considered kinases into different subfamilies and to extract the structural features that are important for a particular kinase subfamily. Additionally, 3D QSAR studies based on CDK1 using purine-based inhibitors revealed essential structural elements required for potent inhibition across this family.[36]

A classification of protein kinases based on small-molecule inhibition data retrieved from the literature was performed by Vieth and co-workers.[40,41] They used structural similarity and binding profiles of small molecules for the classification of protein kinases, and compared this classification with a sequence-based classification scheme. The authors found that for highly homologous kinases close in sequence space, on average, the inhibition profiles of small-molecule inhibitors correspond closely. However, they were also able to identify cases where both classifications suggest distinct clustering.

Recently, experimental approaches have identified new protein targets for small-molecule kinase inhibitors. These experimental results can be used to establish interaction profiles of both small molecules and protein kinases. Affinity chromatography using immobilized kinase inhibitors, in combination with mass spectrometry, allowed detection of previously unknown kinase targets on a proteome-wide scale.[42,43] Another experimental approach using protein kinases expressed by phage display methods in a competition binding assay allowed the establishment of interaction profiles for 20 kinase inhibitors used in clinical studies with respect to 119 kinases.[44] Based on the interaction profiles, a hierarchical cluster analysis was performed, which elucidated family-wide interrelationships across kinases. Obviously this type of experimental profiling of kinases allows the identification of previously unknown targets of given kinase inhibitors and of cross-reactivity with other kinases.

In this paper we address the question of whether such cross-reactivities can be detected using our previously developed Cavbase approach. Cavbase performs a comparison and classification of proteins using the physicochemical properties exposed in their active sites. It does not rely on any sequence similarity information and operates in a fully automated fashion. Furthermore, we use the Cavbase analysis to rationalize known data about cross-reactivities between sequentially unrelated kinases, and present examples of possible new cross-reactivities that are suggested by the analysis.

This contribution is organized as follows: In the Experimental Section, a brief overview of the Cavbase method and the similarity analysis and clustering procedure is given. In the Results section, the classification of a comprehensive kinase dataset is presented. In addition, a more detailed analysis of the MAP and c-Abl kinases is described. The section concludes with the presentation of observed and predicted cross-reactivities between different protein kinases and a comparison of the Cavbase classification with classifications obtained using small-molecule inhibition data.

Experimental Section

Cavity detection and property description

Cavbase is a method for describing and comparing protein binding pockets in terms of exposed physicochemical properties.[45–47] It is a modular extension of the protein–ligand database Relibase+.[48–50] Protein binding sites on the protein surface are automatically detected by applying the Ligsite algorithm.[51] In this algorithm, the protein under consideration is embedded into a regularly spaced Cartesian grid with 0.5 Å spacing. Any grid points that fall either within or less than 1.5 Å outside the van der Waals radius of any protein atom are classified as solvent inaccessible. For the remainder of the grid points, a depth value, representing the degree of burial in the protein binding pocket, is determined. Where there exists a contiguous cluster of deeply buried points of at least a specified minimum size, a cavity is generated, as described in a previous contribution.[46] The surface of the cavity is represented

by the grid points of the cluster that are in contact with the protein surface. Any amino acid with at least one atom closer than 1.1 Å to a surface point is defined as a 'cavity-flanking residue'. This information is used to represent the basic geometric shape of the cavity. In a following step, the physicochemical properties of these cavity-flanking amino acids are represented by 3D descriptors known as pseudocenters. The pseudocenters are defined as points in 3D space that represent properties determinant for molecular recognition of the underlying amino acids, and have one of the following five types: hydrogen bond (HB) donor, HB acceptor, mixed HB donor/acceptor, hydrophobic aliphatic contact, and aromatic contact (Figure 1).

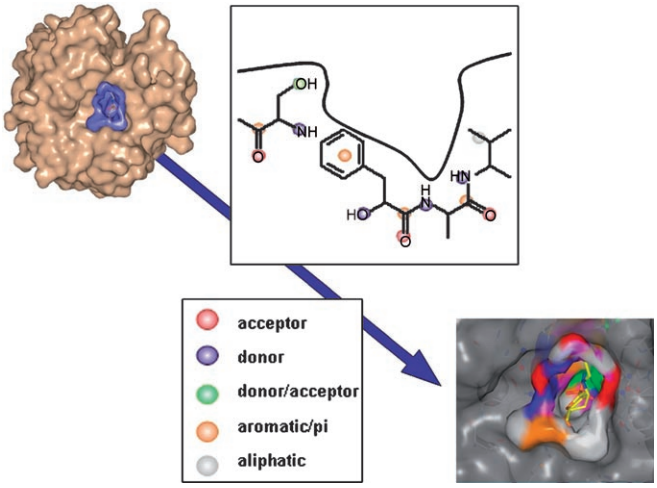

Figure 1. From protein structure to binding site. Cavities are detected as depressions on the protein surface. Physicochemical interaction properties of the amino acids flanking the cavity are encoded into 3D descriptors in the form of pseudocenters. Pseudocenters are displayed as colored spheres (color scheme: hydrogen bond (HB) donor (blue), HB acceptor (red), ambivalent HB donor/acceptor (green), hydrophobic aliphatic (white), aromatic (orange)). Only those pseudocenters that expose their property onto the cavity surface are retained. Binding pockets classified by this scheme are stored in the Cavbase database.

The surface points neighboring each pseudocenter are then assigned the same property type as the pseudocenter in order to indicate what type of interaction that part of the cavity surface is able to form to a ligand. In general, each surface point is assigned the property of the nearest pseudocenter, providing that one exists within 3.0 Å; however, geometric restrictions are applied in order to take into account the directionality of the interactions. Those pseudocenters that represent interactions directed towards the interior of the protein are discarded. The amino acids flanking the binding site, the grid points with their degree of burial, the pseudocenters, as well as the attributed cavity surface points are stored in Cavbase (Figure 1).

Cavities are compared in a pairwise manner using a clique-detection algorithm,[52] which identifies 3D arrangements of pseudocenters that are common to two cavities. For each pairwise comparison, a predefined number (default 100) of the largest clique solutions is taken and further evaluated by scoring them according to the degree of spatial overlap between corresponding cavity surface patches. To score a clique solution, the two cavities are first superimposed based on the mapping of the pseudocenters found in the clique detection. The surface patch overlap for each of the pseudocenter pairs is calculated by summing the relative frequency of surface grid points of both surface patches that fall next to each other below a distance threshold of 1.0 Å. A match is only considered to contribute to the overall binding site similarity if at least 70% of the surface points of the matching surface patches are shared in common to avoid the consideration of strongly fragmented surface patches.[46] Three scoring functions are available in Cavbase. The R_1 scoring is calculated by summing all surface patch overlap values of the pseudocenter pairs that pass the overlap criterion. These pseudocenters represent the physicochemical properties shared by both cavities. Only the most highly scored clique solution is further considered.[46,47] Besides R_1 two alternative scoring schemes R_2 and R_3 can be applied to score a superposition of two cavities. They additionally incorporate information about the RMSD of the detected pseudocenters or the ratio of the matched and the total number of pseudocenters in each cavity, respectively. A detailed description of the binding site comparison and scoring can be found elsewhere.[46,47] The visualization of cavities and superpositions of two cavities is performed using PyMOL.[53]

Cavity clustering and consensus ranking procedures

Each cavity in the dataset is compared to all other cavities applying each of the three scoring schemes available in Cavbase. A similarity matrix is constructed for each scoring scheme, and this serves as the input for various clustering algorithms. The clustering is performed using the clustering toolkit "Cluto",[54] as described in our previous contribution.[47] Four clustering algorithms are used, comprising two partitional (rb and rbr), one agglomerative (agglo), and one graph-partitioning (graph) method. In combination with the different clustering algorithms, agglomerative merging schemes can be applied to obtain a hierarchical tree structure for the different clustering solutions. This allows for an intuitive navigation through the clustered solutions and supports the detection of relationships. All methods were used with the default settings provided by Cluto.[54]

An alternative method of analyzing similarity relationships between kinases is to rank them according to their Cavbase similarity score with a given query. In ligand-based virtual screening, the output of similarity searching using multiple reference structures is often combined using data fusion techniques.[55–57] This approach, also referred to as consensus scoring, generally leads to increased enrichment rates. We follow a similar approach. A Cavbase similarity search of our kinase dataset was performed using all individual cavities from the same subfamily in turn as the query. For each cavity in the dataset, the scores with the different queries are then combined using a simple score-based consensus function, whereby the consensus score for each cavity is simply the mean of the scores with

the individual reference cavities [Eq. (1)]. Subsequently, the entire dataset is re-ranked based on the consensus score.

$$Consensus_x = \frac{\sum_{k=1}^{n_{ref}} Score_k(x)}{n_{ref}} \quad (1)$$

This analysis of the data complements the output from the clustering analysis. Whereas the latter reveals a more global overview of relationships in the data, the ranked lists and the cavity consensus score allow for a fine-grained analysis of single kinase subfamilies. Relationships among the subfamily under investigation and other kinase subfamilies become visible and can be analyzed.

Principal components analysis

Principal components analysis (PCA) is a method for evaluating the overall variation in a dataset, concurrently detecting and rationalizing relationships between the entries of the dataset. It reduces the dimensionality of the variables used to describe the dataset, while preserving characteristics that explain most of the data variance. This is performed by transforming the original set of variables to a new orthogonal set of linear combinations (principal components) of the original variables. For every *k*, the first *k* principal components are chosen in a way as to maximize that part of the total variance in the data which is explained by the corresponding *k*-dimensional representation. A 2D plot of the first two principal components can be useful to visualize relationships in the data that may not be apparent in the original representation. The Cavbase similarity matrix was used as the input for PCA analysis in our work.

Protein kinase dataset

The kinase structures to be included in the dataset were identified and selected by searching the PDB for structural neighbors with the VAST algorithm,[58] using a cAMP-dependent protein kinase (PDB code: 1atp) as the query structure. A total of 864 neighboring protein domains originating from 303 different proteins were found. All non-kinase protein domains and proteins with a resolution above 3 Å were discarded. In the case of multimeric protein structures, only one molecule per protein entry was considered. The selected dataset comprises 258 kinases, including examples from almost all presently known structural subfamilies (PDB version: September 2004).

Table 1 lists the kinase entries used in the present Cavbase similarity analysis together with their corresponding sequence-based group annotation[1] and the SCOP superfamily classification. Because Cavbase usually detects multiple cavities on the surface of a single protein structure, all selected kinase cavities in the dataset were carefully analyzed to confirm that only ATP binding sites were considered.

SCOP classification and sequence-based clustering

The Cavbase classification is compared with a classification based on sequence similarity and with the SCOP classification (SCOP version 1.65) to assess the relevance of our clustering. The SCOP domain entries for each protein kinase are determined and the SCOP superfamily and family annotations are used to assess the Cavbase classification.

To perform a sequence-based classification, the sequences of the catalytic kinase domains were extracted using Relibase+. All kinase sequences were mutually aligned using FASTA 3.5,[59] based on standard settings. The sequence identity values provided by FASTA were normalized to values between 0 and 1. The normalized values were used to construct a similarity matrix that is then clustered in the same way as the Cavbase similarity scores.

Results and Discussion

Kinase binding site and kinase dataset

A typical protein kinase cavity as detected by Cavbase, together with the bilobal protein kinase fold, is displayed in Figure 2. The detected cavity is located between the two lobes of the

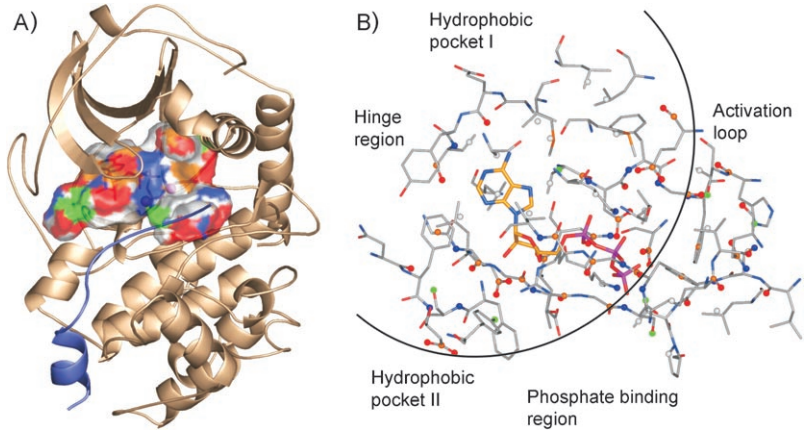

Figure 2. A) Typical bilobal protein kinase fold of a cAMP-dependent protein kinase (PDB code: 1atp) and the ATP binding site as detected by Cavbase. The surface of the ATP cavity as detected by Cavbase is displayed with the bound protein-like substrate in blue. The ATP cavity is rather narrow and deeply buried and has a predominantly hydrophobic character next to the adenine moiety (surface positions in white). B) The functionally important regions of the kinase binding site are displayed: the adenine-binding and the hydrophobic pockets I and II, the sugar and phosphate binding region, the DFG motif, and parts of the glycine-rich, activation, and catalytic loop. In all following figures, this view is used unless noted otherwise. To concentrate on the adenine pocket and to better reflect the selectivity-determining features with respect to small-molecule binding, the cavities of the kinases are limited to the ATP pocket (black line). Pseudocenters are excluded from the cavity comparison if they are more than 16.0 Å away from the hinge region (distance measured to backbone nitrogen atom of hinge residue Val 123).

Table 1. Kinase structures used in the Cavbase similarity analysis.

SCOP Superfamily[a]	Sequence Group[b]	No.[c]	PDB Code
3-Phosphoinositide-dependent PK-1 Pdk1	AGC	6	1h1w 1uu3 1uu7 1uu8 1uu9 1uvr
Abelsone tyrosine kinase (abl)	TK	5	1fpu 1iep 1m52 1opj 1opk
Aurora A (aurora-2)	Other	4	1mq4 1muo 1ol6 1ol7
B-Raf proto-oncogene PK	TKL	1	1uwh
Bruton's tyrosine kinase (Btk)	TK	1	1k2p
Calmodulin-dependent PK	CAMK	1	1a06
cAMP-dependent PK, catalytic subunit	AGC	3	1q24 1q61 1q62
cAMP-dependent PK, catalytic subunit	AGC	26	1apm 1atp 1bkx 1bx6 1cdk 1cmk 1ctp 1fmo 1fot 1j3h 1jbp 1jlu 1L3r 1q8t 1q8u 1q8w 1rdq 1re8 1rej 1rek 1smh 1stc 1ydr 1yds 1ydt 2cpk
Carboxyl-terminal src kinase (csk)	TK	2	1byg 1k9a
Casein kinase-1, CK1	CK1	5	1cki 1ckj 1csn 1eh4 2csn
Cell-cycle checkpoint kinase chk1	CAMK	4	1ia8 1nvq 1nvr 1nvs
Cell division control protein 2 homologue (Pfpk5)	CMGC	4	1ob3 1v0b 1v0o 1v0p
c-jun N-terminal kinase (jnk3s)	CMGC	5	1jnk 1pmn 1pmq 1pmu 1pmv
c-Kit kinase	TK	3	1pkg 1t45 1t46
c-src protein tyrosine kinase	TK	4	1fmk 1ksw 2ptk 2src
Cyclin-dependent PK, CDK2	CMGC	70	1aq1 1b38 1b39 1buh 1ckp 1di8 1dm2 1e1v 1e1x 1e9h 1fin 1fq1 1fvt 1fvv 1g5s 1gih 1gii 1gij 1gy3 1gz8 1h00 1h01 1h07 1h08 1h0v 1h0w 1h1p 1h1q 1h1r 1h1s 1h24 1h25 1h26 1h27 1h28 1hck 1hcl 1jst 1jvp 1ke5 1ke6 1ke7 1ke8 1ke9 1ogu 1oiq 1oir 1oit 1oku 1ol1 1ol2 1p2a 1p5e 1pf8 1pkd 1pw2 1pxi 1pxj 1pxk 1pxl 1pxm 1pxn 1pxo 1pxp 1qmz 1r78 1urw 1v1k 1vyw 1vyz
Cyclin-dependent PK, CDK6	CMGC	1	1blx
Death-associated PK (Dap)	CAMK	4	1g1 1jkk 1jkl 1jks
EGF receptor tyrosine kinase (Erbb-1)	TK	2	1m14 1m17
Ephb2 receptor tyrosine kinase	TK	1	1jpa
Ephrin A2 (Epha2) receptor PK	TK	1	1mqb
Fibroblast growth factor receptor 1	TK	4	1agw 1fgi 1fgk 2fgi
Fibroblast growth factor receptor 2	TK	1	1gjo
Fl cytokine receptor	TK	1	1rjb
Focal adhesion kinase 1 (Fak)	TK	1	1mp8
Glycogen synthase kinase-3β (Gsk3β)	GMGC	11	1gng 1h8f 1i09 1j1b 1j1c 1o9u 1pyx 1q3d 1q3w 1q4l 1uv5
G-protein-coupled receptor kinase 2	AGC	1	1omw
Hematopoietic cell kinase Hck	TK	3	1ad5 1qcf 2hck
Hepatocyte growth factor receptor C-met	TK	2	1r0p 1r1w
Insulin receptor	TK	6	1gag 1i44 1ir3 1irk 1p14 1rqq
Insulin-like growth factor 1 receptor	TK	3	1jqh 1m7n 1p4o
Lymphocyte kinase (lck)	TK	5	1qpc 1qpd 1qpe 1qpj 3lck
MAP kinase activated protein kinase 2 (mapkap2)	CAMK	3	1kwp 1nxk 1ny3
MAP kinase Erk2	CMGC	6	1erk 1gol 1pme 2erk 3erk 4erk
MAP kinase p38α	CMGC	18	1a9u 1bl6 1bl7 1bmk 1di9 1kv1 1kv2 1lew 1lez 1m7q 1ouk 1ouy 1ove 1oz1 1p38 1r39 1r3c 1wfc
MAP kinase p38γ	CMGC	1	1cm8
Musk tyrosine kinase	TK	1	1luf
Mycobacterial protein kinase PknB, catalytic domain	Other	2	1mru 1o6y
pak1, from human	STE	1	1f3m
Pkb kinase (Akt-2)	AGC	7	1gzk 1gzn 1gzo 1mrv 1mry 1o6k 1o6l
Protein kinase CK2, α subunit	Other	14	1daw 1day 1f0q 1j91 1jam 1lp4 1lpu 1lr4 1m2p 1m2q 1m2r 1na7 1om1 1pjk
Sky1p, from baker's yeast	CMGC	5	1how 1q8y 1q8z 1q97 1q99
Tie2 kinase	TK	1	1fvr
Titin, kinase domain	CMGC	1	1tki
Twitchin, kinase domain	CAMK	1	1kob
Vascular endothelial growth factor receptor 2 (kdr)	TK	1	1vr2
γ subunit of glycogen phosphorylase kinase (Phk)	CAMK	3	1phk 1ql6 2phk
Type I TGFβ receptor R4	TKL	2	1b6c 1ias

[a] Superfamily for given catalytic kinase domain extracted from the SCOP database release 1.65. If no SCOP superfamily is given for a particular protein kinase, the information is adapted from the PDB (PK = protein kinase). [b] Sequence-based family classification data taken from reference [1] (http://www.kinase.com/human/kinome). Abbreviations for sequence groups: AGC: containing PKA, PKG, PKC families; CAMK: calcium/calmodulin-dependent protein kinase; CK1: casein kinase 1; CMGC: containing CDK, MAPK, GSK3, CLK families; STE: homologues of yeast sterile 7, sterile 11, sterile 20 kinases; TK: tyrosine kinase; TKL: tyrosine kinase like. [c] Number of cavities in each family.

protein kinase and comprises the ATP binding pocket and extends towards the activation loop. The ATP binding site is categorized into five regions following the pharmacophore model introduced by Traxler, using the ATP binding pose as a reference: the adenine region, the ribose pocket, the hydrophobic regions I (also referred to as the selectivity pocket) and II (also referred to as the specificity surface), and the phosphate binding region[9,60] (Figure 2). Crystal structure analysis reveals that

ATP leaves both hydrophobic regions unoccupied. Consequently, inhibitors addressing the two hydrophobic regions are particularly able to establish selectivity-determining features towards different kinases.

The current dataset comprises 258 kinase cavities (Table 1). The most heavily represented subfamily is the CDK2s (almost 27% of the dataset), followed by protein kinase A, MAP kinases p38 and Erk2, and casein kinases 1 and 2. According to the SCOP classification, 48 different kinase subfamilies are covered in the dataset.

Focusing on the ATP binding site

A kinase ATP binding site as extracted by Cavbase comprises, in addition to the five pharmacophoric regions, parts of the catalytic and activation loop (Figure 2). Aiming for a kinase classification that captures the determining features for kinase selectivity and subfamily characterization, we focus our analysis on residues next to the ATP binding site. Accordingly, the amino acids involved in the hinge hydrogen-bonding network (defined as residues structurally corresponding to the peptide backbone carbonyl (Glu 121) and the peptide amide nitrogen (Val 123) in PKA (1atp)) are identified for each kinase. For our comparative analysis, only those pseudocenters that are located within a predefined sphere around the localized hinge region are considered. The radius of this sphere is defined as 16.0 Å (Figure 2) in order to capture the residues comprising the adenine and both hydrophobic regions, the sugar pocket, the phosphate binding region, and the residues of the DFG motif. (The three amino acids aspartate, phenylalanine, and glycine denote the starting amino acids of the activation loop. The aspartate is involved in magnesium coordination.)

With this preselection of pseudocenters, the approach is more strongly focused on the essential discriminating features exposed to the ATP binding site. Furthermore, the classification is less influenced by the variations in conformation that are exhibited by the activation loop, away from the ATP binding site. This loop is known to undergo major conformational rearrangement, and the different conformations exhibited by it might otherwise dominate the results of our classification analysis. Our approach concentrates on similarities and differences of the ATP binding site region which is typically addressed by ATP-competitive small-molecule inhibitors. Therefore, a classification focusing solely on the ATP pocket is more relevant in terms of inhibitor design and for the understanding of selectivity-determining features.

Overall analysis of the clustering results

In a previous study, the Cavbase classification, applied to a smaller kinase dataset comprising 30 cavities,[47] revealed a clear separation of serine/threonine and tyrosine protein kinase entries into distinct clusters. A clustering on the subfamily level could be achieved. In the present study the performance of Cavbase when applied to a significantly larger kinase dataset and focusing purely on the ATP binding region is investigated.

Recently we evaluated the influence of the choice of the Cavbase scoring scheme and applied clustering algorithms on the Cavbase clustering output. From that study we were able to identify clustering setups that lead to reasonable clustering solutions.[47] All Cavbase scoring schemes give convincing results in terms of cavity clustering, but the combination of the R_1 scoring scheme together with partitional clustering algorithms (rb) perform particularly well. Therefore, this clustering setup is also used in the present study. One crucial parameter is the number of output clusters, which must be predefined. This variable determines the level of detail at which the clustering will classify the data. If the number of clusters is set too low, dissimilar cavities will be grouped into one cluster and relationships present in the dataset will probably be occluded. If the number is set too high, the resulting clustering solution will be too fine-grained, and many singletons will be obtained.[47] We have proposed that for the classification of protein families, the number of output clusters should be of the same order as the expected number of protein subfamilies present in the dataset.[47] Nevertheless, the effect of different numbers of clusters on the clustering results has been evaluated for the present dataset. Cavbase clustering based on different numbers of predefined clusters have been compared with each other. The cluster solutions are robust with respect to the number of output clusters and generally comparable results regarding the classification of kinase subfamilies are obtained. Therefore, in the remainder of this contribution we analyze two settings based on 16 and 48 predefined clusters. Whereas the smaller number provides a general overview of the kinase dataset, the larger number, which equals the amount of SCOP kinase subfamilies in the dataset, allows a more detailed analysis. Textual representations of the obtained clustering can be found in the Supporting Information (Tables 1 and 2). Generally, both clustering solutions are in very good agreement with respect to the grouping of kinase subfamilies; for example, the large kinase subfamilies CDK2, cAMP-dependent protein kinases, Erk, and p38 kinases are grouped together in both clustering schemes. In the case of 48 predefined clusters, these subfamilies are further split into more fine-grained clusters revealing subtypes of these subfamilies. Similarly, cavities that do not show pronounced similarities to any other cavities, such as the calmodulin-dependent protein kinase 1a06, are found as singletons.

Figure 3 shows the clustering obtained using the R_1 scoring function together with rb clustering and 16 clusters specified. This clustering solution provides an overall view of the performance of Cavbase. The number of cavities considered in the clustering procedure is too large for a detailed graphical analysis. However, mapping a sequence-based clustering of the individual kinases onto the Cavbase clustering solution allows a general overview of the classification. The sequence-based group annotation from Manning et al.[1] was used for this purpose and is indicated by applying a color code onto the Cavbase classification shown in Figure 3. Each cavity in the dataset is assigned to one of the following seven sequence groups:[1]

[1] *Abbreviations for sequence groups: AGC: containing PKA, PKG, PKC families; CAMK: calcium/calmodulin-dependent protein kinase; CK1: casein kinase 1; CMGC: containing CDK, MAPK, GSK3, CLK families; STE: homologues of yeast sterile 7, sterile 11, sterile 20 kinases; TK: tyrosine kinase; TKL: tyrosine kinase like.*

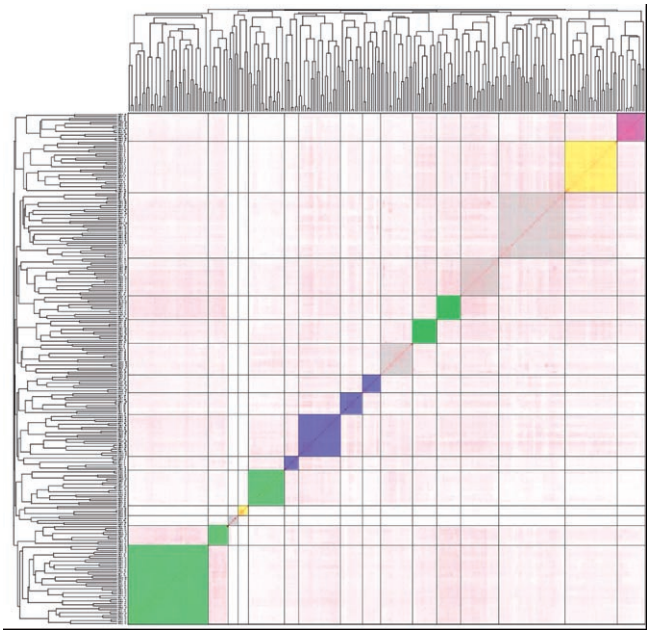

Figure 3. Clustering of the 258 binding cavities using the Cavbase similarity scoring (R_1, rb, 16 clusters). The sequence-based classification reported by Manning and co-workers[1] is mapped onto the Cavbase results by applying the following color code: AGC (yellow), CAMK (orange), CMGC (green), TK (blue), and other (violet). A Cavbase cluster is colored with the respective color if this cluster consists entirely of cavities from a particular group in the sequence-based classification, or contains no more than one outlier from another sequence group. Clusters that contain cavities from multiple sequence groups are colored gray (e.g., clusters containing cavities corresponding to the sequence groups STE, CK1, and TKL). Generally, Cavbase tends to group cavities from one sequence class together. Kinase cavities that have a high degree of sequence similarity are often also considered similar in Cavbase space.

AGC, CAMK, CK1, CMGC, STE, TK, and TKL. The annotation of non-human kinases, which are not included in the Manning classification scheme, is performed manually by attributing them to the human homologue protein kinase. A comparison of the two classification schemes indicates that a high sequence similarity between two protein kinases is generally accompanied by high structural similarity of their binding sites.

Table 2 indicates the protein kinase subfamilies for which a single cluster or multiple adjacent clusters have been generated containing all cavities of that subfamily, but no cavities from other subfamilies. The binding pockets from these subfamilies exhibit sufficient similarities even when the cavities limited to the ATP pocket (Figure 2) are used. Subsequently, they are clustered together with other examples from the same subfamily. This fact supports the hypothesis that even in a homologous series of ATP binding sites, structural differences are sufficient to achieve proper subfamily classification using Cavbase. The fact that the Cavbase clustering places binding sites with high sequence similarity into the same cluster is by no means trivial since different information is evaluated. Cavbase concentrates only on structural similarities across binding sites and neglects the major part of the sequence data of the protein. Nevertheless, sequentially highly related proteins also appear to exhibit similar binding sites.

Table 2. Kinase subfamilies placed by Cavbase clustering in clusters consisting only of cavities from that subfamily.

Kinase SCOP Superfamily
3-Phosphoinositide-dependent protein kinase-1 Pdk1
Cell-cycle checkpoint kinase chk1
MAP kinase Erk2
MAP kinase p38α
Protein kinase CK2, α subunit

Even for the largest kinase subfamily in the dataset (CDK2), the similarity across active sites within the subfamily is detected, and all CDK2 cavities are grouped together. The CDK2s are found in the four clusters in the lower-left corner of Figure 3. Three clusters are colored green and consist only of CDK2 cavities; the forth cluster is colored gray because it also contains entries from two other kinase subfamilies as well as CDK2. However, when the number of clusters generated is increased to 48, these two other kinase families are separated from the CDK2 cavities into distinct clusters. A more detailed analysis of the CDK2 cavities in the respective clusters indicates an obvious splitting into separate subclusters representing CDK2s in active or inactive conformational states. This observation becomes even more opportune, as 70 kinase cavities from this protein subfamily are used for the classification and only information limited to the ATP pocket is evaluated.

MAP kinases

Mitogen-activated protein kinases (MAPKs) are involved in signal-transduction cascades that control complex processes such as cell proliferation, differentiation, and apoptosis. Based on their signature activation sequences, MAPKs can be categorized into at least three broad subfamilies: MAP kinase p38, extracellular-signal-regulated kinases (ERKs), and c-jun amino-terminal kinases (JNKs).[61,62] The MAP kinase p38α subfamily is of particular pharmaceutical interest because these kinases are involved in the production of proinflammatory cytokines.[63] Thus, selective inhibition of the p38α isoform provides a possible therapeutic approach towards a wide range of inflammatory diseases.

The dataset used for the similarity analysis (258 kinase cavities) includes as a subset 30 MAP kinases originating from the p38, Erk2, and JNK3 subfamilies as listed in Table 3. Most MAP kinase crystal structures are determined in their unphosphorylated, inactive state. The cluster analysis using Cavbase separates the 30 respective MAP kinases on a subfamily level into four groups: p38α, p38γ, Erk2, and JNK3 (data not shown).

To further investigate and analyze the cross-relationships between various members of the MAP kinase family, a principal components analysis (PCA) was performed using the Cavbase scores between different cavities of the MAP kinase family only. The first principal component explains 44.7% of the variance, while the second adds another 14.6%. Similarly to the Cavbase clustering, the cavities are also grouped on a subfamily level (p38α, p38γ, Erk2, JNK3). Both approaches distinguish-

Table 3. MAP kinases used in the Cavbase classification analysis.

PDB Code	MAP Subfamily	Complex Structure	Activity State	Reference
1erk	Erk2	apo structure	inactive	[76]
1gol	Erk2	binary MgATP	inactive	[77]
1pme	Erk2	pyridinylimidazole	inactive	[64]
2erk	Erk2	apo structure	active	[78]
3erk	Erk2	pyrimidinylimidazole	inactive	[67]
4erk	Erk2	olomoucine	inactive	[67]
1a9u	p38α	pyridylimidazole	inactive	[67]
1bl6	p38α	pyridylimidazole	inactive	[67]
1bl7	p38α	pyrimidinylimidazole	inactive	[67]
1bmk	p38α	pyrimidinylimidazole	inactive	[67]
1di9	p38α	4-anilinoquinazoline	inactive	[79]
1kv1	p38α	diarylurea	inactive	[80]
1kv2	p38α	diarylurea	inactive	[80]
1lew	p38α	with peptide substrate	inactive	[81]
1lez	p38α	with peptide substrate	inactive	[81]
1m7q	p38α	dihydroquinazolinone	inactive	[82]
1ouk	p38α	pyrimidinylimidazole	inactive	[70]
1ouy	p38α	dihydropyridopyrimidinone	inactive	[70]
1ove	p38α	dihydroquinolinone	inactive	[70]
1oz1	p38α	4-azaindole	inactive	[83]
1p38	p38α	apo structure	inactive	[84]
1r39	p38α	apo structure	inactive	[85]
1r3c	p38α	apo structure	inactive	[85]
1wfc	p38α	apo structure	inactive	[86]
1cm8	p38γ	ATP analogue	active	[87]
1jnk	JNK3	ATP analogue	inactive	[88]
1pmn	JNK3	pyrimidinylimidazole	inactive	[89]
1pmq	JNK3	pyrimidinylimidazole	inactive	[89]
1pmu	JNK3	phenanthroline	inactive	[89]
1pmv	JNK3	dihydroanthrapyrazole	inactive	[89]

ed between the four subfamilies (Figure 4). Furthermore, PCA allows one to correlate the cavity space spanned by the MAP kinase family. In analogy to the concept of the target family landscapes presented by Naumann and Matter,[36] these Cavbase/PCA plots assist in the visualization of cross-relationships.

An interesting example of functional cross-relationships between protein families involves the penta-mutant of a MAP Erk2 kinase (PDB code: 1pme[64]). This mutated Erk2 kinase has been constructed to mimic the binding pocket of MAP p38α based on the Erk2 skeleton. The comparison of this Erk2 kinase with respect to other MAP kinases reveals a transient character of this penta-mutant. When analyzing its closest neighbors in cavity space, Cavbase detects similarities either to Erk2 and p38α examples. In the scatter plot of the first two principal components (Figure 4), the ambiguous character of the penta-mutant (1pme) can be easily recognized. This cavity is found halfway between the two classes.

Two further Cavbase applications are described below, and a graphical analysis of the detected similarities and dissimilarities between binding sites is presented.

As expected, the Cavbase classification separates the entries from the Erk2 and p38α subfamily. However, what structural features determine their difference in function? Inhibition profiles of small molecules reveal that there are several compound classes showing a selective inhibition of p38α, but not of Erk2.

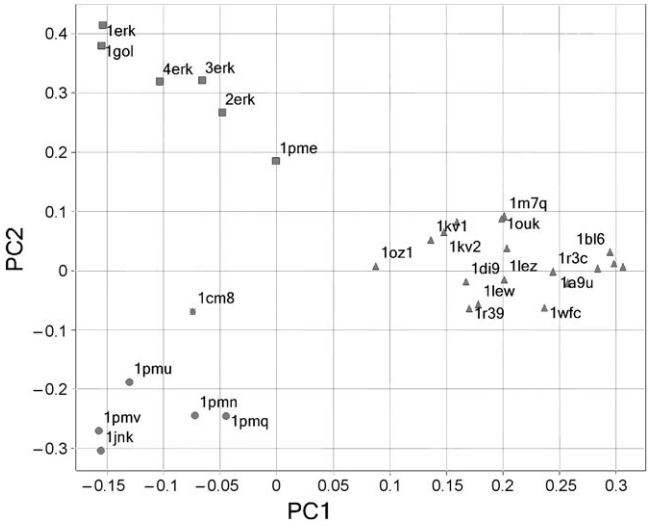

Figure 4. Scatter plot of the first two principal components derived from the comparison of the 30 MAP kinases. The Cavbase/PCA analysis is able to clearly distinguish between the four subfamilies (p38α (▲), p38γ (✱, 1cm8), Erk2 (■), JNK3 (●); labels: PDB codes). The shape of the entries was mapped after the PCA analysis and serves only for illustrative purposes. The Erk2 penta-mutant 1pme mimicking p38α is found in an intermediate position between Erk2 and p38α, reflecting the ambitious character of this compound.

One example is pyridinylimidazole-type inhibitors (Figures 5 and 6) introduced by SmithKline Beecham, such as compound SB203580 (**1**).[65–67] This class shows high selectivity for p38α compared to other MAP kinases. This observation can be explained by the properties of the gatekeeper residue and its influence on the size of the hydrophobic pocket II (Figure 5).[67,68] In p38α, the size of this pocket is controlled by a small threonine residue, whereas other MAP kinases exhibit a larger residue at this position (e.g. a glutamine in Erk2).[68,69] The Cavbase analysis indicates similarities next to the hinge hydrogen-bonding region and the adenine binding site, but both cavities lack any similarity in the hydrophobic region II close to the gatekeeper residue (Figure 5).

As a further example, the pronounced p38α specificity of quinazolinone- and dihydropyridopyrimidinone-based inhibitors (compounds **2** and **3**, respectively) compared to other MAP kinases can be rationalized by Cavbase. Several crystal structures of complexes with these inhibitors have been determined.[70] Besides the effect of the differing gatekeeper residues in Erk2 and p38α described above, a backbone peptide flip between Met 109 and Gly 110 is assumed to be responsible for an unexpected specificity for dihydropyridopyrimidinone-based inhibitors **3**.[70] These inhibitors possess a unique hydrogen bond acceptor functionality (keto group) at the position where most kinase inhibitors exhibit a donor function (e.g. a pyrimidine nitrogen atom), which forms a hydrogen bond to the hinge region. This deviating pattern of donor/acceptor properties exposed to the hinge region is specific to this compound class and is not found in other kinase inhibitors. It particularly addresses the peptide backbone in its flipped orientation in the case of the p38α family, and it is assumed that it selectively provokes this flip. A glycine residue next to Met 109 in

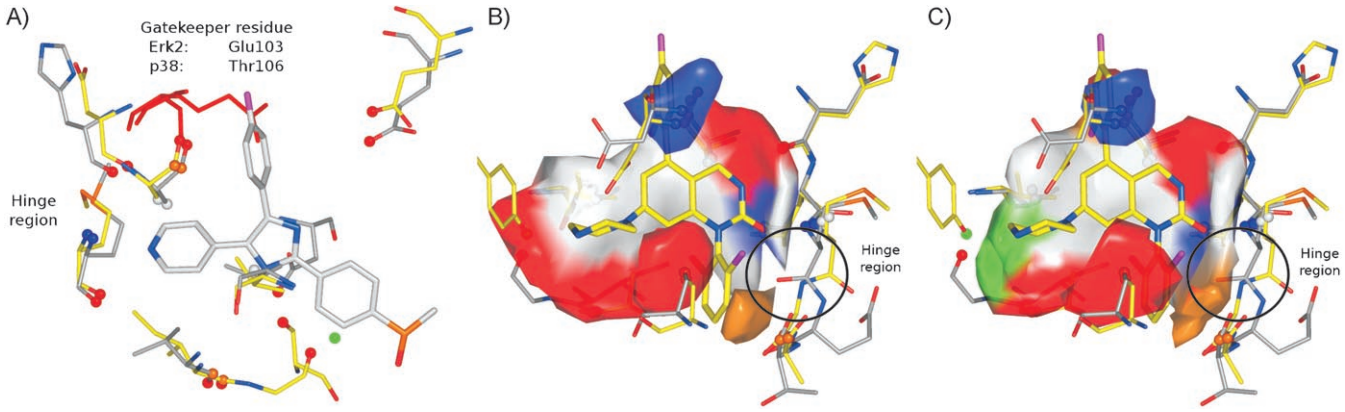

Figure 5. A) Selectivity-determining features in the binding sites of MAP kinases of the p38α and Erk2 subfamilies. The pseudocenters detected as common to both subfamilies are shown together with the amino acids in the binding site of a p38α kinase (PDB code: 1a9u, carbon atoms in white) and an Erk2 kinase (PDB code: 1erk, carbon atoms in yellow). The possible clash of the gatekeeper residue (in red) in Erk2 with the fluorophenyl ring of the p38α selective **1** (Figure 6) is displayed. Cavbase detects no similarity between the two proteins in the region next to the gatekeeper. B), C) Subtle differences of an Erk2 or p38α cavity are apparent. Dihydroquinazolinone-based inhibitors, such as **2** and **3**, show a unique hydrogen-bonding pattern to the hinge region. The amino acids detected as similar are shown together with the pseudocenters of an Erk2 (PDB code: 1erk, carbon atoms in gray) and a p38α kinase (PDB code: 1m7q, carbon atoms in yellow). In B) the surface of Erk2 is mapped onto the superimposed binding pockets; in C) the surface of p38α is shown in similar fashion. The lack of similarity next to the flipped peptide bond can be easily detected; for clarity this bond is indicated by a circle.

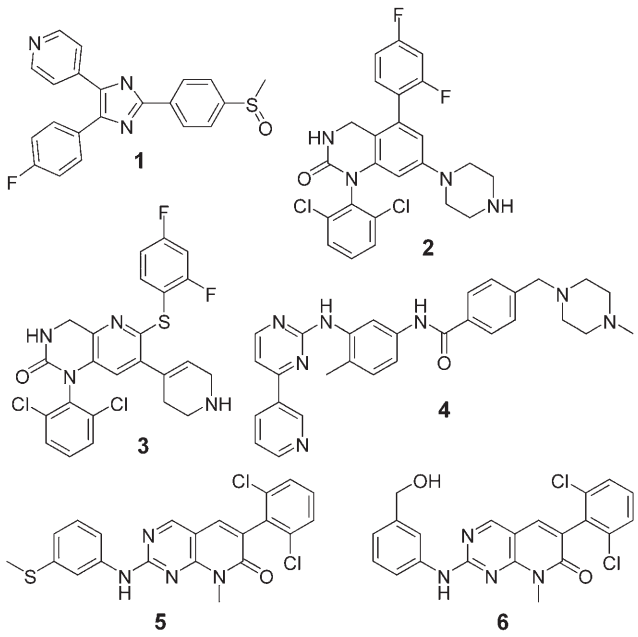

Figure 6. Low-molecular-weight ATP-competitive kinase inhibitors of known binding modes. The SmithKline Beecham compound SB203580 (**1**) belongs to the pyridinylimidazole-based class of inhibitors. They are selective inhibitors of MAP kinase p38α. The dihydroquinazolinone-based **2** and dihydropyridopyrimidinone-based inhibitor **3** are also selective inhibitors of MAP kinase p38α. Imatinib (**4**) is a potent inhibitor of c-Abl kinase, platelet-derived growth factor receptor (PDGF-R) kinase α and β, and the c-Kit kinase. The pyrido[2,3-d]pyrimidine compounds PD173955 (**5**) and **6** are classical hinge binders and have a greater potency towards c-Abl, but are less selective inhibitors.

MAP p38α assists this flip, whereas larger residues present at this position in other MAP kinases are detrimental to support the flipped orientation. Cavbase is able to capture the deviating H-bonding pattern next to the Met 109 backbone region due to the flipped peptide bond in p38α, in comparison to other MAP kinases. The corresponding regions in the active site of two MAP p38α kinases in complex with a pyridopyrimidine and a dihydroquinazolinone ligand are shown in Figure 5B and Figure 5C. In contrast to most pairwise kinase comparisons, no similarity is detected next to the flipped donor/acceptor functionality.

c-Abl tyrosine kinases

The cellular form of the Abelsone leukemia virus tyrosine kinase (c-Abl) belongs to the class of non-receptor tyrosine kinases. Patients suffering from chronic myelogenous leukemia (CML) often possess a reciprocal translocation between chromosomes 9 and 22 that leads to the formation of a chimeric gene (*bcr–abl*) on the so-called Philadelphia chromosome.[71,72] The resulting fusion protein (Bcr–Abl) has an intact Abl kinase domain, but lacks the internal control mechanism that keeps c-Abl in its inactive form.[72] Increased tyrosine kinase activity of the Bcr–Abl protein leads to CML.[73] 2-Phenylaminopyrimidine-type inhibitors such as imatinib (**4**, Gleevec, approved for therapy against CML) have a high affinity towards Abl kinase.[74]

Five c-Abl structures are present in our dataset of 258 kinases, two with pyrido[2,3.d]pyrimidine-type inhibitors (PDB codes: 1m52 and 1opk) and three with 2-phenylaminopyrimidine-type inhibitors (PDB codes: 1iep, 1fpu, 1opj). Cavbase detects the similarity in their ATP binding sites, and all five c-Abl binding sites are clustered together (Figure 7). Interestingly, the structure of a stem-cell factor receptor c-Kit (a receptor tyrosine protein kinase, PDB code: 1t46) with bound imatinib (**4**) is also found in the c-Abl cluster.

A more detailed analysis of the two c-Abl clusters shows that Cavbase groups the c-Abl cavities with bound pyrido-[2,3.d]pyrimidine-type inhibitors **5** and **6** into one cluster (Figure 7). The remaining c-Abl and c-Kit cavities hosting **4** or an analogue of **4** are also grouped together. This discrimina-

tion reflects the different activation states of the corresponding c-Abl kinase entries. The c-Abl structures with the pyrido-[2,3.d]pyrimidine-based inhibitors are in the active state, whereas **4** traps the c-Abl kinase in its inactive form.[75] It is believed that **4** cannot bind c-Abl in its active conformation due to a steric clash with the DFG motif (as discussed below). The smaller **5** is more potent against c-Abl than **4**. One possible explanation could be that **5** binds to both the inactive and active conformations of c-Abl. It would thus recognize multiple conformational states of the protein. However, **5** is a less selective inhibitor. In structural terms, the kinase domain in both complexes with either **4** or **5**[75] is virtually identical, but large deviations in the conformations of the activation loop and some rearrangements of the DFG motif are observed. The c-Abl kinase cavities in the activated state with bound pyrido[2,3.d]pyrimidine-type inhibitors show a substantial degree of similarity to each other (Figure 8 and Table 4). The protein coordinates and the position of the bound inhibitors are virtually superimposable. In contrast, a direct comparison of the active and inactive conformations of c-Abl kinases reveals similarity next to the hinge region and the hydrophobic region II only. There are significant structural differences near the DFG motif, which adopts very different conformations in the active and inactive states. In the active state, the aspartate of this motif points into the cavity towards the magnesium ions (DFG-in conformation), whereas it is oriented away from the ATP binding site in the inactive state (DFG-out conformation). Inhibitors such as **4**, which address the DFG-out conformation have been classified as "DFG-out binders", in contrast to classical ATP-competitive hinge binders. Figure 8C and D show a superposition of both cavities. The areas considered similar by Cavbase in the ATP binding site are located next to the hinge region and the hydrophobic pocket. The difference in the conformation of the DFG motifs, preventing imatinib from binding to the active conformation (DFG-in), can be easily recognized.

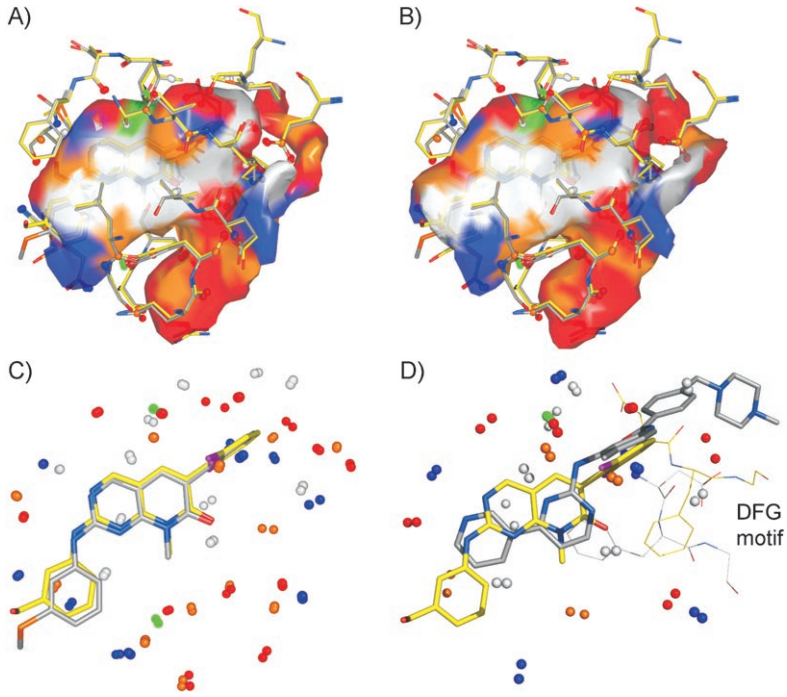

Figure 8. Comparisons of the ATP binding sites of c-Abl kinases in the same activation state and in different activation states. A)–C) Equivalent areas of the active sites of c-Abl kinases exhibiting the same activation state are displayed. The amino acids and ligands from 1opk (carbon atoms in yellow) and 1m52 (carbon atoms in gray) with the matching pseudocenters and cavity surface are shown in A) and B). A) and B) each show the matching surface area from one of the two cavities (1opk and 1m52, respectively). In C) the matching pseudocenters and bound ligands are displayed; the two cavities have a high degree of similarity in the active sites. D) Shows the comparison between inactive c-Abl (PDB code: 1opj, carbon atoms in yellow) and active c-Abl (PDB code: 1opk, carbon atoms in gray) kinases, which exhibit less similarity in the ATP site. The matching areas in the two cavities (indicated by the matched pseudocenters) are found near the adenine binding pocket and the hydrophobic region II. The conformation and spatial orientation of the DFG motif differs drastically between the two states of the c-Abl kinases (DFG-in and DFG-out conformations). The flipped arrangement of the aspartate and phenylalanine can easily be visualized using the superposition based on the similarities in the ATP pocket. This superposition indicates that Gleevec cannot bind to the active conformation, as it would clash with the DFG-in conformation.

Rationalizing the cross-reactivity of Gleevec against other kinases

In vitro screens against a range of different protein kinases have shown that **4** inhibits at least two other kinases besides

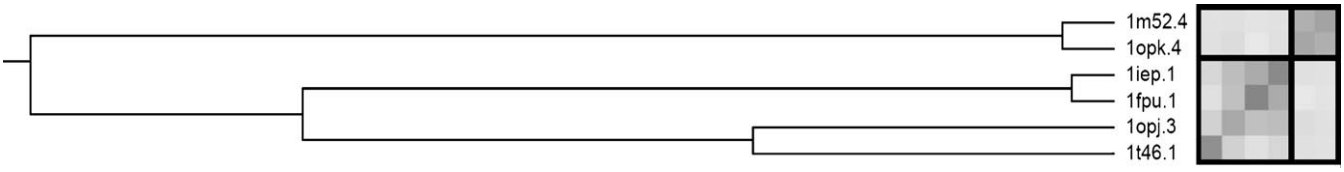

Figure 7. Clustering of c-Abl kinases. Cavbase detects the similarity between the five c-Abl cavities in the present dataset comprising 258 kinase structures and groups them into two adjacent clusters. Furthermore, Cavbase distinguishes between the c-Abl structure in an active conformation (1m52 and 1opk) and those in the inactive conformation (1fpu, 1iep, and 1opj). Interestingly, the c-Kit kinase cavity (1t46) with bound imatinib is found in the same cluster.

Table 4. Matched pseudocenters and amino acids in the binding sites c-Abl and c-Kit.

c-Abl (1opj)			c-Kit (1t46)		
Pseudocenter Type	Corresponding Amino Acid[a]		Pseudocenter Type	Corresponding Amino Acid[a]	
pi	A 288	p	pi	A 621	p
acceptor	A 288	p	acceptor	A 621	p
pi	V 289	p	pi	V 622	p
donor	K 290	p	donor	K 623	p
acceptor	E 305	s	acceptor	E 640	s
pi	E 305	s	pi	E 640	s
donor	V 318	p	donor	V 654	p
acceptor	I 332	p	acceptor	V 668	p
donor	T 334	p	donor	T 670	p
donor/acceptor	T 334	s	donor/acceptor	T 670	s
acceptor	E 335	p	acceptor	E 671	p
pi	F 336	s	pi	Y 672	s
donor	M 337	p	donor	C 673	p
acceptor	M 337	p	acceptor	C 673	p
pi	G 340	p	pi	G 676	p
donor	G 340	p	donor	G 676	p
donor/acceptor	H 380	s	donor/acceptor	H 790	s
pi	V 398	p	pi	I 808	p
acceptor	V 398	p	acceptor	I 808	p
pi	A 399	p	pi	C 809	p
donor	D 400	p	donor	D 810	p
acceptor	D 400	p	acceptor	D 810	p
pi	D 400	s	pi	D 810	s
aliphatic	L 267	s	aliphatic	L 595	s
aliphatic	V 275	s	aliphatic	V 603	s
aliphatic	A 288	s	aliphatic	A 621	s
aliphatic	M 309	s	aliphatic	L 644	s
aliphatic	V 318	s	aliphatic	V 654	s
aliphatic	I 332	s	aliphatic	V 668	s
aliphatic	T 334	s	aliphatic	T 670	s
aliphatic	L 389	s	aliphatic	L 799	s
aliphatic	A 399	s	aliphatic	C 809	s

[a] One-letter code of the amino acid, the number of the amino acid, and origin of pseudocenter: from side-chain (s) or from peptide bond (p).

c-Abl: the stem-cell receptor c-Kit (see above) and the platelet-derived growth factor (PDGF) receptor.[74] Both proteins belong to the type III transmembrane receptor protein tyrosine kinase (RPTK) family. Three c-Kit structures are present in the dataset: one auto-inhibited example (PDB code: 1t45) with no ligand in the ATP pocket, one complex in the activated state with bound ADP (PDB code: 1pkg) and one complex with **4** (PDB code: 1t46). A comparison of the c-Abl cavities against those of all other kinases in the dataset shows that the c-Kit cavities exhibit highest similarity to the c-Abl cavities. The overlapping and similar areas in both binding sites are displayed in Figure 9. Similarities are exhibited by the adenine region and the phosphate binding pocket. Furthermore, consensus scoring of the dataset using c-Abl kinase cavities as queries was performed, and a member of the c-Kit family was the most highly scored cavity apart from other c-Abl cavities (see Table 7 below). The c-Kit cavity with bound ADP (PDB code: 1pkg) is ranked 16th still exhibiting substantial similarities in the ATP binding site. The cavity of the auto-inhibited c-Kit structure (PDB code: 1t45) is not detected as nearest neighbor. In this structure the flexible c-Kit juxtamembrane domain inserts into the ATP pocket between both kinase domains, keeping the kinase in an auto-inhibited state. Due to these conformational changes, similarity to the c-Abl cavities is only found next to the hinge and the adenine binding region. The observed similarity between the ATP binding sites of c-Abl and c-Kit in the activated state provides an explanation in structural terms for the observed cross-reactivity and the similarity of the inhibition profiles found for **4** towards both kinases.

Cross-relationships between unrelated kinases

From a medicinal chemistry point of view, the discovery of structural similarities across kinases that are not closely related in sequence space is of utmost importance. Knowledge about such similarities in structural and physicochemical terms supports the selection of off-target kinases for the profiling of putative lead compounds that are developed with the goal of being highly selective.

In order to identify unexpected similarities and dissimilarities in ATP binding sites of the 258 kinases in our dataset, a sequence-based similarity measure is compared against the Cavbase similarity score (R_1). Two situations are of particular inter-

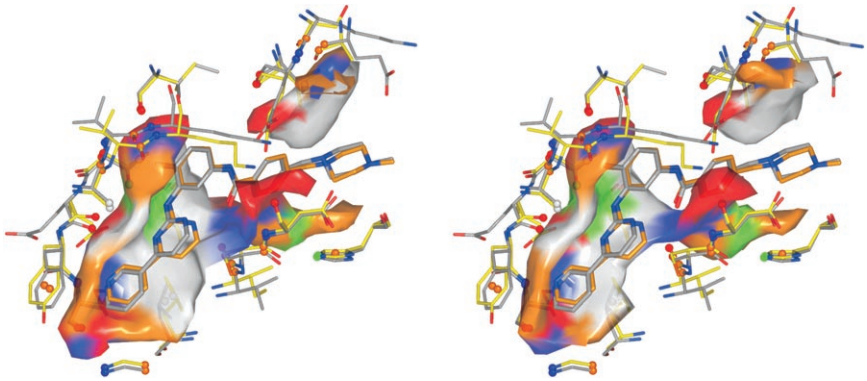

Figure 9. Rationalizing the cross-reactivity of Gleevec against c-Abl and c-Kit. Similar regions in the active sites of c-Abl (PDB code: 1opj) and c-Kit (PDB code: 1t46) are displayed. The matching amino acids and pseudocenters are shown together with the bound ligands. Additionally, each diagram shows the matching cavity surface for one of the two cavities; c-Abl (left) and c-Kit (right). Both kinases show a high degree of similarity in their binding sites. The pseudocenters (RMSD = 0.61 Å) and inhibitor molecules superimpose almost perfectly.

est. Firstly, we are interested in cases where a low sequence similarity is accompanied by a high Cavbase similarity. Kinases belonging to this group are likely to show possible cross-reactivities with bound ligands, because they exhibit similar structural and physicochemical properties in the active site. Secondly, we are interested in cases where a high sequence similarity is accompanied by a low Cavbase similarity. These kinases represent a situation where Cavbase detects unexpected dissimilarity that is not apparent from sequence analysis alone.

Low sequence similarity and high Cavbase similarity

Several kinase pairs exhibit a sequence identity below 50%, yet possess a Cavbase R_1 similarity score greater than 22.0, where the cavity comparison is limited to the ATP region. Values of the R_1 Cavbase score above this threshold indicate that a substantial degree of similarity is present between the two cavities. This similarity score threshold has been empirically validated. Visual inspection of the matched surface patches shows that this similarity corresponds to an area next to the hinge region and the hydrophobic pocket I, extending towards the sugar binding site. For example, two kinase cavities from one subfamily show R_1 scores of about 50 (e.g. the two c-Abl cavities shown in Figure 8A and B), whereas the related kinases c-Abl and c-Kit possess a Cavbase similarity score R_1 of 29.5 (Figure 9). The cutoff value for the corresponding sequence identity is set to 50% in order to focus on cases where a high Cavbase similarity is unexpected. According to experiences in the field of homology modeling, a sequence identity greater than 50% allows for the construction of meaningful homology models. In our application, it is assumed that kinase pairs showing greater than 50% sequence identity across the entire kinase domain denote cases where the similarity between both kinases is obvious. In our case study, the threshold for the sequence identity is set conservatively to 50% to capture a large number of possible cross-reactivities. In Table 5, detected kinase pairs satisfying these criteria are listed. As an additional requirement, a pair of kinases is only considered if it has been found multiply as a hit at least by five mutual cavity matches. This criterion excludes cavity pairs that are only sparsely populated in the dataset. Thus the analysis is focused on entries of subfamily pairs that are more frequently present in the dataset.

For example, the tyrosine kinase of the type I TGFβ receptor exhibits pronounced similarities with kinases from the src family, such as Hck, Csk, and Lck. In Figure 10, the superposition of a TGFβ kinase (PDB code: 1ias) and a hematopoietic cell kinase Hck pocket (PDB code: 1ad5) is shown. A large part of the binding pocket is found to be similar, although this is limited to the ATP neighboring region. This similarity encompasses the adenine binding pocket and both hydrophobic sites, and extends towards the phosphate binding region. The two catalytic protein kinase domains show a sequence identity of only 26%. The achieved Cavbase similarities are listed in Table 6.

Recently, Fabian and co-workers[44] detected a new target for **4** using a competition binding assay. This study revealed that

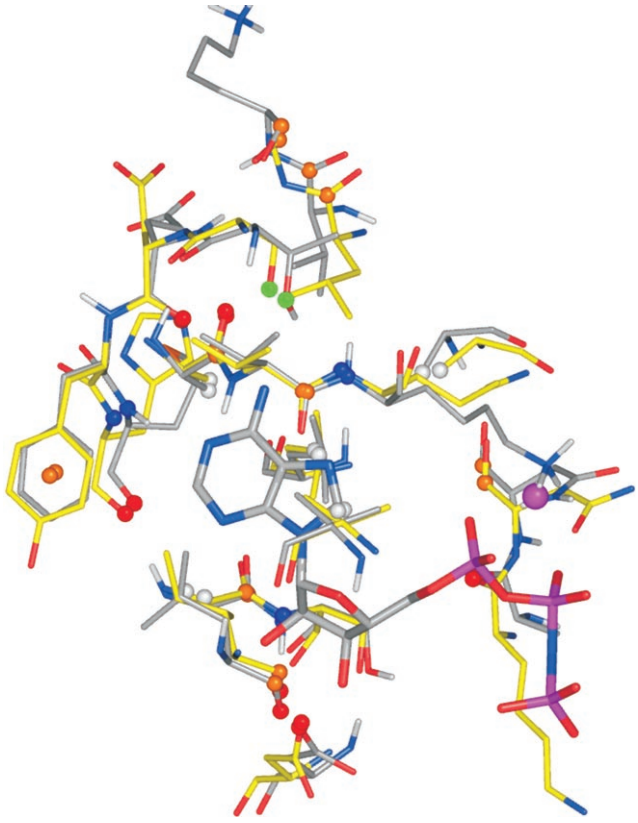

Figure 10. Unexpected high similarities between type I TGFβ and c-src-based kinase binding sites. Large regions of the ATP binding site are detected as similar between type I TGF (PDB code: 1ias, carbon atoms in yellow) and Hck (PDB code: 1ad5, carbon atoms in gray), as indicated by the matching amino acids, pseudocenters, and bound ATP analogue.

lymphocyte kinase (Lck) is also inhibited by **4** in the nanomolar range. The two proteins share only low sequence identity (48%). Interestingly, Cavbase is able to capture these similarities. In Table 7, the results of consensus scoring using five c-Abl cavities as query cavities are listed. As expected, all of the other c-Abl cavities in the dataset are most highly ranked. However, the next most highly ranked cavity is from the c-Kit subfamily, followed by the B-Raf and the Lck subfamily.

The Cavbase analysis also shows Lck as a case exhibiting extensive structural similarities to a large set of other protein kinases (Table 5). As mentioned above, binding of **4** to Lck indicated the relationship to c-Abl. In Figure 11, the superpositions of an Lck cavity on one inactive c-Abl and one active c-Abl cavity are shown. Cavbase detects a similar arrangement between Lck and c-Abl in both activation states next to the hinge region, the adenine binding region, and the sugar pocket. However, in the case of the active c-Abl, additional similarities are found in the hydrophobic pocket I (c-Abl residues Leu 299 and Asp 381) and the hydrophobic pocket II (c-Abl residue Leu 248).

Fabian et al. also constructed a hierarchical clustering of the protein kinases based on specificity profiles.[44] The profiling was performed in two steps. In the first step, the 20 different kinase inhibitors were screened in a competition assay, and in a second step quantitative binding constants (K_d values) were

Table 5. Unexpected similarities and dissimilarities in the binding sites of kinases.

SCOP Superfamily	SCOP Superfamily	Sequence Identity	R_1
Low sequence identity (< 50%) and high Cavbase similarity ($R_1 > 22.0$)			
Abelsone tyrosine kinase (abl)	Lymphocyte kinase (lck)	48	24.54
Aurora A (aurora-2)	Cyclin-dependent PK, CDK2	29	27.89
cAMP-dependent PK, catalytic subunit	Death-associated PK (Dap)	30	28.91
cAMP-dependent PK, catalytic subunit	Pkb kinase (Akt-2)	44	35.92
cAMP-dependent PK, catalytic subunit	γ subunit of glycogen phosphorylase kinase (Phk)	34	26.74
cAMP-dependent PK, catalytic subunit	Glycogen synthase kinase-3β (Gsk3β)	30	25.18
cAMP-dependent PK, catalytic subunit	Mycobacterial protein kinase PknB	26	28.47
cAMP-dependent PK, catalytic subunit	Lymphocyte kinase (lck)	24	24.72
Carboxyl-terminal src kinase (csk)	Lymphocyte kinase (lck)	44	26.66
c-Kit kinase	Lymphocyte kinase (lck)	37	26.42
c-src protein tyrosine kinase	Type I TGFβ receptor R4	26	29.10
Cyclin-dependent PK, CDK2	γ subunit of glycogen phosphorylase kinase (Phk)	29	27.54
Cyclin-dependent PK, CDK2	Lymphocyte kinase (lck)	28	27.11
Cyclin-dependent PK, CDK2	Glycogen synthase kinase-3β (Gsk3β)	34	24.55
Cyclin-dependent PK, CDK2	Sky1p, from baker's yeast	30	25.90
Death-associated PK (Dap)	γ subunit of glycogen phosphorylase kinase (Phk)	39	29.06
Death-associated PK (Dap)	Pkb kinase (Akt-2)	31	27.07
Death-associated PK (Dap)	Mycobacterial protein kinase PknB	27	26.39
EGF receptor tyrosine kinase (Erbb-1)	Lymphocyte kinase (lck)	37	31.82
Ephb2 receptor tyrosine kinase	Lymphocyte kinase (lck)	44	27.87
Ephrin A2 (Epha2) receptor PK	Lymphocyte kinase (lck)	44	28.86
Fibroblast growth factor receptor 1	Lymphocyte kinase (lck)	39	29.35
Insulin receptor	Lymphocyte kinase (lck)	38	24.18
Lymphocyte kinase (lck)	Type I TGFβ receptor R4	26	27.35
Sky1p, from baker's yeast	γ subunit of glycogen phosphorylase kinase (Phk)	26	29.47
High sequence identity (> 50%) and low Cavbase similarity ($R_1 < 13.0$)			
MAP kinase p38α	MAP kinase p38γ	63	11.88
Insulin receptor	Insulin-like growth factor 1 receptor	80	11.81
Cell division control protein 2 homologue (Pfpk5)	Cyclin-dependent PK, CDK2	63	12.98
c-Kit kinase	Vascular endothelial growth factor receptor 2 (kdr)	58	11.79
Fl cytokine receptor	Vascular endothelial growth factor receptor 2 (kdr)	53	11.62
MAP kinase Erk2	MAP kinase p38α	51	12.28

measured for all kinases that were identified as hits in the first screen. According to this setup, they found that Lck and Epha2 receptor tyrosine kinase share similarities. Analysis of the consensus scores for the Lck cavities reveals a high similarity to the Ephrin A2 subfamily (Table 8). Furthermore, the three most similar cavities to the Lck case originate from a Fibroblast growth factor receptor 1, an Ephrin B2 receptor, and an EGF receptor tyrosine kinase (Erbb-1). By comparing the cluster composition and relationships across clusters from the assay profiling with the clustering in Cavbase, a substantial overlap of both classification schemes is suggested (e.g. kinase subfamilies detected as neighbors by both approaches: Ephrin A2 (Epha2) receptor protein kinase, carboxyl-terminal src kinase (csk), Ephrin B2 receptor tyrosine kinase, c-Kit kinase, c-src protein tyrosine kinase, hematopoietic cell kinase Hck, and Abelsone tyrosine kinase). Nevertheless, it must be recognized that not all kinases studied by Fabian et al.[44] are also present in our dataset (as a crystal structure is not available for all of them), and vice versa.

High sequence similarity and low Cavbase similarity

Conversely to the situation described above, there are several subfamilies that are closely related in sequence space but exhibit pronounced differences in cavity space (Table 5). As an example, the MAP kinases p38α and p38γ display high sequence similarity (both kinases are found in one sequence space cluster; sequence identity is 63%), whereas only modest Cavbase similarity is exhibited. In view of this disparity, it is interesting to note that Vieth and co-workers[40] could show that both families differ with respect to their small-molecule inhibition profiles. Binding site similarities exhibited by cavities of two members of these families are shown in Figure 12.

Conclusions

In this contribution, we have presented the clustering and classification of protein kinases using Cavbase. A comprehensive dataset of 258 protein kinases was extracted from Relibase+ and analyzed based on similarities in the spatial arrangement and distribution of physicochemical properties in their binding

Table 6. Areas of the binding sites of Hck and Type 1 TGFβ detected as equivalent by Cavbase.

Hck (1ad5)			Type 1 TGFβ (1ias)		
Pseudocenter Type	Corresponding Amino Acid[a]		Pseudocenter Type	Corresponding Amino Acid[a]	
pi	L 273	p	pi	I 211	p
acceptor	L 273	p	acceptor	I 211	p
pi	A 293	p	pi	A 230	p
acceptor	A 293	p	acceptor	A 230	p
pi	V 294	p	pi	V 231	p
donor	K 295	p	donor	K 232	p
pi	V 323	p	pi	L 260	p
pi	K 324	p	pi	G 261	p
donor/acceptor	T 338	s	donor/acceptor	S 280	s
acceptor	E 339	p	acceptor	D 281	p
pi	F 340	s	pi	Y 282	s
donor	M 341	p	donor	H 283	p
acceptor	M 341	p	acceptor	H 283	p
pi	G 344	p	pi	G 286	p
donor	S 345	p	donor	S 287	p
acceptor	D 348	s	acceptor	D 290	s
acceptor	A 390	p	acceptor	K 337	p
pi	N 391	p	pi	N 338	p
aliphatic	L 273	s	aliphatic	I 211	s
aliphatic	V 281	s	aliphatic	V 219	s
aliphatic	A 293	s	aliphatic	A 230	s
aliphatic	L 393	s	aliphatic	L 340	s
aliphatic	A 403	s	aliphatic	A 350	s

[a] One-letter code of the amino acid, the number of the amino acid, and origin of pseudocenter: from side-chain (s) or from peptide bond (p).

Table 7. Consensus ranking using c-Abl cavities as query cavities.

PDB Code	SCOP Kinase Subfamily	Consensus Score [R_1]
1iep	Abelsone tyrosine kinase (abl)	38.60
1fpu	Abelsone tyrosine kinase (abl)	37.71
1opj	Abelsone tyrosine kinase (abl)	33.72
1m52	Abelsone tyrosine kinase (abl)	32.06
1opk	Abelsone tyrosine kinase (abl)	31.78
1t46	c-Kit Kinase	23.89
1uwh	B-Raf proto-oncogene PK	20.05
1qpd	Lymphocyte kinase (lck)	19.87
1qpc	Lymphocyte kinase (lck)	19.86
1qpj	Lymphocyte kinase (lck)	19.68
3lck	Lymphocyte kinase (lck)	19.54
1qpe	Lymphocyte kinase (lck)	19.46
1byg	Carboxyl-terminal src kinase (csk)	17.82
1m17	EGF receptor tyrosine kinase (Erbb-1)	17.52
2fgi	Fibroblast growth factor receptor 1	17.37
1pkg	c-Kit kinase	17.10
1m14	EGF receptor tyrosine kinase (Erbb-1)	16.21
1ias	Type I TGFβ receptor R4	15.91
1ove	MAP kinase p38α	15.85
1k9a	Carboxyl-terminal src kinase (csk)	15.64

sites. Cavbase provides an automated procedure to obtain a reasonable alignment of kinase cavities focusing on functionally important parts of their active sites. This procedure supports the similarity analysis of protein structures, identifying which structural regions differ between different kinases and which correspond. A clustering approach that allows for a fast and efficient clustering analysis of large datasets in an automated fashion was implemented. Based on the obtained clustering solutions in combination with consensus ranking of the hit lists, it was possible to identify kinase subfamilies that exhibit similar areas in their binding sites. A detailed analysis of the Cavbase clustering results reveals that Cavbase provides a relevant classification of protein kinases. Kinases are clustered on

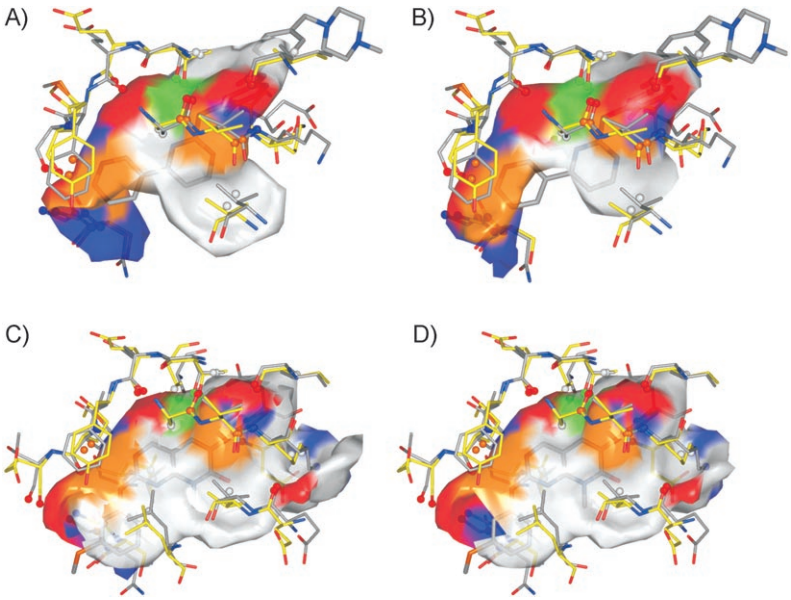

Figure 11. Unexpected high similarities between the binding sites of Lck and c-Abl. The similar regions are shown in two activation states of Lck and c-Abl. The amino acids detected as similar are shown together with the pseudocenters, the bound ligand **4** and the cavity surface of A) inactive c-Abl (PDB code: 1iep) and B) Lck (PDB code: 3lck). The same Lck cavity in comparison with active c-Abl (PDB code: 1m52) and **5** is also shown with the matched cavity surface for C) c-Abl and D) Lck. The similarity in different activation states is substantial and comprises areas next to the hinge, adenine, and the sugar binding region. In the case of active c-Abl, the similarity is even higher; furthermore, areas around the selectivity pocket are detected as similar.

Table 8. Consensus ranking using Lck cavities as query cavities.

PDB Code	SCOP Kinase Subfamily	Consensus Score [R_1]
1qpd	Lymphocyte kinase (lck)	48.05
1qpe	Lymphocyte kinase (lck)	46.97
1qpc	Lymphocyte kinase (lck)	46.97
3lck	Lymphocyte kinase (lck)	44.68
1qpj	Lymphocyte kinase (lck)	42.89
1m14	EGF receptor tyrosine kinase (Erbb-1)	27.19
2fgi	Fibroblast growth factor receptor 1	26.96
1mqb	Ephrin A2 (Epha2) receptor PK	25.98
1m17	EGF receptor tyrosine kinase (Erbb-1)	25.84
1jpa	Ephb2 receptor tyrosine kinase	25.62
2src	c-src protein tyrosine kinase	25.18
1k9a	Carboxyl-terminal src kinase (csk)	24.65
1ias	Type I TGFβ receptor R4	24.61
1pkg	c-Kit kinase	24.11
1ksw	c-src protein tyrosine kinase	23.49
1b6c	Type I TGFβ receptor R4	23.43
1qcf	Hematopoietic cell kinase Hck	23.34
1ad5	Hematopoietic cell kinase Hck	23.18
1m52	Abelsone tyrosine kinase (abl)	23.15
1rqq	Insulin receptor	23.14

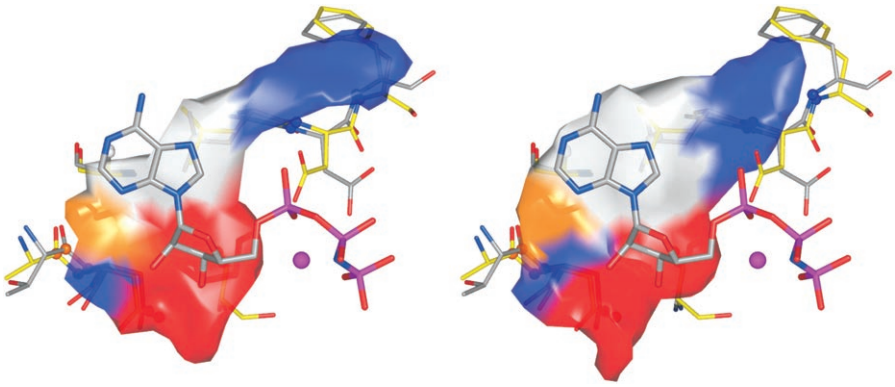

Figure 12. Similar areas in the binding sites of MAP kinase p38γ (PDB code: 1cm8) and MAP kinase p38α (PDB code: 1p38). The matching area of both cavities is rather small and is located next to the hinge region. The amino acids, pseudocenters, and bound ATP are displayed. On the left, the superposition is shown with the color-coded surface of the p38γ pocket (1cm8), and on the right, with the color-coded surface of p38α (1p38).

the superfamily level, and these results are in good agreement with other classification schemes based on sequence or structural information (SCOP classification). Unlike sequence-based methods, Cavbase is also able to distinguish between different activation states of kinases in their subfamilies. Using examples from the CDK2 and c-Abl superfamilies, the grouping of these superfamilies into distinct clusters could be demonstrated.

Furthermore, particular interest was taken in cases where the results of the sequence-based clustering and the Cavbase clustering were different. Cases where high similarity in cavity space is accompanied by low sequence similarity may suggest an unexpected functional similarity and probable cross-reactivity between two kinases. During the optimization of kinase inhibitors, selective inhibition of the kinase under consideration is of utmost importance for the development of effective and safe drugs. The selection of protein kinases to use for selectivity screens can be facilitated using the similarity information from the cavity space, which complements information obtained from sequence alignments or small molecule inhibition profiles.

Several cases of cross-reactivity between kinases confirmed by biochemical assay data could be rationalized using Cavbase. The similarity between c-Abl/c-Kit and c-Abl/Lck becomes evident from the Cavbase analysis.

Protein kinases show high flexibility in their active sites. The present study highlights the importance of the analysis of multiple X-ray structures for each kinase subfamily. The relationship between different kinases becomes more apparent when kinases in distinct activation states are used in the analysis. With an increasing number of available kinase X-ray structures, a better understanding of similarity relationships will be possible, allowing for a more fine-grained analysis.

Acknowledgements

The authors thank Dr. Hans Matter (Sanofi-Aventis, Frankfurt) and Dr. Gerhard Müller (GPC, Munich) for helpful discussions regarding the classification of protein kinases. We also thank Dr. Ulf Börjesson, Dr. Christoph Sotriffer, and Dr. Simon Cotrell for carefully reading and critically reviewing the manuscript.

Keywords: kinase selectivity · protein binding sites · protein kinases · similarity analysis · structural analysis

[1] G. Manning, D. B. Whyte, R. Martinez, T. Hunter, S. Sudarsanam, *Science* **2002**, *298*, 1912 – 1934.
[2] B. Nolen, S. Taylor, G. Ghosh, *Mol. Cell* **2004**, *15*, 661 – 675.
[3] R. A. Engh, D. Bossemeyer, *Pharmacol. Ther.* **2002**, *93*, 99 – 111.
[4] M. Huse, J. Kuriyan, *Cell* **2002**, *109*, 275 – 282.
[5] J. Saklatvala, *Curr. Opin. Pharmacol.* **2004**, *4*, 372 – 377.
[6] A. Matter, *Drug Discovery Today* **2001**, *6*, 1005 – 1024.
[7] J. Dancey, E. A. Sausville, *Nat. Rev. Drug Discovery* **2003**, *2*, 296 – 313.
[8] P. Cohen, *Nat. Rev. Drug Discovery* **2002**, *1*, 309 – 315.
[9] P. M. Traxler, *Expert Opin. Ther. Pat.* **1998**, *8*, 1599 – 1625.
[10] M. E. Noble, J. A. Endicott, L. N. Johnson, *Science* **2004**, *303*, 1800 – 1805.
[11] A. Arora, E. M. Scholar, *J. Pharmacol. Exp. Ther.* **2005**, *315*, 971 – 979.
[12] B. M. Klebl, G. Muller, *Expert Opin. Ther. Targets* **2005**, *9*, 975 – 993.
[13] Z. A. Knight, K. M. Shokat, *Chem. Biol.* **2005**, *12*, 621 – 637.
[14] S. Cheek, H. Zhang, N. V. Grishin, *J. Mol. Biol.* **2002**, *320*, 855 – 881.
[15] S. Cheek, K. Ginalski, H. Zhang, N. V. Grishin, *BMC Struct. Biol.* **2005**, *5*, 6.
[16] I. N. Shindyalov, P. E. Bourne, *Protein Eng.* **1998**, *11*, 739 – 747.
[17] L. Lo Conte, S. E. Brenner, T. J. Hubbard, C. Chothia, A. G. Murzin, *Nucleic Acids Res.* **2002**, *30*, 264 – 267.
[18] F. M. Pearl, D. Lee, J. E. Bray, I. Sillitoe, A. E. Todd, A. P. Harrison, J. M. Thornton, C. A. Orengo, *Nucleic Acids Res.* **2000**, *28*, 277 – 282.
[19] S. Goldsmith-Fischman, B. Honig, *Protein Sci.* **2003**, *12*, 1813 – 1821.
[20] R. V. Spriggs, P. J. Artymiuk, P. Willett, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 412 – 421.
[21] A. C. Wallace, N. Borkakoti, J. M. Thornton, *Protein Sci.* **1997**, *6*, 2308 – 2323.

[22] R. B. Russell, *J. Mol. Biol.* **1998**, *279*, 1211 – 1227.
[23] A. Stark, R. B. Russell, *Nucleic Acids Res.* **2003**, *31*, 3341 – 3344.
[24] G. J. Kleywegt, *J. Mol. Biol.* **1999**, *285*, 1887 – 1897.
[25] P. P. Wangikar, A. V. Tendulkar, S. Ramya, D. N. Mali, S. Sarawagi, *J. Mol. Biol.* **2003**, *326*, 955 – 978.
[26] T. Hamelryck, *Proteins* **2003**, *51*, 96 – 108.
[27] M. Rosen, S. L. Lin, H. Wolfson, R. Nussinov, *Protein Eng.* **1998**, *11*, 263 – 277.
[28] A. Shulman-Peleg, R. Nussinov, H. J. Wolfson, *J. Mol. Biol.* **2004**, *339*, 607 – 633.
[29] K. Kinoshita, J. Furui, H. Nakamura, *J. Struct. Funct. Genomics* **2002**, *2*, 9 – 22.
[30] A. R. Poirrette, P. J. Artymiuk, D. W. Rice, P. Willett, *J. Comput.-Aided Mol. Des.* **1997**, *11*, 557 – 569.
[31] J. V. Lehtonen, K. Denessiouk, A. C. May, M. S. Johnson, *Proteins* **1999**, *34*, 341 – 355.
[32] S. J. Pickering, A. J. Bulpitt, N. Efford, N. D. Gold, D. R. Westhead, *Comput. Chem.* **2001**, *26*, 79 – 84.
[33] M. Jambon, A. Imberty, G. Deleage, C. Geourjon, *Proteins* **2003**, *52*, 137 – 145.
[34] T. A. Binkowski, L. Adamian, J. Liang, *J. Mol. Biol.* **2003**, *332*, 505 – 526.
[35] N. D. Gold, R. M. Jackson, *J. Mol. Biol.* **2006**, *355*, 1112 – 1124.
[36] T. Naumann, H. Matter, *J. Med. Chem.* **2002**, *45*, 2366 – 2378.
[37] P. J. Goodford, *J. Med. Chem.* **1985**, *28*, 849 – 857.
[38] R. C. Wade, P. J. Goodford, *J. Med. Chem.* **1993**, *36*, 148 – 156.
[39] M. A. Kastenholz, M. Pastor, G. Cruciani, E. E. Haaksma, T. Fox, *J. Med. Chem.* **2000**, *43*, 3033 – 3044.
[40] M. Vieth, R. E. Higgs, D. H. Robertson, M. Shapiro, E. A. Gragg, H. Hemmerle, *Biochim. Biophys. Acta* **2004**, *1697*, 243 – 257.
[41] M. Vieth, J. J. Sutherland, D. H. Robertson, R. M. Campbell, *Drug Discovery Today* **2005**, *10*, 839 – 846.
[42] H. Daub, K. Godl, D. Brehmer, B. Klebl, G. Muller, *Assay Drug Dev. Technol.* **2004**, *2*, 215 – 224.
[43] K. Godl, J. Wissing, A. Kurtenbach, P. Habenberger, S. Blencke, H. Gutbrod, K. Salassidis, M. Stein-Gerlach, A. Missio, M. Cotten, H. Daub, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 15434 – 15439.
[44] M. A. Fabian, W. H. Biggs III, D. K. Treiber, C. E. Atteridge, M. D. Azimioara, M. G. Benedetti, T. A. Carter, P. Ciceri, P. T. Edeen, M. Floyd, J. M. Ford, M. Galvin, J. L. Gerlach, R. M. Grotzfeld, S. Herrgard, D. E. Insko, M. A. Insko, A. G. Lai, J. M. Lelias, S. A. Mehta, Z. V. Milanov, A. M. Velasco, L. M. Wodicka, H. K. Patel, P. P. Zarrinkar, D. J. Lockhart, *Nat. Biotechnol.* **2005**, *23*, 329 – 336.
[45] S. Schmitt, M. Hendlich, G. Klebe, *Angew. Chem.* **2001**, *113*, 3237 – 3241; *Angew. Chem. Int. Ed.* **2001**, *40*, 3141 – 3144.
[46] S. Schmitt, D. Kuhn, G. Klebe, *J. Mol. Biol.* **2002**, *323*, 387 – 406.
[47] D. Kuhn, N. Weskamp, S. Schmitt, E. Hullermeier, G. Klebe, *J. Mol. Biol.* **2006**, *359*, 1023 – 1044.
[48] M. Hendlich, *Acta Crystallogr. Sect. D* **1998**, *54*, 1178 – 1182.
[49] M. Hendlich, A. Bergner, J. Gunther, G. Klebe, *J. Mol. Biol.* **2003**, *326*, 607 – 620.
[50] J. Günther, A. Bergner, M. Hendlich, G. Klebe, *J. Mol. Biol.* **2003**, *326*, 621 – 636.
[51] M. Hendlich, F. Rippmann, G. Barnickel, *J. Mol. Graphics J. Mol. Graphics Modell.* **1997**, *15*, 359 – 363, 389.
[52] C. Bron, J. Kerbosch, *Commun. ACM* **1973**, *16*, 575 – 577.
[53] W. L. DeLano, *The PyMOL Molecular Graphics System* **2004**, DeLano Scientific, LLC.
[54] G. Karypis, **2002**, http://glaros.dtc.umn.edu/gkhome/views/cluto.
[55] N. Salim, J. Holliday, P. Willett, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 435 – 442.
[56] J. Hert, P. Willett, D. J. Wilton, P. Acklin, K. Azzaoui, E. Jacoby, A. Schuffenhauer, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1177 – 1185.
[57] M. Whittle, V. J. Gillet, P. Willett, A. Alex, J. Loesel, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1840 – 1848.
[58] J. F. Gibrat, T. Madej, S. H. Bryant, *Curr. Opin. Struct. Biol.* **1996**, *6*, 377 – 385.
[59] W. R. Pearson, D. J. Lipman, *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 2444 – 2448.
[60] M. Cherry, D. H. Williams, *Curr. Med. Chem.* **2004**, *11*, 663 – 673.
[61] G. L. Johnson, R. Lapadat, *Science* **2002**, *298*, 1911 – 1912.
[62] H. Schramek, *News Physiol. Sci.* **2002**, *17*, 62 – 67.
[63] S. Kumar, J. Boehm, J. C. Lee, *Nat. Rev. Drug Discovery* **2003**, *2*, 717 – 726.
[64] T. Fox, J. T. Coll, X. Xie, P. J. Ford, U. A. Germann, M. D. Porter, S. Pazhanisamy, M. A. Fleming, V. Galullo, M. S. Su, K. P. Wilson, *Protein Sci.* **1998**, *7*, 2249 – 2255.
[65] J. C. Lee, J. T. Laydon, P. C. McDonnell, T. F. Gallagher, S. Kumar, D. Green, D. McNulty, M. J. Blumenthal, J. R. Heys, S. W. Landvatter, J. E. Strickler, M. M. McLaughlin, I. R. Siemens, S. M. Fisher, G. P. Livi, J. R. White, J. L. Adams, P. R. Young, *Nature* **1994**, *372*, 739 – 746.
[66] A. Cuenda, J. Rouse, Y. N. Doza, R. Meier, P. Cohen, T. F. Gallagher, P. R. Young, J. C. Lee, *FEBS Lett.* **1995**, *364*, 229 – 233.
[67] Z. Wang, B. J. Canagarajah, J. C. Boehm, S. Kassisà, M. H. Cobb, P. R. Young, S. Abdel-Meguid, J. L. Adams, E. J. Goldsmith, *Structure* **1998**, *6*, 1117 – 1128.
[68] J. C. Lee, S. Kassis, S. Kumar, A. Badger, J. L. Adams, *Pharmacol. Ther.* **1999**, *82*, 389 – 397.
[69] R. J. Gum, M. M. McLaughlin, S. Kumar, Z. Wang, M. J. Bower, J. C. Lee, J. L. Adams, G. P. Livi, E. J. Goldsmith, P. R. Young, *J. Biol. Chem.* **1998**, *273*, 15605 – 15610.
[70] C. E. Fitzgerald, S. B. Patel, J. W. Becker, P. M. Cameron, D. Zaller, V. B. Pikounis, S. J. O'Keefe, G. Scapin, *Nat. Struct. Biol.* **2003**, *10*, 764 – 769.
[71] J. D. Rowley, *Nature* **1973**, *243*, 290 – 293.
[72] B. J. Druker, *J. Clin. Oncol.* **2003**, *21*, 239s – 245s.
[73] T. G. Lugo, A. M. Pendergast, A. J. Muller, O. N. Witte, *Science* **1990**, *247*, 1079 – 1082.
[74] R. Capdeville, E. Buchdunger, J. Zimmermann, A. Matter, *Nat. Rev. Drug Discovery* **2002**, *1*, 493 – 502.
[75] B. Nagar, W. G. Bornmann, P. Pellicena, T. Schindler, D. R. Veach, W. T. Miller, B. Clarkson, J. Kuriyan, *Cancer Res.* **2002**, *62*, 4236 – 4243.
[76] F. Zhang, A. Strand, D. Robbins, M. H. Cobb, E. J. Goldsmith, *Nature* **1994**, *367*, 704 – 711.
[77] M. J. Robinson, P. C. Harkins, J. Zhang, R. Baer, J. W. Haycock, M. H. Cobb, E. J. Goldsmith, *Biochemistry* **1996**, *35*, 5641 – 5646.
[78] B. J. Canagarajah, A. Khokhlatchev, M. H. Cobb, E. J. Goldsmith, *Cell* **1997**, *90*, 859 – 869.
[79] L. Shewchuk, A. Hassell, B. Wisely, W. Rocque, W. Holmes, J. Veal, L. F. Kuyper, *J. Med. Chem.* **2000**, *43*, 133 – 138.
[80] C. Pargellis, L. Tong, L. Churchill, P. F. Cirillo, T. Gilmore, A. G. Graham, P. M. Grob, E. R. Hickey, N. Moss, S. Pav, J. Regan, *Nat. Struct. Biol.* **2002**, *9*, 268 – 272.
[81] C. I. Chang, B. E. Xu, R. Akella, M. H. Cobb, E. J. Goldsmith, *Mol. Cell* **2002**, *9*, 1241 – 1249.
[82] J. E. Stelmach, L. Liu, S. B. Patel, J. V. Pivnichny, G. Scapin, S. Singh, C. E. Hop, Z. Wang, J. R. Strauss, P. M. Cameron, E. A. Nichols, S. J. O'Keefe, E. A. O'Neill, D. M. Schmatz, C. D. Schwartz, C. M. Thompson, D. M. Zaller, J. B. Doherty, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 277 – 280.
[83] A. Trejo, H. Arzeno, M. Browner, S. Chanda, S. Cheng, D. D. Comer, S. A. Dalrymple, P. Dunten, J. Lafargue, B. Lovejoy, J. Freire-Moar, J. Lim, J. McIntosh, J. Miller, E. Papp, D. Reuter, R. Roberts, F. Sanpablo, J. Saunders, K. Song, A. Villasenor, S. D. Warren, M. Welch, P. Weller, P. E. Whiteley, L. Zeng, D. M. Goldstein, *J. Med. Chem.* **2003**, *46*, 4702 – 4713.
[84] Z. Wang, P. C. Harkins, R. J. Ulevitch, J. Han, M. H. Cobb, E. J. Goldsmith, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2327 – 2332.
[85] S. B. Patel, P. M. Cameron, B. Frantz-Wattley, E. O'Neill, J. W. Becker, G. Scapin, *Biochim. Biophys. Acta* **2004**, *1696*, 67 – 73.
[86] K. P. Wilson, M. J. Fitzgibbon, P. R. Caron, J. P. Griffith, W. Chen, P. G. McCaffrey, S. P. Chambers, M. S. Su, *J. Biol. Chem.* **1996**, *271*, 27 696 – 27 700.
[87] S. Bellon, M. J. Fitzgibbon, T. Fox, H. M. Hsiao, K. P. Wilson, *Struct. Fold. Des.* **1999**, *7*, 1057 – 1065.
[88] X. Xie, Y. Gu, T. Fox, J. T. Coll, M. A. Fleming, W. Markland, P. R. Caron, K. P. Wilson, M. S. Su, *Structure* **1998**, *6*, 983 – 991.
[89] G. Scapin, S. B. Patel, J. Lisnock, J. W. Becker, P. V. LoGrasso, *Chem. Biol.* **2003**, *10*, 705 – 712.

Received: March 31, 2007
Revised: June 6, 2007
Published online on August 10, 2007

