Published in final edited form as:

Biomol Eng. 2006 December ; 23(6): 307–315.

Feature-similarity kinase classifier as a ligand engineering tool

Sridhar Maddipati¹ and Ariel Fernández^{2,3,*}

¹ School of Chemical Engineering, Purdue University, West Lafayette, IN 47906

² Department of Bioengineering, Rice University, Houston, TX 77005

³ Department of Computer Science, The University of Chicago, Chicago, IL 60637

Abstract

Kinases have been often targeted in drug therapy aimed at blocking signaling pathways. However, the conservation of protein structure across homologs often leads to uncontrolled cross reactivity. On the other hand, sticky packing defects in proteins are typically *not* conserved across homologs, making them ligand-anchoring sites potentially important to enhance selectivity. Thus, we introduce a hierarchical clustering of PDB-reported kinases according to packing differences. This kinome partitioning is highly correlated with proximity relations arising from the pharmacological profiling of kinases. A variable packing sensitivity is observed for individual drugs, with highly promiscuous ligands being the most insensitive to packing differences. Our classifier enables a strategy to design selective inhibitors.

INTRODUCTION

Kinases are signal transducers in the cell and thus, kinase inhibition has become a central therapeutic strategy to block signaling pathways [1–6]. However, high-throughput screening data reveals that most kinase inhibitors of pharmacological relevance exhibit unexpected cross reactivity, mostly due to the high degree of structural conservation within the kinase superfamily [7]. Cross reactivity does not necessarily hinder drug activity but it is frequently linked to toxic side effects [2].

While folds are typically conserved across homologs, packing defects in the form of incompletely packed hydrogen bonds, the so-called *dehydrons* [8–11], are often *not* conserved [8]. Side effects resulting from off-target ligand binding may be then minimized by selectively targeting nonconserved dehydrons that may in effect function as selectivity filters [9]. These observations prompt us to introduce a “packing metric” to classify kinase space, to infer cross-reactivity regions, to identify drugs most sensitive to packing differences and to guide the impact of drugs towards targets of clinical relevance by focusing on packing differences across targets.

Dehydrons are targetable because they constitute sticky sites on the protein surface, with a propensity to become further dehydrated by promoting protein associations [10–11]. The term *packing* refers to a cluster of nonpolar groups framing an anhydrous microenvironment. Dehydrons are singled out by an insufficient number of nonpolar groups in their microenvironments. Their stickiness arises from the reduction of charge screening as nonpolar

(*) Send Correspondence to arifer@rice.edu Tel.: 713 348 3681; FAX: 713 348 3699

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

groups come into proximity of hydrogen bonds: water exclusion enhances and stabilizes hydrogen bonds, compensating for the dehydration penalty [11].

Dehydrons have been shown to be markers of protein-ligand association [10–11]. Thus, examination of the 814 non-redundant protein-inhibitor PDB complexes revealed that 488 of them have binding cavities with average hydrophobicity, not significantly higher than the rest of the surface. On the other hand, the interfaces of the complexes were reexamined to determine whether inhibitors were “dehydron wrappers”, that is, whether nonpolar groups of inhibitors penetrated the desolvation domain of dehydrons. This feature was found in 631 complexes, and *invariably* found in the 488 complexes whose binding cavities presented average or below average surface hydrophobicity [9]. Thus, in most PDB protein-inhibitor complexes, the ligand is in effect a dehydron wrapper, as illustrated in Figs. 1, 2.

Comparing dehydron patterns of kinases, we now construct a hierarchical clustering of the pharmacokinome. This classifier sets an upper bound to the level of selectivity that may be achieved by a rational structure-based design of compounds that may serve as ATP-displacing ligands. Such inferences require that we assess differences in the dehydron patterns of the ATP-binding sites. By comparing such patterns, we construct a “packing distance” that enables a partition of target space into packing similarity clusters.

The structure-based hierarchical classifier is validated by contrasting it against the pattern of proximity relations arising from the pharmacological profiling of kinases [7]. On the other hand, individual drugs are endowed with highly variable levels of efficacy in achieving selectivity by intermolecular improvement of protein packing.

The packing classifier is experimentally validated by introducing a methodology to engineer drug selectivity and redirect the drug impact by focusing on nonconserved packing defects [12]. Thus, the modulation of target selectivity is assayed on re-designs of lead compounds guided by packing differences among its targets. We focus on an extreme case of clinical relevance of the parental compound: Gleevec (Imatinib, STI571), the powerful anticancer drug [7]. Our aim is to device a selectivity switch by a judicious modification of the parental drug that enables us to sharpen and redirect the inhibitory impact of the ligand.

MATERIALS AND METHODS

Identifying packing defects (dehydrons) in proteins

Packing defects adopt the form of under-wrapped backbone hydrogen bonds, the so-called dehydrons [8], and may be identified from the atomic coordinates of protein structure, according to the following tenets (Fig. 2): 1) The extent of intramolecular hydrogen-bond desolvation, ρ , in a monomeric structure is quantified by determining the number of nonpolar groups (carbonaceous, not covalently bonded to an electrophilic atom) contained within a desolvation domain. 2) The desolvation domain is defined as two intersecting balls of fixed radius centered at the α -carbons of the residues paired by the backbone amide-carbonyl hydrogen bond. The statistics of hydrogen-bond wrapping depend on the choice of desolvation radius, but the tails of the distribution invariably single out the same dehydrons in a given structure over a 6.0–7.1 Å range in the adopted radius. In this work the value 6.2 Å was used throughout.

In folds for soluble proteins at least two thirds of the backbone hydrogen bonds are wrapped on average by $\rho=26.6\pm7.5$ nonpolar groups ([8–11]; for other measures of wrapping see [10]). Dehydrons are here defined as hydrogen bonds whose extent of wrapping lies in the tails of the distribution, i.e. with 19 or fewer nonpolar groups in their desolvation domains, so their ρ -value is below the mean, minus one Gaussian dispersion.

Packing distance

For a protein of length N with reported structure, a matrix of dehydrons or underwrapped H-bonds D_{ij} , $i, j = 1, 2, \dots, N$ is constructed by choosing $D_{ij} = 1$ if residues i and j are paired by a dehydron and $D_{ij} = 0$ otherwise. Then the Hamming distance $M_H(X, Y)$, which serves as an indicator of the packing distance between proteins X and Y is given by:

$$M_H(X, Y) = \sum_{i < j} |D_{i,j}(X) - D_{i,j}(Y)|$$

where $\mathbf{D}(X)$ and $\mathbf{D}(Y)$ represent respectively the dehydron matrices for proteins X and Y .

Pharmacological distance

The pharmacological distance between two kinases is defined as the Euclidean distance between the normalized affinity vectors with entries corresponding to the negative logarithm of the binding constants reported in [7]. The pharmacological matrix (PM) is obtained by calculating the pharmacological distances between all the pairs of 32 kinases :

$$PM_{X,Y} = \|X - Y\| = \sqrt{\sum_{n \in \text{set of inhibitors}} (X_n - Y_n)^2}$$

where X_n , Y_n represent respectively the normalized values of the negative logarithm of binding constants for complexation of kinase X and kinase Y with drug inhibitor n .

Synthesis of methylated Gleevec

The synthesis of the modified Gleevec compound resulting from methylation at position 6 on the piperidine ring recapitulates the Novartis patent WO03027100A1, 2003 (see also [13]), replacing a single reactant, N,N-dimethylformamide dimethylacetal, for N,N-dimethylformamide methyl-dimethylacetal in the first step of synthesis. The original synthesis begins with the treatment of 3-acetylpyridine with N,N-dimethylformamide dimethylacetal to give 3-dimethylamino-1-(3-pyridyl)-2-propene-1-one. This last compound is originally treated with 2-methyl-5-nitroaniline-guanidine nitrate salt and sodium hydroxide in refluxing isopropanol to yield N-(2-methyl-5-nitrophenyl)-4-(3-pyridyl)-2-pyrimidine-amine, the core intermediate in Gleevec synthesis. Instead, the treatment of starting compound 3-acetylpyridine with N,N-dimethylformamide methyl-dimethylacetal leads to 3-methyl,3-dimethylamino-1-(3-pyridyl)-2-propene-1-one with 44% yield, and ultimately to N-(2-methyl-5-nitrophenyl)-4-(3-pyridyl)-2-(6-methyl-pyrimidine)-amine with 51% yield. The latter is the core intermediate in the synthetic pathway that yields Gleevec methylated at position 6 on the piperidine ring, as required for this work.

Kinase production: Cloning of kinase genes

The genes of kinases of interest were amplified from full-length human complementary DNA (cDNA) library (Invitrogen) by polymerase chain reaction (PCR) using specific primers that correspond to the N- and C-termini of the kinase domains, respectively. For example, *Primer 1*: 5'-GTAG gga tcc ATG GCA CAT CAC CAC CAT CAC GTG ACC TAC AAA TAT TTA CAG AAA CC-3' (containing a Bam H I site, underlined) and *Primer 2*: 5'-GTCCC tct aga TCA AAT ATG ATT GGT GCT CTC TGA AAT CTG-3' (containing a Xba I site, underlined) were used to amplify the gene of C-Kit (residues 544–935 with the kinase insertion domain at residues 694–753 replaced with Thr-Ser). The PCR products were digested with appropriate restriction enzymes (Bam H I and Xba I) and inserted into the vector pFastBac-HT that has N-terminal his-tag and TEV site (Invitrogen). The recombinants were confirmed by DNA sequencing.

The kinases were expressed in baculovirus-insect cell system. All recombinants have an N-terminal histidine-tag for ease of purification. Recombinant proteins were purified from cells by metal affinity chromatography and size-exclusion chromatography.

The specifics on molecular cloning, including primer sequences and restriction enzymes, for each particular kinase followed the protocols described in the respective primary sources: Bcr-Abl [12], C-Kit [14] and Lck [15].

Production of recombinant baculovirus

The recombinant plasmids were transformed into *Escherichia coli* (*E. coli*) DH10Bac containing bacmid DNA. The transposition between donor plasmid and bacmid yields recombinant bacmids in white colonies on IPTG/Blue-gal-containing plates. The recombinant bacmids were expanded, purified, and transfected into sf21 cells. Baculovirus obtained from the transfection was used to infect sf21 cells for expression.

Expression and purification of recombinant kinase domains

Recombinant proteins were purified from sf21 cells by Ni-NTA affinity chromatography (Ni-NTA Superflow Agarose, Qiagen). The his-tag was removed from purified proteins by TEV (Tobacco Etch Virus protease) digestion. The recombinant proteins were further purified using anion exchange (macropep high Q-Sepharose, Bio-Rad) and subsequent affinity chromatography. The columns used were 15Phenyl and Sephacryl S200. The proteins were further purified on size-exclusion chromatography. The yields were in the range of 2–10mg per liter culture.

Spectrophotometric kinetic assay

To determine the level of selectivity of drug inhibitors, kinetic assays of the inhibition of multiple kinases have been conducted. To measure the rate of phosphorylation due to kinase activity in the presence of inhibitors, a standard spectrophotometric assay [9,12] has been adopted in which the adenosine diphosphate production is coupled to the NADH oxidation and determined by absorbance reduction at 340nm. Reactions were carried out at 35°C in 500l of buffer (100mM Tris-HCl, 10mM MgCl₂, 0.75mM ATP, 1mM phosphoenol pyruvate, 0.33mM NADH, 95 units/ml pyruvate kinase, pH7.5). Each kinetic run for inhibition was repeated eight times, with three phosphorylation measurements in each run at 400nM differences in inhibitor concentration. The substrate peptides used for kinase phosphorylation (Invitrogen/Biaffin) were: AEEEIYGEFEAKKKKG for unphosphorylated but active Bcr-Abl [12]; KVEEINGNNYVYIDPTQLPY for active C-Kit [16]; GLARLIEDNEYTAREGAKFPI for active Lck [17].

RESULTS

Pharmacological relevance of a packing-based kinase classifier

To determine the boundaries in ligand selectivity based on the molecular engineering of wrappers of packing defects, we first introduce a packing distance between kinases defined by comparing the dehydron patterns of the ATP sites. Thus, the packing distance between two kinases may be determined by aligning and comparing their dehydron matrices (Methods) at the ligand-binding site and quantifying the extent of feature similarity by computing an appropriate normalized distance between the dehydron matrices.

The evolutionary relatedness of kinases ensures a considerable structural similarity, which in turn, facilitates the alignment needed for the feature similarity analysis proposed above. Thus, Fig. 3a shows the structural alignment of the Bcr-Abl kinase (PDB.1FPU) and the C-Kit kinase (PDB.1T46), both Gleevec targets [7]. The packing distance is then defined by comparing the

dehydron patterns within the ligand-binding site Λ . The pattern comparison is illustrated in Fig. 3b which displays the dehydron/hydrogen bond matrices restricted to Λ for the C-Kit kinase (dehydrons as green entries) and Bcr-Abl kinase (dehydrons as red entries). The set Λ consists of ligand-contacting residues identified by the presence of ligand atoms within balls of radius 6.2Å centered at the α -carbons of the residues. This radius is consistent with the defined dehydron microenvironment, as shown in Methods. Since different protein chains typically have different numberings, it is best to retain the numbering in one of them and define the normalized packing distance $d_{\text{pack}}(X, Y)$ as:

$$d_{\text{pack}}(X, Y) = L(X, Y)^{-1} \sum_{i < j; i, j \in \Lambda} |D_{ij}(X) - D_{f(i)f(j)}(Y)|,$$

where D_{ij} is the dehydron characteristic function defined in Methods, the bijective function f is the structure-aided alignment-based mapping of residues of X onto residues in Y (see Fig. 3a), and $L(X, Y)$ =number of residue pairs (i, j) , with $i, j \in \Lambda$ such that either residues i and j are paired by a dehydron in X or $F(i)$ and $F(j)$ are paired by a dehydron in Y . For example, if X =C-Kit and Y =Bcr-Abl, we obtain (cf. Fig. 3b): number of dehydrons in X within Λ = 6; number of dehydrons in Y within Λ = 3; $L(X, Y)$ = 6; $d_{\text{pack}}(X, Y)$ = 1/2. The reader should note that the dehydron/hydrogen bond matrices are symmetric.

To validate our feature-similarity analysis, a hierarchical clustering of kinases based on the packing metric was determined across the 32 kinases reported in PDB for which independent affinity fingerprinting is available [7]. 17 inhibitors were selected from a pool of 20 that have been independently assayed for cross reactivity against a set of 113 kinases [7]. Three inhibitors, staurosporine, SU11248 and EKB569 were excluded from the computation, since their higher level of promiscuity [7] precludes a clear pharmacological assessment. The packing-based hierarchical clustering of the kinases is built so that the minimum number of nodes along a walk from one kinase-node to another is proportional to the normalized packing distance between the two kinases (Fig. 4a). The hierarchical clustering becomes thus a representation of a packing-distance matrix and, as such may be contrasted with a pharmacological distance matrix. The latter is obtained by computing the Euclidean distance between ligand-affinity vectors [7] in $\mathbf{R}^{17}$ with entries given in $-\ln$ scale (or dimensionless $\Delta G/RT$ units, with ΔG =Gibbs free energy change associated with binding, R =universal gas constant, T =absolute temperature).

While visual inspection and comparison of trees is useful, the lack of uniqueness of trees is a cause for concern. A direct comparison of distance matrices provides a more rigorous comparison of measured properties. In general, if a distance matrix is to be represented faithfully by a tree, it must satisfy the so-called ‘four-point condition’ ($d(V, W) + d(X, Y) \leq \max[d(V, X) + d(W, Y), d(V, Y) + d(W, X)]$, where d denotes distance) [18]. This condition implies the triangular inequality condition (satisfied by any distance), since the latter constitutes a special case of the former obtained when $Y=X$: $d(V, W) = d(V, W) + d(X, X) \leq \max[d(V, X) + d(W, X), d(V, X) + d(W, X)] = d(V, X) + d(W, X)$. However, the four-point condition is substantially more restrictive. This requirement is not generally satisfied [18] and implies that typical biological trees will not uniquely represent a given biological distance matrix [18]. Thus direct comparison of distance matrices becomes a more reliable technique.

By plotting packing versus pharmacological distance (Fig. 4b) for each pair of kinases reported in PDB and fingerprinted for affinity against the 17 drug ligands, we establish a strong correlation ($R^2=0.9028$). *This correlation reveals that the pattern of packing defects is actually an operational selectivity filter for drug design, even though drugs were not purposely designed to wrap packing defects in proteins.* In contrast, previous attempts at correlating the sequence based evolutionary distances with the pharmacological distances have not been very successful [19].

Thus, the packing blueprint of the protein target emerges as a selectivity filter only when proteins are fingerprinted against a whole set of drug compounds that covers the entire kinase space sampled (Fig. 4a, b). This picture is not obvious at the level of individual drugs, where the responsiveness to packing differences varies widely, depending on the structural features that promote binding (Fig. 4c). Thus, GW.2016 is the most sensitive to packing differences, while the highly promiscuous staurosporine is the least responsive. The packing sensitivity of a drug may be defined to serve as an indicator of the discriminatory power of the drug based on differences in the dehydron patterns of its purported targets. In this context, a drug n with “perfect” discriminatory power would have a cross reactivity that follows *exactly* the normalized packing distance, according to the correlation defined in Fig. 4b: $d_n(X, Y) = \lambda d_{\text{pack}}(X, Y)$, with $\lambda=0.043$ = proportionality constant from the packing vs. pharmacological distance correlation and d_n =pharmacological distance defined exclusively with respect to drug n , that is, $d_n(X, Y) = |X_n - Y_n|$, with X_n, Y_n respectively the normalized binding constants (Methods) for association of kinase X and kinase Y with drug n . We may then quantify departures from the perfect discriminatory power by introducing a packing sensitivity.

The packing sensitivity of drug n is quantified according to the following operational steps: a) the differences between properly scaled packing distance and pharmacological distance limited to drug n are determined for every kinase pair inhibited by drug n ; b) for the sake of amplification, the average of such differences are weighted inversely so the highest sensitivity corresponds to drugs with the minimal difference between pharmacological and packing distances, or in other words, those for which a uniform nonzero affinity is limited only to kinase clusters with similar packing; and c) the weights are averaged only over the pairs of kinases inhibited by drug n . Thus, the packing sensitivity, $\psi(n)$, of drug n is given by $\psi(n) = \langle |d_n(X, Y) - \lambda d_{\text{pack}}(X, Y)| \rangle_n^{-1}$, where $\langle \rangle_n$ = average over all kinase pairs X, Y such that the respective affinities for n satisfy $X_n \neq 0$ and $Y_n \neq 0$. Thus, if drug cross reactivity would be absolutely determined by packing differences (the ideal case), d_n would be *identical* to λd_{pack} for all kinase target pairs, giving $\psi(n)=\infty$. Taking the inverse of the departure from the ideal cross-reactivity (maximum sensitivity) ensures a clear amplification of the differences between drugs in terms of their sensitivity to packing differences among their targets.

To conclude, the packing blueprint emerges as a selectivity filter in drug design *in an average sense*, not apparent when considering individual drugs. This result is expected because dehydron targeting is not acknowledged to be a component of current drug-design methodologies.

Drug selectivity switch based on packing differences across targets

The evolutionary proximity of kinases fosters side effects arising from off-target ligand binding. Making use of the wrapping technology, we have previously illustrated how to sharpen the binding affinity within the pharmacokinome associated with a specific drug [9].

The selective inhibition of the Bcr-Abl (Abelson tyrosine kinase) was achieved by redesigning the potent inhibitor Gleevec [12]. Of the alternative Gleevec targets with reported structure, the C-kit tyrosine kinase has been recognized as binding partner [20,21]. In addition Gleevec binds tightly to the lymphocyte kinase Lck [7].

Since the packing of a protein serves as a selectivity filter (Fig. 4b), we may take advantage of packing differences across paralogs to redirect and sharpen the inhibitory impact of a drug. We focus on the activation loops for the three paralog kinases C-Kit, Bcr-Abl and Lck, known to have affinity for Gleevec [7]. As we compare the first pair of targets (C-Kit, Bcr-Abl), the packing of the active conformation of the C-Kit activation loop (Fig. 5a) differs significantly from that of the inactive (Gleevec-binding) loop conformation of Bcr-Abl [22]. This fact becomes evident as we align the dehydron patterns of C-Kit (PDB.1T46) with that of Bcr-Abl

(PDB.1FPU), as shown in Fig. 3a,b. There are also major differences as the packing of the activation loop in C-Kit is contrasted with the conformation in Lck (PDB.1QPJ), shown in Fig. 5b. A structural alignment of C-Kit, Bcr-Abl and Lck reveals that dehydron F811-A814 in the active conformation of the activation loop of C-Kit (Figs. 3a, 5a,b) maps onto the nonbonded residues F382; S385 on the inactive conformation of the activation loop of Bcr-Abl in complexation with Gleevec and on the well dehydrated F383-A386 hydrogen bond in the active conformation of the activation loop of Lck (Fig. 5b). Given the relative position of the 811–814 dehydron of C-Kit with respect to Gleevec in the protein-inhibitor complex (PDB.1T46), we may determine which modification of Gleevec would selectively direct its impact towards C-Kit. By methylating Gleevec at position 6 on the pyrimidine ring, highlighted in Fig. 5a, the inhibitor becomes a selective wrapper of the packing defect in C-Kit. In contrast, it interferes sterically with the inactive Bcr-Abl activation loop conformation and hinders the hydration of D382 in Lck, while not serving as dehydrator for the aligned and already well dehydrated F383-A386 hydrogen bond.

The selective inhibitory impact of Gleevec methylated at piperidine position 6 (Fig. 5a) becomes apparent in the phosphorylation rates obtained from a spectrophotometric assay ([7, 13]) designed to test whether the specificity and affinity for C-kit improved. Thus, we conducted an assay (Fig. 5c) to measure the phosphorylation rate of peptide substrates [12, 15–17] in the presence of the kinase inhibitor at different concentrations. The inhibitory impact of the parental compound (Gleevec, open triangles) was contrasted with that of the modified Gleevec (filled squares). Both inhibitors were tested on the paralog kinases reported in PDB with known affinity for Gleevec and for which a peptide substrate ([12,16,17]) was commercially available: Bcr-Abl (red), C-kit (blue) and Lck (green). The inhibitory impact of the Gleevec variant is now sharply concentrated on C-kit, rather than on Bcr-Abl. Thus, by redesigning Gleevec according to the wrapping differences in the activation loops of its purported targets, we redirected the primary inhibitory impact from one kinase to another. This test-tube experiment suggests that the wrapping technology may introduce a selectivity switch in inhibitor design, taking advantage of the packing differences in the activation loops of alternative targets of the parental compound.

DISCUSSION

The drug-based inhibition of protein kinases is a widely used therapeutic strategy. Nevertheless, the structural similarities resulting from the evolutionary proximity of kinases are often responsible for toxic side effects. Such structural similarities are extreme on secondary structure elements but less apparent in loopy regions. In this work we report on a novel selectivity filter, the pattern of packing defects of a kinase, and sharpened this filter by focusing on the activation loops of kinases. We demonstrate that the filter is highly correlated with the affinity profiling of kinases, thus establishing a packing blueprint for drug design. While we found marked differences in the efficacy of available drugs to harness on packing differences, the overall correlation is most striking when taking into account that packing defects have not been purposely targeted in drug design. The relevance of this filter for the development of a new generation of inhibitors of controllable impact stems from the enabled modulation of selectivity based on packing differences across kinases. Thus, this work introduces the foundational steps to shape new strategies for parallel inhibition of multiple targets as well as strategies for sharpening the inhibitory impact of lead compounds.

Acknowledgements

The research of A. F. is supported by NIH grant 1R01 GM072614–01A1, by a grant from the John and Ann Doerr Fund for Computational Biomedicine (Program GC4R 2005), and by an unrestricted grant from Eli Lilly and Company. The authors also thank Drs. Chen Su and Harry Harlow (Lilly) for their valuable input.

References

1. Cohen P. Protein kinases – the major drug targets of the twenty-first century? *Nat Rev Drug Discov* 2002;1:309–315. [PubMed: 12120282]
2. Donato NJ, Talpaz M. Clinical use of tyrosine kinase inhibitors: Therapy for chronic myelogenous leukemia and other cancers. *Clin Cancer Res* 2000;6:2965–66. [PubMed: 10955770]
3. Dancey J, Sausville EA. Issues and progress with protein kinase inhibitors for cancer treatment. *Nat Rev Drug Discov* 2003;2:296–313. [PubMed: 12669029]
4. Fabbro D, Garcia-Echeverria CG. Targeting protein kinases in cancer therapy. *Curr Opin Drug Discov Dev* 2002;5:701–712.
5. Gabriele A, King GL. Protein kinase C inhibitors in the treatment and prevention of diabetic complications. *Curr Opin Endocrinol Diabetes* 2001;8:197–204.
6. Myers MR, He W, Hulme C. Inhibitors of tyrosine kinases involved in inflammation and autoimmune disease. *Curr Pharm Des* 1997;3:473–502.
7. Fabian MA, et al. A small molecule kinase interaction map for clinical kinase inhibitors, *Nat. Biotechnol* 2005;23:329–336.
8. Fernández A, Berry RS. Molecular dimension explored in evolution to promote proteomic complexity. *Proc Natl Acad Sci USA* 2004;101:13460–13465. [PubMed: 15347802]
9. Fernández A. Incomplete protein packing as a selectivity filter in drug design. *Structure* 2005;13:1829–1836. [PubMed: 16338411]
10. Fernández A, Rogale K, Scott RL, Scheraga HA. Inhibitor design by wrapping packing defects in HIV-1 proteins. *Proc Natl Acad Sci USA* 2004;101:11640–11645. [PubMed: 15289598]
11. Fernández A. Keeping dry and crossing membranes. *Nature Biotech* 2004;22:1081–1085.
12. Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B, Kuriyan J. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* 2000;289:1938–1942. [PubMed: 10988075]
13. Li, JJ.; Johnson, DS.; Sliskovic, DR.; Roth, BD. *Contemporary Drug Synthesis*. Wiley-Interscience.; 2004. p. 32-33.
14. Mol CD, Dougan DR, Schneider TR, Skene RJ, Kraus ML, Scheibe DN, Snell GP, Zou H, Sang B, Wilson KP. Structural basis for the auto inhibition and STI-571 inhibition of C-KIT tyrosine kinase. *J Biol Chem* 2004;279:31655–31663. [PubMed: 15123710]
15. Zhu X, Kim JL, Newcomb JR, Rose PE, Stover DR, et al. Structural analysis of the lymphocyte-specific kinase LCK in complex with non-selective and src family selective kinase inhibitors *Struct. Fold Des* 1999;7:651–61.
16. Timokhina I, Kissel H, Stella G, Besmer P. Kit signaling through PI 3-kinase and Src kinase pathways: an essential role for Rac1 and JNK activation in mast cell proliferation. *EMBO J* 1998;17:6250–6262. [PubMed: 9799234]
17. Perlmutter RM, Marth JD, Lewis DB, Peet R, Ziegler SF, Wilson CB. Structure and expression of Lck transcripts in human lymphoid cells. *J Cell Biochem* 1988;38:117–126. [PubMed: 3265417]
18. Gusfield, D. *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*. Cambridge University Press; 1997. 1997
19. Vieth M, Higgs RE, Robertson DH, Shapiro M, Gragg EA, Hemmerle H. Kinomics-structural biology and chemogenomics of kinase inhibitors and targets. *Biochim Biophys Acta* 2004;1697:243–257. [PubMed: 15023365]
20. Attoub S, Rivat C, Rodrigues S, Van Borexlaer S, Bedin M, Buyneel E, Louvet C. The c-kit tyrosine kinase inhibitor STI-571 for colorectal cancer therapy. *Cancer Res* 2002;62:4879–4883. [PubMed: 12208734]
21. Skene RJ, Kraus ML, Scheibe DN, Snell GP, Zou H, Sang BC, Wilson KP. Structural basis for autoinhibition and STI-571 inhibition of C-kit Tyrosine kinase. *J Biol Chem* 2004;279:31655–31663. [PubMed: 15123710]
22. Noble ME, Endicott JA, Johnson LN. Protein kinase inhibitors: Insights into drug design from structure. *Science* 2004;303:1800–1805. [PubMed: 15031492]

Figure 1.

Protein ligand as dehydron wrapper illustrated by the *epidermal growth factor receptor kinase (EGFRK)* in complex with the drug inhibitor *erlotinib* (PDB.1M17). There are four dehydrons in the ATP-binding site that are wrapped by the inhibitor molecule upon association: Gly697-Phe699; Gly695-Val702; Gly772-Met769; Gly772-Val821. The protein chain backbone is represented by blue virtual bonds joining α -carbons, well wrapped backbone hydrogen bonds are shown as light grey segments joining the α -carbons of the paired residues, and dehydrons are shown as green segments. The figure shows the cavity in detail, the pattern of packing defects and the inhibitor positioned as a dehydron wrapper.

Figure 2.

The wrapping concept in drug design. An intramolecular hydrogen bond (HB) in a soluble protein target prevails only if it is protected from water attack. The backbone amide-carbonyl HB is here represented as a thick segment joining the electron-deprived amide proton (open disk) and the carbonyl oxygen (red disk). Thus, the extent of intramolecular wrapping, ρ , by nonpolar “greasy” groups (black balls) becomes central to determine its stability and strength ([8–11]. Poorly wrapped hydrogen bonds (dehydrons) are solvent-exposed and become strengthened and stabilized by removal of surrounding water. Thus, we may design small-molecule drugs that exogenously “wrap” dehydrons upon association by contributing nonpolar groups to their desolvation spheres.

Figure 3b:

Figure 3.

Figure 3a. Structural alignment of C-Kit kinase (PDB.1T46, light blue ribbon) with Bcr-Abl kinase (PDB.1FPU, magenta ribbon). The α -carbons of aligned residues in contact with Gleevec are shown as red balls (C-Kit) and yellow balls (Bcr-Abl). The well wrapped hydrogen bonds between contact residues are marked as white segments joining α -carbons of paired residues, C-Kit dehydrons are marked in green and Bcr-Abl dehydrons are marked in red. Figure 3b. Aligned matrices of dehydrons (green and red entries) and hydrogen bonds (grey entries) for residues in Λ (main text). The packing of the C-Kit kinase is represented by green dehydrons (upper matrix) and that for Bcr-Abl kinase by red dehydrons (lower matrix). The residue numbering is consistent with the PDB entries 1T46 (C-Kit) and 1FPU (Bcr-Abl).

Figure 4.

Figure 4a. Packing-based hierarchical clustering of the 32 PDB-reported kinases that have been independently fingerprinted in drug-affinity assays [7]. The hierarchical tree is built based on the packing distance (Methods), so that the minimum number of nodes that must be crossed to walk from one protein-node to another is proportional to the packing distance between the two proteins. The scale shown on the y axis of the tree represents the dissimilarity measure,

i.e., packing distance, between the merged groups. Packing distances are computed by comparing dehydron patterns of ATP pockets. The packing distance is the Hamming distance between dehydron matrices restricted to the residues forming the hydrogen-bond microenvironments within the ATP pocket. The high-throughput affinity profiling of proteins against 17 nonpromiscuous drugs was obtained from [7]. Figure 4b. Correlation between packing distance and pharmacological distance (Methods) for all pairs constructed from a pool of 32 PDB-reported kinases independently fingerprinted for affinity against 17 drugs [7]. Packing and pharmacological distances for four different pairs of kinases are highlighted by showing them as colored circles (red-SYK and INSR, green- BTK and HCK, yellow-ABL and KIT, blue-SYK and LCK). Figure 4c. Packing sensitivity of individual drugs. The packing sensitivity, $\psi(n)$, of drug n measures the drug responsiveness to packing differences and thus becomes an indicator of its discriminatory power.

Figure 5.

Figure 5a. Selectivity switch determined by packing differences across paralog kinases. The figure shows the position of dehydron F811-A814 in the active conformation of the activation loop for C-Kit and its position relative to Gleevec in the protein-inhibitor complex (PDB.1T46). By methylating Gleevec at the highlighted piperidine 6-position, the inhibitor becomes a selective wrapper of the packing defect in C-Kit. The oxygen position of water molecules (red balls) was resolved in the crystallography and is displayed accordingly. Figure 5b. Activation loop in active conformation for Lck (PDB.1QPJ). The F811-A814 dehydron in C-Kit aligns with a well wrapped F383-A386 hydrogen bond in Lck. Thus, a sticky site in the activation loop is absent in Lck. Figure 5c. Phosphorylation rates from spectrophotometric test-tube assay on the three paralog kinases with PDB reported structure and demonstrated affinity for Gleevec. The kinases are inhibited by the modified Gleevec methylated at the position indicated in Fig. 5a (filled squares, thick lines), and by the parental compound (open triangles, thin lines). Phosphorylation rate plots are given for Bcr-Abl (red), C-kit (blue) and Lck (green). Error bars represent dispersion over 8 runs for each kinetic assay. Each assay run consisted of three phosphorylation measurements at 400nM differences in inhibitor concentration. Notice the selective enhancement of C-kit inhibition by the wrapper of nonconserved dehydrons when contrasted with the parental compound. Within the means of detection, the kinase phosphorylation rates do not vary appreciably in the range 0–100nM inhibitor concentration.