Journal of
Medicinal
Chemistry
Article

J. Med. Chem. **2009,** *52,* 6649–6659 **6649**
DOI: 10.1021/jm9008899

Knowledge-Based Characterization of Similarity Relationships in the Human Protein−Tyrosine Phosphatase Family for Rational Inhibitor Design

Dušica Vidović[†] and Stephan C. Schürer*,[†,‡]

[†]*Center for Computational Science, University of Miami, Miami, Florida 33136, and* [‡]*Department of Pharmacology, Miller School of Medicine, University of Miami, Miami, Florida 33136*

Received June 17, 2009

Tyrosine phosphorylation, controlled by the coordinated action of protein−tyrosine kinases (PTKs) and protein−tyrosine phosphatases (PTPs), is a fundamental regulatory mechanism of numerous physiological processes. PTPs are implicated in a number of human diseases, and their potential as prospective drug targets is increasingly being recognized. Despite their biological importance, until now no comprehensive overview has been reported describing how all members of the human PTP family are related. Here we review the entire human PTP family and present a systematic knowledge-based characterization of global and local similarity relationships, which are relevant for the development of small molecule inhibitors. We use parallel homology modeling to expand the current PTP structure space and analyze the human PTPs based on local three-dimensional catalytic sites and domain sequences. Furthermore, we demonstrate the importance of binding site similarities in understanding cross-reactivity and inhibitor selectivity in the design of small molecule inhibitors.

Introduction

Tyrosine phosphorylation is involved in the regulation of many physiological processes, including growth, proliferation and differentiation, metabolism, cell cycle regulation and cytoskeletal function, cell−cell interactions, neuronal development, gene transcription, and the immune response.[1−6] The levels of cellular protein tyrosine phosphorylation are regulated by the coordinated action of protein−tyrosine phosphatases (PTPs[a]) and kinases (PTKs).[1,5] Until recently, PTKs were considered to be the main enzymes regulating tyrosine phosphorylation and huge progress has been made over the past 20 years in clarifying their significance in signal transduction.[7−9] Today, beyond kinases, PTPs are recognized as critical regulators of signal transduction.[10] The ability of PTPs to dephosphorylate phosphotyrosine residues selectively on their substrates plays an important role in initiating, sustaining, and terminating cellular signaling.[5] Several studies have shown that the diversity of functions for the PTPs matches the diversity of functions for the PTKs.[11,12]

Malfunction of the PTP activity is related to a number of human diseases ranging from cancer to neurological disorders and diabetes. The diversity of cellular functions regulated by PTPs and their implications in human diseases suggest that PTPs are prospective drug targets.[12−14]

The human genome contains 107 PTPs.[15,16] On the basis of the catalytic mechanism of dephosphorylation, the PTPs can

be grouped into two separate families, Cys-based family comprising 103 members and Asp-based family comprising 4 members. The Cys-based PTPs, which are the focus of the present study, can be further divided into four major classes: classical PTPs, dual-specificity PTPs (DUSPs), cdc25 PTPs, and low-molecular weight (LMW) PTPs.

Although protein similarities and classification are generally anticipated by sequence similarity, three-dimensional structures tend to be more conserved than sequences and are essential for the functional properties of proteins.[17−19] In enzymes, the protein substrate recognition occurs at structurally conserved and specific binding sites. Hence, structural features of the catalytic sites define protein function. Several studies show that comparative sequence analyses should be combined with other approaches (such as genomic and proteomic analyses) to fully understand structure, function, and evolution of protein families.[20,21]

PTPs utilize the active site signature (H/V)C(X)5R(S/T) motif in the conserved PTP catalytic domain to hydrolyze phosphoester bonds in protein and non-protein substrates.[22,23] This structure motif is called PTP loop (red loop in Figure 1). Key features of the domain also include the phosphotyrosine recognition loop (blue loop in Figure 1) and the WPD loop that occurs in two conformations, open and closed (Figure 1, yellow and green loops, respectively). In the native form the WPD loop is in an open conformation, and the binding pocket is easily accessible to substrate. Upon substrate binding, the WPD loop closes over the active site, forming a tight binding pocket for the substrate.[24,25] In the active, closed form the Asp residue from WDP loop is in position to act as a general acid/base catalyst in the dephosphorylation reaction.[26] Furthermore, it has been shown that the catalytic activities of the PTPs are influenced by the flexibility and stability of the WPD loop in its active form.[27,28]

---

*To whom correspondence should be addressed. Phone: +1-305-243-4842. Fax: +1-305-243-9304. E-mail: sschurer@med.miami.edu.

[a]Abbreviations: DUSP, dual-specificity phosphatase; LMW, low-molecular weight; MAP, mitogen activated protein; MKP, mitogen activated protein kinase phosphatase; MST, minimum spanning tree; NW, Needleman−Wunsch; PTK, protein−tyrosine kinase; PTP, protein−tyrosine phosphatase; TIP, target-informatics platform.

© 2009 American Chemical Society

Published on Web 10/07/2009   pubs.acs.org/jmc

Figure 1. Structural features of PTP catalytic domain. Substrate recognition loop is in blue, PTP loop in red, flexible WPD loop in either open or closed position in yellow and green, respectively.

The PTP binding site is highly polar with the deprotonated thiol anion of the catalytic cysteine acting as a nucleophile. Such binding environment favors polar binders, and it is therefore one of the challenges in developing useful compounds to balance inhibitory activity with cellular permeability.
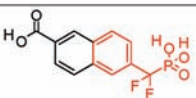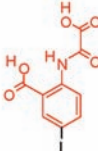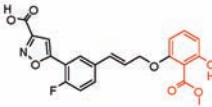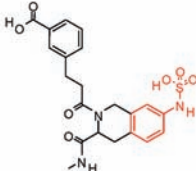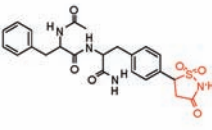
One key component in the design of PTP inhibitors is a hydrolytically stable phosphotyrosine or phosphate mimic as a "head" group. Several classes of mimics have been reported,[29] including the difluoromethylenephosphonates, sulfamic acid, and benzoic acids such as 2-(oxalylamino)benzoic acids, salicylic acids, and its derivatives. Various PTP inhibitor cocrystal structures with these types of head groups have been reported. Table 1 shows potent representative PTP1B inhibitors with different head groups and their corresponding PDB codes.

To date, most of the studies related to PTPs were performed on sequences of classical phosphatases[5,16] and PTP1B in particular.[12,30,35,36] Here we present a comprehensive comparative analysis of the catalytic domain sequences and the three-dimensional catalytic sites of the entire human Cys-based PTP protein family. Experimental small molecule inhibition data illustrate that similarities of the catalytic site can reflect a PTP's propensity for selectivity and promiscuity. Local three-dimensional site similarity can be a first-order structure-based assessment to identify most similar targets, which are likely to show cross-reactivity toward a small molecule inhibitor and therefore should be tested experimentally during lead optimization.

Methods

Human PTP Gene Family. In order to compile a PTP gene list, we searched literature[15] and gene databases[37−39] and retrieved 105 genes encoding human PTPs. Among them, two are pseudogenes, PTPN20C and PTPRV. PTPN20A and PTPN20B are coded by the same chromosome and therefore have the same domain sequence. After elimination of redundant genes, the final list assembles 102 human PTP genes with 37 genes encoding classical PTPs, 61 genes encoding DUSPs, 3 genes for the cdc25 PTPs, and 1 for the LMW PTP. Classical PTPs are further divided into the classical receptor type phosphatases and classical nonreceptor type phosphatases. DUSPs comprise several subtypes: MAP kinase phosphatases (MKPs), atypical DUSP, slingshots, myotubularins, PRLs, CDC14s, and PTENs. The

Table 1. Typical Phophatase Inhibitor "Head" Groups[a]

Head group in co-crystallized ligands	PDB codes of similar ligands
difluoromethylenephosphonate (1bzj)[30]	1bzc, 1kak, 1kav, 1lqf, 1pxh, 1q6j, 1q6m, 1q6n, 1q6p, 1q6s, 1q6t, 2fjn
2-(oxalylamino)-benzoic acid (1ecv)[31]	1c83, 1c84, 1c85, 1c86, 1c88, 1gfy, 1l8g, 1wax
salicylic acid and derivatives (1q1m)[32]	1g7g, 1jf7, 1qxk, 1xbo
phenyl sulfamic acid (2f71)[33]	2f6v, 2f6y, 2f6z, 2f70
isothiazolidinone (2cm7)[34]	2bgd, 2bge, 2cma, 2cmb, 2cm8, 2nta,

[a] Representative potent PTP1B inhibitors and the PDB codes of their cocrystal structures and cocrystal structures of their analogues with the same head group are shown.

list of 102 human PTP genes is given in the Supporting Information along with the protein names (Table S1). Here we refer to a phosphatase by the corresponding gene annotation.

PTP Sequences and Domains. On the basis of the identified gene list, we have constructed a database with the PTP domain sequences from SWISS-PROT.[38] A core SWISS-PROT data record consists of sequence data, citation information, and taxonomic data, while the annotation consists of a description of the function of the protein, post-translational modifications, domains and sites (for example, calcium binding regions, PTP catalytic sites, zinc fingers), similarities to other proteins, diseases associated with deficiency of the protein, etc. Some of these annotations are derived from other databases, which are linked on the basis of controlled vocabulary (and unique IDs) each describing an important piece of biological complexity. Most entries in Swiss-Prot have a cross-reference to Pfam or InterPro.

Pfam and InterPro are large collections of protein families and domains.[40−42] The human PTPs have different domains depending on the PTP types and subtypes. In the present study we use the InterPro domain annotation, since it is the most comprehensive for PTP domains. However, some of the InterPro annotated catalytic domains are remarkably shorter and do not contain the catalytic site (signature motif).

Table 2. List of PDB Structures Used as Templates in Homology Modeling

template PDB code	template PTP	resolution	template PDB code	template PTP	resolution
2f71A	PTPN1	1.55	1wrmA	DUSP22	1.5
2h02A	PTPRB	2.3	1vhrA	DUSP3	2.1
2ooqA	PTPRT	1.8	2imgA	DUSP23	1.9
2g59A	PTPRO	2.19	2pq5A	DUSP13B	2.3
2i1yA	PTPRN	2.23	2esbA	DUSP18	2
1ygrA	PTPRC	2.9	1yz4A	DUSP15	2.4
1wchA	PTPN13	1.85	1oheA	CDC14B	2.2
2g6zA	DUSP5	2.7	1xm2A	PTP4A1	2.7
1d5rA	PTEN	2.1	1fpzA	CDKN3	2
1zzwA	DUSP10	1.6	1wvhA	TNS1	1.5
2nt2A	SSH2	2.1	1zsqA	MTMR2	1.82
2r0bA	STYX	1.6	5pntA	ACP1	2.2
2e0tA	DUSP26	1.67	1c25A	cdc25A	2.3

In these cases we alternatively use the SWISS-PROT annotation where the PTP domains comprise the catalytic site. This was the case for CDC14A, CDC14B, RNGTT, PTPM1, AUXI, TENC1, TPTE, TPTE2, MTM1, MTMR1, MTMR2, MTMR3, MTMR4, MTMR5, MTMR6, MTMR7, MTMR8, MTMR9, MTMRA, MTMRB, MTMRC, and MTMRD. For myotubularin MTMR14 there was no annotated domain in any of the databases. In order to determine a PTP domain for MTMR14, we aligned its sequence to the sequences of other myotubularins using Clustal W.[43] After verifying that the domains and catalytic sites of other myotubularins were properly aligned, we extracted the analogous domain for MTMR14 phosphatase. MTMR15 has one annotated domain, VRR_NUC, which is not a PTP domain, and the alignment to other myotubularins did not indicate the catalytic region. MTMR15 was therefore excluded from further analysis, resulting in a total of 101 PTP genes considered in this study. For tensine-like PTPs, TENS1, and TENS3 (no annotation), the domain sequence was retrieved from the literature.[44]

Some of the classical receptor phosphatases have two cytoplasmic PTP domains, a membrane proximal domain (D1) and a membrane distal domain (D2). In total, we collected 113 PTP annotated domains for this study, one PTP domain for each PTP plus distal PTP domain for classical receptor phosphatases with two PTP domains.

Sequence Similarities and Identities. Pairwise alignments and similarities of the 113 human PTP domain sequences were calculated by the Needleman–Wunsch (NW) algorithm[45] (Blosum62 substitution matrix, gap penalty 10, gap extension penalty 0.5). NW uses dynamic programming to identify an optimum global alignment as the best pathway through a scoring matrix representing the two sequences to be aligned and that is constructed by optimizing the alignment score of successively increasing sequence segments.

Homology Modeling. The STRUCTFAST[46] algorithm implemented in the Target Informatics Platform (TIP) software system[39,47] was used to generate homology models for all PTP sequences against a number of templates (see below). STRUCTFAST is an automated profile–profile database search algorithm capable of detecting weak similarities between protein sequences. Multiple sequence alignment profiles are used for both the query and primary template sequence. The query sequence profiles are generated with a modified version of the PSI-BLAST algorithm. A database of profiles for template representatives from the PDB[48] is generated in a similar manner but incorporating information from structure–structure alignments derived from the template protein's structural family. A query profile is aligned

and scored against the library of structural profile templates, and the alignments are ranked by the significance of their scores using convergent island statistics. STRUCTFAST uses dynamic programming to incorporate gap information from the structural family directly into the alignment process. Because of rigorous analytical treatment of the profile–profile scores, STRUCTFAST scoring function includes no parameters to optimize.

The PTP models were built using a TIP database, including the entire PDB database as of June 15, 2008, including 229 PTP structures. The PTP domain sequences were loaded into the TIP sequence database to generate models corresponding exactly to the defined PTP domains. The primary templates from which to derive the α carbon coordinates of the PTP STRUCTFAST models were selected as described in the following.

PTP Primary Structure Templates. To select suitable templates for homology modeling, we explore the PDB. There are more than 200 structures of the human PTPs deposited in PDB but about 100 for PTP1B alone (corresponding gene annotated as PTPN1).

The primary criterion for selection was the active form of the corresponding PTP crystal structure, which is determined by the closed conformation of the WPD loop (see Figure 1). However, for most of the phosphatases there is no experimental structure available and not all deposited PTP crystal structures are in the active form. Indeed, only 26 PTPs have at least one crystal structure with the active WPD loop conformation. If a phosphatase has several structures in the active form, the one with the highest resolutions was kept as a template.

Some PTP types such as cdc25 phosphatases and myotubularins do not have the WPD loop and therefore exist in only one form. It is not quite clear which residue replaces the aspartic acid from the WPD loop and reacts as a general acid in the catalytic reaction of such PTPs. Some previous studies have shown that the general acid residue perhaps can be a part of the catalytic $(H/V)C(X)5R(S/T)$ loop.[49,50] The side chain conformation of such a residue would therefore define an active form for these phosphatases. However, the PDB crystal structures of MTMR2 and cdc25A are available and were used as templates for modeling PTPs without the WPD loop.

For PTPs that are structurally undetermined we compared their domain sequence similarities to the 26 selected templates. The crystal structure with the highest sequence similarity was defined as the primary template for generating a model of the corresponding phosphatase domain. The primary templates are listed in Table 2 along with their gene symbol and the resolution of the crystal structure.

In addition, Table S2 in the Supporting Information shows model to template identities.

Because we expect the catalytic site in a homology model to be related to the three-dimensional template structure, in many cases we generated several models based on different primary templates. This is to estimate (or minimize) a template bias that may be introduced into the models but also to incorporate structural flexibility defined implicitly by different experimental structures. These additional primary templates were selected from the 26 candidate structures based on succeeding sequence similarity. Since most of the available experimental structures belong to the classical PTPs and DUSPs, we generated more models for these two PTP types.

Definition of the Catalytic (Binding) Sites. Once the homology models were generated, the local sites of interest were defined. For each template model (in "template models" the PTP domain sequence corresponds to the crystal structure template, and they are therefore very close or identical to the original template PDB structure) the initial site was defined as a set of solvent accessible residues within 10 Å around the catalytic cysteine. The site was further manually corrected by adding or removing residues that can still interact with a virtual ligand. For example, residues that belong to the inner area under a tangent on the protein surface are defined as a part of the catalytic site (tangent starts at the binding pocket). Residues outside this area were removed from the site even if they are proximate to the binding pocket. After the template sites were defined, the corresponding models were aligned to their template model and the model binding sites were defined on the basis of their matching residues. The so-defined sites were in addition visually inspected for accuracy.

Calculation of Site Similarities. The SiteSorter algorithm implemented in the TIP software system computes pairwise 3D similarities between sites. This is performed in three steps: (1) the two sites are described as graph representations; (2) the optimal overlay of the two sites is determined by optimizing the overlap score between the two site graphs; (3) the physiochemical similarity of the two optimally overlaid sites is scored. The SiteSorter algorithm is similar to Klebe's approach[51] in which sites are represented as collections of surface points and edges, which are inputs to a clique detection algorithm[52] that determines the best site overlay as the maximum complete subgraph. However, SiteSorter in addition takes into account the orientation of each surface point with respect to the pocket opening. The similarities of each of the matching surface points are described as a continuum of scores, and a weighted clique detection algorithm is used. An overlay score can be derived for any given orientation of the two graph surfaces considering distance and angle constraints of the corresponding surface points. The best overlaid sites are then scored on the basis of chemical group similarity incorporating site chain and backbone atoms. These (raw) chemical scores are further normalized in the Tanimoto-like definition: $S_{ABnormalized} = S_{AB}/(S_{AA} + S_{BB} - S_{AB})$, where S_{AB} is the raw value for the site similarity between sites A and B and where S_{AA} and S_{BB} are self-site similarities for site A and site B, respectively. We use this normalized site similarity in our analyses. For pairs where no site overlay score was generated because of dissimilarity between sites, we assigned a site similarity value of zero.

Cluster Analysis. Domain sequence similarities and local (3D) site similarities were classified by the hierarchical clustering using the Spotfire Decision Site software.[53] Minimum Spanning Trees (MST) were generated by Kruskal's algorithm[54] and visualized by Cytoscape (force-directed layout, weighted by similarity).[55]

Structural Model Alignment. After the template catalytic sites were defined, structural alignment of the PTP models to their corresponding template model was performed using Schrodinger's Protein Structure Alignment program.[56]

Structure Visualization. PyMOL[57] was used for visualizing PDB structures, models, and binding sites and also for defining the template binding sites.

Workflows. We used Scitegic Pipeline Pilot[58] collection of components for data retrieval, filtering, and analysis.

SAR Data. The literature and PDB database were searched for known PTP inhibitors. We collected a moderately large list of small cocrystallized PTP1B inhibitors and their analogues that show reasonable potency against a set of different classical PTPs.[31,35,59−62]

Results and Discussion

Global Trend of Phosphatase Site vs Sequence Similarities. In the present study we generated models for 113 domain sequences representing 101 PTPs (retrieved from the SWISS-PROT database, domains annotated by InterPro or SWISS-PROT) as described in the Methods section. Four-hundred-fifty-five models were generated using as primary templates the 26 different PTP structures in the active conformation that are available in the PDB; at least one model was generated for each of the 113 PTP sequences. The binding sites were defined as a set of residues within 10 Å around catalytic cysteine considering the solvent-accessible surface. Pairwise site similarities were calculated following three-dimensional site overlay using a scoring function based on surface chemical features as described in the Methods. The site similarity value depends on the size of the site because larger sites can have a larger overlaid surface. Although the sites are reasonably similar in size (PTPs, 10 Å around the catalytic residue), we normalized the raw site similarity score using a Tanimoto-type definition after calculating the chemical site similarity of each site against itself. A correlation plot of normalized vs raw site similarities is given in the Supporting Information (Figure S1). We observed more robust clustering of the normalized site similarities compared to the raw chemical scores.

For most PTP domains multiple models were generated on the basis of different primary templates, and therefore, each PTP can be characterized by different (catalytic) sites, which can lead to slightly different site similarity values for any given PTP domain pair. For the analysis presented here we used the sites emerging from best models (highest identity to model template; in case of several template candidates the model based on the highest resolution structure is used). Figure 2 shows a scatter plot of catalytic site similarity vs domain sequence similarity for each PTP pair based on the currently modelable PTPs. All PTP domain sequence pairs were aligned using the NW algorithm to compute sequence similarity and sequence identity values. The average site similarities and the maximum site similarities between PTP pairs (which may provide a conservative estimate of the propensity of a pair of PTPs being similar around their catalytic site) are given in the Supporting Information (Figure S2). We also illustrate site similarities vs sequence identities and histograms of the different similarity and identity measures to visualize their global distributions (supporting Figures S3 and S4).
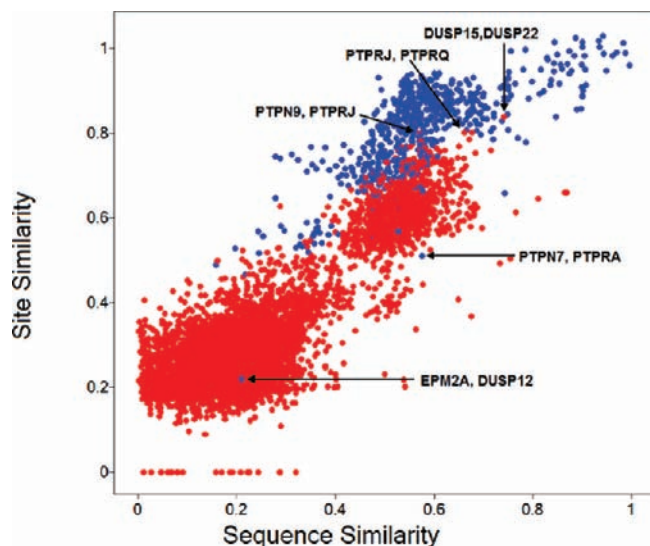
Figure 2. Correlation of normalized site similarities and sequence similarities for all PTP combinations. Blue dots represent site pairs corresponding to models based on the same primary template, while for red data points the templates are different.

Figure 3. Hierarchical clustering of human PTP domain sequences. Highest similarity is colored in red, medium in yellow, and low similarity in green.

Qualitatively, Figure 2 illustrates that PTPs of high sequence similarity also have very similar catalytic sites while there is more variability of site similarities among pairs of lower and average similar sequences. This general trend also holds for average and maximum site similarities and in particular for sequence identities (Supporting Information Figures S2 and S3). From Figure 2 and the histograms (Supporting Information Figure S4) one can identify highly site-similar PTP pairs that correspond to lower sequence similarities; this is relative to the mode of the main subpopulations of the two similarity distributions and thus not an artifact of the different scaling of the measures. PTP pairs modeled from the same primary template structure show on average higher site similarities compared to those based on different templates. This is expected because pairs of models based on the same template also have on average higher sequence similarities (sharing higher identity to their template means that they are more similar to each other as well; the average identity of PTP pairs is shown in supporting Figure S5). But also there may be a bias introduced by the specific conformation of the template structure. However, the results should be viewed as our current stage of knowledge, based on the available structural body of the PDB. The vast majority of models are well within or above the required template identity to generate reliable homology models for the purpose of comparing sites,[18] in particular for the profile−profile STRUCTFAST method used here.[46,63] We can expect this picture to be further refined as more structures become available. Nevertheless, in a more detailed analysis we can identify highly site-similar PTP pairs based on different templates (for example, PTPRJ and PTPRQ, PTPN9 and PTPRJ, DUSP15 and DUSP22) and also pairs of low site similarity modeled from the same template (for example, PTPN7 and PTPRA, EPM2A and DUSP12). We therefore conclude that the template bias is relatively small.

Categorization of PTPs Based on Site and Sequence Similarities. In addition to a global trend and direct comparison of individual PTPs we were also interested in identifying major and local groupings. We performed hierarchical
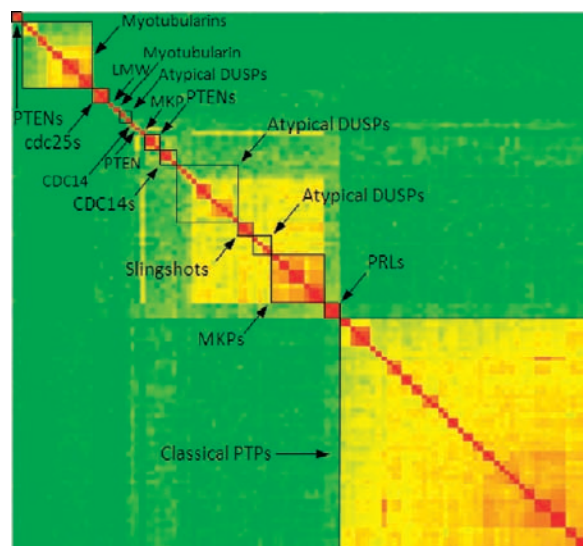
clustering of the 113 PTP domains based on sequence- and site similarity matrices.

Figure 3 shows the PTP domain groups in sequence space. The PTPs were hierarchically clustered using single linkage and the Euclidean distance of the sequence similarity vectors.

Several large clusters are evident. As expected, all classical PTPs group together (lower right corner). The DUSPs are separated into two large groups. One (in the central part of the heat map) comprises atypical DUSP, MKPs, PRLs, slingshots, CDC14s, and several PTENs. The other DUSP group (in the upper left corner) comprises all myotubularins and two remaining PTENs. These two DUSP groups are separated by two small clusters, one containing three cdc25s and one with the single LMW PTP. On the basis of sequences, the DUSPs thus represent a very diverse group of PTPs with some discontinuous subtypes. Within the large cluster of DUSPs the different subtypes are mostly grouped together. The exceptions are STYXL1, which does not cluster with the rest of MKPs, CDKN3, which falls outside of the CDC14 cluster, and the atypical DUSPs, which are separated into two fairly close subgroup. The relatively small sets of three slingshots and three PRLs form clusters including all their respective members.

A similar clustering analysis was performed using site similarity vectors as a distance measure. As before, from the ensemble of catalytic sites corresponding to each domain sequence we selected the one that corresponds to the best model based on template identity. Results are shown in Figure 4.
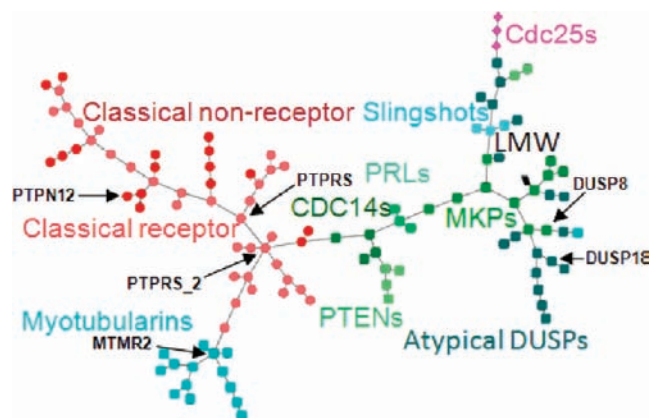
The major groupings that are obtained on the basis of site similarities closely reflect sequence-based clustering. The same major clusters emerge when using ensemble average or maximum pairwise site similarities. However, in contrast to sequence space the members within the major groups, and in particular the classical PTPs, appear much closer. This is consistent with our earlier observation of highly site-similar PTP pairs that correspond to lower sequence similarity; but here we can clearly identify the clusters of the closest PTPs based on site similarities. Similar to sequence space, in site space the DUSPs occupy the middle section of the map but its subgroups are less continuous. The MKPs form two
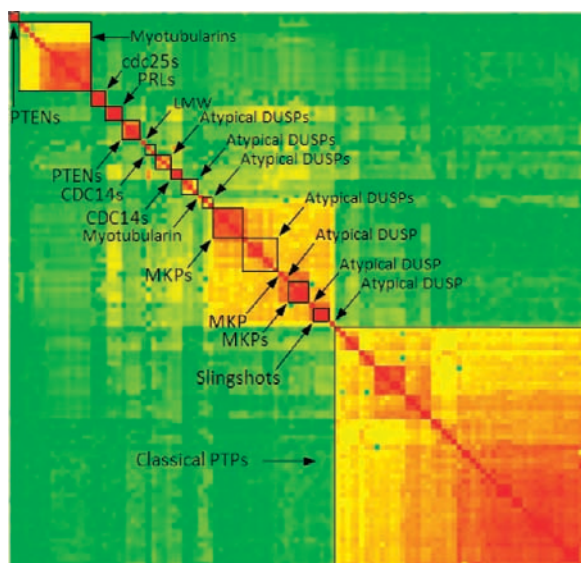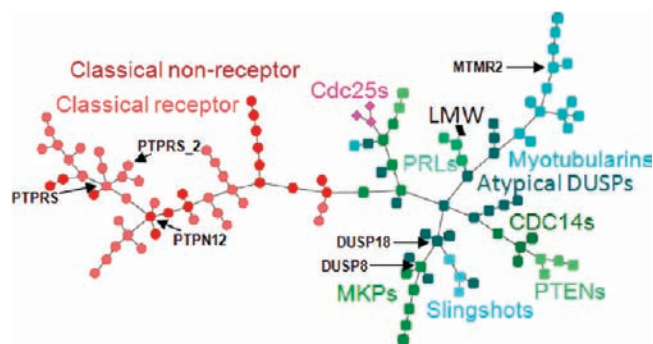
Figure 4. Hierarchical clustering of catalytic site similarities of human PTPs. Highest similarity is colored in red, medium in yellow, and low similarity in green.

Figure 5. Sequence similarity network of human PTP domains. Nodes are colored by PTP subtypes.

Figure 6. Site similarity network of human PTPs. Nodes are colored by PTP subtypes.

distinct clusters and the atypical DUSPs are split into many separate groups. However, the central cluster of MKPs and some of the atypical DUSPs is more pronounced compared to site space. Although the major PTP groupings in site space are comparable to those in sequence space, the subgroupings among classical PTPs and DUSPs are different. Here we provide several examples for DUSPs while classical PTPs are discussed in more detail in the next section. Several larger clusters evident from the DUSP sequence clustering are not present in the DUSP site clustering. For example, in sequence space all MKPs belong to one cluster except STYXL, which forms a singleton. In site space STYXL is also isolated from the subtype members, but the other MKPs split into two clusters, containing DUSP1, DUSP2, DUSP4, DUSP5, and DUSP6 in one cluster and containing DUSP7, DUSP8, DUSP9, DUSP10, DUSP16 in the other. Two large atypical DUSP clusters (separated by slingshots) in sequence space exchange their members split into several smaller clusters and singletons in site space. For the CDC14 subtype, in sequence space CDC14A, CDC14B, PTPDC1 form one cluster and CDKN3 is isolated from the group, while in site space CDC14A and CDC14B belong to one cluster and CDKN3 and PTPDC1 form a separate cluster. The PTENs are grouped better in site space where PTEN, TENC1, TPTE, and TPTE2 form one cluster, while TENC1 is isolated in sequence space. Therefore, PRLs and slingshots present the only DUSP subtypes with preserved grouping (all members group together) in both spaces while all other (and in particular the atypical) DUSPs are separated into smaller clusters or individual targets. This fragmentation of the subtype groups is more distinct in site space. While the major PTP groupings that emerge in sequence vs site space are closely related, our detailed analysis shows that domain sequence-based categorization does not reflect the similarity relationships derived from comparing three-dimensional catalytic sites. Principle component analysis (Supporting Information Figure S6) suggests the same conclusion (similar major but different local groupings).

In addition to hierarchical clustering we also visualized the similarity relationships among the PTP family members as

networks. MSTs for both sequence and site similarities were computed and visualized as described in Methods (Figures 5 and 6). In contrast to hierarchical clustering where the distance of two PTPs is measured on the basis of their similarities to all other PTPs, the MST is constructed on the basis of the individual similarity of a PTP and its joining neighbor. It is therefore in particular suitable for analysis of local relationships.

The network tree representations intuitively illustrate how the majority of the members of each of the PTP subtypes (except the atypical DUSP) group together in sequence as well as site space. Despite the similar groupings by subtype, the sequence and site similarity network trees reveal differences, which may have important implications for the development of selective inhibitors. In particular local neighborhoods, node connectivity, and hubs (nodes with many neighbors) are different in sequence vs site space. For example, there are only two nodes with at least five neighbors in the sequence MST (MTMR2 and the distal domain of PTPRS annotated by PTPRS_2); in the site MST these correspond to nodes with three and one neighbors. In contrast there are four PTPs with at least five neighbors (PTPRS, PTPN12, DUSP8, and DUSP18) in the site MST (Figure 6) but with fewer neighbors in the sequence MST (three, one, two, and one, respectively, Figure 5). Phosphatases with many neighbors (specifically in the site similarity tree) may be particularly challenging drug targets because the development of selective inhibitors can be complicated by the presence of many closely related PTP "off-targets".

Detailed Analysis of Classical PTPs. Previous studies related to PTPs were focused primarily on the classical type.
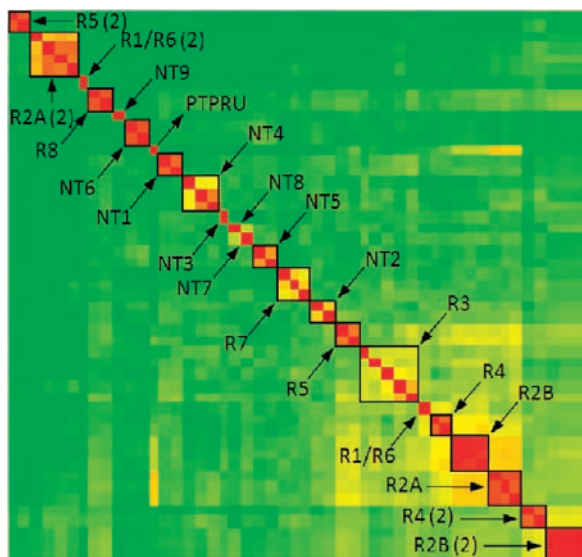
Figure 7. Hierarchical clustering of all human classical PTP domain sequences (for both membrane proximal and distal domains). More detailed clustering presentation can be found in Supporting Information Figure S7.
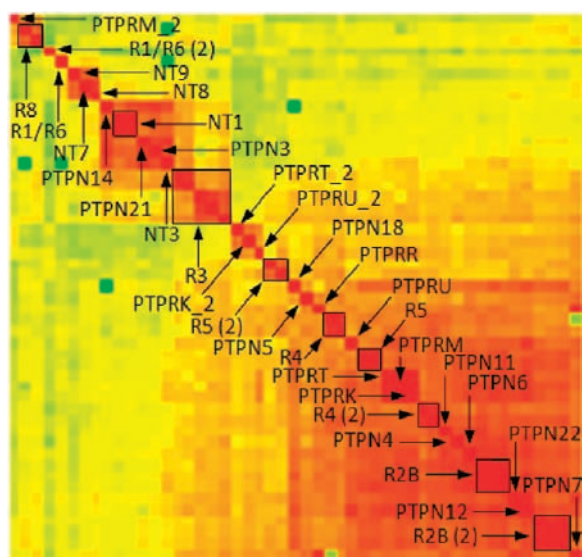

Figure 8. Hierarchical clustering based on catalytic site similarities of the classical PTPs. More detailed clustering presentation can be found in Supporting Information Figure S8.

Andersen et al.[5] have shown clustering of vertebrate classical PTP domains into 17 subtypes based on sequence alignment. Our hierarchical clustering of the domain sequence similarity matrix of the classical PTPs reproduced the identical subtypes with only one exception. PTPRU had previously been categorized as R2A subtype but here does not cluster within this group (Figure 7). However, the distal domains of the subtype R2B members group together into the R2B (2) cluster, which includes the distal domain of PTPRU. The other membrane-distal domains cluster in the same way as their membrane-proximate domains without any exception.

The corresponding analysis based on catalytic site similarities (Figure 8) shows a different picture.
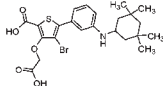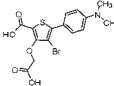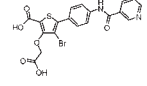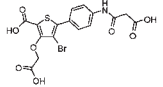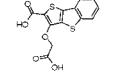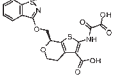
While the subtypes of classical PTPs are defined on the basis of sequence similarity, different clustering results are obtained from site similarities. The conserved groupings are
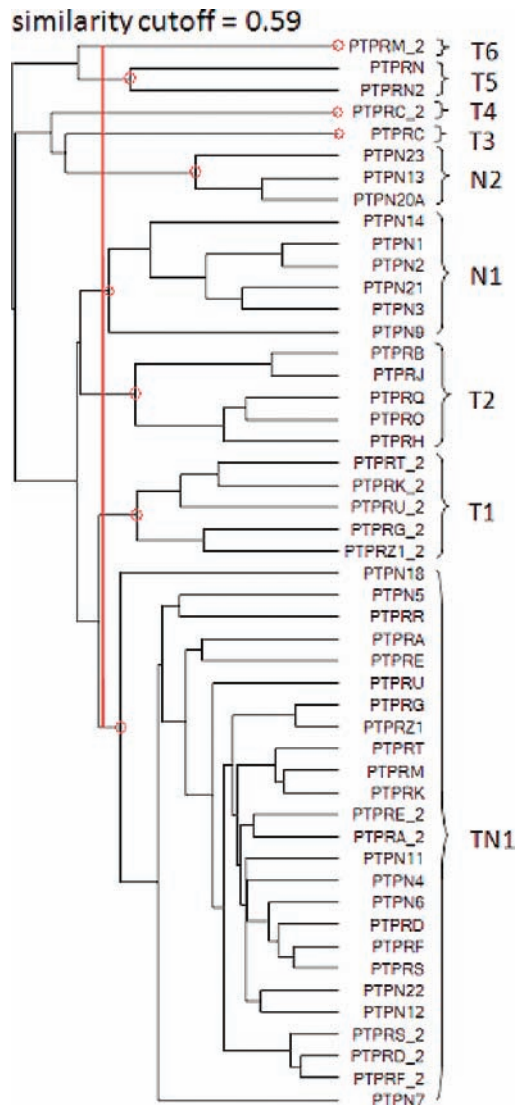
Figure 9. Suggested main groupings of 49 classical PTP domains based on similarity of the catalytic sites.

NT1, R3, R4, R4(2) (distal domains of R4 subtype), R5 and R5(2) (R5 distal domains), R8, R2B, and distal domains of R2B(2) subtype. However, in site similarity space most of the (small) sequence-based subtype groupings are not conserved and different clusters are formed. While sequence similarity clustering primarily defines numerous small groups, site similarity clustering (Figure 8) suggests a few larger groups depending on the similarity cutoff. On the basis of the similarities of the catalytic sites, we suggest a different categorization of the classical PTPs. The dendogram (with corresponding gene annotation) and the clusters formed using a site similarity cutoff of 0.59 are shown in Figure 9. The largest group includes the majority of the transmembrane classical PTPs (22 receptor domains from subtypes R2A, R2B, R4, R5, R7). We name this group TN1 (T representing transmembrane and N nonreceptor classical PTPs). The group in the center of the site similarity dendrogram includes the distal domain of subtype R5 (PTPRG_2 and PTPRZ1_2) and distal domains of three members of R2A (PTPRU_2, PTPRK_2, and PTPRT_2). This cluster is annotated as T1. Two small but distinct clusters are positioned in the upper middle part. The first one, T2, contains the R3 subtype (PTPRB, PTPRJ, PTPRQ, PTPRO, and

Table 3. Experimental Activities of PTPN1 Inhibitors across Different Classical PTPs[a]

Molecule		PTPN2	PTPN1	PTPRC	PTPRF	PTPN6	PTPRB	PTPRE	PTPRA
	a1	1.3	3.2	280	>500	-	-	-	-
	a2	0.026	0.036	151	>1000	-	-	-	-
	b1	1	1.6	64	>1000	-	-	-	-
	b2	0.68	0.82	55	320	-	-	-	-
	b3	0.32	0.3	26	>1000	-	-	-	-
	b4	0.18	0.14	33	>1000	-	-	-	-
	c1	4.1	9.2	>1250	1100	-	-	-	-
	d1	-	23	160	>2000	510	33	130	870
	d2	-	9.9	37	68	94	14	45	190
	d3	-	14	56	550	16	6	66	800
	d4	-	8	20	450	28	3.6	33	180
	d5	-	14	49	420	84	17	33	89
	e1	-	58	260	>2000	>2000	160	360	>2000
	e2	-	62	210	>2000	60	18	740	1700
	e3	-	8.1	41	410	100	19	45	290
	e4	-	0.29	59	>2000	>2000	560	1100	>2000
	e5	-	14	53	360	350	11	20	170
	f1	1.1	0.6	489	176	289	21	>500	>500

[a] Activity given as inhibition constant K_i in μM.
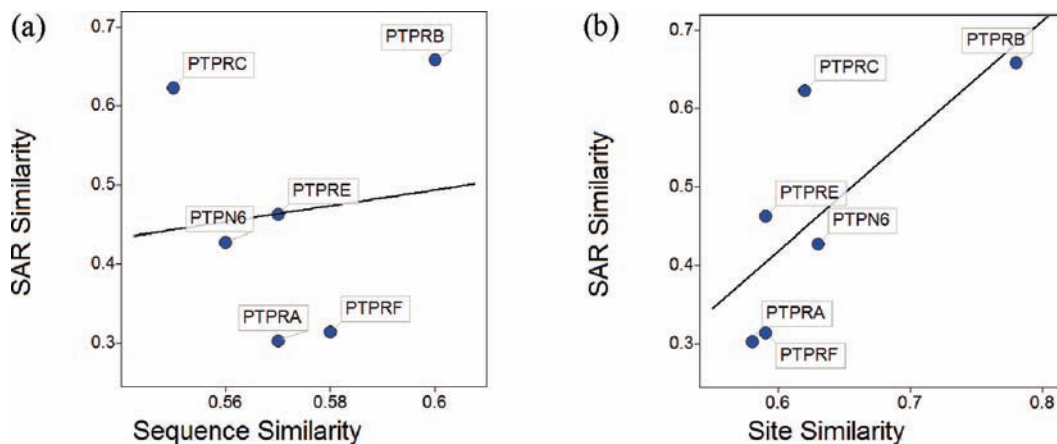
Figure 10. Correlation of SAR-based similarity (calculated from small molecule activity data) to sequence similarities (a) and catalytic site similarities (b) of PTPN1 and the PTPs.

PTPRH). The second cluster includes all members of NT1 (PTPN1 and PTPN2), NT3 (PTPN9), and NT6 (PTPN14 and PTPN21) and one member of NT5 (PTPN3) subtype and is annotated as N1 (nonreceptor classical PTPs, Figure 9). Cluster N2 contains three nonreceptor PTPs (PTPN13, PTPN20A, and PTPN23). Proximal and distal domains of PTPRC are singletons and denoted as T3 and T4, respectively. The last group with just two members (PTPRN and PTPRN2) is T5, while the distal domain of PTPRM_2 forms another singleton annotated as T6.

The average proximity of the classical PTP appears much closer in site vs sequence space (relative to the distance to all the other PTPs). The differences in grouping and in particular a few larger clusters in site similarity space are also illustrated in the PCA plots in the Supporting Information (Figures S9 and S10). The global trend of site vs sequence similarity of the classical PTPs (Figure S11 Supporting Information) also suggests that the classical PTPs are on average much more similar by their catalytic sites compared to their domain sequences. This observation is substantiated by the experienced difficulty to develop highly selective inhibitors for classical PTPs.[64,65] The comparison of the catalytic sites of one such example, PTPN1 (the most studied PTP) and PTPN3, is shown in Figure S12 in the Supporting Information, illustrating better evolutionary preservation of the catalytic sites compared to sequences alone.

Mapping of Small Molecule Inhibition Data to PTP Site Similarities. We explored the PDB in order to collect small molecule PTP inhibitors with binding modes in accordance with our site definition (active PTP form, 10 Å radius around the catalytic cysteine). As expected, the majority of published structures belong to PTPN1. We retrieved experimental activity data for PTPN1 inhibitors screened across a panel of PTPs.[31,35,59−62] Selected PTPN1 ligands and their experimental activities are given in Table 3.

To evaluate target similarities based on these small molecule activity data, we calculated pK_i values and assigned pK_i of 2 to inactive compounds (for example, activity of >1000 in Table 3). We individually looked at the two subsets of full SAR matrices, one including selectivity data of compounds **a1**, **a2**, **b1−b4**, and **c1** against PTPN1, PTPN2, PTPRC, and PTPRF, and the second set of compounds **d1−d5**, **e1−e5**, and **f1** screened against PTPN1, PTPRA, PTPRB, PTPRC, PTPRE, PTPRF, and PTPN6. For each data set the small molecule activity-based similarities between PTPN1 and the

other PTPs were calculated from the Euclidean distances of the corresponding activity vectors defined as the pK_i values of all compounds tested against the respective PTP. The so-determined activity-based similarities of PTP pairs are much better correlated to catalytic site similarities than to sequence similarities. For example, Figure 10 illustrates the correlation of activity-based (SAR) similarity against sequence (a) and site similarity (b), respectively, for the second data set above including PTPN1, PTPRA, PTPRB, PTPRC, PTPRE, PTPRF, and PTPN6. While SAR similarity is uncorrelated to sequences ($r^2 = 0.013$), the square correlation coefficient to site similarity is 0.54. Correlations for the first data set including PTPN1, PTPN2, PTPRC, and PTPRF are shown in the Supporting Information (Figure S13). Again, SAR similarity is better correlated to site similarity ($r^2 = 0.88$) compared to sequences ($r^2 = 0.76$).

SAR-based similarities among the PTPs were further evaluated by hierarchical clustering of the two experimental data sets based on the correlation of the activity vectors (Figure S14, Supporting Information). The proximity of each PTP to PTPN1 (activity data sets of PTPN1 inhibitors are used) based on clustering the experimental pK_i values again is in better agreement with site than with sequence similarities (Table S3, Supporting Information). A complete list of sequence and site similarities for all classical PTPs relative to PTPN1 is given in Table S4 in the Supporting Information. The promiscuity of inhibitors for PTPN1 and PTPN2 can be explained by the high similarity of their catalytic sites (Supporting Information Figure S15). To develop (PTPN1/PTPN2) selective inhibitors, it is therefore required to utilize interactions with residues outside the binding sites considered in this analysis, for example, by designing bidentate inhibitors that bind to the catalytic and a so-called second binding site.[66]

Since our binding sites are defined as a set of residues within 10 Å radius around the catalytic cysteine, we considered PTPN1 inhibitors that do not overextend that volume. To compare larger inhibitors that reach residues beyond 10 Å radius, the binding sites would have to be further extended to include these additional residues. As a consequence, the PTP site similarities would also be different. Systematic analysis of binding site similarity relationships as a function of cutoff radii around the catalytic residues may reveal differences and similarities among PTPs that are particularly relevant in the context of ligands of

specific size. However, we consider the binding site definition we applied here as appropriate for most small molecule inhibitors.

In summary, the analysis of activity data of small molecule PTP inhibitors leads to the conclusion that local site similarities correspond much better to experimental observations than sequence similarities. In developing selective inhibitors, binding site similarity as described here may therefore be useful as a first-order assessment to identify similar targets, which should be tested experimentally.

Conclusions

We have performed the most comprehensive analysis of the human PTP family based on domain sequences and for the first time evaluated the three-dimensional binding site similarities of the entire family. Using a parallel modeling approach, we can amplify the currently existing PTP structural space covering all 113 PTP domains in their active conformation. We observe a global (and expected) trend that PTPs are generally more similar on the basis of the functionally relevant three-dimensional sites around the catalytic residues compared to their overall domain sequences. This is in particular the case for the classical PTPs. The analysis of site vs sequence similarity space confirms comparable major global groupings by PTP subtypes. However, clustering details and analysis of local neighborhoods reveal significant differences within the subtypes and how they are connected. Focusing on classical PTPs, we suggest a novel categorization based on local site similarities as an alternative to the sequence-based categorization.

On the basis of available experimental data, we show that cross-reactivity and selectivity, two critical criteria in lead optimization, can be better understood in the context of site similarity compared to sequence similarity alone. Examples of PTPs that are more closely related by their binding sites compared to sequences illustrate that site similarity may be a useful measure to aid in the development of inhibitors targeting the catalytic domain. We conclude that local site similarities better than sequence similarities reflect the propensity of a PTP for promiscuity or selectivity of small molecule inhibitors.

This work is a relevant starting point to improve our understanding of substrate specificity, selectivity, and cross-reactivity among PTPs, and it provides a first-order structural basis for the development of specific and strongly binding PTP inhibitors. It also gives a new insight into global and local relationships among all members of the human PTP family.

Supporting Information Available: List of human PTP gene family, sequence to model identities, global trend figures for different site and sequence similarities, and additional figures for analysis of classical PTPs. This material is available free of charge via the Internet at http://pubs.acs.org. Generated homology models are available from the author in an upcoming Web site.

References

(1) Zhang, Z. Y. Protein-tyrosine phosphatases: biological function, structural characteristics, and mechanism of catalysis. *Crit. Rev. Biochem. Mol. Biol.* **1998**, *33*, 1–52.

(2) Tonks, N. K. Introduction: protein tyrosine phosphatases. *Semin. Cell Biol.* **1993**, *4*, 373–377.

(3) Barford, D.; Jia, Z.; Tonks, N. K. Protein tyrosine phosphatases take off. *Nat. Struct. Biol.* **1995**, *2*, 1043–1053.

(4) Tonks, N. K.; Neel, B. G. From form to function: signaling by protein tyrosine phosphatases. *Cell* **1996**, *87*, 365–368.

(5) Andersen, J. N.; Mortensen, O. H.; Peters, G. H.; Drake, P. G.; Iversen, L. F.; Olsen, O. H.; Jansen, P. G.; Andersen, H. S.; Tonks, N. K.; Moller, N. P. Structural and evolutionary relationships among protein tyrosine phosphatase domains. *Mol. Cell. Biol.* **2001**, *21*, 7117–7136.

(6) Burke, T. R., Jr.; Zhang, Z. Y. Protein-tyrosine phosphatases: structure, mechanism, and inhibitor discovery. *Biopolymers* **1998**, *47*, 225–241.

(7) Hanks, S. K.; Quinn, A. M. Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. *Methods Enzymol.* **1991**, *200*, 38–62.

(8) Hubbard, S. R.; Till, J. H. Protein tyrosine kinase structure and function. *Annu. Rev. Biochem.* **2000**, *69*, 373–398.

(9) Hunter, T. A thousand and one protein kinases. *Cell* **1987**, *50*, 823–829.

(10) Hunter, T. Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. *Cell* **1995**, *80*, 225–236.

(11) Hooft van Huijsduijnen, R. Protein tyrosine phosphatases: counting the trees in the forest. *Gene* **1998**, *225*, 1–8.

(12) Zhang, Z. Y. Protein tyrosine phosphatases: structure and function, substrate specificity, and inhibitor development. *Annu. Rev. Pharmacol. Toxicol.* **2002**, *42*, 209–234.

(13) Tonks, N. K. Protein tyrosine phosphatases: from genes, to function, to disease. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 833–846.

(14) Tautz, L.; Pellecchia, M.; Mustelin, T. Targeting the PTPome in human disease. *Expert Opin. Ther. Targets* **2006**, *10*, 157–177.

(15) Alonso, A.; Sasin, J.; Bottini, N.; Friedberg, I.; Osterman, A.; Godzik, A.; Hunter, T.; Dixon, J.; Mustelin, T. Protein tyrosine phosphatases in the human genome. *Cell* **2004**, *117*, 699–711.

(16) Barr, A. J.; Ugochukwu, E.; Lee, W. H.; King, O. N.; Filippakopoulos, P.; Alfano, I.; Savitsky, P.; Burgess-Brown, N. A.; Muller, S.; Knapp, S. Large-scale structural analysis of the classical human protein tyrosine phosphatome. *Cell* **2009**, *136*, 352–363.

(17) Lesk, A. M.; Chothia, C. How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J. Mol. Biol.* **1980**, *136*, 225–270.

(18) Chothia, C.; Lesk, A. M. The relation between the divergence of sequence and structure in proteins. *EMBO J.* **1986**, *5*, 823–826.

(19) Bajaj, M.; Blundell, T. Evolution and the tertiary structure of proteins. *Annu. Rev. Biophys. Bioeng.* **1984**, *13*, 453–492.

(20) Sankar, N.; Machado, J.; Abdulla, P.; Hilliker, A. J.; Coe, I. R. Comparative genomic analysis of equilibrative nucleoside transporters suggests conserved protein structure despite limited sequence identity. *Nucleic Acids Res.* **2002**, *30*, 4339–4350.

(21) Kinnings, S. L.; Jackson, R. M. Binding site similarity analysis for the functional classification of the protein kinase family. *J. Chem. Inf. Model.* **2009**, *49*, 318–329.

(22) Guan, K. L.; Dixon, J. E. Evidence for protein-tyrosine-phosphatase catalysis proceeding via a cysteine-phosphate intermediate. *J. Biol. Chem.* **1991**, *266*, 17026–17030.

(23) Pot, D. A.; Woodford, T. A.; Remboutsika, E.; Haun, R. S.; Dixon, J. E. Cloning, bacterial expression, purification, and characterization of the cytoplasmic domain of rat LAR, a receptor-like protein tyrosine phosphatase. *J. Biol. Chem.* **1991**, *266*, 19688–19696.

(24) Barford, D.; Flint, A. J.; Tonks, N. K. Crystal structure of human protein tyrosine phosphatase 1B. *Science* **1994**, *263*, 1397–1404.

(25) Jia, Z.; Barford, D.; Flint, A. J.; Tonks, N. K. Structural basis for phosphotyrosine peptide recognition by protein tyrosine phosphatase 1B. *Science* **1995**, *268*, 1754–1758.

(26) Zhang, Z. Y. Chemical and mechanistic approaches to the study of protein tyrosine phosphatases. *Acc. Chem. Res.* **2003**, *36*, 385–392.

(27) Yang, J.; Liang, X.; Niu, T.; Meng, W.; Zhao, Z.; Zhou, G. W. Crystal structure of the catalytic domain of protein-tyrosine phosphatase SHP-1. *J. Biol. Chem.* **1998**, *273*, 28199–28207.

(28) Yang, J.; Niu, T.; Zhang, A.; Mishra, A. K.; Zhao, Z. J.; Zhou, G. W. Relation between the flexibility of the WPD loop and the activity of the catalytic domain of protein tyrosine phosphatase SHP-1. *J. Cell. Biochem.* **2001**, *84*, 47–55.

(29) Burke, T. R., Jr.; Lee, K. Phosphotyrosyl mimetics in the development of signal transduction inhibitors. *Acc. Chem. Res.* **2003**, *36*, 426–433.

(30) Groves, M. R.; Yao, Z. J.; Roller, P. P.; Burke, T. R., Jr.; Barford, D. Structural basis for inhibition of the protein tyrosine phosphatase 1B by phosphotyrosine peptide mimetics. *Biochemistry* **1998**, *37*, 17773–17783.

(31) Andersen, H. S.; Iversen, L. F.; Jeppesen, C. B.; Branner, S.; Norris, K.; Rasmussen, H. B.; Moller, K. B.; Moller, N. P. 2-(Oxalylamino)-benzoic acid is a general, competitive inhibitor of protein-tyrosine phosphatases. *J. Biol. Chem.* **2000**, *275*, 7101–7108.

(32) Liu, G.; Xin, Z.; Pei, Z.; Hajduk, P. J.; Abad-Zapatero, C.; Hutchins, C. W.; Zhao, H.; Lubben, T. H.; Ballaron, S. J.; Haasch, D. L.; Kaszubska, W.; Rondinone, C. M.; Trevillyan, J. M.; Jirousek, M. R. Fragment screening and assembly: a highly efficient approach to a selective and cell active protein tyrosine phosphatase 1B inhibitor. *J. Med. Chem.* **2003**, *46*, 4232–4235.

(33) Klopfenstein, S. R.; Evdokimov, A. G.; Colson, A. O.; Fairweather, N. T.; Neuman, J. J.; Maier, M. B.; Gray, J. L.; Gerwe, G. S.; Stake, G. E.; Howard, B. W.; Farmer, J. A.; Pokross, M. E.; Downs, T. R.; Kasibhatla, B.; Peters, K. G. 1,2,3,4-Tetrahydroisoquinolinyl sulfamic acids as phosphatase PTP1B inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 1574–1578.

(34) Ala, P. J.; Gonneville, L.; Hillman, M. C.; Becker-Pasha, M.; Wei, M.; Reid, B. G.; Klabe, R.; Yue, E. W.; Wayland, B.; Douty, B.; Polam, P.; Wasserman, Z.; Bower, M.; Combs, A. P.; Burn, T. C.; Hollis, G. F.; Wynn, R. Structural basis for inhibition of protein-tyrosine phosphatase 1B by isothiazolidinone heterocyclic phosphonate mimetics. *J. Biol. Chem.* **2006**, *281*, 32784–32795.

(35) Iversen, L. F.; Andersen, H. S.; Branner, S.; Mortensen, S. B.; Peters, G. H.; Norris, K.; Olsen, O. H.; Jeppesen, C. B.; Lundt, B. F.; Ripka, W.; Moller, K. B.; Moller, N. P. Structure-based design of a low molecular weight, nonphosphorus, nonpeptide, and highly selective inhibitor of protein-tyrosine phosphatase 1B. *J. Biol. Chem.* **2000**, *275*, 10300–10307.

(36) Yue, E. W.; Wayland, B.; Douty, B.; Crawley, M. L.; McLaughlin, E.; Takvorian, A.; Wasserman, Z.; Bower, M. J.; Wei, M.; Li, Y.; Ala, P. J.; Gonneville, L.; Wynn, R.; Burn, T. C.; Liu, P. C.; Combs, A. P. Isothiazolidinone heterocycles as inhibitors of protein tyrosine phosphatases: synthesis and structure–activity relationships of a peptide scaffold. *Bioorg. Med. Chem.* **2006**, *14*, 5833–5849.

(37) Eyre, T. A.; Ducluzeau, F.; Sneddon, T. P.; Povey, S.; Bruford, E. A.; Lush, M. J. The HUGO gene nomenclature database, 2006 updates. *Nucleic Acids Res.* **2006**, *34*, D319–D321.

(38) Bairoch, A.; Apweiler, R. The SWISS-PROT protein sequence data bank and its new supplement TREMBL. *Nucleic Acids Res.* **1996**, *24*, 21–25.

(39) *Target Informatics Platform* (*TIP*); Eidogen-Sertanty, Inc.: Oceanside, CA; http://eidogen-sertanty.com/.

(40) Sonnhammer, E. L.; Eddy, S. R.; Durbin, R. Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* **1997**, *28*, 405–420.

(41) Bateman, A.; Coin, L.; Durbin, R.; Finn, R. D.; Hollich, V.; Griffiths-Jones, S.; Khanna, A.; Marshall, M.; Moxon, S.; Sonnhammer, E. L.; Studholme, D. J.; Yeats, C.; Eddy, S. R. The Pfam protein families database. *Nucleic Acids Res.* **2004**, *32*, D138–D141.

(42) Finn, R. D.; Mistry, J.; Schuster-Bockler, B.; Griffiths-Jones, S.; Hollich, V.; Lassmann, T.; Moxon, S.; Marshall, M.; Khanna, A.; Durbin, R.; Eddy, S. R.; Sonnhammer, E. L.; Bateman, A. Pfam: clans, Web tools and services. *Nucleic Acids Res.* **2006**, *34*, D247–D251.

(43) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680.

(44) McCleverty, C. J.; Lin, D. C.; Liddington, R. C. Structure of the PTB domain of tensin1 and a model for its recruitment to fibrillar adhesions. *Protein Sci.* **2007**, *16*, 1223–1229.

(45) Needleman, S. B.; Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **1970**, *48*, 443–453.

(46) Debe, D. A.; Danzer, J. F.; Goddard, W. A.; Poleksic, A. STRUCTFAST: protein sequence remote homology detection and alignment using novel dynamic programming and profile–profile scoring. *Proteins* **2006**, *64*, 960–967.

(47) Hambly, K.; Danzer, J.; Muskal, S.; Debe, D. A. Interrogating the druggable genome with structural informatics. *Mol. Diversity* **2006**, *10*, 273–281.

(48) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(49) Kolmodin, K.; Aqvist, J. Prediction of a ligand-induced conformational change in the catalytic core of Cdc25A. *FEBS Lett.* **2000**, *465*, 8–11.

(50) Begley, M. J.; Taylor, G. S.; Brock, M. A.; Ghosh, P.; Woods, V. L.; Dixon, J. E. Molecular basis for substrate recognition by MTMR2, a myotubularin family phosphoinositide phosphatase. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 927–932.

(51) Schmitt, S.; Kuhn, D.; Klebe, G. A new method to detect related function among proteins independent of sequence and fold homology. *J. Mol. Biol.* **2002**, *323*, 387–406.

(52) Brint, A. T.; Willett, P. Upperbound procedures for the identification of similar three-dimensional chemical structures. *J. Comput.-Aided Mol. Des.* **1989**, *2*, 311–320.

(53) *Spotfire*, version 9.0; TIBCO: Palo Alto, CA, 2007.

(54) Kruskal, J. B. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. Am. Math. Soc.* **1956**, *7*, 48–50.

(55) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504.

(56) *Schrodinger, Suite 2008*; Schrodinger, LLC.: Portland, OR, 2008.

(57) *The PyMOL Molecular Graphics System*; DeLano Scientific LLC: Palo Alto, CA, 2008.

(58) *SciTegic Pipeline Pilot*, version 7.0; Accelrys Software: San Diego, CA, 2008.

(59) Wilson, D. P.; Wan, Z. K.; Xu, W. X.; Kirincich, S. J.; Follows, B. C.; Joseph-McCarthy, D.; Foreman, K.; Moretto, A.; Wu, J.; Zhu, M.; Binnun, E.; Zhang, Y. L.; Tam, M.; Erbe, D. V.; Tobin, J.; Xu, X.; Leung, L.; Shilling, A.; Tam, S. Y.; Mansour, T. S.; Lee, J. Structure-based optimization of protein tyrosine phosphatase 1B inhibitors: from the active site to the second phosphotyrosine binding site. *J. Med. Chem.* **2007**, *50*, 4681–4698.

(60) Wan, Z. K.; Lee, J.; Xu, W.; Erbe, D. V.; Joseph-McCarthy, D.; Follows, B. C.; Zhang, Y. L. Monocyclic thiophenes as protein tyrosine phosphatase 1B inhibitors: capturing interactions with Asp48. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 4941–4945.

(61) Moretto, A. F.; Kirincich, S. J.; Xu, W. X.; Smith, M. J.; Wan, Z. K.; Wilson, D. P.; Follows, B. C.; Binnun, E.; Joseph-McCarthy, D.; Foreman, K.; Erbe, D. V.; Zhang, Y. L.; Tam, S. K.; Tam, S. Y.; Lee, J. Bicyclic and tricyclic thiophenes as protein tyrosine phosphatase 1B inhibitors. *Bioorg. Med. Chem.* **2006**, *14*, 2162–2177.

(62) Iversen, L. F.; Andersen, H. S.; Moller, K. B.; Olsen, O. H.; Peters, G. H.; Branner, S.; Mortensen, S. B.; Hansen, T. K.; Lau, J.; Ge, Y.; Holsworth, D. D.; Newman, M. J.; Hundahl Moller, N. P. Steric hindrance as a basis for structure-based design of selective inhibitors of protein-tyrosine phosphatases. *Biochemistry* **2001**, *40*, 14812–14820.

(63) Hillisch, A.; Pineda, L. F.; Hilgenfeld, R. Utility of homology models in the drug discovery process. *Drug Discovery Today* **2004**, *9*, 659–669.

(64) Guo, X. L.; Shen, K.; Wang, F.; Lawrence, D. S.; Zhang, Z. Y. Probing the molecular basis for potent and selective protein-tyrosine phosphatase 1B inhibition. *J. Biol. Chem.* **2002**, *277*, 41014–41022.

(65) Wiesmann, C.; Barr, K. J.; Kung, J.; Zhu, J.; Erlanson, D. A.; Shen, W.; Fahr, B. J.; Zhong, M.; Taylor, L.; Randal, M.; McDowell, R. S.; Hansen, S. K. Allosteric inhibition of protein tyrosine phosphatase 1B. *Nat. Struct. Mol. Biol.* **2004**, *11*, 730–737.

(66) Puius, Y. A.; Zhao, Y.; Sullivan, M.; Lawrence, D. S.; Almo, S. C.; Zhang, Z. Y. Identification of a second aryl phosphate-binding site in protein-tyrosine phosphatase 1B: a paradigm for inhibitor design. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 13420–13425.