

In Silico Screening for PTPN22 Inhibitors: Active Hits from an Inactive Phosphatase Conformation

Shuangding Wu, Massimo Bottini, Robert C. Rickert, Tomas Mustelin, and Lutz Tautz*[a]

A gain-of-function mutant of the lymphoid phosphatase Lyp (PTPN22) has recently been implicated in type 1 diabetes and other autoimmune diseases, suggesting that small-molecule inhibitors of Lyp could be useful for the treatment of autoimmunity. Virtual ligand screening (VLS) was applied in the search for hit compounds. Two different docking algorithms, FlexX and ICM, were used to screen a library of 'drug-like' molecules against two different 3D structures, representing the catalytic site of Lyp in both the inactive 'open' and active 'closed' conformations. The top-scoring compounds of each VLS run were tested for their in-

hibitory activity against recombinant Lyp. Interestingly, VLS with both active and inactive conformations yielded very potent hits, with IC_{50} values in the sub- and low-micromolar range. Moreover, many of these hits showed high docking scores only with one conformation. For instance, this was the case with several 2-benzamidobenzoic acid derivatives, which specifically docked into the inactive open form. Tryptophan fluorescence measurements further support a binding mode in which these compounds seem to stabilize the phosphatase in its inactive conformation.

Introduction

Tyrosine phosphorylation^[1] plays an extremely important role in many processes that are characteristic of higher eukaryotes, such as cell-to-cell communication and coordination of the behavior of cell populations within multicellular organisms.^[2] This rapidly reversible post-translational modification is catalyzed by protein tyrosine kinases (PTKs) and is reversed by protein tyrosine phosphatases (PTPs). Thereby, PTPs often play very specific, non-redundant, and highly regulated roles. The human genome contains at least 107 PTP genes, with 81 genes encoding active protein phosphatases. Although PTKs have been considered potential drug targets for some 20 years and PTK inhibitors are already on the market (e.g. Gleevec, Iressa, Tarceva), PTPs have only recently been implicated in human diseases, including cancer and diabetes.^[3,4,5] The lymphoid tyrosine phosphatase Lyp,^[6] which is encoded by the *ptpn22* gene, has a critical negative regulatory role in T cell receptor signaling. Recently, a single-nucleotide polymorphism in Lyp was discovered to correlate strongly with the incidence of type 1 diabetes^[7] and other autoimmune diseases such as rheumatoid arthritis,^[8] juvenile rheumatoid arthritis,^[9] systemic lupus erythematosus,^[10] and Grave's disease.^[7] Because the autoimmunity-predisposing allele is a gain-of-function mutant,^[11] a specific small-molecule inhibitor of Lyp could be beneficial in treating these diseases.

Based on the increasing number of available three-dimensional structures of PTPs in recent years, in silico methods have become increasingly popular as hit/lead discovery tools for tyrosine phosphatases.^[12] Depending on the conformation of the WPD loop, which contains the catalytically essential general acid/base aspartic acid residue, two types of PTP structures can be typically found: The inactive 'open' conformation refers to the WPD loop in a position distal to the catalytic pocket. Substrate or ligand binding to the bottom of the catalytic

pocket causes the loop to shift by ~ 8 Å, forming the active 'closed' conformation.^[13] Usually, only structures in the closed conformation are considered as suitable receptors for in silico screens of inhibitors thought to target the active site. Herein we present how virtual ligand screening (VLS) with a structure that contains the WPD loop in the open conformation can also lead to unique and potent hits. Analysis of the docking poses for these compounds as well as tryptophan fluorescence measurements suggest a binding mode that is very specific and that appears to stabilize Lyp in its inactive conformation.

Results and Discussion

In silico screening

To identify hit compounds for Lyp by VLS, two docking algorithms, FlexX^[14] and ICM,^[15] were employed to screen a library of 27030 compounds. A high-resolution crystal structure of Lyp's catalytic domain in the open conformation (LypO, PDB code: 2P6X) was used, as well as a homology model of Lyp in the closed conformation (LypC), as a crystal or NMR structure was not available for the closed form. The modeled LypC could be structurally aligned with LypO with an RMSD of 1.52 Å when WPD loop atoms were omitted (Figure 1 A). How-

[a] Dr. S. Wu, Dr. M. Bottini, Prof. Dr. R. C. Rickert, Prof. Dr. T. Mustelin, Dr. L. Tautz

Infectious and Inflammatory Disease Center
Burnham Institute for Medical Research
10901 N. Torrey Pines Rd., La Jolla, CA 92037 (USA)
Fax: (+1) 858-713-9925
E-mail: tautz@burnham.org

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/cmdc.200800375>.

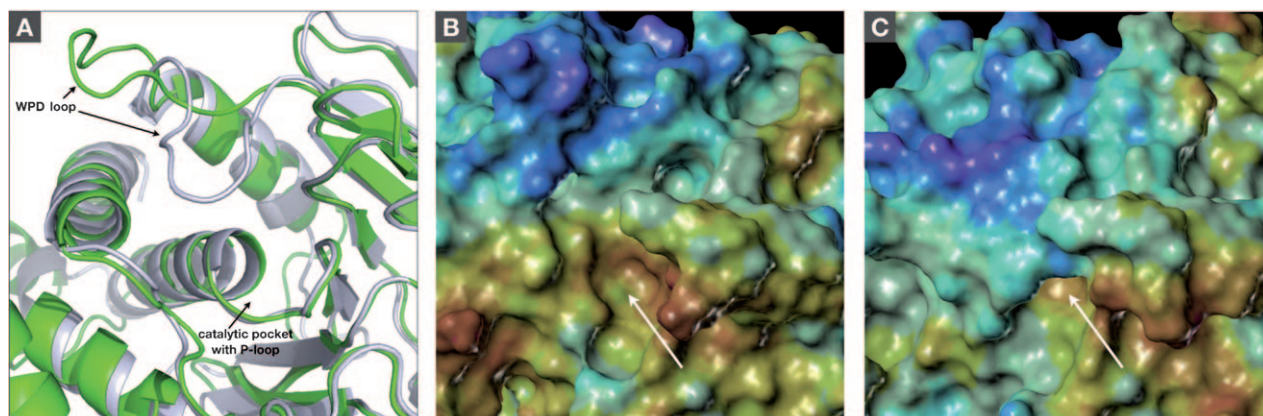


Figure 1. A) Alignment of the crystal structure of Lyp with the WPD loop in the open conformation (PDB code: 2P6X, green) and the homology model of Lyp with the WPD loop in the closed conformation (blue–white). B/C) Active site surface representation of the Lyp crystal structure with the WPD loop in the open conformation (B) and the Lyp homology model with the WPD loop in the closed conformation (C). Structures were aligned, and pictures represent the exact same view. The color code of the MOLCAD surfaces represents the normalized electrostatic potential (red: most positive, purple: most negative). The white arrows indicate the catalytic pocket with the P-loop.

ever, the surface topology around the catalytic pocket, especially toward the WPD loop, differed dramatically between the two structures (Figure 1 B,C), suggesting that VLS should yield distinct hits for each receptor conformation. Four VLS experiments were carried out (LypC with FlexX/ICM and LypO with FlexX/ICM), and compounds were ranked according to their docking scores. For each VLS run, the 20 best-ranked compounds were then chosen for evaluation. Because nine compounds were among more than one top-20 sets, a total of 71 compounds were purchased. Clustering these hits by Tanimoto distance revealed 10 different compound classes with at least two members and 14 singletons at a distance of 0.4 (Supporting Information table S1). As suspected, some of the clusters were very specific to only one receptor conformation.

Evaluation of screening hits

To evaluate the inhibitory activity of the 71 hits, a 96-well plate phosphatase assay was applied. At a compound concentration of 40 μM , inhibitory activity was determined as the percent inhibition relative to a dimethyl sulfoxide (DMSO) control (Supporting Information table S1). The overall performance in generating active compounds was very similar among the four VLS runs. This result was completely unexpected for the two runs using LypO, because this structure does not resemble the active receptor conformation. Interestingly, there was only little overlap of high docking scores for each hit among the various VLS conditions. Moreover, VLS with Lyp in the open conformation yielded several potent inhibitors that completely failed to dock into the closed conformation. For instance, the 2-benzamidobenzoic acid (2-BBA) derivatives, which all ranked among the top-21 hits (inhibition >90%), exhibited high docking scores only with LypO. Their IC_{50} values for Lyp ranged from 0.947 to 22.9 mM (Table 1).

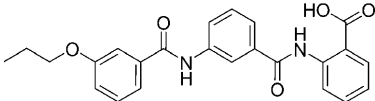
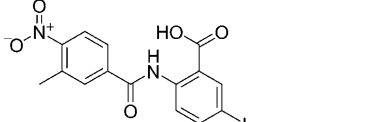
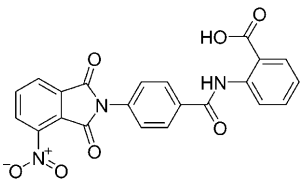
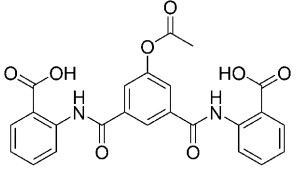
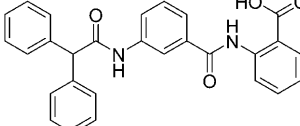
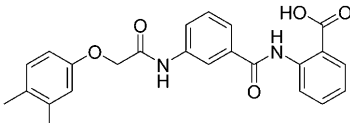
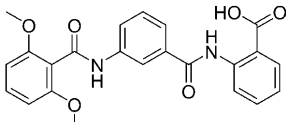
When we evaluated the binding mode of these compounds, all 2-BBAs, except **7**, occupied surface areas in LypO that are not accessible with the WPD loop in the closed conformation

(Figure 2). The docking poses suggested that each molecule binds to both the phosphate binding loop (P-loop) at the bottom of the catalytic pocket as well as the distant WPD loop—a binding mode that may stabilize the WPD loop in the open conformation. Only compound **7** occupied a position within the active site of LypO that resembles that found in LypC with the benzoic acid moiety unable to bind to the P-loop due to steric hindrance of the iodine substituent. Instead, the nitrophenyl group was found to bind into the catalytic pocket; this was confirmed by testing 2-benzamido-5-iodobenzoic acid, which lacks the nitro group and is 157-fold less potent in inhibitory activity toward Lyp. For all other 2-BBAs, the benzamidobenzoic acid group was found to form a complex network of hydrogen-bonding interactions with residues within the P-loop, including the invariant Arg 233, as shown for compound **12** (Figure 3). In addition, **12** was also found to interact specifically with amino acid residues within the WPD loop region. In particular, one of the two phenyl groups seemed to fit perfectly into a hydrophobic pocket, which is formed by parts of the WPD loop and the $\beta 3$ strand in LypO. These hydrophobic interactions of compound **12** with the WPD loop seem to contribute a great deal of binding energy and therefore should be important for inhibitory activity. Because **12** did not dock into the closed receptor conformation with reasonable scores, the observed docking pose may well reflect the actual binding mode, in which an active site inhibitor stabilizes the WPD loop in the open conformation. In fact, during the preparation of this manuscript, Zhong-Yin Zhang's research group reported a co-crystal structure of Lyp and an active site inhibitor, in which Lyp was also found in an open conformation.^[16]

Tryptophan fluorescence measurements

To give further evidence that the 2-BBAs may stabilize the inactive conformation, we used fluorescence spectroscopy to measure the tryptophan emission of recombinant Lyp. Although

Table 1. Docking scores and inhibitory activity of 2-benzamidobenzoic acids sorted by their IC₅₀ values against recombinant Lyp.

Compd	Structure	IC ₅₀ [μM]	LypC FlexX ^[a]	LypC ICM ^[b]	LypO FlexX ^[c]	LypO ICM ^[d]
8		0.947	−29.56	−15.94	−32.04	−37.51
7		0.966	−30.16	−10.69	−32.55	−29.21
6		1.69	−28.37	−17.14	−28.42	−41.73
14		4.85	−27.95	−12.75	−45.68	−33.68
12		7.75	−25.09	−7.09	−37.23	−23.90
11		11.0	−22.71	−5.82	−29.77	−39.93
21		22.9	−18.40	−9.81	−30.96	−43.33

[a] FlexX docking score with LypC. [b] ICM docking score with LypC. [c] FlexX docking score with LypO. [d] ICM docking score with LypO.

there are a total of six tryptophan residues within the catalytic domain, none of them is near the active site except Trp 193 of the WPD loop. At 100, 25, 6.25, or 1.56 μM, all 2-BBAs quenched the tryptophan emission to various extents in a concentration-dependent manner relative to a control, with compound **7** having the highest quenching effect (Figure 4). In addition, a red shift of the emission maximum was observed in the presence of compounds, relative to DMSO alone. This shift was between 3 and 7 nm for all compounds except for **7**, which caused a significantly greater shift of 15 nm. As a substantial conformational change in the WPD loop causes a more profound effect on the fluorescence emission of Trp 193, these data support the notion that all 2-BBAs except compound **7** stabilize Lyp in an open conformation.

Conclusions

We have shown how in silico screening with a tyrosine phosphatase structure in its inactive 'open' conformation can yield very potent inhibitors. These compounds still bind to amino acid residues inside the catalytic pocket with functional groups that mimic the phosphotyrosine group of the natural substrate. However, substantial portions of the molecules seem to interact with a surface area outside the catalytic pocket that would normally be occupied by the WPD loop in the closed conformation. This observation strongly suggests a binding mode different from natural substrates or other known PTP inhibitors. Notably, these interactions seem to be very strong and specific in nature, and most likely stabilize the open conformation of the WPD loop. This principle may apply more broadly to PTPs and other enzymes that undergo substantial conformational cycling during catalysis. Our best compounds

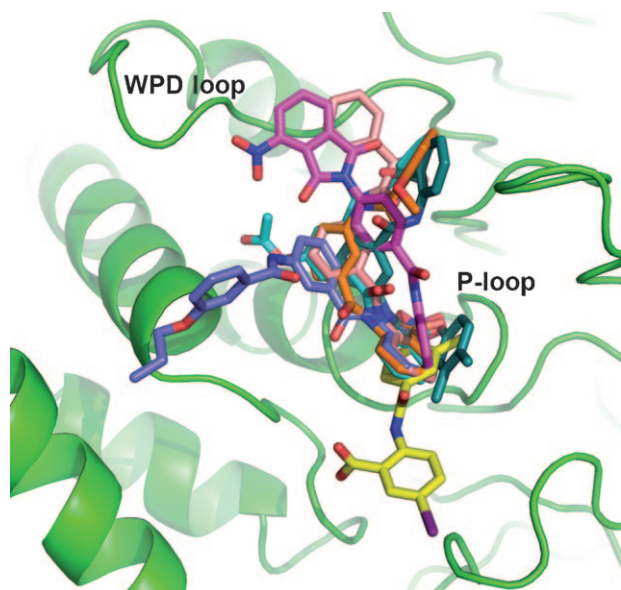


Figure 2. Ribbon diagram of Lyp in the inactive open conformation with FlexX docking solutions of compounds **6** (magenta), **7** (yellow), **8** (blue), **11** (green), **12** (pink), **14** (cyan), and **21** (orange). Compounds are shown in stick representation.

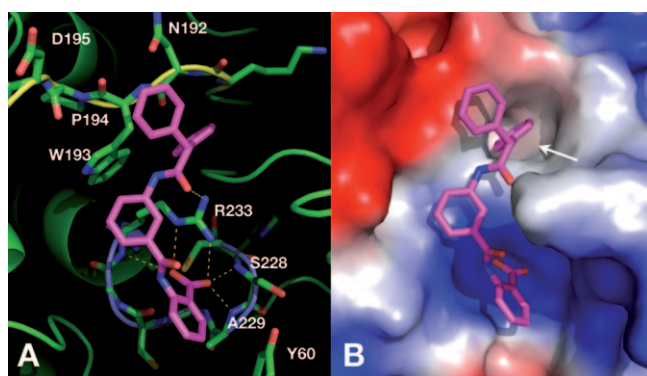


Figure 3. A) Diagram of compound **12** docked to Lyp (ribbon representation) in the open conformation using FlexX. Amino acid residues of the catalytic pocket and the WPD loop, as well as compound **12** are shown in stick representation. The P-loop is colored in light blue; the WPD loop is colored in yellow. Dashed yellow lines indicate hydrogen-bonding interactions between **12** and Lyp. B) Surface diagram of the same docking solution as shown in (A); the white arrow indicates a hydrophobic pocket that is formed by parts of the WPD loop and the β 3 strand.

represent a pool of inhibitors that will be further tested for selectivity and cell-based activity, and may serve as starting points for developing Lyp inhibitors that have the potential for use in the treatment of autoimmune diseases.

Experimental Section

Compound library. To generate a virtual library of 'drug-like' compounds, the DIVERSet library from ChemBridge (ChemBridge, Inc.), containing 50 000 molecules, was decreased by removing compounds that did not meet certain criteria, which were calculated with the ICM software package (Molsoft, LLC). First, compounds

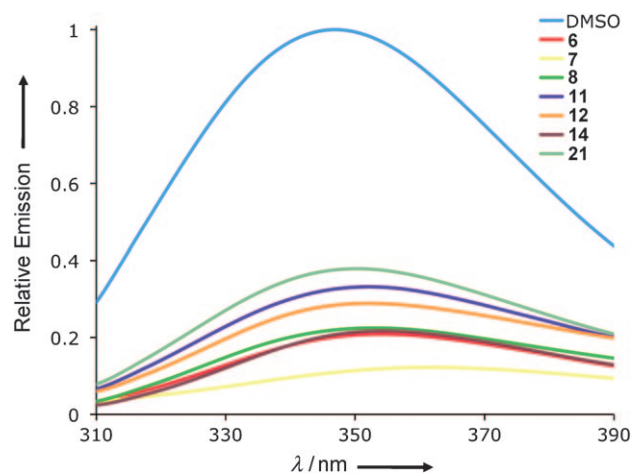


Figure 4. Normalized tryptophan fluorescence emission of the Lyp catalytic domain (1 μ M) in the presence of 2% DMSO or 100 μ M inhibitors with an excitation wavelength of 285 nm.

with so-called 'bad groups'—unwanted functional groups in drugs—were removed from the set.^[17] Second, compounds with a calculated 'drug-likeness' score < -0.5 were removed. Third, compounds with ClogP values < -0.5 and > 5 were removed. And fourth, compounds with polar surface area (PSA) values $> 140 \text{ \AA}^2$ were removed from the set, resulting in a library of 27 030 compounds with molecular weights $< 600 \text{ Da}$. The library was saved as an SD file.

LypC homology modeling. The SWISS-MODEL comparative protein modeling server^[18] (<http://swissmodel.expasy.org/>) was used to generate a homology model for Lyp in the closed conformation (LypC). A high-resolution structure of the close homologue PTP1B (sequence identity 38.6%) in the closed conformation (PDB code: 1PTY) was used as a specific template structure. SwissModel Automatic ModellingMode^[18] was used to generate the model of LypC, which consisted of amino acids 44–288. The model structure was assessed by three different methods, Anolea, Gromos, and Verify3D, which all suggested a reliable model prediction for most parts of the structure, including the active site.

VLS with FlexX. The FlexX algorithm was applied as part of the SYBYL software package (version 7.2, Tripos, Inc.), and calculations were run on a Linux workstation. The homology model of the closed conformation and the crystal structure of the open conformation (PDB code: 2P6X) were prepared with the structure preparation tool as implemented in SYBYL. AMBER charges were assigned, orientations of side chain amides were corrected, and hydrogen atoms were added and their positions optimized by energy minimization using the AMBER7 FF99 force field. The receptor site was defined as a sphere of 8 \AA radius with the catalytic Cys 227 as the center. Three-dimensional coordinates of the compounds were generated using CONCORD as implemented in SYBYL and saved as MOL2 files. For docking, acids were deprotonated, whereas amines were protonated to account for their charged state at pH 6. The 30 lowest-energy poses for each compound and their corresponding Total-Scores (FlexX scores) were calculated.

VLS with ICM. The ICM docking algorithm was applied as part of the ICM software package (version 3.4, Molsoft, LLC), and calculations were run on a PowerPC G5 workstation. The protein structures were converted into ICM objects, charges were assigned, orientations of side chain amides were corrected, and hydrogen

atoms were added and their positions optimized by energy minimization using the MMFF force field. The receptor site was defined as a sphere of 8 Å radius with the catalytic Cys227 as the center. Three-dimensional coordinates of the compounds were generated, and charges were assigned using the implemented compound preparation tool.

Recombinant Lyp. The Lyp catalytic domain (residues 2–309) was cloned into a pGEX-4T-1 vector, and the plasmid was transformed into *E. coli* BL21 DE3 competent cells (Stratagene). For expression, an overnight culture was diluted in LB/kanamycin medium and grown at 37 °C to a cell density of $A_{600}=0.6$. Isopropyl- β -D-1-thiogalactopyranoside (0.4 mM) was used to induce protein overexpression for 4 h at room temperature. Harvested cells were resuspended in 10 mL Extraction Solution #1 (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 5 mM β -mercaptoethanol, 0.005 % NaN_3 , 0.08 mg mL⁻¹ lysozyme, 1 mM phenylmethylsulfonyl fluoride) per gram of cells and stirred at 4 °C for 30 min. Then, 1 mL of 10 \times Extraction Solution #2 (1.5 M NaCl, 100 mM CaCl_2 , 100 mM MgSO_4 , 20 $\mu\text{g mL}^{-1}$ DNase, 50 μg ovomucoid) per 10 mL cell suspension was added. After centrifugation at 17000 rpm (centrifuge: Sorvall RC5C Plus; rotor: Sorvall SS-34), the clear supernatant was loaded onto an affinity column with 5 mL Glutathione Sepharose 4B beads (GE Healthcare), which were pre-equilibrated with wash buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 5 mM β -mercaptoethanol). GST-Lyp was eluted with 10 mM glutathione in wash buffer. The active fraction was collected and further purified on a Superdex S-200 column that was pre-equilibrated with 20 mM Tris-HCl pH 8.0 and 1 mM dithiothreitol (DTT, BioVectra). The purified GST-Lyp was homogeneous as determined by a single band in SDS-PAGE.

Lyp enzymatic assay. Lyp-catalyzed hydrolysis of 6,8-difluoro-4-methylumbelliferyl phosphate (DiFMUP, Invitrogen Inc.) in the presence of compounds was assayed at 30 °C in a 60 μL 96-well format reaction system in 20 mM Bis-Tris (pH 6.0) assay buffer with an ionic strength of 20 mM (adjusted with NaCl) containing 1 mM DTT, 1 % poly(ethylene glycol) (PEG 8000, Sigma) and 5 % DMSO. At a compound concentration of 40 μM , the initial rate at 48 μM DiFMUP (equal to the corresponding K_M value) was determined by using a FLx800 micro-plate reader (Bio-Tek Instruments, Inc.; $\lambda_{\text{ex}}=360$ nm) and by measuring the emission of the fluorescent reaction product 6,8-difluoro-7-hydroxy-4-methylcoumarin (DiFMU) at 460 nm. The non-enzymatic hydrolysis of the substrate was corrected by measuring the control without addition of enzyme.

IC_{50} measurements. With the same buffer system, substrate, and reader as above, the initial rate at 48 μM DiFMUP was determined at various compound concentrations, ranging from 1 mM to 1 nM. The non-enzymatic hydrolysis of the substrate was corrected by measuring the control without addition of enzyme. The IC_{50} value was determined by plotting the relative activity versus inhibitor concentration and fitting to Equation 1 using GraphPad Prism software (GraphPad Software, Inc.).

$$V_i/V_0 = \text{IC}_{50}/(\text{IC}_{50}+[I]) \quad (1)$$

In this case, V_i is the reaction velocity at the inhibitor concentration $[I]$, V_0 is the reaction velocity with no inhibitor, and $\text{IC}_{50}=K_i+K_i[S]/K_M$.

Fluorescence emission measurements. Tryptophan emission spectra were recorded at room temperature on a MOS-250 spectrophotometer (BioLogic, Claix, France) at a scan speed of 60 nm min⁻¹

and an emission bandwidth of 10 nm, using a 250 μL quartz cuvette. The concentration of recombinant Lyp catalytic domain was 1 μM in 20 mM Bis-Tris (pH 6.0), containing 5 mM DTT and 1 % PEG 8000. DMSO or compound in DMSO were added and mixed to a final concentration of 100, 25, 6.25, or 1.56 μM and 2 % DMSO. After incubation for 1 min, the fluorescence emission was measured at an excitation wavelength of 285 nm.

Acknowledgements

This work was supported by grants from the US National Institutes of Health (AI053585 to T.M. and CA132121 to L.T.) and the Oxnard Foundation.

Keywords: autoimmunity • drug design • inhibitors • phosphatases • virtual screening

- [1] T. Hunter, B. M. Sefton, *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 1311–1315.
- [2] A. Alonso, J. Sasin, N. Bottini, I. Friedberg, I. Friedberg, A. Osterman, A. Godzik, T. Hunter, J. Dixon, T. Mustelin, *Cell* **2004**, *117*, 699–711.
- [3] L. Bialy, H. Waldmann, *Angew. Chem.* **2005**, *117*, 3880–3906; *Angew. Chem. Int. Ed.* **2005**, *44*, 3814–3839.
- [4] L. Tautz, M. Pellicchia, T. Mustelin, *Expert Opin. Ther. Targets* **2006**, *10*, 157–177.
- [5] T. Vang, A. V. Miletic, Y. Arimura, L. Tautz, R. C. Rickert, T. Mustelin, *Annu. Rev. Immunol.* **2008**, *26*, 29–55.
- [6] S. Cohen, H. Dadi, E. Shaoul, N. Sharfe, C. M. Roifman, *Blood* **1999**, *93*, 2013–2024.
- [7] N. Bottini, L. Musumeci, A. Alonso, S. Rahmouni, K. Nika, M. Rostamkhani, J. MacMurray, G. F. Meloni, P. Lucarelli, M. Pellicchia, G. S. Eisenbarth, D. Comings, T. Mustelin, *Nat. Genet.* **2004**, *36*, 337–338.
- [8] A. B. Begovich, V. E. Carlton, L. A. Honigberg, S. J. Schrodri, A. P. Chokkalingam, H. C. Alexander, K. G. Ardlie, Q. Huang, A. M. Smith, J. M. Spoecker, M. T. Conn, M. Chang, S. Y. Chang, R. K. Saiki, J. J. Catanese, D. U. Leong, V. E. Garcia, L. B. McAllister, D. A. Jeffery, A. T. Lee, F. Batliwalla, E. Remmers, L. A. Criswell, M. F. Seldin, D. L. Kastner, C. I. Amos, J. J. Sninsky, P. K. Gregersen, *Am. J. Hum. Genet.* **2004**, *75*, 330–337.
- [9] M. B. Ladner, N. Bottini, A. M. Valdes, J. A. Noble, *Hum. Immunol.* **2005**, *66*, 60–64.
- [10] C. Kyogoku, C. D. Langefeld, W. A. Ortmann, A. Lee, S. Selby, V. E. Carlton, M. Chang, P. Ramos, E. C. Baechler, F. M. Batliwalla, J. Novitzke, A. H. Williams, C. Gillett, P. Rodine, R. R. Graham, K. G. Ardlie, P. M. Gaffney, K. L. Moser, M. Petri, A. B. Begovich, P. K. Gregersen, T. W. Behrens, *Am. J. Hum. Genet.* **2004**, *75*, 504–507.
- [11] T. Vang, M. Congia, M. D. Macis, L. Musumeci, V. Orru, P. Zavattari, K. Nika, L. Tautz, K. Tasken, F. Cucca, T. Mustelin, N. Bottini, *Nat. Genet.* **2005**, *37*, 1317–1319.
- [12] L. Tautz, T. Mustelin, *Methods* **2007**, *42*, 250–260.
- [13] Z. Y. Zhang, *Annu. Rev. Pharmacol. Toxicol.* **2002**, *42*, 209–234.
- [14] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, *J. Mol. Biol.* **1996**, *261*, 470–489.
- [15] M. Totrov, R. Abagyan, *Proteins* **1997**, *Suppl. 1*, 215–220.
- [16] X. Yu, J. P. Sun, Y. He, X. Guo, S. Liu, B. Zhou, A. Hudmon, Z. Y. Zhang, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19767–19772.
- [17] M. Hann, B. Hudson, X. Lewell, R. Lifely, L. Miller, N. Ramsden, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 897–902.
- [18] K. Arnold, L. Bordoli, J. Kopp, T. Schwede, *Bioinformatics* **2006**, *22*, 195–201.

Received: November 11, 2008

Revised: December 11, 2008

Published online on January 28, 2009