

In silico Spleen Tyrosine Kinase Inhibitor Screening by ChooseLD

HIDEAKI UMEYAMA^{1,a)} MITSUO IWADATE^{1,b)} Y-H. TAGUCHI^{2,c)}

Received: February 19, 2015, Accepted: May 13, 2015, Released: August 19, 2015

Abstract: Background: Spleen tyrosine kinase (SYK) is a protein related to various diseases. Aberrant SYK expression often causes the progression and initiation of several diseases including cancer and autoimmune diseases. Despite the importance of inhibiting SYK and identifying candidate inhibitors, no clinically effective inhibitors have been reported to date. Therefore, there is a need for novel SYK inhibitors. Results: Candidate compounds were investigated using *in silico* screening by chooseLD, which simulates ligand docking to proteins. Using this system, known inhibitors were correctly recognized as compounds with high affinity to SYK. Furthermore, many compounds in the DrugBank database were newly identified as having high affinity to the ATP-binding sites in the kinase domain with a similar affinity to previously reported inhibitors. Conclusions: Many drug candidate compounds from the DrugBank database were newly identified as inhibitors of SYK. Because compounds registered in the DrugBank are expected to have fewer side effects than currently available compounds, these newly identified compounds may be clinically useful inhibitors of SYK for the treatment of various diseases.

Keywords: FAMS, chooseLD, SYK, *in silico* drug discovery

1. Introduction

Spleen tyrosine kinase (SYK) has been a drug target since it was identified as a disease-related non-receptor kinase [1], [2]. SYK regulates many key proteins that are involved in the initiation or progression of various diseases. The Gendoo server [3], [4] lists many diseases reported to be related to SYK (Table 1). The deletion of SYK has been reported to suppress the formation of immune complex arthritis [5]. SYK has also been reported to be activated in diffuse large B-cell lymphoma [6]. Hypermethylation of the SYK gene promoter region has been reported to be associated with oncogenesis and metastasis of gastric carcinoma [7]. Furthermore, specific inhibition of SYK has been reported to suppress leukocyte immune function and inflammation in animal models of rheumatoid arthritis [8]. We have recently observed that the SYK gene promoter is often aberrantly methylated in three autoimmune diseases [9]. Thus, effective SYK inhibitors are urgently required for the treatment of numerous diseases.

There are several targets for SYK inhibition. SYK consists of a C-terminal kinase domain and two Src homology 2 (SH2) domains separated by a linker domain [10]. Inhibitors that target the kinase domain mainly target the ATP-binding sites. For example, R112, R406, R788 and R343 are structurally related pyrimidine analogs that compete with ATP binding [10].

Alternatively, some inhibitors target the SH2 domains [11]. Many compounds have been reported to inhibit SYK [12]. The inhibition of protein complex formation has also been proposed [13].

Despite these reported studies, no clinically effective SYK inhibitors have been established to date. Based on recent developments in computational methods, many studies have identified drug candidate compounds computationally. Li et al. [14] have tried to identify SYK inhibitors using machine learning methods. Kaur et al. [15] have investigated SYK inhibitors using 3D-Quantitative Structure Activity Relationship (QSAR), and Xie et al. [16] have used chemical features based on 3D pharmacophore models. Although several listed drug candidate compounds have been identified, they were based on features extracted from candidate compounds, thus, the estimation of drug activity was indirect. To our knowledge, there have been no comprehensive screens of compounds that target the kinase domain using docking-based prediction.

This study evaluated the docking affinity between SYK and over 1,000 compounds extracted from the DrugBank database and ranked these based on their binding affinity using chooseLD, a docking-based *in silico* drug-screening software. Top-ranked compounds were considered promising SYK inhibitor drug candidate compounds.

2. Materials and Methods

2.1 Tertiary Structure Prediction of SYK

Tertiary structure prediction of SYK was performed by Full Automatic Modeling System (FAMS) [17], [18]. The amino acid sequence of SYK (uniprot ID P43405.1) in fasta format

¹ Department of Biological Science, Chuo University, Bunkyo, Tokyo 112–8551, Japan

² Department of Physics, Chuo University, Bunkyo, Tokyo 112–8551, Japan

^{a)} umeyama@bio.chuo-u.ac.jp

^{b)} iwadate@bio.chuo-u.ac.jp

^{c)} tag@graular.com

Table 1 Diseases reported to be related to SYK by the Gendoo server.

Diseases	P-values
Breast Neoplasms	3.92×10^{-9}
Arthus Reaction	5.09×10^{-7}
Lymphoma, B-Cell	1.79×10^{-6}
Neoplasm Metastasis	7.67×10^{-6}
Inflammation	8.28×10^{-5}
Agammaglobulinemia	2.38×10^{-4}
Lymphoma, Extranodal NKT-Cell	4.91×10^{-4}
Leukemia	5.00×10^{-4}
Neoplasm Invasiveness	5.38×10^{-4}
Purpura	9.13×10^{-4}
Gonorrhoea	1.33×10^{-3}
Lymphoma, Large B-Cell, Diffuse	1.38×10^{-3}
Ehrlichiosis	1.47×10^{-3}
Lymphoma	3.86×10^{-3}
Leukemia, Lymphocytic, Chronic, B-Cell	3.33×10^{-3}
Lymphedema	3.86×10^{-3}
Lymphoma, T-Cell, Peripheral	5.46×10^{-3}
Leukemia, Basophilic, Acute	6.09×10^{-3}
Mediastinal Neoplasms	7.41×10^{-3}
Carcinoma, Ductal	7.56×10^{-3}
Urticaria	7.77×10^{-3}
Autoimmune Diseases	7.84×10^{-3}
Rhinitis	9.02×10^{-3}
Lymphoma, T-Cell, Cutaneous	9.23×10^{-3}
Synovitis	9.37×10^{-3}
Breast Neoplasms, Male	9.92×10^{-3}
Lymphoma, Large-Cell, Anaplastic	1.14×10^{-2}
Precursor B-Cell Lymphoblastic Leukemia-Lymphoma	1.31×10^{-2}
Leukemia, B-Cell	1.38×10^{-2}
Peritonitis	1.62×10^{-2}
Lymphatic Metastasis	1.69×10^{-2}
Nasal Polyps	1.80×10^{-2}
Bronchial Hyperreactivity	1.81×10^{-2}
Mammary Neoplasms, Animal	2.00×10^{-2}
Carcinoma, Intraductal, Noninfiltrating	2.12×10^{-2}
Arthritis, Rheumatoid	2.25×10^{-2}
Edema	2.34×10^{-2}
Arthritis, Experimental	2.42×10^{-2}
Vasculitis	2.67×10^{-2}
Wiskott-Aldrich Syndrome	2.80×10^{-2}
Stomach Neoplasms	3.09×10^{-2}
Melanoma	3.11×10^{-2}
Melanoma, Experimental	3.55×10^{-2}
Immunologic Deficiency Syndromes	3.63×10^{-2}
Hodgkin Disease	4.28×10^{-2}
Bacterial Infections	4.40×10^{-2}
Shock, Septic	4.70×10^{-2}

obtained from uniprot [19] was uploaded to an isolated FAMS server. Then, the obtained top-ranked model proteins modeled using the Protein Data Bank (PDB) structure 3VF8_A (SYK) were regarded as drug discovery template candidates because of the following reasons.

3VF8_A includes OJE (3-[5-(5-ETHOXY-6-FLUORO-1H-BENZIMIDAZOL-2-YL)-1H-PYRAZOL-4-YL]-1,1-DIETHYLUREA). Each of other enzymes which have almost same sequences and other binding ligands in the same ATP binding sites was superimposed using the Combinatorial Extension (CE) [20] fit program to a temporally reference protein. Each binding ligand was simultaneously transferred using the same transformation matrix used in the CE fit. The enzyme structures having fewer short contact atoms between the enzyme and the transferred ligand set were remarked. As the results, we selected 3VF8_A as the reference protein in this paper.

2.2 Drug Compound Candidates

A total of 6,583 compounds included in the DrugBank [21],

Table 2 Ten template proteins with ligands used for *in silico* screening.

protein		ligand	
PDB	name	ID	name
4DFN	SYK	OK1	3-amino-6-[3-(1-methyl-1H-pyrazol-4-yl)phenyl]-N-[(1R,2r,3S,5s,7s)-5-hydroxyadamantan-2-yl]pyrazine-2-carboxamide
3FQE	SYK	P5C	2-[[[(1R,2S)-2-aminocyclohexyl]amino]-4-[(3-methylphenyl)amino]pyrimidine-5-carboxamide
1XBB	SYK	STI	IMATINIB
1XBC	SYK	STU	STAUROSPORINE
3VF8	SYK	OJE	3-[5-(5-ethoxy-6-fluoro-1H-benzimidazol-2-yl)-1H-pyrazol-4-yl]-1,1-diethylurea
3VF9	SYK	477	3-[2-[5-(difluoromethyl)-2H-thieno[3,2-c]pyrazol-3-yl]-1H-indol-6-yl]pentan-3-ol
3SRV	SYK	S19	GSK143
4DFL	SYK	OK0	3-amino-6-[3-[(methylsulfonyl)amino]phenyl]-N-(piperidin-4-ylmethyl)pyrazine-2-carboxamide
3FQH	SYK	057	N-(2-hydroxy-1,1-dimethylethyl)-1-methyl-3-(1H-pyrrolo[2,3-b]pyridin-2-yl)-1H-indole-5-carboxamide
3EMG	SYK	685	2-[2-[(3,5-dimethylphenyl)amino]pyrimidin-4-yl]-N-[(1S)-2-hydroxy-1-methylethyl]-4-methyl-1,3-thiazole-5-carboxamide

[22] were downloaded. Of these, 6,510 tertiary structures were produced using Babel software [23]. Then, 1,043 compounds with a Tanimoto index greater than 0.25, with at least one of 10 ligands that have been reported to bind to one of 10 template proteins included in the PDB, were selected as drug compound candidates. Tanimoto indices were computed by Babel 2.2.3 with the default options [24]. The 10 template proteins' PDB IDs together with ligand IDs are listed in **Table 2**.

2.3 Execution of Ligand FAMS

Although many tertiary structures of SYK have been reported in the PDB (Table 2), it is impossible to align them with each other, because distinct ligands were bound to each of them, and thus their structures were not completely identical. Ligand FAMS [25] tended to map multiple ligands, whose binding structures to homologous proteins have been reported in the PDB, to one structure predicted by homology modeling. In mapping multiple ligands process CE fit program was used. In ligand FAMS, after structure alignments of multiple tertiary structures of SYK reported in PDB, main and sub chains are arranged in a way that multiple ligands can bind to one unique protein structure. The obtained tertiary protein structure of SYK, as well as ligands that bind to SYK, were used for chooseLD.

2.4 In silico Screening

In silico screening was performed using the template-based ligand docking simulation program, chooseLD [26]. In chooseLD, the ligand affinity to SYK was evaluated based on comparisons with 10 known ligand compounds (Table 2). If the ligands tested were well aligned with known ligand compounds, the ligand was given a high ranked score, i.e., FingerPrint Alignment Scores (FPAScores). Then, all tested compounds were ranked based on their attributed FPAScores. During this simulation, atom types were modified to achieve more accurate FPAScores. Three independent trials were performed for each compound, and the mean

Table 3 14 compounds from ChEMBL used to evaluate the chooseLD performance to estimate ligands' binding affinity to SYK, in the descending order of Ki or log₁₀Ki.

CHEMBL ID	Ki (nM)	log ₁₀ Ki
CHEMBL553	1584.89	3.20
CHEMBL340384	1584.89	3.20
CHEMBL196363	1258.93	3.10
CHEMBL211378	1258.93	3.10
CHEMBL422897	794.33	2.90
CHEMBL213505	501.19	2.70
CHEMBL262433	398.11	2.60
CHEMBL1421	199.53	2.30
CHEMBL49120	79.43	1.89
CHEMBL243088	63.1	1.80
CHEMBL7064	63.1	1.80
CHEMBL396523	31.62	1.50
CHEMBL244378	10	1.00
CHEMBL379975	6.31	0.80

FPAScores were used to rank the compounds.

2.5 Selection of Compounds in the Evaluation Set

To evaluate the performance of chooseLD in estimating the binding affinity to SYK, 14 compounds, whose absolute inhibition constant Ki values are listed in ChEMBL [27], [28], were downloaded (see Table 3).

The selection of the 14 compounds was as follows: Five-hundred-seventy-five binding experiments associated with the Ki values of compounds for the tyrosine-protein kinase SYK of *Homo sapiens* are included as ChEMBL2599 in the ChEMBL Database. In the database they are shown as “ChEMBL Bioactivity Search Results: 575.” The first column shows the ingredients for the compound. The fifth column shows the standard value for the Ki value. The standard unit of the Ki value is shown in the sixth column using the nM unit.

To check whether the FPAScore values negatively correlate with the logarithmic Ki value or not, a training set of 14 compounds was selected by eye from ChEMBL2599 in the ChEMBL Database so that the sampling interval of the logarithmic Ki values between the 14 compounds were taken to be as equal as possible (see Table 3). Although this intentional selection criteria may affect the quality of validation, it is unavoidable to some extent, because of the relatively small number of listed compounds with higher affinity (smaller Ki).

2.6 Validation of Evaluation Set Using SwissDock

To evaluate the performance of chooseLD, compounds in the evaluation set were also tested by SwissDock [29], [30]. Compound structures were computed by the canonical Simplified Molecular-Input Line-Entry System (SMILES) using open Babel [23], and were uploaded to SwissDock as the ligand structure. For target protein structures, model protein structures inferred by FAMS using 3VF8_A as a reference protein were uploaded to SwissDock. Minimum dG_s for each compound were used for evaluation.

3. Results

To perform *in silico* drug screening for SYK, the tertiary structure of the SYK protein must be determined. To infer the SYK tertiary structure, we used FAMS [17], [18]. SYK (uniprot ID

P43405.1) has 635 amino acids. Using 2OZO_A (tyrosine-protein kinase ZAP-70) as a reference protein, 625 amino acids of SYK were successfully modeled (*E*-value obtained by BLAST search was 1×10^{-170}), and the sequence similarity between 2OZO and SYK was 50%. Amino acids 363-635 of SYK were modeled using 3VF8_A (SYK) (*E*-value obtained by BLAST search was 1×10^{-95}). Comparison of the model structures based on 2OZO_A or 3VF8_A showed no significant difference within the commonly predicted regions of the protein. Because the ATP-binding region was included in both models, we used the model protein structure based on 3VF8_A for the *in silico* screening.

In addition, the binding ligand, OJE (for more details, see Table 2), has been described for 3VF8_A. To use a tertiary structure as a template for ligand docking, a reference protein must have a ligand that binds to it.

Following the procedure described in the Materials and Methods, we successfully obtained ranking for 1,043 compounds based on FPAScores (the 20 top-ranked compounds are listed in Table 4. The full list of the ranked compounds is available in additional file 1).

4. Discussion

4.1 Evaluation of Top-ranked Compounds

The 20 top-ranked compounds shown in Table 4 are reported as kinase inhibitors in either the DrugBank or ChEMBL databases. Table 4 also includes four SYK inhibitors (DB07194, ChEMBL512172 (cmpd_648) ranked 3rd; DB04739, ChEMBL56904 (cmpd_507) ranked 4th; DB06834, ChEMBL1229525 (cmpd_550) ranked 12th; DB07545, ChEMBL383899 (cmpd_744) ranked 17th), excluding the SYK inhibitor included in the template ligands, imatinib (see Table 2), ranked 2nd. As a result, five of the 20 top-ranked compounds were identified as SYK inhibitors. Thus, the remaining 15 compounds were also expected to be SYK inhibitors.

4.2 Comparisons with Known SYK Inhibitors' Binding Affinity

Although the top-ranked compounds were candidate SYK inhibitors, it is important to validate the FPAScores-based ranking using independent samples. For this purpose, we prepared a validation set of compounds (see Materials and Methods). Figure 1 shows comparisons between the FPAScores and Ki values. Because smaller Ki values indicate a larger binding affinity, the significant negative correlation observed shows that chooseLD correctly determined the binding affinity of compounds to SYK in the validation set (we regarded negative correlations associated with *P*-values less than 0.05 as significant); the FPAScores had correlation coefficients of -0.58 with Ki ($P = 0.0278$) and -0.58 with log₁₀ Ki ($P = 0.030$). Although one may wonder whether outliers affect the significance, it is unlikely that outliers affect the significance much, because the Spearman's rank correlation coefficient (Fig. 1 (b)), which is supposed to be robust to outliers [31], was still significant.

In addition, the largest FPAScores attributed to compounds with the smallest Ki in the validation set (Fig. 1, vertical axis) were at most 1,400 to 1,500. As seen in Table 4, there were at

Table 4 Twenty top-ranked compounds based on FPAScores.

Rank	FPAScore	“Drug Name (cmpd No.)”	“DrugBank No.; ChEMBL No.”	Target/Activity *
1	1583.5	“N-[(1S)-2-amino-1-phenylethyl]-5-(1H-pyrrolo[2,3-b]pyridin-4-yl)thiophene-2-carbox amide (cmpd.807)”	“DB07812; ChEMBL471034”	“—”;“RAC-beta serine/threonine-protein kinase, Glycogen synthase kinase-3 beta, Inhibition of AKT1”
2	1481	“Imatinib (cmpd.55)”	“DB00619; ChEMBL941”	“—”;“Inhibitor of BCR/ABL fusion protein isoform X9, Antagonist of Alpha and Beta platelet-derived growth factor receptor, Inhibitor of Proto-oncogene tyrosine-protein kinase ABL1, Inhibition of PTK, Abl, carbonic anhydrase, CSF1R, PDGFR α , LYN, LCK, FRK, 5HT2A, MAPK10, and BLK”
3	1468	“2-{2-[(3,5-dimethylphenyl)amino]pyrimidin-4-yl}-N-[(1S)-2-hydroxy-1-methylethyl]-4-methyl-1,3-thiazole-5-carboxamide (cmpd.648)”	“DB07194; ChEMBL512172”	“—”;“Tyrosine-protein kinase SYK , Inhibition of SYK , ZAP70, ROCK, SRC, and CDK2”
4	1456.3	“4-[(4-methyl-1-piperazinyl)methyl]-N-[3-[[4-(3-pyridinyl)-2-pyrimidinyl]amino]phenyl]-benzamide (cmpd.507)”	“DB04739; ChEMBL56904”	“—”;“Proto-oncogene tyrosine-protein kinase Src, Inhibition of SYK , v-Abl tyrosine kinase, c-Src-tyrosine kinase, and platelet-derived growth factor”
5	1436.4	“N-[4-Methyl-3-[[4-(3-pyridinyl)-2-pyrimidinyl]amino]phenyl]-3-pyridinecarboxamide (cmpd.413)”	“DB03878; —”	“—”;“Proto-oncogene tyrosine-protein kinase ABL, —”
6	1436	“4-[4-(1-amino-1-methylethyl)phenyl]-5-chloro-N-[4-(2-morpholin-4-ylethyl)phenyl]pyrimidin-2-amine (cmpd.242)”	“DB02491; ChEMBL233209”	“—”;“Fibroblast growth factor receptor 2, Inhibition of VEGFR2, CDK1, and HeLa, A375, HCT116 cells”
7	1432.1	“4-[4-(4-methyl-2-methylamino-thiazol-5-Yl)-pyrimidin-2-ylamino]-phenol (cmpd.472)”	“DB04407; ChEMBL47590”	“—”;“Cell division protein kinase 2, Inhibition of Cyclin-dependent kinase 2 (CDK2), CDK4, and Plk1”
8	1418.1	“N-(2-methoxyethyl)-4-((4-[2-methyl-1-(1-methylethyl)-1H-imidazol-5-YL]pyrimidin-2-YL)amino)benzenesulfonamide (cmpd.799)”	“DB07790; ChEMBL478409”	“—”;“Cell division protein kinase 2, Inhibition of Cyclin-dependent kinase 2 (CDK2), CyclinE, CDK4, MCF7 and LoVo cells”
9	1405.4	“4-(4-chlorobenzyl)-1-(7H-pyrrolo[2,3-d]pyrimidin-4-yl)piperidin-4-aminium (cmpd.908)”	“DB08150; —”	“cAMP-dependent protein kinase inhibitor alpha”;“ —”
10	1391.6	“4-(2,4-dimethyl-thiazol-5-Yl)-pyrimidin-2-Yl]-[4-(trifluoromethyl-phenyl)-amine (cmpd.294)”	“DB02915; ChEMBL48109”	“—”;“Cell division protein kinase 2 and Cyclin-A2, Inhibition of Cyclin-dependent kinase 2 (CDK2), CDK4, and A549, HT-29, SaOS-2 tumor cells”
11	1379.7	“4-[[4-imidazo[1,2-a]pyridin-3-Yl]pyrimidin-2-Yl]amino]benzenesulfonamide (cmpd.214)”	“DB02197; ChEMBL73303”	“—”;“Cell division protein kinase 2, Inhibition of Cyclin-dependent kinase 2 (CDK2), IGF1R, and MCF7 cells”
12	1352.2	“N-(2-hydroxy-1,1-dimethylethyl)-1-methyl-3-(1H-pyrrolo[2,3-b]pyridin-2-yl)-1H-indole-5-carboxamide (cmpd.550)”	“DB06834, ChEMBL1229525”	“—”;“Tyrosine-protein kinase SYK , Inhibition of SYK ”
13	1335.4	“3-[4-(2,4-dimethyl-thiazol-5-Yl)-pyrimidin-2-ylamino]-phenol (cmpd.486)”	“DB04518; ChEMBL47527”	“—”;“Cell division protein kinase 2, Inhibition of Cyclin-dependent kinase 2 (CDK2) and CDK4”
14	1334	“(2R)-1-[(5,6-diphenyl-7H-pyrrolo[2,3-D]pyrimidin-4-YL)amino]propan-2-ol (cmpd.770)”	“DB07647; ChEMBL371415”	“—”;“Serine/threonine-protein kinase Chk1, Inhibition of serine/threonine-protein kinase Chk1, cyclin-dependent kinase 1 (CDK 1), and protein kinase A (PKA)”
15	1312.7	“(2R)-3-[[4Z]-5,6-diphenyl-6,7-dihydro-4H-pyrrolo[2,3-D]pyrimidin-4-ylidene]amino]propane-1,2-diol (cmpd.771)”	“DB07648, ChEMBL372247”	“—”;“Serine/threonine-protein kinase Chk1, Inhibition of serine/threonine-protein kinase Chk1, cyclin-dependent kinase 1 (CDK 1), and protein kinase A (PKA)”
16	1311.1	“[4-(2-amino-4-methyl-thiazol-5-Yl)-pyrimidin-2-Yl]-[3-nitro-phenyl]-amine (cmpd.281)”	“DB02833; ChEMBL298445”	“—”;“Cell division protein kinase 2 and Cyclin-A2, Inhibition of Cyclin-dependent kinase 2 (CDK2), CDK9, CDK4, CDK7, CDK1, GSK3-beta, Aurora A/B, and Abl Kinase”
17	1301.4	“N-[3-[[4-[[3-(trifluoromethyl)phenyl]amino]pyrimidin-2-YL]amino]phenyl]cyclopropanecarboxamide (cmpd.744)”	“DB07545; ChEMBL383899”	“—”;“Serine/threonine-protein kinase 6, Inhibition of Aurora Kinase A, Lck, Bmx, IGF1R, SYK , and EGFR”
18	1287.7	“K-252a (cmpd.209)”	“DB02152; ChEMBL281948”	“—”;“Hepatocyte growth factor receptor and Dual specificity mitogen-activated protein kinase kinase 1, Inhibition of trka, VEGFR, protein kinase C, and myt1 kinase”
19	1279.9	“1-(dimethylamino)-3-(4-((4-(2-methylimidazo[1,2-A]pyridin-3-YL)pyrimidin-2-YL)amino)phenoxy)propan-2-ol (cmpd.828)”	“DB07889, ChEMBL102926”	“—”;“Cell division protein kinase 2, Inhibition of Cyclin-dependent kinase 1 (CDK1), CDK2, and CDK4”
20	1276.9	“2-{4-[4-((4-[2-methyl-1-(1-methylethyl)-1H-imidazol-5-yl]pyrimidin-2-yl)amino)phenyl]piperazin-1-yl}-2-oxoethanol (cmpd.854)”	“DB07982; ChEMBL477786”	“—”;“Cell division protein kinase 2, Inhibition of CDK2, MCF7 cells and ERG”

* “—” means that we cannot find a description for the corresponding compound in the database of ChEMBL or DrugBank.

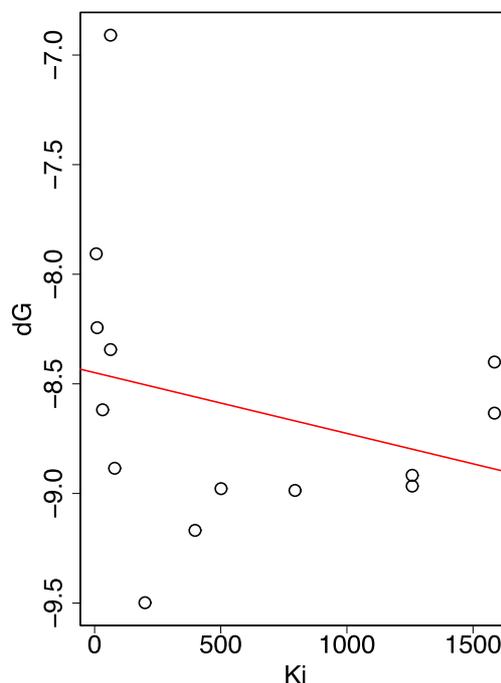
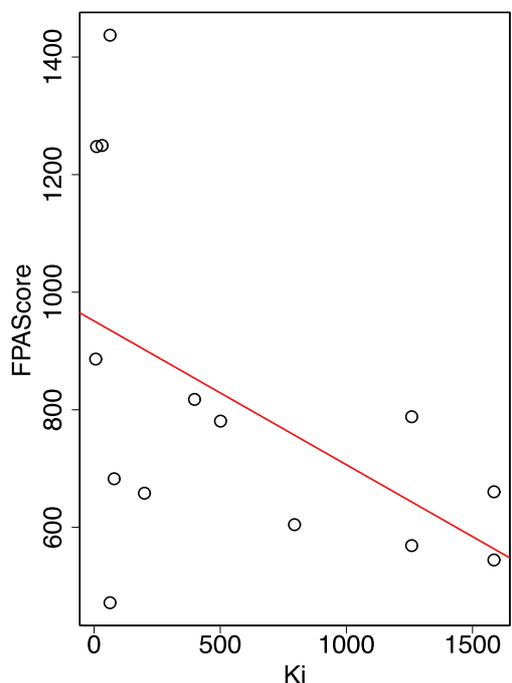
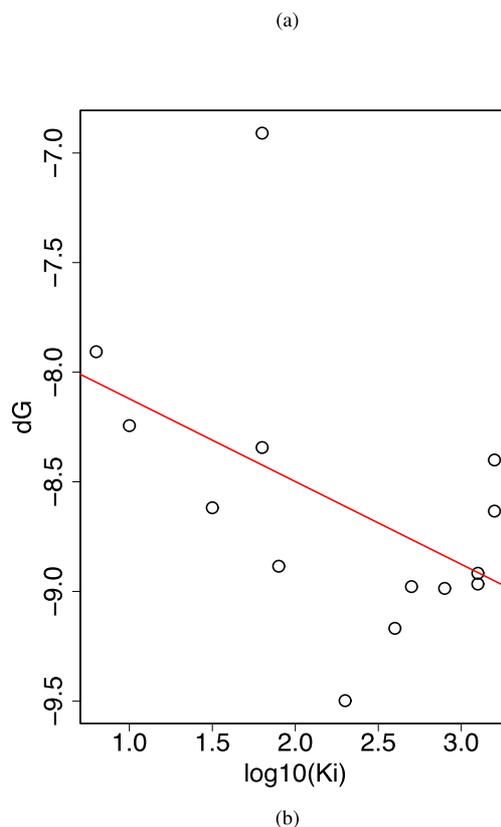
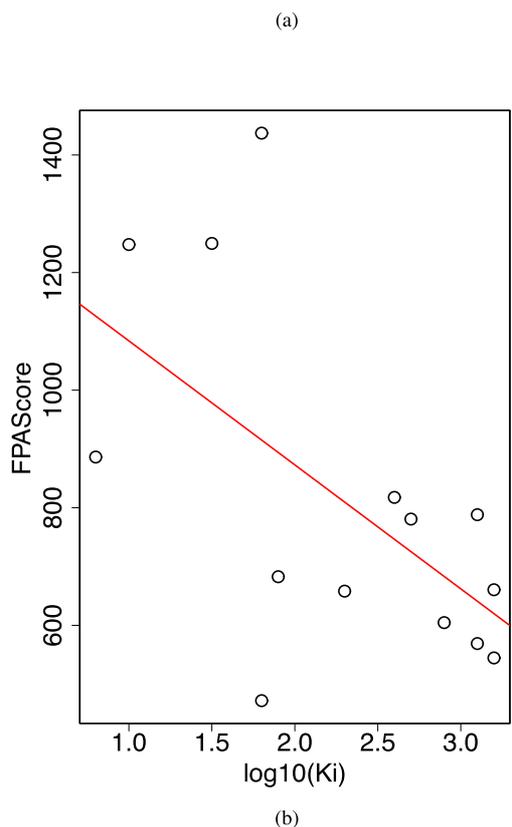


Fig. 1 Comparison between K_i and FPAScores (a) Vertical: FPAScores, horizontal: $\log_{10} K_i$. Pearson's correlation coefficient was -0.58 ($P = 0.0278$). (b) Vertical: FPAScores, horizontal: K_i . Spearman's rank correlation coefficient was -0.58 ($P = 0.030$). Solid lines indicate regression line.

Fig. 2 Comparison between K_i and dG (a) Vertical: dG , horizontal: $\log_{10} K_i$. Pearson's correlation coefficient was -0.48 ($P = 0.079$). (b) Vertical: dG , horizontal: K_i . Spearman's rank correlation coefficient was -0.49 ($P = 0.077$). Solid lines indicate regression line.

least 10 compounds with FPAScores in this range, suggesting that the top-ranked compounds listed in Table 4 are promising SYK inhibitors.

Finally, to verify and demonstrate the advantage of chooseLD,

SwissDock [29], [30] was used for validation (**Fig. 2**). Although the dG s (the amount of Gibbs free energy reduction due to ligand binding) negatively correlated with the K_i as expected, the correlation coefficients were not significant; the dG s had correlation coefficients of -0.48 with the K_i ($P = 0.079$) and -0.49 with

\log_{10} Ki ($P = 0.077$). This suggests that candidate compounds identified by chooseLD were superior compared with those identified by SwissDock.

Although we have successfully demonstrated the advantage of chooseLD over SwissDock, because the former needs (or can make use of) known binding ligands to homologous proteins while the latter does not (or cannot make use of), one should recognize that the advantage is more or less context dependent.

4.3 Conclusion

This study performed comprehensive *in silico* drug screening for SYK using chooseLD software. Top-ranked drug candidate compounds were kinase inhibitors that included several reported SYK inhibitors. The performance of chooseLD was evaluated using an independent evaluation set, and the FPAScores obtained by chooseLD significantly and negatively correlated with experimentally reported Ki values as expected. The significance of chooseLD was higher than that of SwissDock, another *in silico* screening software. Thus, the predicted drug candidate compounds are promising new SYK inhibitors for the treatment of several diseases.

Acknowledgments We would like to thank Dr. Katsuichiro Komatsu who helped with the *in silico* drug screening using chooseLD. This research was supported by Kakenhi, 23300357, 26120528 and a Chuo University Joint Research Grant.

References

[1] Taniguchi, T., Kobayashi, T., Kondo, J., Takahashi, K., Nakamura, H., Suzuki, J., Nagai, K., Yamada, T., Nakamura, S. and Yamamura, H.: Molecular cloning of a porcine gene *syk* that encodes a 72-kDa protein-tyrosine kinase showing high susceptibility to proteolysis, *J. Biol. Chem.*, Vol.266, No.24, pp.15790–15796 (1991).

[2] Geahlen, R.L.: Getting Syk: spleen tyrosine kinase as a therapeutic target, *Trends Pharmacol. Sci.*, Vol.35, No.8, pp.414–422 (2014).

[3] Nakazato, T., Bono, H., Matsuda, H. and Takagi, T.: Gendoo: Functional profiling of gene and disease features using MeSH vocabulary, *Nucleic Acids Res.*, Vol.37, No.Web Server issue, pp.W166–169 (2009).

[4] Nakazato, T. et al.: Gendoo, Database Center for Life Science (online), available from <http://gendoo.dbcls.jp/> (accessed 2015-02-19).

[5] Elliott, E.R., Van Ziffle, J.A., Scapini, P., Sullivan, B.M., Locksley, R.M. and Lowell, C.A.: Deletion of Syk in neutrophils prevents immune complex arthritis, *J. Immunol.*, Vol.187, No.8, pp.4319–4330 (2011).

[6] Cheng, S., Coffey, G., Zhang, X.H., Shakhovich, R., Song, Z., Lu, P., Pandey, A., Melnick, A.M., Sinha, U. and Wang, Y.L.: SYK inhibition and response prediction in diffuse large B-cell lymphoma, *Blood*, Vol.118, No.24, pp.6342–6352 (2011).

[7] Wang, S., Ding, Y.B., Chen, G.Y., Xia, J.G. and Wu, Z.Y.: Hypermethylation of Syk gene in promoter region associated with oncogenesis and metastasis of gastric carcinoma, *World J. Gastroenterol.*, Vol.10, No.12, pp.1815–1818 (2004).

[8] Coffey, G., DeGuzman, F., Inagaki, M., Pak, Y., Delaney, S.M., Ives, D., Betz, A., Jia, Z.J., Pandey, A., Baker, D., Hollenbach, S.J., Phillips, D.R. and Sinha, U.: Specific inhibition of spleen tyrosine kinase suppresses leukocyte immune function and inflammation in animal models of rheumatoid arthritis, *J. Pharmacol. Exp. Ther.*, Vol.340, No.2, pp.350–359 (2012).

[9] Ishida, S., Umeyama, H., Iwamoto, M. and Taguchi, Y.H.: Bioinformatic screening of autoimmune disease genes and protein structure prediction with FAMS for drug discovery, *Protein Pept. Lett.*, Vol.21, No.8, pp.828–839 (2014).

[10] Kaur, M., Singh, M. and Silakari, O.: Inhibitors of switch kinase 'spleen tyrosine kinase' in inflammation and immune-mediated disorders: a review, *Eur. J. Med. Chem.*, Vol.67, pp.434–446 (2013).

[11] Niimi, T., Orita, M., Okazawa-Igarashi, M., Sakashita, H., Kikuchi, K., Ball, E., Ichikawa, A., Yamagiwa, Y., Sakamoto, S., Tanaka, A., Tsukamoto, S., Fujita, S., Tatsuta, K., Maeda, Y. and Chikachi, K.: Design and synthesis of non-peptidic inhibitors for the Syk C-terminal

SH2 domain based on structure-based *in-silico* screening, *J. Med. Chem.*, Vol.44, No.26, pp.4737–4740 (2001).

[12] Kraskouskaya, D., Duodu, E., Arpin, C.C. and Gunning, P.T.: Progress towards the development of SH2 domain inhibitors, *Chem. Soc. Rev.*, Vol.42, No.8, pp.3337–3370 (2013).

[13] Villoutreix, B.O., Laconde, G., Lagorce, D., Martineau, P., Miteva, M.A. and Dariavach, P.: Tyrosine kinase *syk* non-enzymatic inhibitors and potential anti-allergic drug-like compounds discovered by virtual and *in vitro* screening, *PLoS ONE*, Vol.6, No.6, p.e21117 (2011).

[14] Li, B.K., Cong, Y., Yang, X.G., Xue, Y. and Chen, Y.Z.: *In silico* prediction of spleen tyrosine kinase inhibitors using machine learning approaches and an optimized molecular descriptor subset generated by recursive feature elimination method, *Comput. Biol. Med.*, Vol.43, No.4, pp.395–404 (2013).

[15] Kaur, M., Kumari, A., Bahia, M.S. and Silakari, O.: Designing of new multi-targeted inhibitors of spleen tyrosine kinase (Syk) and zeta-associated protein of 70kDa (ZAP-70) using hierarchical virtual screening protocol, *J. Mol. Graph. Model.*, Vol.39, pp.165–175 (2013).

[16] Xie, H.Z., Li, L.L., Ren, J.X., Zou, J., Yang, L., Wei, Y.Q. and Yang, S.Y.: Pharmacophore modeling study based on known spleen tyrosine kinase inhibitors together with virtual screening for identifying novel inhibitors, *Bioorg. Med. Chem. Lett.*, Vol.19, No.7, pp.1944–1949 (2009).

[17] Iwamoto, M., Umeyama, H. et al.: FAMS, Chuo University (online), available from (<http://fams.bio.chuo-u.ac.jp/fams/>) (accessed 2015-02-19).

[18] Umeyama, H. and Iwamoto, M.: FAMS and FAMSBASE for protein structure, *Curr Protoc Bioinformatics*, Vol.Chapter 5, p.Unit5.2 (2004).

[19] consortium, U.: uniprot, UniProt consortium (online), available from (<http://www.uniprot.org/>) (accessed 2015-02-19).

[20] Shindyalov, I.N. and Bourne, P.E.: Protein structure alignment by incremental combinatorial extension (CE) of the optimal path, *Protein Eng.*, Vol.11, No.9, pp.739–747 (1998).

[21] Law, V. et al.: DrugBank, University of Alberta (online), available from (<http://www.drugbank.ca/>) (accessed 2015-02-19).

[22] Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., Pon, A., Banco, K., Mak, C., Neveu, V., Djoumbou, Y., Eisner, R., Guo, A.C. and Wishart, D.S.: DrugBank 3.0: A comprehensive resource for 'omics' research on drugs, *Nucleic Acids Res.*, Vol.39, No.Database issue, pp.D1035–1041 (2011).

[23] O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T. and Hutchison, G.R.: Open Babel: An open chemical toolbox, *J. Cheminform*, Vol.3, p.33 (2011).

[24] O'Boyle et al.: Open Babel 2.2.3, Open babel (online), available from (http://openbabel.org/wiki/Open_Babel.2.2.3) (accessed 2015-03-28).

[25] Umeyama, H., Watanabe, Y., Iwamoto, M. and Shitaka, M.: Machinery and method that predict protein structures, program, and recording media (in Japanese), Patent: JP 4368594 B2 (2009.11.18).

[26] Takaya, D., Takeda-Shitaka, M., Terashi, G., Kanou, K., Iwamoto, M. and Umeyama, H.: Bioinformatics based Ligand-Docking and *in-silico* screening, *Chem. Pharm. Bull.*, Vol.56, No.5, pp.742–744 (2008).

[27] Gaulton, A., Bellis, L.J., Bento, A.P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B. and Overington, J.P.: ChEMBL: A large-scale bioactivity database for drug discovery, *Nucleic Acids Res.*, Vol.40, No.Database issue, pp.D1100–1107 (2012).

[28] Bento, A.P. et al.: ChEMBL, EMBL-EBI (online), available from (<https://www.ebi.ac.uk/chembl/>) (accessed 2015-02-19).

[29] Grosdidier, A., Zoete, V. and Michielin, O.: SwissDock, a protein-small molecule docking web service based on EADock DSS, *Nucleic Acids Res.*, Vol.39, No.Web Server issue, pp.W270–277 (2011).

[30] Grosdidier, A. et al.: SwissDock, Swiss Institute of Bioinformatics (online), available from (<http://www.swissdock.ch/>) (accessed 2015-02-19).

[31] Gideon, R.A. and Hollister, R.A.: A rank correlation coefficient resistant to outliers, *Journal of the American Statistical Association*, Vol.82, No.398, pp.656–666 (1987).

Appendix

Additional Files

Additional file 1 — Full list of FPAScores of drug candidate compounds

Mean FPAScores of three independent trials for 1,043 drug

candidate compounds taken from the DrugBank. <http://dx.doi.org/10.6084/m9.figshare.1312839>



Hideaki Umeyama was born in 1944. He has received his B.S. from the Faculty of Pharmaceutical Sciences, The University of Tokyo, M.S. and Ph.D. from the Graduate School of Pharmaceutical Sciences, The University of Tokyo in 1968, 1970, and 1973, respectively. He has moved to Kitasato University as an assistant professor. Then he became a full professor there in 1988. He has become an honorary professor of Kitasato University in 2010. And he has moved to Chuo University as a joint research member. His research interests include bioinformatics, molecular mechanics and quantum chemistry.

Then he became a full professor there in 1988. He has become an honorary professor of Kitasato University in 2010. And he has moved to Chuo University as a joint research member. His research interests include bioinformatics, molecular mechanics and quantum chemistry.



Mitsuo Iwadate was born in 1973. He has received his B.S. in Biotechnology, M.S. and Dr.Sci. in Biotechnology from Tokyo University of Agriculture and Technology, Japan in 1995, 1997, and 2000. From 2000 to 2008, he was an assistant professor in Kitasato University. Then he has moved to Chuo University as an associate professor in 2008. His research interests include bioinformatics and structural biology.

Then he has moved to Chuo University as an associate professor in 2008. His research interests include bioinformatics and structural biology.



Y-h. Taguchi was born in 1961. He has received his B.S. in Applied Physics, M.S. and Dr.Sci. in Physics from Tokyo Institute of Technology, Japan in 1984, 1986, and 1988, respectively. From 1988 to 1997, he was an assistant professor in Tokyo Institute of Technology. Then he has moved to Chuo University as an associate professor in 1997. He has been a full professor there from 2006 to the present. His research interests include bioinformatics and nonlinear physics.

Then he has moved to Chuo University as an associate professor in 1997. He has been a full professor there from 2006 to the present. His research interests include bioinformatics and nonlinear physics.

(Communicated by *Masakazu Sekijima*)