0022-3565/09/3301-304–315$20.00
THE JOURNAL OF PHARMACOLOGY AND EXPERIMENTAL THERAPEUTICS
Vol. 330, No. 1
149955/3482255
JPET 330:304–315, 2009
Printed in U.S.A.

What Are Next Generation Innovative Therapeutic Targets? Clues from Genetic, Structural, Physicochemical, and Systems Profiles of Successful Targets[S]

Feng Zhu, LianYi Han, ChanJuan Zheng, Bin Xie, Martti T. Tammi, ShengYong Yang, YuQuan Wei, and YuZong Chen

Bioinformatics and Drug Design Group, Center for Computational Science and Engineering, Departments of Pharmacy (F.Z., L.Y.H., C.J.Z., B.X., Y.Z.C.) and Biological Science (M.T.T.), National University of Singapore, Singapore; and State Key Laboratory of Biotherapy, Sichuan University, Chengdu, People's Republic of China (S.Y.Y., Y.Q.W., Y.Z.C.)

Received December 16, 2008; accepted April 7, 2009

ABSTRACT

Low target discovery rate has been linked to inadequate consideration of multiple factors that collectively contribute to druggability. These factors include sequence, structural, physicochemical, and systems profiles. Methods individually exploring each of these profiles for target identification have been developed, but they have not been collectively used. We evaluated the collective capability of these methods in identifying promising targets from 1019 research targets based on the multiple profiles of up to 348 successful targets. The collective method combining at least three profiles identified 50, 25, 10, and 4% of the 30, 84, 41, and 864 phase III, II, I, and nonclinical trial targets as promising, including eight to nine targets of positive phase III results. This method dropped 89% of the 19 discontinued clinical trial targets and 97% of the 65 targets failed in high-throughput screening or knockout studies. Collective consideration of multiple profiles demonstrated promising potential in identifying innovative targets.

The majority of clinical drugs achieve their therapeutic effects by binding and modulating the activity of protein targets (Ohlstein et al., 2000; Zambrowicz and Sands, 2003). Intensive efforts in target search (Chiesi et al., 2001; Matter, 2001; Walke et al., 2001; Ilag et al., 2002; Zheng et al., 2006b) have led to the discovery of >1000 research targets (targeted by investigational agents only) (Zheng et al., 2006b). These targets have been derived from analysis of disease relevance, functional roles, expression profiles, and loss-of-function genetics between normal and disease states (Ryan and Patterson, 2002; Nicolette and Miller, 2003; Kramer and Cohen, 2004; Austen and Dohrmann, 2005; Jackson and Harrington, 2005; Lindsay, 2005; Sams-Dodd, 2005). Many of them have been targeted by target-selective leads (Simmons, 2006; Zheng et al., 2006b). Despite heavy spending and exploration of new technologies (Booth and Zemmel, 2004), fewer innovative targets have emerged (Lindsay, 2005), and it typically takes ~8 to 20 years to derive a marketed drug against these innovative targets (Zheng et al., 2006a). Innovative targets refer to the targets with no other subtype of the same protein successfully explored before.

Low productivity of innovative targets (Lindsay, 2005) has been attributed to problems in target selection and validation (Smith, 2003; Lindsay, 2005; Sams-Dodd, 2005). A particular problem is inadequate physiological and clinical investigations (Rosenberg, 1999; Lindsay, 2005; Sams-Dodd, 2005). Drug effects are due to interactions with various sites of human physiological systems and pathways as well as its intended target, which collectively determine the success of target exploration (Zheng et al., 2006a,b). Current efforts have been focused on target-selective agents minimally interacting with other human members of the target family (Drews, 1997; Ohlstein et al., 2000). However, their possible interactions with other human proteins, pathways, and tissues have not been fully considered, leading to frequent failures in subsequent developmental stages. Therefore, a target cannot be fully validated by considering disease relevance and target selectivity alone (Lindsay, 2005; Sams-Dodd, 2005).

Integrated target and physiology-based approaches have been proposed for target identification and validation (Lindsay, 2005; Sams-Dodd, 2005). Different in silico approaches have been explored for target prediction based on sequence similarity (Hopkins and Groom, 2002; Zheng et al., 2006b), structural similarity and binding-site geometric and energetic features (Hajduk et al., 2005), target physicochemica-

This work was supported by the Academic Research Funds Singapore [Grant R-148-000-081-112]; the National Natural Science Foundation of China [Grant 30772651]; and the Ministry of Science and Technology, 863 Hi-Tech Program China [Grant 2006AA020400].
Article, publication date, and citation information can be found at http://jpet.aspetjournals.org.
doi:10.1124/jpet.108.149955.
[S] The online version of this article (available at http://jpet.aspetjournals.org) contains supplemental material.

land other characteristics detected by machine learning (Zheng et al., 2006b; Han et al., 2007; Xu et al., 2007), and systems profiles (similarity to human proteins, pathway and tissue distribution) (Zheng et al., 2006a,b; Sakharkar et al., 2008; Yao and Rzhetsky, 2008). We evaluated whether target prediction can be improved by combinations of these approaches, which were tested against 155 clinical trial targets (data are collected from CenterWatch Drugs in Clinical Trials Database 2008, http://www.centerwatch.com/professional/cwpipeline/), 864 nonclinical trial research targets (Chen et al., 2002), 19 difficult targets currently discontinued in clinical trials (with clinical trial drug discontinued and no new drug entered clinical trial at the moment) (data collected from CenterWatch Drugs in Clinical Trials Database), and 65 nonpromising targets failed in large-scale HTS campaigns (Payne et al., 2007) or found nonviable in knockout studies (Mdluli and Spigelman, 2006).

Materials and Methods

Sequence Similarity Analysis between Drug-Binding Domain of Studied Target and That of Successful Target. BLAST (Altschul et al., 1997) was applied to determine the level of similarity between the sequence of the drug-binding domain of each studied

ABBREVIATIONS: HTS, high-throughput screening; BLAST, Basic Local Alignment Search Tool; SVM, support vector machine(s); NK, neurokinin; MMP, matrix metalloproteinase; PI-88, phosphomannopentaose sulfate; AMD-3100, 1,1′-[1,4-phenylenebis(methylene)]bis [1,4,8,11-tetraazacyclotetradecane] octohydrobromide dihydrate; PXD101, belinostat; SNS-032, *N*-(5-(((5-(1,1-dimethylethyl)-2-oxazolyl)methyl)thio)-2-thiazolyl)-4-piperidinecarboxamide; UCN-01, 7-hydroxystaurosporine; HMN-214, (*E*)-4-(2-(2-(*N*-acetyl-*N*-(4-methoxybenzenesulfonyl)amino)stilbazole)) 1-oxide; AT7519, 4-(2,6-dichlorobenzoylamino)-1*H*-pyrazole-3-carboxylic acid piperidin-4-ylamide; SNS-032, *N*-(5-(((5-(1,1-dimethylethyl)-2-oxazolyl)methyl)thio)-2-thiazolyl)-4-piperidinecarboxamide; TAK-475, 1-((1-(3-acetoxy-2,2-dimethylpropyl)-7-chloro-5-(2,3-dimethoxyphenyl)-2-oxo-1,2,3,5-tetrahydro-4,1-benzoxazepin-3-yl)acetyl)piperidine-4-acetic acid; R115777, tipifarnib; IPI-504, 17-(allylamino)-17-demethoxygeldanamycin; LY335979, zosuquidar trihydrochloride; CGP71683A, *N*-[[4-[[(4-aminoquinazolin-2-yl)amino]methyl]cyclohexyl]methyl]naphthalene-1-sulfonamide; ABT-239, 4-(2-{2-[(2*R*)-2-methylpyrrolidinyl]ethyl}-benzofuran-5-yl)benzonitrile; LY293111, 2-[2-propyl-3-[3-[2-ethyl-4-(4-fluorophenyl)-5-hydroxyphenoxy]-propoxy]phenoxyl]benzoic acid; LY2140023, (1*R*,4*S*,5*S*,6*S*)-2-thiabicyclo[3.1.0]hexane-4,6-dicarboxylic acid, 4-[(2*S*)-2-amino-4-(methylthio)-1-oxobutyl]amino-, 2,2-dioxide monohydrate; LY354740, (2*S*,4*S*)-2-amino-4-(4,4-diphenylbut-1-yl)-pentane-1,5-dioic acid; NSCLC, non–small-cell lung carcinoma; BMS-275291, (*S*)-*N*-[2-mercapto-1-oxo-4-(3,4,4-trimethyl-2,5-dioxo-1-imidazolidinyl)butyl]-L-leucyl-*N*,3-dimethyl-L-valinamide; SCH-530348, (9-{2-[5-(3-fluorophenyl)-pyridin-2-yl]-vinyl}-1-methyl-3-oxo-dodecahydro-naphtho[2,3-*c*]furan-6-yl)-carbamic acid ethyl ester; AMD-070, *N*1-(1*H*-benzoimidazol-2-ylmethyl)-*N*1-(5,6,7,8-tetrahydro-quinolin-8-yl)-butane-1,4-diamine; DX-88, ecallantide; CI-1033, *N*-[4-(3-chloro-4-fluoro-phenylamino)-7-(3-morpholin-4-yl-propoxy)-quinazolin-6-yl]-acrylamide; XL999, 5-(1-ethyl-piperidin-4-ylamino)-3-[(3-fluorophenyl)-(4-methyl-1*H*-imidazol-2-yl)-methylene]-1,3-dihydro-indol-2-one; CHIR-258, 4-amino-5-fluoro-3-[5-(4-methylpiperazin-1-yl)-1*H*-benzimidazol-2-yl]quinolin-2(1*H*)-one; MS-275, *N*-(2-aminophenyl)-4-[*N*-(pyridin-3-yl-methoxycarbonyl)aminomethyl]benzamide; KD3010, 4-[2,6-dimethyl-4-(4-trifluoromethoxy-phenyl)-piperazine-1-sulfonyl]-indan-2-carboxylic acid; RX-0201, 5′-GCTGCATGATCTCCTTGGCG-3′; DG031, 2-(4-(quinolin-2-yl-methoxy)-phenyl)-2-cyclopentylacetic acid; CNF1010, carbamic acid 19-allylamino-13-hydroxy-8,14-dimethoxy-4,10,12,16-tetramethyl-3,20,22-trioxo-2-aza-bicyclo[16.3.1]docosa-1(21),4,6,10,18-pentaen-9-yl ester; SNX-5422, amino-acetic acid 4-[2-carbamoyl-5-(6,6-dimethyl-4-oxo-3-trifluoromethyl-4,5,6,7-tetrahydro-indazol-1-yl)-phenylamino]-cyclohexyl ester; PG-530742, 2-[4-(4-methoxy-benzoylamino)-benzenesulfonylamino]-6-morpholin-4-yl-hex-4-ynoic acid; GD0039, octahydro-indolizine-1,2,8-triol; BB-3644, *N*1-[2,2-dimethyl-1-(pyridin-2-ylcarbamoyl)-propyl]-*N*4-hydroxy-2-isobutyl-3-methoxy-succinamide; AZD 7545, (2*R*)-*N*-{4-[4-(dimethylcarbamoyl)phenylsulfonyl]-2-chlorophenyl}-3,3,3-trifluoro-2-hydroxy-2-methylpropanamide; CAP-232, (1*R*,4*S*,7*R*,10*S*,13*R*)-4-(4-aminobutyl)-*N*-[(2*S*,3*R*)-1-amino-3-hydroxy-1-oxobutan-2-yl]-13-[[(2*R*)-2-amino-3-phenylpropanoyl]amino]-10-[(4-hydroxyphenyl)methyl]-7-(1*H*-indol-3-ylmethyl)-3,6,9,12-tetraoxo-15,16-dithia-2,5,8,11-tetrazacycloheptadecane-1-carboxamide; C1-INH,MASRLTLLTLLLLLLAGDRASSNPNATSSS-SQDPESLQDRGEGKVATTVISKMLFVEPILEVSSLPTTNSTTNSATKITANTTDEPTTQPTTEPTTQPTIQPTQPTTQLPTDSPTQPTTGSFCPGPVTLCSDLESHSTEAVLGDALVDFSLKLYHAFSAMKKVETNMAFSPFSIASLLTQVLLGAGENTKTNLESILSYPKDFTCVHQALKGFTTKGVTSVSQIFHSPDLAIRDTFVNASRTLYSSSPRVLSNNSDANLELINTWVAKNTNNKISRLLDSLPSDTRLVLLNAIYLSAKWKTTFDPKKTRMEPFHFKNSVIKVPMMNSKKYPVAHFIDQTLKAKVGQLQLSHNLSLVILVPQNLKHRLEDMEQALSPSVFKAIMEKLEMSKFQPTLLTLPRIKVTTSQDMLSIMEKLEFFDFSYDLNLCGLTEDPDLQVSAMQHQTVLELTETGVEAAAASAISVARTLLVFEVQQPFLFVLWDQQHKFPVFMGRVYDPRA; INCB3284, 1-hydroxy-4-[3-isopropyl-3-(3-trifluoromethyl-7,8-dihydro-5*H*-[1,6]naphthyridine-6-carbonyl)-cyclopentylamino]-cyclohexanecarbonitrile; TM30339, APLEPVYPGDNATPEQMAQYAADLRRYINMLTRPRY; KAI-9803, H2N-Cys-Ser-Phe-Asn-Ser-Tyr-Glu-Leu-Gly-Ser-Leu-COOH; XL647, (3,4-dichloro-phenyl)-{6-methoxy-7-[5-(4-trifluoromethyl-phenyl)-[1,2,4]oxadiazol-3-ylmethoxy]-quinazolin-4-yl}-amine; KOS-2187, 7,10,12,13-tetrahydroxy-6-[3-hydroxy-4-(isopropyl-methyl-amino)-6-methyl-tetrahydro-pyran-2-yloxy]-4-(5-hydroxy-4-methoxy-4,6-dimethyl-tetrahydro-pyran-2-yloxy)-3,5,7,9,11,13-hexamethyl-14-phenyl-oxacyclotetradecan-2-one; CPG 52364, *N*′-[6,7-dimethoxy-2-(4-phenyl-piperazin-1-yl)-quinazolin-4-yl]-*N*,*N*-dimethyl-ethane-1,2-diamine; REG1, a two-component system consisting of a single-stranded nucleic acid aptamer RB006 3′-idT-UACCCCUCCGUCCUAAUGCGCCAUAUCAGGGGUA-Ch-5′ and a complementary antidote nucleic acid RB007 3′-uaccccugauauggcgc-5′; MBX-8025, formerly RWJ-800025, {2-methyl-4-[5-methyl-2-(4-trifluoromethyl-phenyl)-2*H*-[1,2,3]triazol-4-ylmethylsulfanyl]-phenoxy}-acetic acid; XL844, 1-[2-(3-amino-propoxy)-phenyl]-3-pyrazin-2-yl-urea; XL880, cyclopropane-1,1-dicarboxylic acid {3-fluoro-4-[6-methoxy-7-(3-morpholin-4-yl-propoxy)-quinolin-4-yloxy]-phenyl}-amide (4-fluoro-phenyl)-amide; AM803, [3-hydroxy-2-methylsulfanylmethyl-5-(pyridin-2-ylmethoxy)-pyrrolo[2,3-*b*]pyridin-1-yl]-acetaldehyde; AM103, 2-[2-(2-oxo-propyl)-5-(quinolin-2-ylmethoxy)-pyrrolo[2,3-*b*]pyridin-1-yl]-acetamide; 659032, 2-[[(2,3-difluorophenyl)methyl]thio]-*N*-[1-(2-methoxyethyl)-4-piperidinyl]-4-oxo-*N*-[[4′-(trifluoromethyl)[1,1′-biphenyl]-4-yl]methyl]-1(4*H*)-quinolineacetamide; AE-941, an analog of squalamine 3β-*N*-1-[*N*-[3-(4-aminobutyl)]-1,3-diaminopropane]-7-α-cholestane 24-sulfate; PSN357, 5-chloro-1*H*-pyrrolo[2,3-*c*]pyridine-2-carboxylic acid [2-[4-(2-dimethylamino-ethyl)-piperazin-1-yl]-1-(4-fluoro-benzyl)-2-oxo-ethyl]-amide; RC-8800, 5-(2-{1-[3-(3,4-dichloro-benzenesulfonyl)-1-*methyl*-propyl]-7α-methyl-octahydro-inden-4-ylidene}-ethylidene)-4-methylene-cyclohexane-1,3-diol; MLN222, CEEPPTFEAMELIGKPKPYYEIGERVDYKCKKGYFYIPPLATHTICDRNHTWLPVSDDACYRETCPYIRDPLNGQAVPANGTYEFGYQMHFICNEGYYLIGEEILYCELKGSVAIWSGKPPICEKVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYCGDNSVWSRAAPECKVVKCRFPVVENGKQISGFGKKFYYKATVMFMTVARPSVPAALPLLGELPRLLLLVLLCLPAVWGDCGLPPDVPNAQPALEGRTSFPEDTVITYKCEESFVKIPGEKDSVICLKGSQWSDIEEFCNRSCEVPTRLNSASLKQPYITQNYFPVGTVVEYECRPGY-RREPSLSPKLTCLQNLKWSTAVEFCKKKSCPNPGEIRNGQIDVPGGILFGATISFSCNTGYKLFGSTSSFCLISGSSVQWSDPLPECREIYCPAPPQIDN-GIIQGERDHYGYRQSVTYACNKGFTMIGEHSIYCTVNNDEGEWSGPPPECRGKSLTSKVPPTVQKPTTVNVPTTEVSPTSQKTTTKTTTP; XL418, 3-bromo-4-{4-[5-chloro-2-methyl-3-(3-pyrrolidin-1-yl-propyl)-phenyl]-piperazin-1-yl}-1*H*-pyrazolo[3,4-*d*]pyrimidine.

TABLE 1

List of phase III targets identified by combinations of at least three of the methods A, B, C, and D used in this study

Targets marked by an asterisk (*) are innovative targets without a protein subtype as a successful target. Tissue distribution "P" represents cases where the target is distributed in more than five tissues, but the disease-relevant targets are located within blood vessels or cells lining the arteries where they have higher priority to bind drugs.

Target	Predicted as Promising by Combination of Methods	No. of Target-Affiliated Pathways	No. of Human Similarity Proteins outside Target Family	No. of Tissues Target Is Primarily Distributed	Targeted Disease Conditions	Target Exploration Status (Tested Drug)	Positive Results in Phase III Trial Reported in Company Website (Year of Report)
Cholecystokinin receptor type A*	Combination of A, B, C, D	2	1	1	Irritable bowel syndrome	Phase III (dexloxiglumide)	Favorable topline results in patients with constipation—predominant irritable bowel syndrome (2007)
Coagulation factor IIa*	Combination of A, B, C, D	3	0	5	Venous thromboembolism	Phase II/III (SR-123781A)	
Neurotrophic tyrosine kinase receptor 1*	Combination of A, B, C, D	3	6	2	Acute myeloid leukemia	Phase II/III (lestaurtinib)	
5-Hydroxytryptamine 3 receptor	Combination of A, C, D	1	0	2	Irritable bowel syndrome	Phase III (cilansetron)	Positive data for treating irritable bowel syndrome with diarrhea predominance (2004)
Heparanase*	Combination of A, C, D	2	0	2	Hepatocellular cancer	Phase III (PI-88)	
Multidrug resistance protein 3	Combination of A, C, D	1	0	3	Acute myeloid leukemia	Phase III (LY335979)	
Orexin-OX1/OX2 receptor*	Combination of A, C, D	1	0	2	Sleep disorders	Phase III (almorexant)	
Somatostatin receptor 1	Combination of A, C, D	1	0	5	Cushing's disease, renal cell carcinoma	Phase III (pasireotide), phase II (CAP-232)	
NK-2 receptor*	Combination of A, C, D	2	0	3	Depression	Phase III (saredutant)	Overall statistically significant efficacy versus placebo, well tolerated (2007)
B2 bradykinin receptor*	Combination of A, B, C	4	0	P	Hereditary angioedema, traumatic brain injuries	Phase III (icatibant), phase II (anatibant)	Positive results for the treatment of hereditary angioedema (2006)
Thrombin receptor*	Combination of A, B, C	4	0	5	Cardiovascular disorders	Phase III (SCH-530348)	
C-X-C chemokine receptor 4	Combination of A, B, D	3	2	P	Non-Hodgkin's lymphoma, late-stage solid tumors	Phase III (AMD-3100), phase I/II (AMD-070), phase I (MSX-122)[a]	Positive results for the treatment of multiple myeloma (2007)
C1 esterase*	Combination of A, B, D	1	3	P	Hereditary angioedema	Phase III (C1-INH)	Positive results for treating hereditary angioedema, significantly decreases the number of attacks in patients (2007)
Sphingosine 1-phosphate receptor 1 (S1PR1)*	Combination of A, B, D	1	0	5	Multiple sclerosis	Phase III (FTY720)	Initial Phase III results show that FTY720 (fingolimod) has superior efficacy to a current standard of care for patients with relapsing-remitting multiple sclerosis (MS) (2008)
Neuropeptide Y receptor 5	Combination of A, B, D	1	0	2	Obesity	Phase III (CGP71683A)	
Plasma kallikrein*	Combination of A, B, D	1	0	5	Hereditary angioedema	Phase III (DX-88)	Positive topline results for treating hereditary angioedema, well tolerated (2007)

[a] Wong and Korz (2008).

TABLE 2

List of phase II and phase I targets identified by combinations of at least three of the methods A, B, C, and D used in this study

Targets marked by an asterisk (*) are innovative targets without a protein subtype as a successful target. Tissue distribution P represents cases where target is distributed in more than five tissues, but the disease-relevant targets are located within blood vessels or cells lining the arteries where they have higher priority to bind drugs.

Research Target	Identified by Combination	No. of Target-Affiliated Pathways	No. of Human Similarity Proteins outside Target Family	No. of Tissues Target Is Primarily Distributed	Targeted Disease Conditions	Target Exploration Stage (Testing Drug)
C-C chemokine receptor 2*	Combination of A, B, C, D	1	0	1	Rheumatoid arthritis, multiple sclerosis	Phase II (INCB3284), phase I (CCX915)
ErbB-4	Combination of A, B, C, D	3	4	2	Breast cancer	Phase II (CI-1033)
Fibroblast growth factor receptor-3	Combination of A, B, C, D	3	0	4	Solid tumors, multiple myeloma	Phase II (XL999), phase I (CHIR-258)
Guanylate cyclase B*	Combination of A, B, C, D	3	0	1	Heart disease	Phase IIa (chimeric natriuretic peptide), preclinical (guanilib)
Histone deacetylase 4	Combination of A, B, C, D	1	1	P	Basal cell carcinoma, melanoma, cancer	Phase II (avugane, romidepsin, MS-275, PXD101)
Neuropeptide Y receptor 2	Combination of A, B, C, D	1	0	4	Obesity	Phase II (obinepitide)
Neuropeptide Y receptor 4	Combination of A, B, C, D	1	0	3	Schizophrenia, schizoaffective disorders	Phase I/II (TM30339)
Toll-like receptor 3	Combination of A, B, C, D	1	0	2	Human papillomavirus infections	Phase II (HspE7)
Fibroblast growth factor receptor-1	Combination of A, B, C	5	0	>10	Coronary heart disease, solid tumors	Phase II (XL999), phase II (fibroblast growth factor-1)
Protein kinase C-γ	Combination of A, B, C	16	0	4	Acute myocardial infarction	Phase II (midostaurin), phase I/II (KAI-9803)
Tyrosine-protein kinase receptor HTK*	Combination of A, B, C	1	4	>10	Lung cancer, solid tumors	Phase II (XL647)
Histamine H3 receptor	Combination of A, C, D	1	0	4	Attention-deficit hyperactivity disorder, Alzheimer's disease, schizophrenia	Phase II (cipralisant), phase I (ABT-239)
Leukotriene B4 receptor 1*	Combination of A, C, D	1	0	4	Cancer, renal cell carcinoma	Phase II (LY293111), phase I (Biomed 101)
Motilin receptor*	Combination of A, C, D	1	0	1	Irritable bowel syndrome, gastrointestinal disorders	Phase IIb (mitemcinal), phase I (KOS-2187)
NK-3 receptor*	Combination of A, C, D	2	0	1	Schizophrenia, schizoaffective disorders	Phase IIb (osanetant), phase II (talnetant)
Somatostatin receptor type 4	Combination of A, C, D	1	2	3	Solid tumors	Phase II (CAP-232)
Tissue kallikrein-2*	Combination of A, C, D	1	0	2	Atopic dermatitis	Phase II (dermolastin)
Toll-like receptor 8	Combination of A, C, D	1	0	5	Genital warts, systemic lupus erythematosus	Phase II (resiquimod), phase I (CPG 52364)
Cell division protein kinase 7	Combination of A, B, D	1	1	P	B-cell malignancies	Phase I (SNS-032)
Coagulation factor IX*	Combination of A, B, D	1	5	1	Thrombosis, venous thromboembolism	Phase IIa (REG1), phase I completed (TTP889)[a]
Melanocortin receptor*	Combination of A, B, D	1	0	3	Sexual (female) and erectile dysfunction	Phase II b (bremelanotide)
Metabotropic glutamate receptor 2/3*	Combination of A, B, D	1	0	1	Psychosis	Phase II (LY2140023, LY354740)
Peroxisome proliferator-activated receptor δ	Combination of A, B, D	3	0	P	Obesity	Phase II (MBX-8025), phase I (KD3010)
Serine/threonine-protein kinase Chk2	Combination of A, B, D	2	0	4	Solid tumors	Phase I (XL844), phase I (UCN-01)
Serine/threonine-protein kinase PLK*	Combination of A, B, D	1	1	P	Pancreatic, prostate and a number of other cancers	Phase I (HMN-214)

[a] Howard et al., 2007; Eriksson et al., 2008; Tomillero and Moral, 2008.

research target and the sequence of the drug-binding domain of each of the 168 successful targets with identifiable drug-binding domains. The BLAST program was downloaded from National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/download.shtml). A stricter BLAST cut-off, E-value = 0.001, was used for selecting the research targets similar to a successful target, i.e., the E-value of the drug-binding domains is ≤0.001. The details of the analysis are described in Supplemental Data 1.

Structural Comparison between Drug-Binding Domain of Studied Target and That of Successful Target. The ligand-binding or catalytic sites are the most relevant subsets of a domain, which are normally located within the so-called ligand sensing core where actual catalytic conversion of enzyme substrates, or the binding event of small-molecule ligands, occurs. It has been suggested that structural similarity considerations should be confined to ligand-sensing cores, instead of whole domains, according to three-dimensional similarities with respect to so-called protein structure similarity clusters (Koch and Waldmann, 2005). In this study, ligand sensing or catalytic cores of drug-binding domain of the studied research target were clustered against those of 129 successful targets with available three-dimensional structure based on visual inspection and structural superimposition and alignment tools in SYBYL (SYBYL 6.7; Tripos, St. Louis, MO) and Insight II (Accelrys, San Diego, CA) following the same procedure used for generating SCOP structural folds (Murzin et al., 1995). The details of this analysis are described in the Supplemental Data 2.

Target Classification Based on Characteristics of Successful Targets Detected by a Machine Learning Method. Promising targets can be separated from other proteins based on the structural and physicochemical characteristics of successful targets detected by a machine learning method. By using sequence-derived structural and physicochemical descriptors of the successful targets and those of other proteins, a machine learning algorithm attempts to separate successful targets from other proteins by searching for a projection function that maps the descriptors of successful targets and those of other proteins into separate regions in a high-dimensional feature space, and these regions are separated by easily defined borders. A research target is classified as promising if it is located in the region of successful targets, which is not necessarily similar in sequence to any successful target because the mapping to the feature space is typically nonlinear and the proteins are characterized by structural and physicochemical descriptors rather than sequence.

The machine learning method used in this work is support vector machines (SVM), which is a supervised learning method used for classification of objects (e.g., proteins) into two classes (e.g., promising targets and other proteins) and has been applied to target prediction (Zheng et al., 2006b). The details of SVM algorithm and computational procedures can be found in Supplemental Data 3. In this work, a nonlinear SVM was used with the following kernel function:

$$K(\mathbf{x}_i,\mathbf{x}_j) = e^{-\|\mathbf{x}_j-\mathbf{x}_i\|^2/2\sigma^2} \quad (1)$$

The nonlinear SVM projects feature vectors into a high-dimensional feature space using the kernel function defined above. The linear SVM was then applied to produce a single hyperplane that separates targets from nontargets. A SVM prediction system was developed by using the feature vectors of the structural and physicochemical properties of 348 successful targets and 24,066 putative nontargets generated by a procedure described in our previous study (Han et al., 2007), which was used to screen the 1019 research targets for identifying potential promising targets. The sequence-derived structural and physicochemical descriptors used in SVM include amino acid composition; dipeptide composition; sequence autocorrelation descriptors; sequence coupling descriptors; and the descriptors for the composition, transition, and distribution of hydrophobicity, polarity, polarizability, charge, secondary structures, surface tension, and normalized van der Waals volumes (Cai et al., 2003).

TABLE 3
Statistics of the promising targets selected from the 1019 research targets by combinations of the methods A, B, C, and D and clinical trial target enrichment factors
Targets that have drugs tested in multiple phases are only included in the highest phase category.

Method or Combination	No. and Percentage of the 30 Phase III Targets Predicted by Method	No. (%) of the 84 Phase II Targets Predicted by Method	No. (%) of the 41 Phase I Targets Predicted by Method	No. and Percentage of the 864 Nonclinical Trial Targets Predicted as Target by Method	Target Prediction Enrichment Factor for Phase II and III Targets	Target Prediction Enrichment Factor for All Clinical Trial Targets
Combination of A, B, C, D	3 (10.0)	7 (8.3)	1 (2.4)	4 (0.5)	6.0	4.8
Any 3-combination of A, B, C, D	15 (50.0)	21 (25.0)	4 (9.8)	31 (3.6)	4.5	3.7
Combination of A, B, C	5 (16.7)	10 (11.9)	1 (2.4)	8 (0.9)	5.6	4.4
Combination of A, B, D	7 (23.3)	11 (13.1)	4 (9.8)	18 (2.1)	4.0	3.6
Combination of A, C, D	9 (30.0)	14 (16.7)	1 (2.4)	14 (1.6)	5.4	4.2
Combination of B, C, D	3 (10.0)	7 (8.3)	1 (2.4)	6 (0.7)	5.3	4.3
Any of A, B, C, D	28 (93.3)	51 (60.7)	25 (61.0)	283 (32.8)	1.8	1.8
A	18 (60.0)	39 (46.4)	17 (41.5)	125 (14.5)	2.6	2.4
B	11 (36.7)	26 (31.0)	8 (19.5)	95 (11.0)	2.4	2.1
C	13 (43.3)	21 (25.0)	3 (7.3)	75 (8.7)	2.7	2.2
D	23 (76.7)	31 (36.9)	13 (31.7)	138 (16.0)	2.4	2.1

TABLE 4

List of phase III targets dropped by combinations of at least three of the methods A, B, C, and D used in this study

The target marked by # has a positive phase III result reported in 2004, but since then there has been no report about the further progress of the phase III drug.

Research Target	Identified as Promising by Method or Combination	No. of Target-Affiliated Pathways	No. of Human Similarity Proteins outside Target Family	No. of Tissues Target Is Primarily Distributed	Targeted Disease Conditions	Target Exploration Status (Tested Drug)
RAC serine/threonine-protein kinase	Combination of A, B	25	1	P	Non-Hodgkin's lymphoma, multiple myeloma, renal cell carcinoma	Phase III (enzastaurin), phase II (perifosine), phase II (XL880), phase I completed (RX-0201), phase I (XL418)
Cell division protein kinase 2	Combination of A, B	4	0	P	Lymphocytic leukemia, lung cancer (NSCLC), non-Hodgkin's lymphoma	Phase III (flavopiridol), phase II completed (seliciclib), phase I/II (AT7519), phase I (SNS-032), preclinical (capridine-β)
α-Glucosidase	Combination of A, D	2	0	P	Cardiovascular disorders	Phase III (acarbose), phase II (celgosivir)
Squalene synthetase	Combination of C, D	2	0	4	Hyperlipidemia	Phase III (TAK-475)
Arachidonate 5-lipoxygenase-activating protein	Only D	1	0	1	Coronary artery disease, heart attack, cardiovascular disorders	Phase III (DG031), phase I (AM803, AM103)
Heme oxygenase	Only D	1	0	1	Neonatal Hyperbilirubinemia, Jaundice	Phase III (stannsoporfin)
Farnesyl protein transferase	Only D	2	0	P	Myeloid Leukemia	Phase III (R115777)
Lipoprotein-associated phospholipase A2	Only D	1	0	P	Atherosclerosis, cardiovascular disorders	Phase II/III (darapladib), phase I (659032)
MMP-12	Only D	1	0	4	Lung cancer (NSCLC)	Phase III (AE-941)
Myophosphorylase	Only D	2	0	1	Lymphocytic leukemia, diabetes mellitus	Phase III (flavopiridol), phase IIa (PSN357)
Neutral endopeptidase	Only D	3	0	P	Hypertension, congestive heart failure	Phase II/III (Ilepatril), phase II (SLV 306)
Sphingosine kinase	Only D	3	0	4	Ovarian cancer	Phase III (phenoxodiol)
Heat shock protein HSP90	Only C	1	0	>10	Multiple myeloma, metastatic breast cancer, prostate cancer	Phase III (tanespimycin), phase II (alvespimycin hydrochloride, IPI-504), phase I (CNF1010, SNX-5422, STA-9090)[a], IND filed (AT13387)[b]
Cathepsin K	None	No-Info	0	4	Osteoporosis, bone metastases	Phase III (odanacatib), phase II (relacatib), phase I/II (MIV-701)[c],
MMP-2/MMP-9	None	3	0	6	Lung cancer (NSCLC), osteoarthritis	Phase III (neovastat), phase II (PG-530742)

[a] Lin et al. (2008).
[b] Hadjuk and Greer (2007).
[c] Mucke et al. (2008).

TABLE 5
List of the difficult targets currently discontinued in clinical trials and having no new drug entering clinical trials, and prediction results by individual or combinations of the methods A, B, C, and D

Currently Discontinued Target	Predicted as Promising by Method or Combination	Targeted Disease Conditions	Discontinued Drug (Company)	Time of Discontinuation	Reason for Discontinuation
Gastrin/cholecystokinin B receptor	Combination of A, B, C, D	Sleep disorders	GW150013 (GlaxoSmithKline, Uxbridge, Middlesex, UK)	December 2001	Not clear
Prolactin receptor (PRLR)	Combination of A, B, D	Cancer/tumors	Endostatin (EntreMed, Rockville, MD)	February 2004	Not clear
B-cell surface antigen CD40	Combination of A, B	Cancer/tumors	Avrend (Amgen, Thousand Oaks, CA)	January 2002	In 1998, phase I results showed few changes in circulating leukocyte subsets after a 5-day course of treatment. In January 2002, Immunex announced that it was no longer developing Avrend.
C3/C5 convertase	Combination of A, B	Coronary artery disease	MLN2222 (Millennium Pharmaceuticals, Cambridge, MA)	August 2005	Not clear
Fungal 14-α demethylase	Combination of A, C	Fungal infections, onychomycosis	Ravuconazole (Eisai, Research Triangle Park, NC)	November 2005	In November 2005, Eisai stated that ravuconazole had been superseded; hence, development was discontinued.
Cytochrome P450 24A1	Only A	Prostate cancer	RC-8800 (Sapphire Therapeutics, Inc., Bridgewater, NJ)	August 2006	Not clear
α-Mannosidase 2	Only D	Cancer/tumors	GD0039 (Inflazyme Pharmaceuticals, Richmond, BC, Canada)	May 2002	In May 2002, GlycoDesign discontinued the phase II clinical trials of GD0039 for the treatment of metastatic renal cancer, because tumor response and adverse events did not meet clinical expectations.
Acyl coenzyme A:cholesterol acyltransferase 1	Only D	Peripheral vascular disease	Avasimibe (Pfizer, Inc., New York, NY)	October 2003	Not clear
Carnitine *O*-palmitoyltransferase I	Only D	Congestive heart failure	Etomoxir (MediGene AG)	April 2003	In April 2003, MediGene terminated phase II trials for etomoxir based on data suggesting an increase in side effects in treated subjects.
MMP-7	Only D	Pancreatic and lung cancer, cancer/tumors	Marimastat (Schering Plough, Kenilworth, NJ) BB-3644 (Vernalis, Cambridge, UK)	June 2003 before 2006	Results of a phase I trial in cancer subjects showed that it caused musculoskeletal pain like marimastat did. At its maximum tolerated dose of 20 mg twice daily, BB-3644 does not show any advantage over marimastat. Due to these results, further trials were not initiated.

Acyl coenzyme A:cholesterol acyltransferase 2	None	Peripheral vascular disease	Avasimibe (Pfizer Central Research (Sandwich, Kent, UK)	October 2003	Not clear
Calcitonin gene-related peptide 2	None	Migraine and cluster headaches	Olcegepant (Boehringer Ingelheim GmbH, Ingelheim, Germany)	March 2007	Not clear
Cell surface glycoprotein MUC18	None	Melanoma	ABX-MA1 (Amgen)	March 2005	Not clear
Hexokinase	None	Prostatic hyperplasia	Lonidamine (Threshold Pharmaceuticals, Redwood City, CA)	July 2006	In July 2006, threshold reported negative results from both a phase II and phase III trial of lonidamine for the treatment of benign prostatic hyperplasia (BPH). Based on these results, threshold announced plans to discontinue development of drug for BPH.
MMP-8	None	Non–small-cell lung cancer (NSCLC)	BMS 275291 (UCB Celltech, Brussels, Belgium)	November 2004	In June 2003, UCB Celltech and Bristol-Myers Squibb Co. (Stamford, CT) announced they were discontinuing the development of BMS 275291 in its current indications due to a general lack of efficacy in phase II.
Pyruvate dehydrogenase kinase	None	Diabetes mellitus	AZD 7545 (AstraZeneca Pharmaceuticals LP, Wilmington, DE)	November 2002	Not clear
Ribonucleoside-diphosphate reductase	None	Various types of cancer	Tezacitabine (Chiron Mimotypes, Clayton, Australia)	March 2004	In March 2004, Chiron Mimotypes announced they were discontinuing development of tezacitabine due to a lack of efficacy in phase II. The compound did not demonstrate sufficient antitumor activity.
Sodium/hydrogen exchanger 1	None	Cardiac surgery	Cariporide (sanofi-aventis, Bridgewater, NJ)	July 2002	Not clear
Sodium/hydrogen exchanger 3	None	Ovarian and lung cancer	Squalamine (Genaera, New York, NY)	January 2007	Not clear

TABLE 6
List of unpromising targets failed in HTS campaigns or found nonviable in knockout studies, and prediction results by combinations of at least three of the methods A, B, C, and D

Target Failed in HTS Campaigns or Found Nonviable in Knockout Studies	Predicted as Promising by Method or Combination	Exploration Results	Target Failed in HTS Campaigns or Found Nonviable in Knockout Studies	Predicted as Promising by Method or Combination	Exploration Results
DNA gyrase subunit A	Combination of A, C, D	Not viable	MabA	Combination of A, B, D	Not viable
Acyl carrier protein synthase	Combination of A, D	No hits	L tRNA synthetase	Combination of A, D	No hits
Penicillin-binding protein-2′	Combination of A, D	No hits	Ribonucleotide reductase	Combination of A, D	Not viable
V tRNA synthetase	Combination of A, D	No hits	AccD5	Combination of B, D	Not viable
Alanine racemase	Combination of C, D	Not viable	AroA	Combination of C, D	Not viable
D-Ala-D-Ala ligase	Combination of C, D	Not viable	FabH	Combination of C, D	Not viable
Thymidine monophosphate kinase	Combination of C, D	Not viable	A tRNA synthetase	Only D	No hits
AcpM	Only D	Not viable	AftA	Only D	Not viable
AroB	Only D	Not viable	AroC	Only D	Not viable
AroE	Only D	Not viable	ArgF	Only D	Not viable
AroG	Only D	Not viable	AroK	Only D	Not viable
AroQ	Only D	Not viable	Biotin ligase (BirA)	Only D	Not viable
Branched chain amino acid aminotransferase	Only D	Not viable	C tRNA synthetase	Only D	No hits
Chorismate synthase	Only D	No hits	CoA (PanK)	Only D	Not viable
D tRNA synthetase	Only D	No hits	DNA polymerase IIIα	Only D	No hits
E tRNA synthetase	Only D	No hits	FtsH ATP-dependent protease	Only D	No hits
G tRNA synthetase	Only D	No hits	Galactofuraosyl transferase	Only D	Not viable
GlmU	Only D	No hits	GlnE	Only D	Not viable
H tRNA synthetase	Only D	No hits	IdeR	Only D	Not viable
K tRNA synthetase	Only D	No hits	LigA	Only D	Not viable
LS, riboflavin synthase	Only D	Not viable	MenA	Only D	Not viable
MenB	Only D	Not viable	MenC	Only D	Not viable
MenD	Only D	Not viable	MenE	Only D	Not viable
MenH	Only D	Not viable	MtrA	Only D	Not viable
MurB	Only D	No hits	N tRNA synthetase	Only D	No hits
P tRNA synthetase	Only D	No hits	Peptidyl tRNA hydrolase	Only D	No hits
Phosphopantetheine adenylyl transferase	Only D	No hits	R tRNA synthetase	Only D	No hits
Ribonuclease P	Only D	No hits	S tRNA synthetase	Only D	No hits
Signal peptidases	Only D	No hits	T tRNA synthetase	Only D	No hits
Transketolase	Only D	No hits	UDP-*N*-acety muramyl:L-alanine ligase (MurC)	Only D	No hits
UMP kinase inhibitor	Only D	No hits	Undecaprenyl (UDP) pyrophosphate synthase	Only D	No hits
Metallo β-lactamase	Non	No hits	SecA subunit of preprotein translocase	Non	No hits
Serine β-lactamase	Non	No hits			

Computation of Number of Human Similarity Proteins, Number of Affiliated Human Pathways, and Number of Human Tissues of a Target. These quantities are needed for determining whether a studied target obeys the simple systems-level druggability rules. Human similarity proteins of a target are those human proteins whose drug-binding domain is similar to that of the studied target by using the same BLAST method as that for analyzing sequence similarity between drug-binding domain of studied target and that of successful target (Altschul et al., 1997). Information about the affiliated pathways of a target was obtained from KEGG database (http://www.genome.jp/kegg/). In estimating the number of human tissues in which each target is distributed, relevant data from the Swiss-Prot database were used. We were able to find the published literature for 92% of these data, and a random check of these publications confirms the quality of the data. We have also used the level 4 tissue distribution data from another database, TissueDistributionDBs (http://genome.dkfz-heidelberg.de/menu/tissue_db/index.html), to derive the tissue distribution pattern of the same set of 158 successful targets. A target is assumed to be primarily distributed in a tissue if no less than 8% of the total protein contents are distributed in that tissue. Approximately 28, 24, 19, 10, 6, 6, 5, and 1% of these targets were found to be affiliated with one to eight tissues, respectively, which are roughly similar to those derived from Swiss-Prot data (Zheng et al., 2006b), although the definition and content of these databases are somehow different. Therefore, our estimated tissue distribution profiles are quite stable, even though the exact percentages may differ by some degrees. The details of this analysis are described in the Supplemental Data 4.

Results

Target Identification by Collective Analysis of Sequence, Structural, Physicochemical, and Systems Profiles of Successful Targets. Each in silico target prediction approach has its unique advantages and limitations. Sequence similarity to the drug-binding domain of a successful target may indicate druggability, which has been extensively explored for target identification (Hopkins and Groom, 2002; Hajduk et al., 2005). However, it cannot fully capture druggable features not reflected by homology (Hajduk et al., 2005) and tends to indiscriminately select homologous proteins. Targets can be identified by structural similarity to drug-binding domain and binding site geometric and energetic features (Hajduk et al., 2005), which are less effective for covering proteins of unknown structure and for describing systems profiles.

Druggability is collectively determined by target structural and physicochemical properties, ability to conduct certain interactions and functions, and patterns of pathway, subcellular, and tissue distributions (Zheng et al., 2006b). Many of these individual properties can be predicted by machine learning (Han et al., 2006), which have been explored for target prediction (Zheng et al., 2006a,b; Han et al., 2007; Xu et al., 2007). This approach cannot fully capture such systems profiles as pathway affiliation and may disproportionately interpret certain physicochemical properties due to biases in protein descriptors or training data sets. Simple systems-level druggability rules have been derived previously (Zheng et al., 2006a,b) and are summarized as follows: targets are similar to fewer (<15) human proteins of nontarget family and associated with fewer (≤ 3) human pathways tend to bind drugs with reduced side effects, and high-efficacy drugs may be more easily derived from targets expressed in fewer tissues (≤ 5) or located within blood vessels or cells lining the arteries where they have higher priority to bind drugs than targets in other tissues. These systems-level rules are not intended for describing structural, physicochemical, and functional aspects of druggability.

These limitations may be reduced if these approaches are combined. Four in silico methods were developed from the relevant profiles of up to 348 successful targets in Therapeutic Target Database (Chen et al., 2002). Method A measures drug-binding domain sequence similarity against those of 168 successful targets with identifiable drug-binding domains. Method B studies drug-binding domain structural similarity against those of 129 successful targets with available structures. Method C predicts druggable proteins from a machine learning model trained by 348 successful targets (Han et al., 2007). Method D evaluates whether the systems-level druggability rules (Zheng et al., 2006a,b) are satisfied. More detailed descriptions about these methods are given in supplemental data.

Performance of Target Identification on Clinical Trial, Nonclinical Trial, Difficult, and Nonpromising Targets. The collective predictive performance of the four methods was tested against clinical trial (from CenterWatch Drugs in Clinical Trials Database) and nonclinical trial research targets (Chen et al., 2002). Clinical trial targets that have drugs in multiple phases are only included in the highest phase category. The best overall performance was produced by the combination of at least three methods, which maximize the collective predictive capability of the methods and minimize the impact of limited structural availability. This combination identified 50% of the 31 phase III (Table 1), 25 and 10% of the 84 phase II and 41 phase I (Table 2), and 4% of the 864 nonclinical trial research targets as promising. We were unable to find a report about target success rates in different developmental stages. It is noted that the reported probabilities of successes in developing systemic broad-spectrum antibacterials are 67, 50, 25, and 3% in phase III, phase II, phase I, and preclinical stages (Payne et al., 2007). The percentages of the identified promising clinical trial targets are lower than but roughly follow a similar descending trend as the reported drug developmental rates. The overall performance of different combinations is given in Table 3. These combinations enriched phase II and phase III target identification rate by ~4- to 6-fold over random selection, with the combination of all four methods producing the highest enrichment.

The 16 identified promising phase III targets include eight to nine targets with positive phase III results. These include six innovative targets without a protein subtype as a successful target (B2 bradykinin receptor, C1 esterase, cholecystokinin receptor type A, NK-2 receptor, sphingosine 1-phosphate receptor 1, and plasma kallikrein) and two conventional targets having a different protein subtype as a successful target (5-hydroxytryptaime 3 receptor and C-X-C chemokine receptor 4). Overall, 60, 43, and 50% of the predicted phase III, phase II, and phase I targets are innovative, which seems to indicate substantial level of successes in exploring novel targets. Most of the identified promising clinical trial targets are from the highly successful G protein-coupled receptor, tyrosine kinase, serine protease, and ATP-binding cassette transporter families for the treatment of cancers, cardiovascular diseases, neural disorders, arthritis, diabetes, and obesity, which suggests

that these families continue to be attractive sources for target discovery (Zambrowicz and Sands, 2003; Overington et al., 2006; Zheng et al., 2006a,b).

The 15 phase III targets dropped by the combination method (Table 4) include MMPs, kinases of cyclin-dependent kinases/mitogen-activated protein kinases/glycogen synthase kinases/CDK-like kinases, cAMP-dependent protein kinase/protein kinase G/protein kinase C extended family and diacylglycerol kinase classes, farnesyltransferases, oxygenase, phospholipase, and others. Only one of these, heme oxygenase, has a positive phase III result reported in 2004. It is noted that this protein is important for attenuating oxidative stress and inflammation and that its inhibition may lead to some adverse effects (Angermayr et al., 2006). The difficulty in exploring some of these targets has been reported previously (Zheng et al., 2006a,b). MMP inhibitors have been explored since the early 1990s, but their trials have not yielded good results due primarily to the lack of subtype selectivity, bioavailability, and efficacy as well as to inappropriate study design (Ramnath and Creaven, 2004). Despite successes in developing several tyrosine kinase inhibitors, kinase inhibitor discovery remains difficult, particularly for nontyrosine kinase classes in part due to broad promiscuity that causes off-target side effects (Fedorov et al., 2007) and network compensatory actions (Sergina et al., 2007).

The combination method dropped 17 of the 19 difficult targets currently discontinued in clinical trials (Table 5) and 63 of the 65 nonpromising targets failed in HTS campaigns or were found nonviable in knockout studies (Table 6). Twelve of the 17 unpredicted difficult targets have been discontinued since 2004 without another drug entering clinical trial. In the HTS campaigns for testing 70 antibacterial targets, up to ~500,000 compounds have been screened at a concentration of 10 μM, 33 of which have yielded no hit and can thus be considered to be highly unpromising (Payne et al., 2007). Target knockout, extensively explored for target validation, has been applied to the validation of 55 targets in *Mycobacterium tuberculosis*, 32 of which have been found to be nonviable for developing drugs (Mdluli and Spigelman, 2006). The low rate in selecting these difficult and unpromising targets suggests that combinations of target prediction methods are capable of eliminating unpromising as well as selecting promising targets.

Discussion

In conclusion, collective use of multiple in silico methods is capable of identifying high percentages of phase III targets, including most of the targets of positive phase III results, and of eliminating difficult and unpromising targets. Our study suggests that comparative analysis of multiple profiles of successful targets provides useful clues to the identification of promising targets. Overall, 71 targets were predicted as promising from a pool of 1019 targets. This number is probably constrained by the limited knowledge from the 348 known successful targets and limited structural information for a large percentage of targets. Rapid progress in genomics (Kramer and Cohen, 2004), structural genomics (Hajduk et al., 2005), and proteomics (Ryan and Patterson, 2002) is revolutionizing target discovery. In addition to high-throughput technologies (Ilag et al., 2002) and cellular (Jackson and Harrington, 2005) and physiological studies (Lindsay, 2005; Sams-Dodd, 2005), various in silico methods are being developed. These methods explore comparative sequence analysis (Hopkins and Groom, 2002), structural analysis (Hajduk et al., 2005), ligand-protein inverse docking (Chen and Zhi, 2001), machine learning of druggability characteristics (Zheng et al., 2006b), and system-related druggability profiles (Zheng et al., 2006a,b) for recognizing target-like and druggable proteins. These progresses combined with increased molecular understanding of diseases and their corresponding targets (Zheng et al., 2006b) enable the development of efficient tools for identifying innovative targets of new therapies and personalized medicine.

Acknowledgments

We thank all former and current Bioinformatics and Drug Design group members for assistance on this project.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:**3389–3402.

Angermayr B, Mejias M, Gracia-Sancho J, Garcia-Pagan JC, Bosch J, and Fernandez M (2006) Heme oxygenase attenuates oxidative stress and inflammation, and increases VEGF expression in portal hypertensive rats. *J Hepatol* **44:**1033–1039.

Austen M and Dohrmann C (2005) Phenotype-first screening for the identification of novel drug targets. *Drug Discov Today* **10:**275–282.

Booth B and Zemmel R (2004) Prospects for productivity. *Nat Rev Drug Discov* **3:**451–456.

Cai CZ, Han LY, Ji ZL, Chen X, and Chen YZ (2003) SVM-Prot: web-based support vector machine software for functional classification of a protein from its primary sequence. *Nucleic Acids Res* **31:**3692–3697.

Chen X, Ji ZL, and Chen YZ (2002) TTD: therapeutic target database. *Nucleic Acids Res* **30:**412–415.

Chen YZ and Zhi DG (2001) Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins* **43:**217–226.

Chiesi M, Huppertz C, and Hofbauer KG (2001) Pharmacotherapy of obesity: targets and perspectives. *Trends Pharmacol Sci* **22:**247–254.

Drews J (1997) Strategic choices facing the pharmaceutical industry: a case for innovation. *Drug Discov Today* **2:**72–78.

Eriksson BI, Dahl OE, Lassen MR, Ward DP, Rothlein R, Davis G, and Turpie AG (2008) Partial factor IXa inhibition with TTP889 for prevention of venous thromboembolism: an exploratory study. *J Thromb Haemost* **6:**457–463.

Fedorov O, Marsden B, Pogacic V, Rellos P, Müller S, Bullock AN, Schwaller J, Sundström M, and Knapp S (2007) A systematic interaction map of validated kinase inhibitors with Ser/Thr kinases. *Proc Natl Acad Sci U S A* **104:**20523–20528.

Hajduk PJ and Greer J (2007) A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov* **6:**211–219.

Hajduk PJ, Huth JR, and Fesik SW (2005) Druggability indices for protein targets derived from NMR-based screening data. *J Med Chem* **48:**2518–2525.

Han L, Cui J, Lin H, Ji Z, Cao Z, Li Y, and Chen Y (2006) Recent progresses in the application of machine learning approach for predicting protein functional class independent of sequence similarity. *Proteomics* **6:**4023–4037.

Han LY, Zheng CJ, Xie B, Jia J, Ma XH, Zhu F, Lin HH, Chen X, and Chen YZ (2007) Support vector machines approach for predicting druggable proteins: recent progress in its exploration and investigation of its usefulness. *Drug Discov Today* **12:**304–313.

Hopkins AL and Groom CR (2002) The druggable genome. *Nat Rev Drug Discov* **1:**727–730.

Howard EL, Becker KC, Rusconi CP, and Becker RC (2007) Factor IXa inhibitors as novel anticoagulants. *Arterioscler Thromb Vasc Biol* **27:**722–727.

Ilag LL, Ng JH, Beste G, and Henning SW (2002) Emerging high-throughput drug target validation technologies. *Drug Discov Today* **7:**S136–S142.

Jackson PD and Harrington JJ (2005) High-throughput target discovery using cell-based genetics. *Drug Discov Today* **10:**53–60.

Koch MA and Waldmann H (2005) Protein structure similarity clustering and natural product structure as guiding principles in drug discovery. *Drug Discov Today* **10:**471–483.

Kramer R and Cohen D (2004) Functional genomics to new drug targets. *Nat Rev Drug Discov* **3:**965–972.

Lin TY, Bear M, Du Z, Foley KP, Ying W, Barsoum J, and London C (2008) The novel HSP90 inhibitor STA-9090 exhibits activity against Kit-dependent and -independent malignant mast cell tumors. *Exp Hematol* **36:**1266–1277.

Lindsay MA (2005) Finding new drug targets in the 21st century. *Drug Discov Today* **10:**1683–1687.

Matter A (2001) Tumor angiogenesis as a therapeutic target. *Drug Discov Today* **6:**1005–1024.

Mdluli K and Spigelman M (2006) Novel targets for tuberculosis drug discovery. *Curr Opin Pharmacol* **6:**459–467.

Mucke HAM, Norman P, Whelan C, and Yeates C (2008) Patent alert. *Curr Opin Investig Drugs* **9:**552–561.

Murzin AG, Brenner SE, Hubbard T, and Chothia C (1995) SCOP: a structural

classification of proteins database for the investigation of sequences and structures. *J Mol Biol* **247:**536–540.
Nicolette CA and Miller GA (2003) The identification of clinically relevant markers and therapeutic targets. *Drug Discov Today* **8:**31–38.
Ohlstein EH, Ruffolo RR Jr, and Elliott JD (2000) Drug discovery in the next millennium. *Annu Rev Pharmacol Toxicol* **40:**177–191.
Overington JP, Al-Lazikani B, and Hopkins AL (2006) How many drug targets are there?. *Nat Rev Drug Discov* **5:**993–996.
Payne DJ, Gwynn MN, Holmes DJ, and Pompliano DL (2007) Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat Rev Drug Discov* **6:**29–40.
Ramnath N and Creaven PJ (2004) Matrix metalloproteinase inhibitors. *Curr Oncol Rep* **6:**96–102.
Rosenberg L (1999) Physician-scientists—endangered and essential. *Science* **283:**331–332.
Ryan TE and Patterson SD (2002) Proteomics: drug target discovery on an industrial scale. *Trends Biotechnol* **20:**S45–S51.
Sakharkar MK, Li P, Zhong Z, and Sakharkar KR (2008) Quantitative analysis on the characteristics of targets with FDA approved drugs. *Int J Biol Sci* **4:**15–22.
Sams-Dodd F (2005) Target-based drug discovery: is something wrong? *Drug Discov Today* **10:**139–147.
Sergina NV, Rausch M, Wang D, Blair J, Hann B, Shokat KM, and Moasser MM (2007) Escape from HER-family tyrosine kinase inhibitor therapy by the kinase-inactive HER3. *Nature* **445:**437–441.
Simmons DL (2006) What makes a good anti-inflammatory drug target? *Drug Discov Today* **11:**210–219.
Smith C (2003) Drug target validation: hitting the target. *Nature* **422:**341, 343, 345 passim.
Tomillero A and Moral MA (2008) Gateways to clinical trials. *Methods Find Exp Clin Pharmacol* **30:**383–408.
Walke DW, Han C, Shaw J, Wann E, Zambrowicz B, and Sands A (2001) In vivo drug target discovery: identifying the best targets from the genome. *Curr Opin Biotechnol* **12:**626–631.
Wong D and Korz W (2008) Translating an antagonist of chemokine receptor CXCR4: from bench to bedside. *Clin Cancer Res* **14:**7975–7980.
Xu H, Xu H, Lin M, Wang W, Li Z, Huang J, Chen Y, and Chen X (2007) Learning the drug target-likeness of a protein. *Proteomics* **7:**4255–4263.
Yao L and Rzhetsky A (2008) Quantitative systems-level determinants of human genes targeted by successful drugs. *Genome Res* **18:**206–213.
Zambrowicz BP and Sands AT (2003) Knockouts model the 100 best-selling drugs–will they model the next 100? *Nat Rev Drug Discov* **2:**38–51.
Zheng C, Han L, Yap CW, Xie B, and Chen Y (2006a) Progress and problems in the exploration of therapeutic targets. *Drug Discov Today* **11:**412–420.
Zheng CJ, Han LY, Yap CW, Ji ZL, Cao ZW, and Chen YZ (2006b) Therapeutic targets: progress of their exploration and investigation of their characteristics. *Pharmacol Rev* **58:**259–279.

Address correspondence to: Prof. YuZong Chen, Department of Pharmacy, National University of Singapore, 18 Science Dr. 4, Singapore 117543. E-mail: phacyz@nus.edu.sg.