Supplemental Material can be found at:
http://pharmrev.aspetjournals.org/content/suppl/2006/05/26/58.2.259.DC1.html

0031-6997/06/5802-259–279$7.00
PHARMACOLOGICAL REVIEWS
Vol. 58, No. 2
60201/3116251
Pharmacol Rev 58:259–279, 2006
Printed in U.S.A
Copyright © 2006 by The American Society for Pharmacology and Experimental Therapeutics

PHARM REV

Therapeutic Targets: Progress of Their Exploration and Investigation of Their Characteristics

C. J. ZHENG, L. Y. HAN, C. W. YAP, Z. L. JI, Z. W. CAO, AND Y. Z. CHEN

Bioinformatics and Drug Design Group, Department of Computational Science, National University of Singapore, Singapore, Singapore

***Abstract*——Modern drug discovery is primarily based on the search and subsequent testing of drug candidates acting on a preselected therapeutic target. Progress in genomics, protein structure, proteomics, and disease mechanisms has led to a growing interest in and effort for finding new targets and more effective exploration of existing targets. The number of reported targets of marketed and investigational drugs has significantly increased in the past 8 years. There are 1535 targets collected in the therapeutic target database compared with ~500 targets reported in a 1996 review. Knowledge of these targets is helpful for molecular dissection of the mechanism of action of drugs and for predicting features that guide new drug design and the search for new targets. This article summarizes the progress of target exploration and investigates the characteristics of the currently explored targets to analyze their sequence, structure, family representation, pathway association, tissue distribution, and genome location features for finding clues useful for searching for new targets. Possible "rules" to guide the search for druggable proteins and the feasibility of using a statistical learning method for predicting druggable proteins directly from their sequences are discussed.**

I. Introduction

Theparadigm of modern drug discovery has primarily been based on the search for drug leads against a preselected therapeutic target followed by subsequent testing of the derived drug candidates (Drews, 1997b, 2000; Ohlstein et al., 2000). Continuous effort has been made to explore the targets of highly successful drugs, and increasing interest has been directed to the identification of new targets (Drews, 1997a,b, 2000; Ohlstein et al., 2000; Terstappen and Reggiani, 2001). Rapid advances in genomics (Debouck and Metcalf, 2000; Peltonen and McKusick, 2001), protein structures (Sali, 1998), proteomics (Dove, 1999), and molecular mecha-

Address correspondence to: Dr. Chen Yu Zong, Department of Pharmacy, Science Faculty, National University of Singapore, Blk S16, Level 8, 08-14, 3 Science Drive 2, Singapore 117543, Singapore. E-mail: phacyz@nus.edu.sg

Article, publication date, and citation information can be found at http://pharmrev.aspetjournals.org.
doi:10.1124/pr.58.2.4.

An erratum has been published:
http://pharmrev.aspetjournals.org/content/58/3/682.full.pdf

PHARMACOLOGICAL REVIEWS

Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

PHARM REV

PHARMACOLOGICAL REVIEWS

nisms of diseases (Macdonald, 2000; Baker and Wood, 2001) not only enable the search for new targets, but also facilitate the study of existing targets for finding clues to new target identification and for probing the molecular mechanisms of drug actions, adverse drug reactions, and the pharmacogenetic implication of variations in gene sequences and in the profiles of expression and post-transcriptional processing (Macdonald, 2000; Cotsarelis and Millar, 2001; Evans and Johnson, 2001; Nicholls, 2003).

These advances (Macdonald, 2000; Baker and Wood, 2001; Cotsarelis and Millar, 2001; Hoffman and Dressman, 2001) and the development of target identification and validation technologies (Drews, 2000; Lizotte-Waniewski et al., 2000; Walke et al., 2001; Ilag et al., 2002) have led to the discovery of a growing number of new and novel targets (Chiesi et al., 2001; Kumar et al., 2001; Matter, 2001; Greenfeder and Anthes, 2002; Helmuth, 2002; Lark and Morrison, 2002). A study undertaken in 1996 showed that there were ~500 targets (Drews, 1997b, 2000), 120 of which have been reported to be the identifiable targets of currently marketed drugs (Hopkins and Groom, 2002). The latest number of reported targets collected in the Therapeutic Target Database (Chen et al., 2002) (http://bidd.nus.edu.sg/group/ttd/ttd.asp) is 997 distinct proteins (undivided into subtypes), 1494 distinct protein subtypes, and 41 nucleic acids. These include 268 successful targets, which are targeted by at least one marketed drug, and 1267 research targets, which are only targeted by investigational agents not approved for clinical use at present. A relatively small percentage of research targets are known to have become successful targets since 1996 (Zambrowicz and Sands, 2003). The significant increase in the number of successful and research targets is probably due in large part to a combination of increasing exploration of disease-specific protein subtypes of existing targets and new information about previously unknown or unreported targets of existing drugs and investigational agents (Leurs et al., 1998; Vane et al., 1998; Kennedy and Ramachandran, 2000; Torphy and Page, 2000).

Statistical analysis of disease genes and related proteins suggested that the total number of the estimated potential targets in the human genome ranges from 600 to1500 (Hopkins and Groom, 2002). Investigation of the yeast genome found that antifungal targets constitute 2 to 5% of the genome (Hopkins and Groom, 2002). With the assumption of a similar percentage of targets, the number of potential targets in disease-related microbial genomes can be roughly estimated to be >1000. A typical viral genome contains one to four targets (Miller and Hazuda, 2001; Wen et al., 2003), which gives a crude estimate of >100 potential targets in disease-related viral genomes. Therefore, the total number of distinct targets is probably in the range of 1700 to 3000. Identification and exploration of these targets are important for the drug discovery communities to find new therapeutic agents and more effective treatment options (Chaix-Couturier et al., 2000).

Knowledge of existing targets is useful for finding clues to new target identification. It is also important for the molecular dissection of the mechanism of action of drugs, the prediction of features that guide new drug design, and the development of tools for these tasks (Kennedy, 1997; Lizotte-Waniewski et al., 2000; Walke et al., 2001; Ilag et al., 2002; van de Waterbeemd and Gifford, 2003). Analysis of these targets also provides useful information about general trends, current focuses of research, and areas of successes and difficulties in the exploration of therapeutic targets for the discovery of drugs against specific diseases. This article is intended to provide an overview of the progress in the exploration of therapeutic targets and to investigate the characteristics of these targets for providing useful clues to search new targets. On the basis of information from the Therapeutic Target Database (Chen et al., 2002), sequence, structure, family representation, pathway association, tissue distribution, and genome location features of both successful and research targets are analyzed. Possible rules to guide the search for druggable proteins and the feasibility of using a statistical learning method, support vector machines, for predicting druggable proteins directly from their sequences are discussed.

II. Distribution of Therapeutic Targets with Respect to Disease Classes

A. General Distribution Pattern

Distribution of successful targets with respect to different disease classes is given in Table 1. Disease classes are based on the international statistical classification of diseases of the World Health Organization (1992). Neoplasms, infectious and parasitic diseases, nervous system and sense organs disorders, circulatory system diseases, and mental disorders constitute the groups with the largest number of targets. Other groups consisting of larger number of targets are respiratory system diseases, genitourinary system diseases, musculoskeletal system and connective tissue diseases, and endocrine disorders. The numbers of targets for each of these classes are 78, 78, 56, 54, 46, 35, 24, 23, and 21, respectively.

Examples of successful targets in the class of neoplasms are estrogen receptors and aromatase (breast cancer), thymidylate synthase and DNA topoisomerase I (colorectal cancer), luteinizing hormone-releasing hormone (prostate cancer), and *BCR-ABL* (chronic myeloid leukemia). Examples in the class of infectious and parasitic diseases are HIV-1 protease (AIDS), influenza A virus M2 protein (influenza A), hepatitis B virus poly-

Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

TABLE 1

Number of successful targets in different disease classes

The total number of nonredundant successful targets is 268, 120 of which are for more than one disease classes. Because of this redundancy of targets, the sum of the number of targets in these classes is greater than 268. The number of targets shared between different disease classes is also given in the table.

Indications	Disease Classes	No. of Therapeutic Targets		Shared Therapeutic Targets																	
		All Related Targets	Non-redundant Targets	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r
a	Blood and blood-forming organ diseases	13	2		8	1	1	1	2	2	0	0	1	0	2	0	4	2	3	1	1
b	Circulatory system diseases	54	9	8		11	10	10	24	15	6	7	6	2	6	12	19	6	8	8	2
c	Digestive system diseases	19	4	1	11		5	3	8	9	4	3	5	1	2	5	5	3	6	1	1
d	Genitourinary system diseases	24	0	1	10	5		6	11	7	3	6	1	1	2	6	12	1	2	2	3
e	Musculoskeletal system and connective tissue diseases	23	4	1	10	3	6		10	6	2	2	5	4	6	6	12	1	5	2	3
f	Nervous system and sense organ diseases	56	7	2	24	8	11	10		17	4	6	3	2	7	27	13	3	14	7	2
g	Respiratory system diseases	35	5	2	15	9	7	6	17		5	3	8	2	5	12	10	2	8	4	1
h	Skin and subcutaneous tissue diseases	13	2	0	6	4	3	2	4	5		3	3	1	1	2	7	2	2	2	1
i	Endocrine disorders	21	6	0	7	3	6	2	6	3	3		3	0	3	3	8	4	1	1	1
j	Immunity disorders	18	2	1	6	5	1	5	3	8	3	3		3	6	2	9	2	3	2	1
k	Infectious and parasitic diseases	78	57	0	2	1	1	4	2	2	1	0	3		4	1	17	4	1	1	2
l	Inflammation	15	1	2	6	2	2	6	7	5	1	3	6	4		2	8	1	4	1	1
m	Mental disorders	46	10	0	12	5	6	6	27	12	2	3	2	1	2		5	3	10	2	0
n	Neoplasms	78	29	4	19	5	12	12	13	10	7	8	9	17	8	5		5	5	6	4
o	Nutritional and Metabolic diseases	21	5	2	6	3	1	1	3	2	2	4	2	4	1	3	5		1	0	0
p	Symptoms, signs, and ill-defined conditions	22	2	3	8	6	2	5	14	8	2	1	3	1	4	10	5	1		1	2
q	Injury and poisoning	15	3	1	8	1	2	2	7	4	2	1	2	1	1	2	6	0	1		0
r	Congenital anomalies	4	0	1	2	1	3	3	2	1	1	1	1	2	1	0	4	0	2	0	
Total successful therapeutic targets based on disease classes		555 (duplicate); 268 (distinct)	148	Redundancy of therapeutic targets = 120; nonredundancy of therapeutic targets = 148																	

merase (hepatitis B), penicillin-binding proteins and DD-carboxypeptidase (bacterial infections), hexamethylenetetraamine and dihydropteroate synthase (malaria), and 1,3-β-glucan synthase and lanosterol-14-α-demethylase (fungal diseases). Those in the class of nervous system and sense organs disorders are acetylcholinesterase and *N*-methyl-D-aspartate (NMDA[1]) receptors (Alzheimer's disease), catechol-*O*-methyltransferase and D2 dopamine receptors (Parkinson's disease), α_2- and β_1-adrenoceptors (glaucoma and ocular hypertension), 5-HT 1D receptor (migraine), and μ/κ opioid receptor (drug dependence).

Additional examples of successful targets are platelet glycoprotein IIb/IIIa receptors (acute coronary syndrome), angiotensin-converting enzyme, angiotensin receptor AT1, and β-1 and α adrenoceptors (hypertension, cardiac failure, and arrhythmias), Endothelin receptor (primary pulmonary hypertension) for circulatory system diseases; monoamine oxidase A and serotonin transporter (depression), D2 dopamine receptor (schizophrenia), GABA receptor and β-adrenergic receptor (insomnia and anxiety) for mental disorders; β_2-adrenergic receptor, 5-lipoxygenase, and leukotriene receptor (asthma), and σ-type opioid receptor (cough) for respiratory system diseases; phosphodiesterase type 5 (erectile disfunction) and muscarinic receptor M3 (overactive bladder) for genitourinary system diseases; cyclooxygenase 2, tumor necrosis factor-α, interleukin-1 receptor (rheumatoid arthritis, osteoarthritis), and farnesyl diphosphate synthase (osteoporosis) for musculoskeletal system and connective tissue diseases; gastrointestinal lipases, fatty acid synthase (obesity), and farnesyl diphosphate synthase (hypercalcemia) for nutritional and metabolic diseases; and insulin receptor and peroxisome proliferator-activated receptor-γ (diabetes) for endocrine disorders.

Since 1996, a number of innovative targets that are based on new mechanisms or new targets for treating diseases have emerged, which usually have large markets and become highly successful (Zambrowicz and Sands, 2003). These targets [with the year of the first Food and Drug Administration (FDA) approval and the name of the approved drug in parentheses] are vascular endothelial growth factor (2004, Avastin) for the treatment of colorectal cancer, NMDA receptor (2003, Namenda) for Alzheimer's disease, HIV gp41 (2003, Fuzeon) for HIV infection, hepatitis B virus DNA polymerase (2002, Hepsera) for hepatitis B, mineralocorticoid receptor (2002, Eplerenone) for hypertension, endothelin receptor (2001, Tracleer) for primary pulmonary hypertension, *BCR-ABL* (2001, Gleevec) for chronic myeloid leukemia, retinoid receptors (1999, Targretin) for cutaneous T-cell lymphoma,

[1] Abbreviations: NMDA, *N*-methyl-D-aspartate; DSPase, dual-specificity protein phosphatase; 5-HT, 5-hydroxytryptamine; FDA, Food and Drug Administration; PDE4, phosphodiesterase-4; MMP, matrix metalloproteinase; ABT-518, [*S*-(*R**,*R**)]-*N*-[1-(2,2-dimethyl-1,3-dioxol-4-yl)-2-[[4-[4-(trifluoromethoxy)-phenoxy]phenyl]sulfonyl]ethyl]-*N*-hydroxyformamide; β_3AR, β_3-adrenergic receptor; SCOP, Structural Classification of Proteins; EC, Enzyme Commission; SVM, support vector machine.

PHARM REV
PHARMACOLOGICAL REVIEWS
aspet
Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

PHARM REV

PHARMACOLOGICAL REVIEWS

gastrointestinal lipase (1999, Xenical) for obesity, FK-binding protein 12 (1999, Rapamune) for the prevention of organ rejection after renal transplantation, HER2/nue (1998, Herceptin) for HER2 positive metastatic breast cancer, phosphodiesterase 5 (1998, Viagra) for erectile dysfunction, platelet glycoprotein IIb/IIIa receptor (1998, Aggrastat, Integrilin) for severe chest pain and small heart attacks, cyclooxygenase 2 (1998, Celebrex) for arthritis, peroxisome proliferator activated receptor (1997, Rezulin) for type 2 diabetes mellitus, and platelet P2Y12 receptor (1997, Plavix) for stroke and heart attack.

B. Targets for the Treatment of Diseases in Multiple Classes

Some targets are used for the treatment of diseases from more than one class. Disease classes with higher concentration of shared targets are circulatory system diseases, neoplasms, and nervous system and sense organs disorders. For instance, there are 24, 19, and 15 targets for circulatory system diseases that are shared with those of nervous system and sense organ disorders, neoplasms, and respiratory diseases, respectively. The high concentration of shared targets in this class is partly attributed to the involvement of the circulatory system in various disease conditions. There are strong interactions between the nervous and cardiovascular systems, and it is not surprising that targets involved in the cross-talk between these systems are used for both diseases (Luchner and Schunkert, 2004). Tumor growth relies on the formation of new blood vessels, and proteins involved in angiogenesis have been targeted for anticancer drug development as well as circulatory system diseases (Matter, 2001). Sensory receptors in the respiratory system are known to respond to irritants and subsequently induce cardiovascular responses, and targets involved in these responses are used for symptom relief of respiratory diseases as well as for the treatment of cardiovascular diseases (Widdicombe and Lee, 2001).

An example of a shared target is the β-adrenoceptor for circulatory system diseases, nervous system disorders, and respiratory system diseases. Heart failure is known to harmfully activate sympathetic nervous system as well as the renin-angiotensin system, and these circulatory system disease-associated disorders can be treated by β-adrenoceptor antagonists (Toda, 2003). β-Adrenoceptor antagonists have been used for the treatment of tremor and reduce the physical symptoms of anxiety (e.g., tremor and palpitations), two nervous system disorders, by blocking peripheral sympathetic responses (Emilien and Maloteaux, 1998). β-Adrenoceptor agonists have been used for the treatment of asthma, a typical respiratory system disease, by dilating bronchial smooth muscle (Emilien and Maloteaux, 1998).

Another example of a shared target is dual-specificity protein phosphatases (DSPases), which represent a subclass of the protein tyrosine phosphatases with highly conserved phosphatase active site motifs. DSPases dephosphorylate serine, threonine, and tyrosine residues in the same protein substrate, and they play important roles in multiple signaling pathways and seem to be deregulated in cancer and Alzheimer's disease (Ducruet et al., 2004). Because of their roles and properties, there has been increasing effort to identify DSPase inhibitors that are more potent and selective than the general tyrosine phosphatase inhibitor sodium orthovanadate, for the treatment of both diseases, which has led to the discovery of several promising leads (Lyon et al., 2002).

C. Research Targets

The number of research targets of each disease class is given in Fig. 1 along with that of successful targets. With the exception of the class of congenital anomalies, there seems to be a significant increase in the level of exploration of targets for every disease class, as evidenced by the significantly larger number of research targets than that of successful targets, which reflects intensive efforts to find effective treatment options for all diseases. Little success seems to have been made in the identification of useful targets for congenital anomalies due partly to the use of surgical therapies as the primary treatment option (Lin et al., 2002; Scheinfeld et al., 2004) and partly to the lack of knowledge of the mechanism of the relevant diseases (Kobayashi and Stringer, 2003). The disease classes with the largest increases of targets are neoplasms with 468 research targets versus 78 successful targets, infectious and parasitic diseases with 287 research targets versus 78 successful targets, nervous system and sense organs disorders with 171 research targets versus 56 successful targets, circulatory system diseases with 168 research targets versus 54 successful targets, nutritional and metabolic disorders with 120 research targets versus 21 successful targets, inflammation with 111 research targets versus 15 successful targets, musculoskeletal system and connective tissue diseases with 92 research targets versus 23 successful targets, and endocrine disorders with 91 research targets versus 21 successful targets.

The majority of the research targets are distributed in six classes. There are 37, 23, 13, 13, 9, and 9% of the research targets distributed in the classes of neoplasms, infectious and parasitic diseases, nervous system and sense organs disorders, circulatory system diseases, nutritional and metabolic disorders, and inflammation, respectively. Overall, the number of nonredundant research targets in these six disease classes is 708, which accounts for 56% of the total number of research targets. This number reflects the intensive efforts directed to the search for effective therapeutic agents for cancer treatment and prevention (Buolamwini, 1999; Dubowchik and Walker, 1999; Elsayed and Sausville, 2001), cardiovascular diseases (Persidis, 1999; Bicknell et al., 2003),

Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

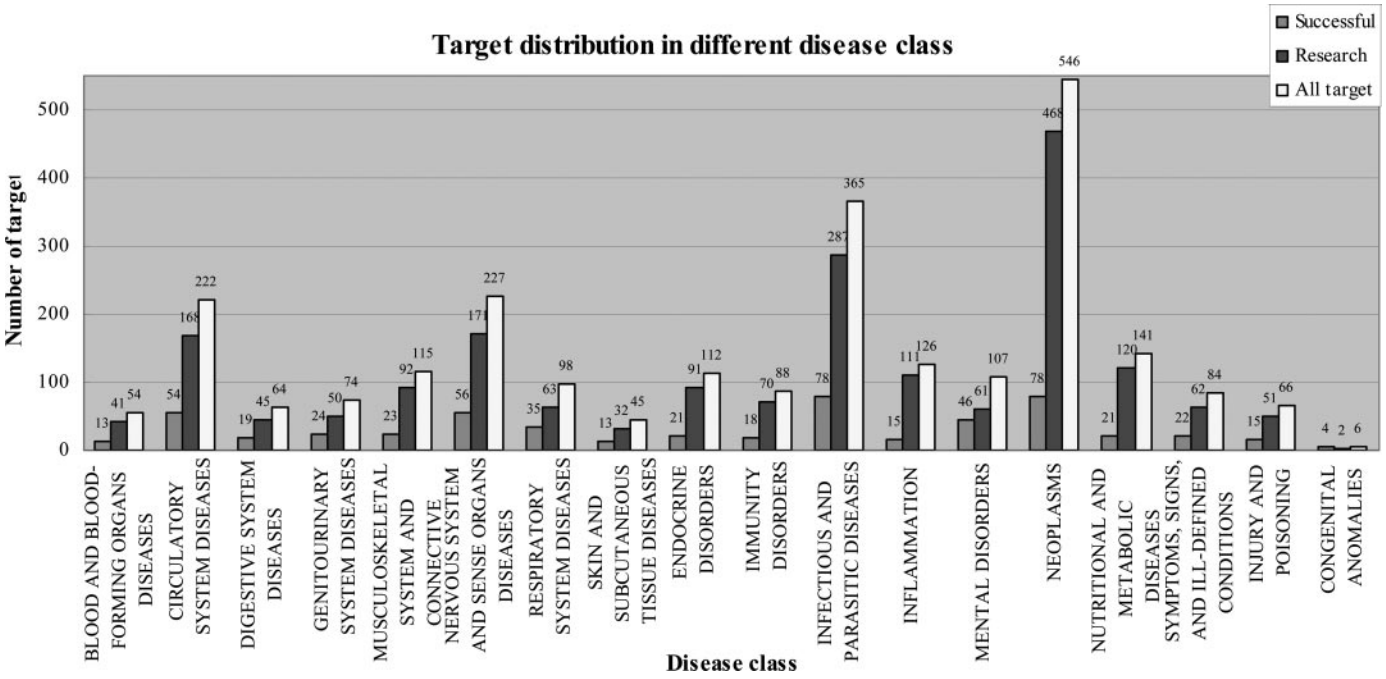

FIG. 1. Distribution of therapeutic targets against disease classes. The gray, black, and white bars represent successful, research, and all targets, respectively.

inflammatory diseases (Lewis and Manning, 1999), obesity (Campfield et al., 1998; Bray and Tartaglia, 2000; Ahima and Osei, 2001; Clapham et al., 2001), and high cholesterol (Chong and Bachenheimer, 2000; Best and Jenkins, 2001),

Examples of specific diseases in these key classes that have a substantial number of research targets are various cancers with 468 targets (Buolamwini, 1999; Dubowchik and Walker, 1999; Elsayed and Sausville, 2001), cardiovascular diseases with 120 targets (Persidis, 1999; Bicknell et al., 2003), diabetes with 65 targets (Wagman and Nuss, 2001), arthritis with 64 targets (Blake and Swift, 2004), obesity with 57 targets (Campfield et al., 1998; Bray and Tartaglia, 2000; Macdonald, 2000; Ahima and Osei, 2001; Clapham et al., 2001), Alzheimer's disease with 44 targets (Irizarry and Hyman, 2001; Windisch et al., 2002), and high cholesterol with 12 targets (Chong and Bachenheimer, 2000; Best and Jenkins, 2001). These diseases affect a significant number of patients and thus have substantial interest has been shown in the development of new therapeutic agents for their treatment.

Another class with a high ratio of research versus successful targets is infectious and parasitic diseases, which has a ratio of 287:78. The significant increase in the number of research targets for this disease class primarily stems from the pursuit for new generations of antibiotics (Bush and Macielag, 2000), antifungal agents (Hossain and Ghannoum, 2000), and anti-HIV drugs (De Clercq, 2001) as well as for the development of effective drugs for malaria (Olliaro and Yuthavong, 1999) and a variety of viral infections such as hepatitis, herpes simplex virus, and respiratory syncytial virus (De Clercq, 2001).

III. Current Trends in Exploration of Therapeutic Targets

A. Targets of Investigational Agents in United States Patents Approved in 2000 through 2004

Clues about the current trends in target exploration can be obtained from the targets described in the recently approved patents of investigational agents. Most of these patents describe molecular mechanisms, and many of them provide the identifiable target for each group of patented agents. Tables 2 and 3 give some of the successful targets and research targets described in the U.S. patents approved between January 2000 and September 2004. A total of 2080 U.S. patents of investigational agents have been approved during this period, 1606 or 77.2% of which have an identifiable target.

There are 395 identifiable targets described in these 1606 patents. Of these targets, 264 have been found in more than one patent and 50 appear in more than 10 patents. The number of patents associated with a target can be considered to partly correlate with the level of effort and intensity of interest currently being directed to it. Approximately one third of the patents with an identifiable target were approved in the past year. This suggests that the effort for the exploration of these targets is ongoing, and there has been steady progress in the discovery of new investigational agents directed to these targets.

PHARM REV
PHARMACOLOGICAL REVIEWS
aspet
Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

TABLE 2

Some of the successful targets explored for the new investigational agents described in the U.S. patents approved in 2000 through 2004

Therapeutic Target		No. of U.S. Patents	Targeted Diseases
Protein	Subgroup		
Adrenergic receptors (63)	α-Adrenergic receptor	1	Nasal congestion, glaucoma, asthma, migraine, diarrhea
	α_1-Adrenergic receptor	4	Congestive heart failure, hypertension, benign prostatic hyperplasia, eye disorders
	α_{1D}-Adrenergic receptor	5	Benign prostatic hyperplasia, peripheral vascular disease, congestive heart failure, hypertension
	α_{1B}-Adrenergic receptor	3	CNS disorders, anxiety, sleep disorders, schizophrenia, hypertension, sexual dysfunction
	α_2-Adrenergic receptor	8	Nasal congestion, glaucoma, asthma, migraine, diarrhea
	α_{2C}-Adrenergic receptor	1	Mental illnesses
	β-Adrenergic receptor	6	Airway inflammatory disorders, asthma, obstructive lung disease, ocular hypertension, glaucoma
	β_2-Adrenergic receptor	5	Pulmonary disorders, asthma, chronic bronchitis, emphysema, neurological disorders, cardiac disorders
	β_3-adrenergic receptor	30	Metabolic disorders, atherosclerosis, gastrointestinal disorders, type 2 diabetes
HIV protease (58)			Retroviral infection, viral infections (HIV),viral infections (EHV)
Serotonin receptors (43)	5-HT receptor	1	Headaches
	5-HT 1 receptor	7	Depression, anxiety, eating disorders, obesity, drug abuse, cluster headache, migraine, pain
	5-HT 1A receptor	4	Mood disorders, pain, neuronal disorders
	5-HT 1B receptor	4	Migraine, depression, psychological disorders
	5-HT 1D receptor	6	Depression, psychological disorders
	5-HT 1F receptor	2	Headaches
	5-HT 2 receptor	3	Cardiovascular disorders, CNS disorders, gastrointestinal disorders, glaucoma
	5-HT 2A receptor	5	Psychotic disorders, schizophrenia, sleep-disordered breathing, sleep apnea syndrome
	5-HT 2B receptor	2	Irritable bowel syndrome
	5-HT 2C receptor	6	Obesity, obsessive-compulsive disorder, depression
	5-HT 3 receptor	8	Gastrointestinal motility disorders, headache, anxiety, depression, psychosis, rheumatoid disease
	5-HT 6 receptor	2	Hyperactivity disorders, attention deficit hyperactivity disorder
	5-HT 7 receptor	4	CNS disorders, aforementioned disorders, disorders of the bladder, urinary retention
Factor Xa (47)			Thrombotic disorders, coronary artery, cerebrovascular disease, inflammatory diseases, cancers
Substance P receptor (39)			Asthma, cough, bronchospasm, depression, emesis, inflammatory diseases, gastrointestinal disorders
Tyrosine kinases (39)	Tyrosine-protein kinase	28	Cancers, atherosclerosis, restenosis, endometriosis, psoriasis
	Tyrosine-protein kinase Src	5	Immune diseases, cancers, atheroscelerosis, graft rejection, rheumatoid arthritis
	Tyrosine-protein kinase JAK3	3	Allergic disorders
	Tyrosine-protein kinase SYK	1	Inflammatory diseases, obstructive airways disease
	Tyrosine-protein kinase BTK	2	Cancers, immune diseases
Cyclooxygenase 2 (38)			Alzheimer's disease, osteoporosis, glaucoma, inflammation, asthma, cancers, heart diseases
Thrombin (36)			Blood coagulation, cardiovascular disorders, thrombosis, ischemia, stroke, restenosis, inflammation
NMDA receptor (27)	NMDA receptor	14	CNS disorders, inflammatory diseases, allergic diseases, depression, drug abuse
	NMDA receptor NR2B	13	Pain, migraine, depression, anxiety, schizophrenia, Parkinson's disease, stroke
Opioid receptors (25)	Opioid receptor	4	Eating disorders, narcotic dependence, alcoholism, pain, drug dependence
	μ-Type opioid receptor	2	Constipation, vomiting and/or nausea, pain, anxiety
	δ-Type opioid receptor	3	CNS disorders, peripheral nervous system diseases, pain
	κ-Type opioid receptor	16	Depression, headaches, inflammation, arthritis, stroke, functional bowel disease, abdominal pain, pruritus
Inducible NOS (24)			CNS disorders, inflammation, shock, immune disorders, disorders of gastrointestinal motility
Muscarinic receptors (22)	Muscarinic receptor	10	Cognitive disorders, Alzheimer's disease, neurologic, psychiatric disorders, pain
	M1 receptor	3	Cognitive disorders, Alzheimer's disease, glaucoma
	M2 receptor	8	Cognitive disorders, Alzheimer's disease, smooth muscle disorders
	M3 receptor	5	Smooth muscle disorders
	M4 receptor	4	Mental disorders, Parkinson's disease, glaucoma
Adenosine receptors (22)	Adenosine receptor	1	Cardiac and circulatory disorders, CNS disorders, respiratory disorders
	Adenosine A1 receptor	6	Allergic disorders, CNS disorders, asthma
	Adenosine A2a receptor	10	CNS disorders, Parkinson's disease
	Adenosine A2b receptor	3	Airway diseases, asthma, inflammation, diabetes mellitus
	Adenosine A3 receptor	6	Bronchus disorders, inflammation, allergosis

PHARM REV
PHARMACOLOGICAL REVIEWS
aspet
Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

TABLE 2—*continued*

Therapeutic Target		No. of U.S. Patents	Targeted Diseases
Protein	Subgroup		
HIV reverse transcriptase (20)			Viral infections (HIV)
PDE5 (19)			Cardiovascular and cerebrovascular disorders, disorders of urogenital system, erectile dysfunction
Histamine receptors (16)	Histamine H1 receptor	8	Allergy, rhinitis, congestion, inflammation, CNS diseases, respiratory disorders, viral infections
	Histamine H2 receptor	6	Dry eye, duodenal ulcer, gastroesophogeal reflux disease, gastrointestinal disorders
	Histamine H3 receptor	5	Allergy, congestion, inflammation, CNS-related diseases
Tumor necrosis factor (16)			Inflammatory diseases, allergic diseases, cytokine-induced toxicity, muscular disorders
Serotonin reuptake (16)			CNS-related diseases, anxiety
Gonadotropin-releasing hormone receptor (16)			Sex-hormone-related disorders, steroid-dependent tumors, prostate cancer
Endothelin receptors (15)	Endothelin receptor	11	Angina, pulmonary hypertension, Raynaud's disease, migraine, blood vessel disorders, renal diseases
	Endothelin A receptor	4	Hypertension, acute myocardial infarction, Raynaud's syndrome, atherosclerosis, asthma, prostate cancer
	Endothelin B receptor	2	Hypertension, acute myocardial infarction, stroke, benign prostate hypertrophy, atherosclerosis, asthma
HMG-CoA reductase (13)			Atherosclerosis, lipid disorders, hypercholesterolemia, hypertriglyceridemia, combined hyperlipidemia
Gastric H^+/K^+-ATPase (12)			Bacterial infections, gastric acid-related diseases, nasal disorders, bronchus disorders, osteoporosis
U-plasminogen activator (12)			Angiogenic disorders, arthritis, inflammation, osteoporotic, cancers, lymphomas, chronic dermal ulcers
LHRH receptor (12)			Hormone-dependent tumors, hormone-influenced disorders, benign prostate hyperplasia, endometriosis
LHRH (12)			Hormone-dependent tumors, hormone-influenced disorders, benign prostate hyperplasia, endometriosis
RARs (11)	Retinoic acid receptor	5	Acne, psoriasis, rheumatoid arthritis, viral infections
	RAR-α	1	Systemic erythematosus, glomerulonephritis, lupus nephritis, autoimmune anemia
	RAR-γ	5	Emphysema and associated pulmonary diseases, dermatological disorders, epithelial lesions, tumors
PPARs (10)	PPAR-α	4	Abnormality of lipid metabolism, type 2 diabetes
	PPAR-γ	5	Diabetes, obesity, metabolic syndrome, cardiovascular diseases, dyslipidemia, cancers
	PPAR-δ	2	Dyslipidemia, syndrome X, cardiovascular diseases, diabetes, obesity, anorexia bulimia
Glycogen synthase kinase-3 (9)			Cancers, diabetes, Alzheimer's disease
Prostanoid FP receptor (9)			Bone disorders, glaucoma, ocular hypertension
Calcium channel (9)			Cardiovascular disorders, angina, hypertension, ischemia
5-Lipoxygenase (8)			Asthma, atherosclerosis, cancers
Glycoprotein IIb/IIIa receptor (8)			Cancers, osteoporosis, arteriosclerosis, restenosis, ophthalmia
DNA topoisomerases (7)	DNA topoisomerase I	3	Cancers
	DNA topoisomerase II	6	Bacterial infections, cancers
Angiotensin-converting enzyme (7)			Diabetic complications, diabetic retinopathy, diabetic neuropathy, diabetic nephropathy
Glucocorticoid receptor (7)			Cocaine addiction , depression, Alzheimer's disease, aforementioned diseases, Diabetes
Serine protease (7)			Cardiovascular disorders, thrombosis, asthma
Angiotensin II receptor (7)	AT1	6	Acute myocardial infarction, cancers, hypertension, QT dispersion
	AT2	6	Acute myocardial infarction, cancers, hypertension, QT dispersion, wounds healing
Estrogen receptors (6)	Estrogen receptor	2	Breast cancer, inflammatory diseases, sepsis, viral infections, cardiovascular diseases
	Estrogen receptor a	4	Uterine cancer, adjuvant breast cancer, prostate cancer, benign prostatic hyperplasia, ovarian cancers
Tryptase (6)			Cardiovascular disorders, inflammatory diseases, cancers
Dopamine receptors (5)	Dopamine receptor	2	Cancers, Parkinson's disease
	D2 dopamine receptor	2	Fibromyalgia, musculoskeletal pain symptoms associated with fibromyalgia
	D3 dopamine receptor	2	Fibromyalgia, musculoskeletal pain symptoms associated with fibromyalgia
	D4 dopamine receptor	1	Central nervous system disorders, psychotic disorders, schizophrenia
Interleukin-1 receptor (5)	IL-1R	4	Hypotension, tachycardia, lung edema, renal failure
	IL-1R-a	1	Allergic rhinitis, allergic asthma, allergic Inflammatory diseases
Neuraminidase (5)			Influenza A, influenza B, viral infections, bacterial infections
Histone deacetylase (5)			Cancers, hematological disorders, metabolic disorders, cystic fibrosis, adrenoleukodystrophy

CNS, central nervous system; NOS, nitric-oxide synthase; LHRH, luteinizing hormone-releasing hormone; RAR, retinoic acid receptor; PPAR; peroxisome proliferator-activated receptor; IL-1R, interleukin-1 receptor.

PHARM REV
PHARMACOLOGICAL REVIEWS
aspet
Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

TABLE 3

Research targets explored for the new investigational agents described in U.S. patents approved in 2000 through 2004

Therapeutic Target		No. of U.S. Patents	Targeted Diseases
Protein	Subgroup		
MMPs (79)	Matrix metalloproteinase	62	Arthritis, cancers, tissue ulceration, periodontal disease, bone disease, diabetes
	MMP-1	1	Pulmonary emphysema
	MMP-2	12	Cancers
	MMP-3	9	Multiple sclerosis, heart failure, cancers, inflammation, arthritis, autoimmune disorders
	MMP-4	1	Arthritis, cancers
	MMP-7	1	Inflammatory diseases, rheumatoid arthritis, tumors
	MMP-8	1	Inflammatory diseases, cancers
	MMP-9	5	Cancers, arthritis
	MMP-11	1	Cancers
	MMP-12	1	Ulcerative colitis, Crohn's disease, atherosclerosis, gastrointestinal ulcers, emphysema
	MMP-13	9	Osteoarthritis, rheumatoid arthritis, cancers, inflammation, heart failure
PDEs (78)	Phosphodiesterase	3	Erectile dysfunction, sexual dysfunction
	PDE1A	2	Cardiovascular and cerebrovascular disorders, erectile dysfunction
	PDE2A	2	Cardiovascular and cerebrovascular disorders, disorders of urogenital system
	PDE3	5	Airway obstructions, inflammatory diseases, premature ejaculation, sexual dysfunction
	PDE4	49	Airway obstructions, inflammatory diseases, allergic disorders
	PDE4A	1	Respiratory disorders, asthma
	PDE7	4	Asthma, rheumatoid arthritis, psoriasis, atopic dermatitis, chronic bronchitis
αv Integrin receptors (40)	αva3 Integrin receptor	39	Cancers, arteriosclerosis, restenosis, osteolytic disorders, osteoporosis, ophthalmic diseases
	αva5 Integrin receptor	16	Cancers, osteoporosis, arteriosclerosis,r estenosis, ophthalmia
Farnesyl-protein transferase (26)			Cancers, restenosis, atherosclerosis
Tumor necrosis factor-α-converting enzyme (25)			Arthritis, tumor metastasis, tissue ulceration, bone disease, diabetes, HIV infection
Cathepsin K (23)			Autoimmune diseases, cartilage degradation, osteoporosis, pulmonary disorders
Substance receptor (22)			Asthma, cough, bronchospasm, depression, inflammation, gastrointestinal disorders
Tachykinin NK_3 receptor (19)			CNS disorders, inflammation, pain, migraine, asthma, emesis, gastrointestinal disorders
Neuropeptide Y receptor (18)	Neuropeptide Y receptor	10	Eating disorders, feeding disorders, cardiovascular disorders, physiological disorders
	Neuropeptide Y5 receptor	8	Eating disorders, diabetes, nutritional disorders, obesity
Cyclin-dependent kinases (17)	Cyclin-dependent kinase	12	Cancers, inflammation, arthritis, Alzheimer's disease, cardiovascular disorders
	Cell division protein kinase 2	4	Alopecia, cancers
	Cell division protein kinase 4	1	Cancers
Stress kinase p38 (17)			Chronic inflammatory, autoimmune diseases, hypercholesterolemia
Hexokinase D (17)			Type 2 diabetes
Phospholipase A_2 (16)	Phospholipase A_2	12	Inflammatory diseases, allergic diseases, pancreatitis, septic shock
	Cytosolic phospholipase A_2	4	Inflammation, asthma, arthritis, inflammatory bowel disease, neurodegenerative diseases
Cytochrome P450RAI (15)			Skin diseases, cancers, cardiovascular diseases, inflammation, neurodegenerative diseases
Cathepsin S (14)			Osteoporosis, autoimmune disorders
Vasopressin receptor (13)	Vasopressin receptor	4	Cerebrovascular disease, cerebral edema, cerebral infarction, depressant, anxiety
	Vasopressin V1a receptor	4	Obsessive-compulsive disorder, aggressive disorders, depression, anxiety
	Vasopressin V2 receptor	7	Diabetes insipidus, nocturnal enuresis, nocturia, urinary incontinence, coagulation disorders
Trypsin-like serine protease (13)			Thrombosis, ischemia, stroke, restenosis, inflammation
Interleukin-8 receptor (13)			Inflammation
Corticoliberin (12)			Circadian rhythm disorders, congestive heart failure, hypertension, metabolic disorders, stroke
Cathepsin B (12)			Autoimmune diseases, pancreatitis, inflammatory airway disease, bone and joint disorders
Cathepsin L (12)			Autoimmune diseases, myocardial infarction, inflammation, muscular dystrophies, Alzheimer's disease
Caspases (12)	Caspase	4	Cancers

PHARM REV
PHARMACOLOGICAL REVIEWS
aspet
Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

TABLE 3—*continued*

Therapeutic Target		No. of U.S. Patents	Targeted Diseases
Protein	Subgroup		
	Caspase-8	8	Inflammation, cancers, autoimmune disorders, neuronal disorders
	Caspase-9	1	Inflammation, cancers, autoimmune diseases, ischemic diseases, neurodegenerative disorders
Chemokine receptors (12)	CCR1	2	Inflammation, immune diseases
	CCR2	2	Atherosclerosis, inflammatory diseases, immune disorders, transplant rejection, aids
	CCR3	3	Respiratory disorders, bronchus disorders, inflammatory diseases, allergy
	CCR5	5	Inflammatory diseases, viral infections (HIV)
Prenyl-protein transferase (12)			Cancers
Prostaglandin E receptor (11)	Prostaglandin E receptor	3	Dry eye, keratoconjunctivitis, Sjögren's syndrome, ocular surface diseases, glaucoma
	Prostanoid EP2 receptor	4	Ocular hypotensive, glaucoma, mesangial proliferative glomerulonephritis
	Prostanoid EP4 receptor	4	Renal failure, dry eye
PTP-1B (10)			Diabetes, obesity, autoimmune diseases, acute and chronic inflammation, osteoporosis, cancers
Serine/threonine protein kinase (10)	Serine/threonine protein kinase	2	Tumor growth, restenosis, atherosclerosis, cancers
	Serine/threonine protein kinase 12	8	Cancers, diabetes, Alzheimer's disease
Endothelin (9)	Endothelin	7	Angina, pulmonary hypertension, Raynaud's disease, migraine, heart failure, pain, respiratory disorders
	Endothelin-1	2	Pulmonary hypertension, cerebral infarction, cerebral ischemia, congestive heart failure
β-Lactamase (9)			Bacterial antibiotic resistance, bacterial infections
Metabotropic glutamate receptor (9)	mGLUR	7	Neurological disorders, psychosis, schizophrenia, Alzheimer's disease, cognitive and memory disorders
	mGLUR1	1	Neurological diseases, neurodegenerative diseases, psychotic diseases
	mGLUR5	1	Neurological disorders, psychiatric disorders
Interleukin-1β convertase (8)			Inflammatory and autoimmune diseases, bone disorders, proliferative disorders, infectious diseases
Glutamate receptors (8)	Glutamate receptor, ionotropic kainate 1	3	Headaches, neuronal disorders
	Glutamate receptor AMPA	5	Epilepsy, diseases resulting in muscle spasm, various neurodegenerative diseases, stroke
Aldose reductase (8)			Diabetic neuropathy, diabetic nephropathy, diabetic retinopathy, diabetic cardiomyopathy
Protease activated receptor 1 (8)			Aggregation of blood platelets, thrombosis, thromboembolism, myocardial infarction

PTP-1B, protein tyrosine phosphatase 1B.

Many of the highly explored targets (those described in a large number of patents) are successful targets, which seems to indicate continuous effort and prolonged interest in the exploration of the targets of highly successful drugs for deriving new therapeutic agents. Successful targets that are described in a higher number of patents are adrenoceptor subtypes (63 distinct patents, 41 β- and 22 α-subtypes, for cardiovascular diseases, depression, hypertension, asthma, diabetes, obesity, and others), HIV protease (58 patents, for HIV infections), 5-HT receptor subtypes (43 distinct patents, 23 5-HT1, 16 5-HT2, 8 5-HT3, 2 5-HT6, and 4 5-HT7 subtypes, for depression, anxiety, eating disorders, obesity, irritable bowel syndrome, attention deficit hyperactivity disorder, bladder disorder, and others), coagulation factor Xa (47 patents, for thromboembolic disorders), substance P receptor (39 targets, for asthma, bronchitis, migraine, and others), tyrosine kinases (39 patents, for angiogenic disorders, cancer, inflammatory diseases, allergic diseases, and others), cyclooxygenase 2 (38 patents, for inflammation, senile dementia, cancer, asthma, and congestive heart failure), thrombin (36 patents, for thrombosis, myocardial ischemia, myocardial infarction, and others), NMDA receptors (27 patents, for central nervous system disorders), opioid receptors (25 patents, for depression, pain, inflammation, arthritis, pruritus, alcohol and drug dependence, and others), inducible nitric oxide synthase (24 patents, for inflammation, pain, arthritis, asthma, bronchitis, and others), muscarinic receptors (22 patents, for Alzheimer's disease, pain, glaucoma, and others), and adenosine receptors (22 patents, for asthma, inflammation, diabetes, coronary artery disease, hepatic fibrosis, renal dysfunction, and others).

Research targets that are described in a higher number of patents are matrix metalloproteinase (79 patents, for cancers, tissue ulceration, abnormal wound healing, periodontal disease, bone disease, diabetes, arthritis, atherosclerosis, inflammation, and others), phosphodiesterase 4 (49 patents, for inflammation, asthma, prostate diseases, osteoporosis, and others), $\alpha v\beta_3$ integrin receptor (39 patents, for angiogenic disorders, inflammation, bone degradation, cancer, diabetic retinopathy, thrombosis, and others), farnesyl-protein transferase (26 patents, for arthropathies, arthritis, gout, cancers,

PHARM REV
PHARMACOLOGICAL REVIEWS
aspet
Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

PHARM REV

PHARMACOLOGICAL REVIEWS

restenosis, and others), tumor necrosis factor-α-converting enzyme (25 patents, for arthritis, cancers, tissue ulceration, abnormal wound healing, periodontal disease, bone disease, and others), cathepsin K (23 patents, for autoimmune diseases, cartilage degradation, osteoporosis, and pulmonary disorders), and substance K receptor (22 patents, for asthma, cough, bronchospasm, inflammatory diseases, arthritis, central nervous system disorders, and others).

B. Progress and Difficulties in Target Exploration

Some of these highly explored research targets were used for drug development well before 2000. Various degrees of progress have been made toward discovery and testing of agents directed at these targets. However, for some of these targets, many difficulties remain to be resolved before viable drugs can be derived. The appearance of a high number of patents associated with these targets partly reflects the intensity of efforts for finding effective drug candidates against these targets.

Farnesyl-protein transferase inhibitors have been designed and tested as novel agents for the treatment of myeloid malignancies since the early 1990s (Gibbs et al., 1993). Initially developed to inhibit the prenylation necessary for Ras activation, their mechanism of action seems to be more complex, involving other proteins unrelated to Ras. Preliminary results from clinical trials demonstrated inhibition of enzyme target, a favorable toxicity profile and promising efficacy (Jabbour et al., 2004). This led to the initiation of phase II trials in a variety of hematologic malignancies and disease settings (Karp and Lancet, 2004).

Phosphodiesterase 4 (PDE4) has been explored as the target of novel anti-inflammatory agents since the mid-1990s (Barnette et al., 1996). The rationale for selecting this target comes, in part, from the clinical efficacy of theophylline, an orally active nonselective PDE inhibitor. It has been found that intracellular cyclic adenosine monophosphate levels regulate the function of many of the cells thought to contribute to the pathogenesis of respiratory diseases such as asthma and chronic obstructive pulmonary disease, and these cells also selectively express PDE4 (Spina, 2003). Recent clinical studies of selective PDE4 inhibitors such as cilomilast and roflumilast for the treatment of inflammatory lung disease showed positive results that offer some optimism, and efforts are being made to reduce the side effect of these drug candidates (Spina, 2003).

Matrix metalloproteinases (MMPs) have been targeted for cancer and other diseases since the early 1990s (Docherty et al., 1992). MMPs degrade the extracellular matrix, promote tumor invasion and metastasis, and regulate host defense mechanisms and normal cell function. Blocking all MMPs may not lead to a positive therapeutic outcome. So far, most clinical trials of MMP inhibitors have not yielded good results, due primarily to the lack of subtype selectivity, bioavailability, and efficacy and in some cases inappropriate study design (Ramnath and Creaven, 2004). Intensive efforts are being directed at the discovery of potent, selective, orally bioavailable MMP inhibitors for the treatment of cancer. There has been encouraging news about some inhibitors, such as ABT-518, that have entered into phase I clinical trials in cancer patients (Wada, 2004).

Intensive research efforts have been directed at development of β_3-adrenergic receptor (β_3-AR) selective agonists for the treatment of type 2 diabetes and obesity in humans since early 1990s (Howe et al., 1992). These agonists have been observed to simultaneously increase lipolysis, fat oxidation, energy expenditure and insulin action leading to the belief that this receptor might serve as an attractive target for the treatment of diabetes and obesity. However, drug design efforts have been hindered by the obstacles in the pharmacological differences between rodent and human β_3-AR, the lack of selectivity of leads, and unsatisfactory oral bioavailability and pharmacokinetic properties of tested agents (de Souza and Burkey, 2001). A recent test of β_3-AR agonists directed at the human receptor showed promising results in their ability to increase energy expenditure in humans after a single dose. However, they do not seem to be able to sustain their effects when administered chronically. Further clinical testing will be necessary, using compounds with improved oral bioavailability and potency, to help assess the physiology of the β_3-AR in humans and its attractiveness as a potential therapeutic for the treatment of type 2 diabetes and obesity (de Souza and Burkey, 2001).

Inspection of the targets reported in these patents also provides useful information about the progress for the search of new targets. Examples of newly explored targets are 88-kDa glycoprotein growth factor for the treatment of cancer (Serrero, 2001), anandamide amidase for pain (Makriyannis et al., 2002), FK506-binding protein 4 for neurological disorders (Wythes et al., 2000), galanin receptor type 2 for central nervous system disorders (Scott et al., 2000), γ-secretase for Alzheimer's disease (Teall, 2001), glycogen synthase kinase-3β for diseases characterized by an excess of Th2 cytokine (Gong et al., 2001), orexin receptor 1 for obesity (Branch et al., 2002), and tripeptidyl-peptidase II for eating disorders and obesity (Schwartz et al., 2000). Most of these new research targets are being explored for the treatment of high-impact diseases needing effective or more treatment options.

C. Targets of Subtype-Specific Drugs

There are 62 targets being explored for the design of subtype-specific drugs, which represents 15.7% of the 395 identifiable targets in U.S. patents approved in 2000 through 2004. Compared with the 11 targets of FDA-approved subtype-specific drugs during the same period, a significantly larger number of targets are being explored for the design of subtype-specific drugs. However,

aspet

Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

PHARM REV

PHARMACOLOGICAL REVIEWS

the percentage of these targets with respect to the total number of targets in U.S. patents is smaller than that of the FDA-approved drugs during the same period, which seems to indicate the level of difficulty of finding subtype-specific agents directed at a variety of targets. For instance, although there are 79 patents for MMP, only three patents describe subtype-specific investigational drugs. These are MMP-9 inhibitors (Bein and Simons, 2001), MMP-4 inhibitors (Greene and Rosen, 2001), and MMP-13 inhibitors (Picard and Wilson, 2002).

The targets with a higher number of patents of subtype-specific investigational drugs are phosphodiesterase 4 with 49 patents (for the treatment of asthma, inflammation, and osteoporosis), cyclooxygenase 2 with 38 patents (inflammation, cancer, and others), adrenoceptor β with 41 patents (hyperglycemia, obesity, gastrointestinal disorders, and others), adrenoceptor α with 22 patents (hypertension, pain, gastric ulcers, vascular diseases, and others), phosphodiesterase 5 with 19 patents (sexual dysfunction), cytochrome P450RAI with 15 patents (diseases responsive to retinoid treatment), 5-HT1 receptor with 17 patents (depression, eating disorders, obesity, headache, and others), 5-HT2 receptor with 12 patents (irritable bowel syndrome), 5-HT3 receptor with 8 patents (blood glucose control), and 5-HT7 receptor with four patents (bladder disorder and urinary retention).

IV. Characteristics of Therapeutic Targets

A. What Constitutes a Therapeutic Target?

The majority of clinical drugs achieve their effect by binding to a cavity and regulating the activity, of its protein target. Specific structural and physicochemical properties, such as the "rule of five" (Lipinski et al., 2001), are required for these drugs to have sufficient levels of efficacy, bioavailability, and safety, which define target sites to which drug-like molecules can bind. In most cases, these sites exist out of functional necessity, and their structural architectures accommodate target-specific drugs that minimally interact with other functionally important but structurally similar sites. These constraints limit the types of proteins that can be bound by drug-like molecules, leading to the introduction of the concept of druggable proteins (Hopkins and Groom, 2002; Hardy and Peet, 2004). Druggable proteins do not necessarily become therapeutic targets (Hopkins and Groom, 2002); only those that play key roles in diseases can be explored as potential targets. Nonetheless, analysis of the characteristics of these druggable proteins is useful for facilitating molecular dissection of the mechanism of drug targeting and for guiding the search for new targets.

Certain characteristics are expected for therapeutic targets (Hopkins and Groom, 2002). These targets play critical and preferably unsubstitutable roles in disease processes. They have certain level of functional and structural novelty to allow for drug specificity. They are not significantly involved in other important processes in humans to limit potential side effects. Expression of these targets is either at a constrained level or tissue selective to allow for sufficient drug efficacy. Drug-binding sites are expected to have certain structural and physicochemical properties to accommodate high-affinity site-specific binding and subsequent regulation of protein activity by drug-like molecules. These characteristics probably define the sequence features, structural architectures, genomic signatures, and proteomic profiles of therapeutic targets and their roles at the pathway, cellular, and physiological levels.

Useful hints about some of the characteristics of therapeutic targets may be probed by analyzing their sequence properties, protein families, structural folds, biochemical classes, similarity proteins, gene locations in the human genome, and associated pathways. These hints may be potentially used for deriving rules and developing predictive tools for searching druggable proteins from genomic data. As part of the effort for supporting such a goal, relevant features of 268 successful targets and 1267 research targets have been described.

B. Protein Families Represented by Therapeutic Targets

The sequence and functional similarities within a protein family usually indicate general conservation of binding site architecture between family members. If a drug can specifically target one member of a family, then it is possible to design molecules of similar physicochemical properties for specific binding to some of the other members of the family, and multiple members of a family have been explored for developing drugs with different therapeutic applications (Chantry, 2003; Gronemeyer et al., 2004). A recent analysis of the identifiable drug-binding domains of 399 targets (including 120 successful targets) suggested that these targets are represented by 130 protein families, nearly half of which are represented by six families (Chantry, 2003), which indicate the level of extensive exploration of multiple members of specific families as therapeutic targets.

With the availability of the information of a significantly higher number of targets than that used in the recent analysis, it is of interest to reinvestigate family representations of therapeutic targets. There are 190 successful targets and 1035 research targets with identifiable drug-binding domain. Analysis of the Pfam (Bateman et al., 2004) protein family of these domains found that these targets are represented by 88 and 357 families, respectively.

Approximately 47% of the 190 successful targets fall into 10 families. These, in terms of Pfam family names, are 7-transmembrane receptor rhodopsin family (32 targets), nuclear hormone receptor (11 targets), zinc finger (11 targets), ion transport protein (seven targets), protein kinase (five targets), short-chain dehydrogenase

aspet

Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

PHARM REV

PHARMACOLOGICAL REVIEWS

(four targets), amino acid permease (four targets), cytochrome P450 (four targets), neurotransmitter-gated ion channel 1 (four targets), fibronectin type III domain (four targets), and sodium/neurotransmitter symporter (three targets).

Approximately 41% of the 1035 research targets fall into 24 families, which include 7-transmembrane receptor rhodopsin family (94 targets), protein kinase (61 targets), immunoglobulin (25 targets), trypsin (21 targets), ion transport protein (18 targets), SH2 domain (17 targets), nuclear hormone receptor (16 targets), zinc finger (15 targets), fibronectin type III domain (15 targets), receptor family ligand binding region (12 targets), phorbol ester/diacylglycerol binding domain (12 targets), leucine-rich repeat (12 targets), ankyrin repeat (11 targets), papain family cysteine protease (10 targets), lectin C-type domain (10 targets), matrixin (nine targets), small cytokines (nine targets), 3′5′-cyclic nucleotide phosphodiesterase (eight targets), hemopexin (eight targets), ATP-binding cassette transporter (seven targets), hormone receptor (seven targets), eukaryotic-type carbonic anhydrase (seven targets), short-chain dehydrogenase (six targets), and neurotransmitter-gated ion channel (six targets).

Overall, 42% or 518 of the 1225 successful and research targets are distributed in 26 protein families, which include all of the six top target-representing families found in the recent study (Chantry, 2003). The remaining 58% or 707 targets are distributed in 358 families. There are seven families both in the top 10 families of successful targets and top 22 families of the research targets. These are 7-transmembrane receptor rhodopsin family, ligand-binding domain of nuclear hormone receptor, protein kinase domain, short-chain dehydrogenase, neurotransmitter-gated ion-channel ligand binding, ion transport protein, and zinc finger.

Two parallel lines of target exploration are indicated. One is the extensive use of successful targets and additional members of a relatively small group of protein families. On average, 20 targets from each of the 26 heavily used families have been explored. The other is the exploration of a diverse range of proteins in a variety of families. On average, only one or two targets from each of the other 358 protein families have been explored or are being evaluated. It is expected that more members from some of these families may be used as viable targets.

It is of interest to estimate the total number of families that represent all of the 3000 targets that are postulated to exist. If we assume that all of the 1535 currently explored targets are viable ones, which is doubtful but does not significantly affect our estimate, there are ~1500 undiscovered targets. If these undiscovered targets roughly follow the same pattern of protein family representation as the currently explored targets, it is expected that 40% of them are from a relatively small group of families, probably no more than a few dozen. Moreover, the bulk, say 60%, of the remaining 60% of these targets is probably from the 358 families that represent 60% of the currently explored targets. Therefore, there are no more than 24% of the undiscovered targets that are from protein families not represented by the known targets, and these targets are represented by no more than 400 families. This gives a crude estimate of no more than 800 target-representing protein families, which is likely to be substantially less, for all of the therapeutic targets. The total number of protein families in the Pfam database is 7677 (Bateman et al., 2004). Thus, target-representing families account for <11% of all protein families, and 40% of the targets are expected to be represented by just a few dozen families.

C. Structural Folds

A common feature of targets in a particular family is the general conservation of binding site architecture. Binding sites of drugs are usually located within a specific cavity of their target proteins, and drug binding is primarily facilitated by hydrophobic, aromatic stacking, hydrogen bonding, and van der Waals interactions (Yu et al., 2003). Certain constrains on the architectures of drug-binding domains are expected for accommodating the binding of the target-specific rule-of-five small molecules that minimally interact with other functionally important but structurally similar sites. There have been reports about specific drug-domain architecture (Benke et al., 1997; Poulos, 1988; Striessnig et al., 1998).

Because of the distribution of therapeutic targets in a relatively small number of protein families, it is expected that these targets are represented by a relatively small number of structural folds. Examination of the structural folds of the drug-binding domains can therefore shed light on the structural characteristics of therapeutic targets. Structural folds of proteins can be obtained from the SCOP database (Andreeva et al., 2004), which contains 1133 structural folds (Release 1.69) generated from the analysis of 25,973 protein entries from the Protein Data Bank database (Sussman et al., 1998).

There are 52 successful targets that have both an available three-dimensional structure and an identifiable drug-binding domain. Analysis of the SCOP structural folds of these targets shows that they are represented by 29 folds, which are given in Table 4. Approximately 60% of these targets are represented by just eight folds. These eight folds, given by SCOP fold names, are nuclear receptor ligand-binding domain (eight targets), triosephosphate isomerase β/α-barrel (six targets), protein kinase-like (four targets), 4-helical cytokines (three targets), NAD(P)-binding Rossmann-fold domains (three targets), trypsin-like serine proteases (three targets), α/β-hydrolases (two targets), and galactose-binding domain-like (two targets).

There are 283 research targets that have both an available three-dimensional structure and an identifiable drug-binding domain, which are represented by 107

Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

TABLE 4
Structural folds represented by successful targets
Structural folds are from the SCOP database. Data are based on 113 successful targets that have available three-dimensional structures.

SCOP Fold Identification	Fold Description	No. of Targets
a0.123	Nuclear receptor ligand-binding domain	8
c0.1	TIM β/α-barrel	6
d0.144.1	Protein kinase-like	4
a0.26.1	4-Helical cytokines	3
b0.47.1	Trypsin-like serine proteases	3
c0.2	NAD(P)-binding Rossmann-fold domains	3
b0.18.1	Galactose-binding domain-like	2
c0.69	α/β-Hydrolases	2
c0.65.10.1	Formyltransferase	1
c0.19.1	FabD/lysophospholipase-like	1
g0.39.1	Glucocorticoid receptor-like (DNA-binding domain)	1
c0.71.1	Dihydrofolate reductases	1
a0.104.10.1	Cytochrome P450	1
b0.74.1	Carbonic anhydrase	1
c0.82.1	ALDH-like	1
b0.68	6-Bladed β-propeller	1
d0.163.1	DNA breaking-rejoining enzymes	1
d0.32	Glyoxalase/bleomycin resistance protein/dihydroxybiphenyl dioxygenase	1
d0.68	IF3-like	1
d0.174.10.1	NO synthase oxygenase domain	1
d0.6.10.1	Prion-like	1
d0.110	Profilin-like	1
c0.66.1	*S*-adenosyl-L-methionine-dependent methyltransferases	1
a0.126	Serum albumin-like	1
d0.179.10.1	Substrate-binding domain of HMG-CoA reductase	1
d0.168.1	Succinate dehydrogenase/fumarate reductase flavoprotein, catalytic domain	1
d0.117.1	Thymidylate synthase/dCMP hydroxymethylase	1
b0.22	Tumor necrosis factor-like	1
j0.61.10.1	Human glutathione reductase inhibitor	1

ALDH, aldehyde reductase (dehydrogenase); IF3, imitation factor 3.

folds. Of these targets 60% are represented by 21 folds. These include protein kinase-like (21 targets), 4-helical cytokines (14 targets), trypsin-like serine proteases (14 targets), P-loop-containing nucleoside triphosphate hydrolases (12 targets), zincin-like (12 targets), triosephosphate isomerase β/α-barrel (11 targets), interleukin 8-like (nine targets), cysteine proteinases (eight targets), cystine-knot cytokines (eight targets), nuclear receptor ligand-binding domain (eight targets), C-type lectin-like (seven targets), NAD(P)-binding Rossmann-fold domains (seven targets), immunoglobulin-like β-sandwich (six targets), caspase-like (five targets), flavodoxin-like (five targets), acid proteases (four targets), α/β-hydrolases (four targets), concanavalin A-like lectins/glucanases (four targets), knottins (four targets), phosphorylase/hydrolase-like (four targets), and PLP-dependent transferases (four targets).

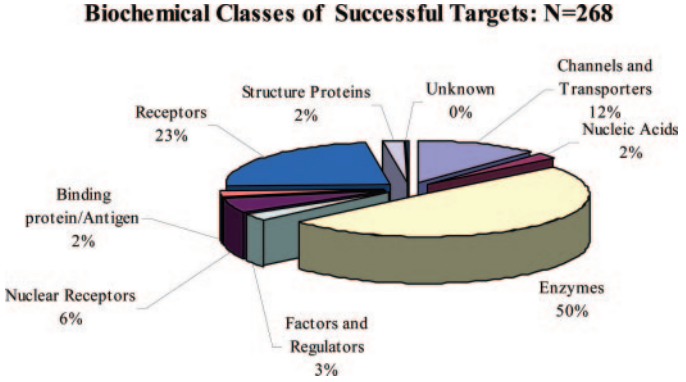

FIG. 2. Distribution of successful targets with respect to different biochemical classes.

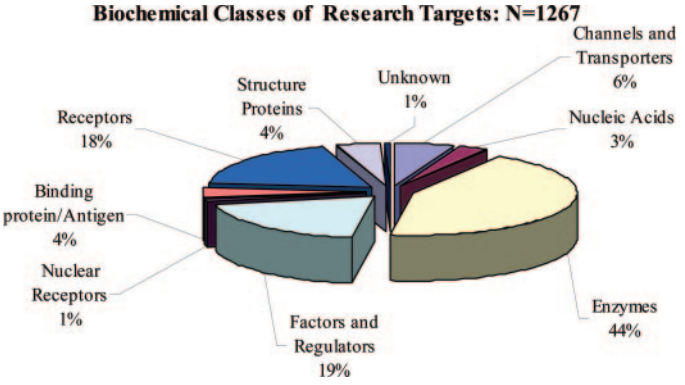

FIG. 3. Distribution of research targets with respect to different biochemical classes.

D. Biochemical Classes

Distribution of successful and research targets with respect to biochemical classes is given in Figs. 2 and 3 respectively. Biochemical classes include enzymes, receptors, nuclear receptors, channels, and transporters, factors and regulators (factors, hormones, regulators, modulators, and receptor-binding proteins involved in a disease process), antigens, and the remaining binding proteins not covered in other classes, structural proteins (nonreceptor membrane proteins, adhesion molecules, envelop proteins, capsid proteins, motor proteins, and other structural proteins), and nucleic acids (Drews,

PHARM REV
PHARMACOLOGICAL REVIEWS
aspet
Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

PHARM REV

PHARMACOLOGICAL REVIEWS

2000). The targets unable to be assigned into any of these biochemical classes are tentatively grouped into a separate "unknown" class.

The overall distribution pattern of successful targets and that of research targets are roughly similar to the pattern of the 120 successful targets (Hopkins and Groom, 2002) and that of the targets with drug-like leads (Drews, 1997b, 2000). The class with the largest number of targets is enzymes, which includes 134 successful and 551 research targets representing 50 and 44% of the total number of successful and research targets, respectively. The second largest group of successful targets is receptors with 61 targets representing 23% of successful target population. The second largest group of research targets is factors and regulators with 242 targets representing 18% of the research target population, which is compared with the corresponding group of eight successful targets that represents only 3% of the total successful target population. Thus, there seems to be a dramatic increase in the number of factors and regulators being explored for the treatment of a variety of diseases including cancers (Darnell, 2002), autoimmune diseases (Eggert et al., 2004), inflammation, diabetes, and neurodegenerative diseases (Collins, 2004).

Target distribution profiles of the groups with a substantial number of successful targets are channels and transporters with 32 targets representing 12% of the successful target population, nuclear receptors with 15 targets representing 6% of the successful target population, and factors and regulators with eight targets representing 3% of the successful target population. The distribution patterns of the research target groups are receptors with 230 targets representing 18% of the research target population, channels and transporters with 75 targets representing 6% of the research target population, structural protein with 56 targets representing 4.4% of the research target population, antigens and other substrate-binding proteins with 50 targets representing 4% of the research target population, nucleic acids with 36 targets representing 3% of the research target population, and nuclear receptors with 19 targets representing 1% of the research target population.

E. Human Proteins Similar to Therapeutic Targets

In the present day drug development processes, drug candidates have frequently been intentionally designed to bind to their target specifically and to avoid strong interactions with other human protein members of the same protein family to which the target belongs (Drews, 1997a,b, 2000; Ohlstein et al., 2000; Terstappen and Reggiani, 2001). The successfully designed agents are thus less likely to significantly interfere with the function of human proteins of the same family, reducing the risk of some of the potential unwanted effects. However, their possible interactions with human proteins outside the family are not intentionally avoided at the design stage, and the potential unwanted effects associated with some of these interactions can only be detected at the later testing stages. Therefore, it tends to be easier to find successful drugs for those targets that have fewer human similarity proteins outside of their family. One can then speculate that targets with fewer human similarity proteins outside their family tend to be more likely to be explored for drug development.

Some crude estimates about the number of human similarity proteins outside the family of each individual target can be provided by conducting a sequence similarity search against the 59,618 proteins in the human genome that are currently available in protein databases. The derived target characteristics depend on the choice of parameters of bioinformatic tools and the quality of data sources. In estimating the number of similar proteins for each target, a stricter Position-Specific Iterated-Basic Local Alignment Search Tool cutoff *e* value = 0.001 was used. This value has been reported to reliably predict homologous relationships (George and Heringa, 2002), and it can be used to find 16% more structural relationships in the SCOP database than that using standard sequence similarity with a 40% sequence-identity threshold (Gerstein, 1998). Most protein pairs that share 40~50% or higher sequence identity differ by less than 1Å RMS deviations (Wood and Pearson, 1999; Koehl and Levitt, 2002), and a larger structural deviation likely alters drug binding properties. Therefore, the adopted *e* value seems to be reasonable for selecting those similarity proteins relevant to the binding of a common set of drugs. Nonetheless, a small percentage of protein pairs of higher sequence identity have been found to differ by larger RMS deviations (Wood and Pearson, 1999), and some protein pairs of low sequence identity may also have high structural similarity, which likely affects the accuracy of our analysis to some extent.

Table 5 summarizes the results of a Basic Local Alignment Search Tool search of the drug-binding domain of each of the 190 targets with identifiable drug-binding domain against available human proteins. Approximately 51% of the targets have <6 human similarity proteins outside their respective family, and a further 19% of the targets have 6 to 10 similarity proteins. This finding seems to support the postulation that targets with fewer human similarity proteins outside their family tend to be more likely to be explored for drug development.

However, a smaller number of human similarity proteins outside the family of a target is not a necessary condition for finding successful drugs. It merely makes the tasks for finding successful drugs against these targets easier as the probability of unwanted interactions with human proteins outside the family is reduced. For targets with a higher number of similarity proteins, it is still possible to find agents that can specifically bind to a particular target and has no significant interactions with human proteins both inside and outside of the family to which the target belongs. This theory is sup-

aspet

Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

TABLE 5

Statistics for the number of similarity human similarity proteins of successful targets that are outside the protein family of the respective target

A total of 190 successful targets with identifiable drug-binding domain are included in the analysis.

No. of Similarity Proteins	No. of Targets with This No. of Similarity Proteins	Targets with This No. of Similarity Proteins	Examples of Targets
		%	
0–5	97	51	5-Hydroxytryptamine 3 receptor, acetylcholinesterase, adenosine A2b receptor, ATP-sensitive K^+ channel
6–10	36	19	α-1D adrenergic receptor, dopamine D1 receptor, histamine H1 receptor, HIV-1 reverse transcriptase, muscarinic acetylcholine receptor M1
11–20	15	10	Coagulation factor VIIIa, epidermal growth factor receptor, HIV-1 protease, insulin receptor, κ-type opioid receptor
21–40	26	11	Androgen receptor, estrogen receptor, γ-aminobutyric acid B receptor, peroxisome proliferator activated receptor α
41–80	9	5	Lutropin-choriogonadotropic hormone receptor, sulfonylurea receptor 2B, thrombin, urokinase-type plasminogen activator
>80	7	4	Human keratin, receptor-type protein-tyrosine phosphatase S, thyroid peroxidase, Toll-like receptor 7

ported by the existence of several successful targets with more than 80 human proteins outside the family of the respective target.

F. Associated Pathways

Association of a target with a fewer number of pathways tends to reduce the chance of unwanted interference with other processes, and these targets are more likely to be successfully discovered and explored for generating a higher number of clinical drugs. This theory can be tested by studying the 132 successful targets that have available pathway information in the KEGG database (Kanehisa, 2002). Table 6 gives the statistics for the number of pathways in which these targets are involved. There are 64 (49%), 36 (27%), and 15 (11%) targets found to be associated with 1, 2, and 3 pathways, respectively. Each of the remaining targets is involved in >3 pathways. Some indications about the success rate of the exploration of the targets in each group can be probed by looking at the highest number of clinical drugs directed at any single target in each group. From Table 6, it is found that the groups of targets associated with <3 pathways have a substantially higher number of clinical drugs than those associated with >3 pathways, which seems to support the hypothesis that targets associated with a fewer number of pathways tend to be more successfully explored.

TABLE 6

Statistics for the number of pathways of similarity proteins of successful targets

A total of 132 successful targets that have available pathway information in the KEGG database are included in the analysis.

No. of Pathways	No. of Targets with Similarity Proteins in This No. of Pathways	Targets with Similarity Proteins in This No. of Pathways	Highest No. of Drugs for a Target
		%	
1	64	49	8
2	36	27	8
3	15	11	5
4	3	2	1
5	4	3	2
6	3	2	3
8	4	3	4
9	1	1	2
>10	2	2	1

G. Tissue Distribution

Some therapeutic targets have been chosen primarily because of their high and selective expression in specific tissues, despite the existence of unfavorable conditions such as high expression abundance (Debouck and Metcalf, 2000). Efforts have been made to more broadly use tissue-selective strategies (Blagosklonny, 2003). This raises an interest for studying tissue distribution patterns of the successful targets to find out to what extent tissue specificity has already been used in existing therapeutics. There are 158 successful targets with available information about tissue distribution in human. Their tissue distribution patterns are given in Table 7. Of these targets 53% are distributed in less than three tissues, which seems to indicate that tissue selectivity may be an important factor for the successful exploration of some of these targets.

In estimating the number of affiliated tissues of each target, relevant data from the Swissprot 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 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 1 to 8 tissues, respectively, which are roughly similar to those derived from Swissprot data, 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.

PHARM REV
PHARMACOLOGICAL REVIEWS
aspet
Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

TABLE 7
Statistics for the human tissue distribution pattern of successful targets
A total of 158 targets that have human tissue distribution information are included in the analysis.

No. of Tissues	No. of Targets Predominantly Distributed in This No. of Tissues	Targets Predominantly Distributed in This No. of Tissues	Examples of Targets
		%	
1	45	28	D3 dopamine receptor, potassium-transporting ATPase α chain 1, solute carrier family 12 member 3
2	39	25	Lutropin-choriogonadotropic hormone receptor, potassium voltage-gated channel subfamily H member 2, ryanodine receptor 1
3	23	15	Acetyl-CoA carboxylase 2, fatty acid synthase, pregnane X receptor
4	12	8	Inducible nitric oxide synthase, peroxisome proliferator activated receptor α
5	5	3	Catechol-*O*-methyl-transferase, CGMP-specific 3′,5′-cyclic phosphodiesterase
6	2	1	Fibroblast growth factor 2, fatty acid hydrolase
7	3	2	Aldehyde oxidase, Toll-like receptor 7
8	6	4	Peroxisome proliferator-activated receptor γ, P2Y purinoceptor 12, insulin receptor
9	1	1	Voltage-gated calcium channel
10	1	1	Inhibitor of nuclear factor κB kinase
Many tissues	21	12	Adenosine deaminase, Na^+-K^+-2Cl cotransporter, receptor-type protein-tyrosine phosphatase S

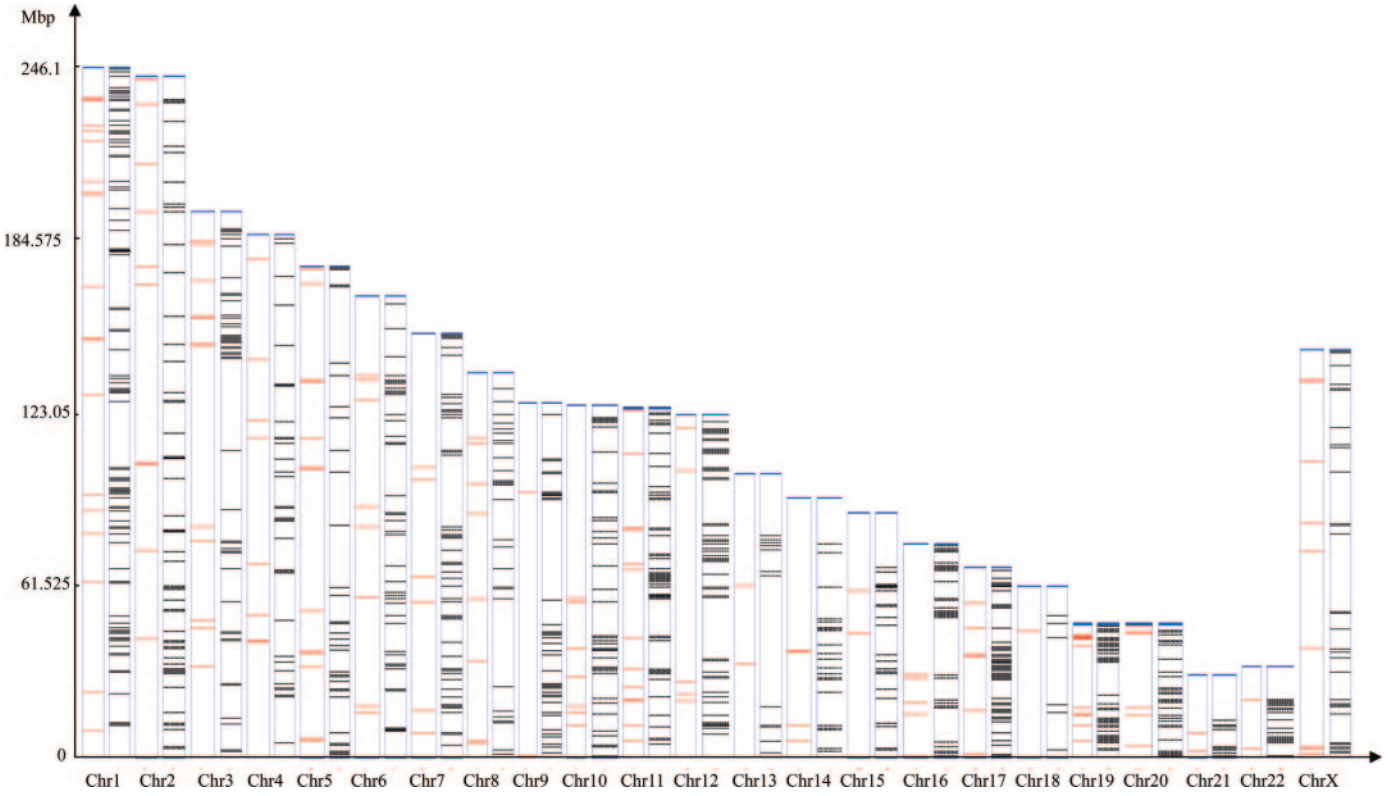

FIG. 4. Distribution patterns of human therapeutic targets in 23 human chromosomes. These patterns are arranged from the left to right for chromosome 1, 2, . . . , 22, and X, respectively. For each chromosome, the pattern of successful targets is given on the left and that of research targets is given on the right. The location of each target in a chromosome is marked by a line, with a red line for a successful target and a black line for a research target.

H. Chromosome Locations

Members of a protein family are known to be distributed in specific clusters in genomes (Yanase et al., 2004; Zhang et al., 2004). Functionally similar but nonhomologous proteins have also been found to be located at specific regions of genomes, which allows these proteins to be similarly regulated (Feldman and Segal, 2004). A large percentage of therapeutic targets are from multiple members of specific protein families or nonhomologous proteins of similar function of other targets. It is thus of interest to study the distribution pattern of existing human targets in the human genome to determine whether there is any level of clustering of these targets in specific regions of the chromosomes.

PHARM REV
PHARMACOLOGICAL REVIEWS
aspet
Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

PHARM REV

PHARMACOLOGICAL REVIEWS

Distribution patterns of the human successful and research targets in each of the 23 chromosomes are given in Fig. 4. These patterns are arranged from the left to right for chromosome 1, 2 . . . 22, and X, respectively. For each chromosome, the pattern of successful targets is given on the left and that of research targets is given on the right. The location of each target in a chromosome is marked by a line, with a red line for a successful target and a black line for a research target. It seems that a substantial percentage of research targets are more densely distributed in or near the regions of higher concentration of successful targets. Thus, there seems to be some level of clustering of targets at specific regions where successful targets are located.

The chromosomes with larger numbers of targets are chromosome 1, 3, 11, and 17. Chromosomes 2, 7, 12, and 19 also contain relatively higher concentrations of targets. Distribution of targets in certain chromosomes seems to be less even than that in other chromosomes. In particular, there are specific sections of larger numbers of targets in chromosomes 1, 3, 5, 9, 12, 17, and 19. Targets in the rest of chromosomes are relatively evenly distributed.

V. Can Druggable Proteins Be Predicted from Their Sequence?

Advances in high-throughput gene sequencing have led to rapid identification of thousands of novel genes, mostly without a known function. For the pharmaceutical industry, the sequencing of the human genome and the genomes of disease species proved to be both a blessing and a curse. Where potential targets were once hard to come by, the industry is now awash with them. This has left drug discovery communities with the difficult task of shifting through the gene data to find novel targets (Debouck and Metcalf, 2000; Smith, 2003). Genomics approaches such as large-scale gene expression analysis, functional screens in model organisms, genome scans for disease susceptibility genes, and the search for new members of effective drug target classes have enabled the finding of countless candidates for many diseases (Sanseau, 2001; Desany and Zhang, 2004; Dohrmann, 2004). Determination of which of these candidates are druggable still relies on experimental studies. Methods that facilitate the identification of druggable proteins from these candidates or directly from genomes are thus particularly useful for target identification.

Investigations of the features of known therapeutic targets from earlier studies (Hopkins and Groom, 2002; Hardy and Peet, 2004) and in the previous sections suggest that targets have certain common characteristics, which may be used as the basis for deriving rules for identification of druggable proteins from their sequence in a manner to that of rule-based methods (such as the rule of five) for predicting "drug-like" compounds from their structures (Lipinski et al., 2001; Baurin et al., 2004). Statistical learning methods have also been successfully applied for developing tools for predicting drug-like molecules from their structures on the basis that they have common structural and physicochemical features (Byvatov et al., 2003; Zernov et al., 2003). It is expected that these statistical learning methods are equally applicable for predicting druggable proteins from their sequences on the basis that druggable proteins share common characteristics.

A. "Rules" for Guiding the Search for Druggable Proteins

Based on the characteristics of therapeutic targets described in earlier studies (Hopkins and Groom, 2002; Hardy and Peet, 2004) and in the previous sections, it seems that the following rules can be proposed for guiding the search of druggable proteins:

- The protein is from one of the target-representing protein families. The number of these families is currently estimated to be no more than 800. So far, 88 confirmed families (each containing at least one successful target) and 357 likely families (each containing at least one research target) have been found.
- Sequence variation between the drug-binding domain of a protein and those of the other human members of its protein family needs to allow a sufficient degree of differential binding of a rule-of-five molecule to the common binding site.
- The protein preferably has <6 human similarity proteins outside its family. Although existence of a higher number of human similarity proteins does not rule out a protein as druggable, it generally increases the chance of unwanted interferences and thus the level of difficulty for finding viable drugs (51% of the successful targets with identifiable drug-binding domain have <6 human similarity proteins).
- The protein is preferably involved in no more than two pathways in humans. Although association with a higher number of human pathways does not rule out a protein as druggable, it generally increases the chance of unwanted interferences with other human processes and thus the level of difficulty for finding a viable target (76% of the successful targets with pathway information are associated with no more than two pathways).
- For organ- or tissue-specific diseases, the protein is preferably distributed in no more than two tissues in humans. Although distribution in a higher number of tissues does not rule it out a protein as druggable, it generally increases the chance of unwanted interferences with other tissues and thus the level of difficulty for finding a viable target (53% of the successful targets with tissue distribution

Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

PHARM REV

PHARMACOLOGICAL REVIEWS

information are distributed in no more than two tissues).

B. Prediction of Druggable Proteins by a Statistical Learning Method

New targets might not bear sequence similarity to known targets or known proteins. Consequently, a straightforward sequence similarity search against effective drug target classes (Sanseau, 2001) and known disease genes (Desany and Zhang, 2004) may not always be useful for identification of novel targets. Although targets seem to have common characteristics that are reflected in their sequences, they are from a diverse range of different families and structural folds. Thus, methods that do not rely on sequence and structure similarity are needed for facilitating the prediction of druggable proteins directly from their sequences.

Statistical learning methods, such as support vector machines and neural networks, have emerged in the last few years as attractive methods for the prediction of protein functional classes (des Jardins et al., 1997; Jensen et al., 2002; Karchin et al., 2002; Cai et al., 2003a, 2004; Bhasin and Raghava, 2004; Han et al., 2004) and structural classes (Zhou and Assa-Munt, 2001; Cai et al., 2003b) without the use of sequence similarity. These classes contain proteins of diverse functions and structures. Examples of some of these classes are RNA-binding proteins, EC2.7 transferases of phosphorus-containing groups, EC3.4 peptidases, and TC1.A α-type channels. It seems that the prediction accuracy of these methods has reached a level sufficient for facilitating the prediction of the functional and structural classes of proteins. For instance, the overall accuracy of support vector machine prediction of the functional family of 13,891 enzymes and 447 RNA-binding proteins is 86 and 98%, respectively. Thus, it is of interest to investigate the feasibility of using statistical learning methods for predicting druggable proteins from their sequences.

Currently, the support vector machine (SVM) method seems to be the most accurate statistical learning method for protein predictions (Karchin et al., 2002; Cai et al., 2003a,b, 2004; Bhasin and Raghava, 2004; Han et al., 2004). Therefore, only this method is investigated here. SVM is based on the structural risk minimization principle from statistical learning theory (Burges, 1998). Known proteins are divided into druggable and nondruggable classes; each of these proteins is represented by their sequence-derived physicochemical features (Cai et al., 2003a). These features are then used by the SVM to construct a hyperplane in a higher dimensional hyperspace that maximally separates druggable proteins and nondruggable ones. By projecting the sequence of a new protein onto this hyperspace, it can be determined whether this protein is druggable from its location with respect to the hyperplane. It is a druggable protein if it is located on the side of druggable class.

The accuracy of SVM depends on the diversity of the protein samples used for finding the hyperspace and its hyperplane, the quality of the representation of protein features, and the efficiency of the SVM algorithm. To a certain extent, no sequence and structural similarity are required per se. Thus, SVM is an attractive approach for facilitating the prediction of classes of proteins of diverse sequences and structures, and thus the prediction of druggable proteins.

A total of 1368 sequence entries of 1535 successful and research targets are used to construct the druggable class, and 12,956 representative proteins from 6856 Pfam (Bateman et al., 2004) protein families (with all of the known target-representing families excluded from these families) are used to construct the nondruggable class. Multiple sequence entries of some viral protein targets are included in the druggable class because of significant sequence variations across strains. Proteins in each class are randomly divided into five subsets of approximately equal size. Four subsets are selected as the training set and the fifth as the testing set. This process is repeated five times such that every subset is selected as a testing set once.

The average prediction accuracy from this 5-fold cross validation study is 69.8% for druggable proteins and 99.3% for nondruggable proteins. The accuracy for nondruggable proteins is comparable but that of druggable proteins is somehow lower than those of protein functional and structural families (Karchin et al., 2002; Cai et al., 2003a,b, 2004; Bhasin and Raghava, 2004; Han et al., 2004), which is expected because of the significantly higher level of sequence and structural diversity of therapeutic targets. Nonetheless, these accuracies are at a meaningful level for facilitating the prediction of druggable proteins.

To test its potential for practical applications, the constructed SVM prediction system is used to scan the human genome for identifying potential druggable proteins that are not in the training and testing sets. A total of 1102 human proteins are predicted to be druggable, which includes 153 G-protein coupled receptors, 65 other receptors, 333 enzymes, and 56 channels. These numbers are within the estimated numbers of druggable proteins and therapeutic targets in the human genome. For instance, the total number of druggable proteins and actual targets in the human genome has been estimated to be ~3000 and ~1500, respectively (Hopkins and Groom, 2002), and the total number of 400 G-protein coupled receptors has been suggested to be potential targets (Wise et al., 2002).

This SVM prediction system is further tested by comparison of its predicted druggable proteins in an HIV genome with known HIV targets. This genome is selected because it is one of the most extensively explored genomes for finding therapeutic targets, and it is highly likely that all of the potential targets in this genome have been identified (Turpin, 2003). The National Cen-

Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

TABLE 8
Comparison of the known HIV-1 protein targets and the SVM-predicted druggable proteins in the NCBI HIV-1 genome entry NC_001802

Protein in HIV-1 Genome	NCBI Protein Accession No.	Target Status	SVM Prediction Status
Gag-Pol	NP_057849.4		
Gag-Pol Transframe peptide	NP_787043.1		
Pol	NP_789740.1		
Protease	NP_705926.1	Successful target	Druggable
Reverse transcriptase	NP_705927.1	Successful target	Druggable
Reverse transcriptase p51 subunit	NP_789739.1		
Integrase	NP_705928.1	Research target	Druggable
Gag	NP_057850.1	Research target	Druggable
Matrix	NP_579876.2		
Capsid	NP_579880.1		
p2	NP_579882.1		
Nucleocapsid	NP_579881.1	Research target	Druggable
p1	NP_787042.1		
p6	NP_579883.1		
Vif	NP_057851.1	Research target	Druggable
Vpr	NP_057852.2		
Tat	NP_057853.1	Successful target	
Rev	NP_057854.1		
Vpu	NP_057855.1		
Envelope surface glycoprotein gp160	NP_057856.1	Research target	Druggable
Envelope signal peptide	NP_579893.2		
Envelope surface glycoprotein gp120	NP_579894.2	Research target	
Envelope transmembrane glycoprotein gp41	NP_579895.1	Successful target	
Nef	NP_057857.2	Research target	Druggable

ter for Biotechnology Information (Wheeler et al., 2004) HIV-1 genome entry NC_001802, with none of its encoded protein sequences used in the SVM training and testing sets, is used for this test, and the results are given in Table 8. There are four successful and seven research targets in the HIV-1 genome. The SVM is able to predict two successful and six research targets as druggable. Overall, 72% of the known successful and research targets and 100% of the nontargets are correctly predicted. This prediction accuracy is consistently similar to that of the 5-fold cross-validation study.

These three tests seem to indicate that the SVM has some potential for facilitating the identification of druggable proteins from genomic data. The prediction accuracy for druggable proteins needs to be improved. One reason for the lower accuracy of druggable proteins is the large imbalance between the number of druggable and nondruggable proteins. Such a large imbalance is known to affect the accuracy of a SVM prediction system and methods for solving these problems are being developed (Bhasin and Raghava, 2004).

REFERENCES

Ahima RS and Osei SY (2001) Molecular regulation of eating behavior: new insights and prospects for therapeutic strategies. *Trends Mol Med* **7:**205–213.

Andreeva A, Howorth D, Brenner SE, Hubbard TJ, Chothia C, and Murzin AG (2004) SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Res* **32:**D226–D229.

Baker MD and Wood JN (2001) Involvement of Na^+ channels in pain pathways. *Trends Pharmacol Sci* **22:**27–31.

Barnette MS, Bartus JO, Burman M, Christensen SB, Cieslinski LB, Esser KM, Prabhakar US, Rush JA, and Torphy TJ (1996) Association of the anti-inflammatory activity of phosphodiesterase 4 (PDE4) inhibitors with either inhibition of PDE4 catalytic activity or competition for [^{3}H]rolipram binding. *Biochem Pharmacol* **51:**949–956.

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, et al. (2004) The Pfam protein families database. *Nucleic Acids Res* **32:**D138–D141.

Baurin N, Baker R, Richardson C, Chen I, Foloppe N, Potter A, Jordan A, Roughley S, Parratt M, Greaney P, et al. (2004) Drug-like annotation and duplicate analysis of a 23-supplier chemical database totalling 2.7 million compounds. *J Chem Inf Comput Sci* **44:**643–651.

Bein K and Simons M (2001) inventors, Beth Israel Deaconess Medical Center, assignee. Peptide inhibitor of MMP activity and angiogenesis. U.S. patent 6,667,388. 2001 Jan 22.

Benke D, Michel C, and Mohler H (1997) $GABA_A$ receptors containing the α4-subunit: prevalence, distribution, pharmacology and subunit architecture in situ. *J Neurochem* **69:**806–814.

Best JD and Jenkins AJ (2001) Novel agents for managing dyslipidaemia. *Expert Opin Investig Drugs* **10:**1901–1911.

Bhasin M and Raghava GP (2004) Classification of nuclear receptors based on amino acid composition and dipeptide composition. *J Biol Chem* **279:**23262–23266.

Bicknell KA, Surry EL, and Brooks G (2003) Targeting the cell cycle machinery for the treatment of cardiovascular disease. *J Pharm Pharmacol* **55:**571–591.

Blagosklonny MV (2003) Tissue-selective therapy of cancer. *Br J Cancer* **89:**1147–1151.

Blake SM and Swift BA (2004) What next for rheumatoid arthritis therapy? *Curr Opin Pharmacol* **4:**276–280.

Branch CL, Johnson CN, Stemp G, and Thewlis K (2002) inventors, SmithKline Beecham p.l.c., assignee. Piperidines for use as orexin receptor antagonists. U.S. patent 6,677,354. 2002 Dec 16.

Bray GA and Tartaglia LA (2000) Medicinal strategies in the treatment of obesity. *Nature (Lond)* **404:**672–677.

Buolamwini JK (1999) Novel anticancer drug discovery. *Curr Opin Chem Biol* **3:**500–509.

Burges C (1998) A tutorial on support vector machine for pattern recognition. *Data Mining Knowl Discov* **2:**121–167.

Bush K and Macielag M (2000) New approaches in the treatment of bacterial infections. *Curr Opin Chem Biol* **4:**433–439.

Byvatov E, Fechner U, Sadowski J, and Schneider G (2003) Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *J Chem Inf Comput Sci* **43:**1882–1889.

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

Cai CZ, Han LY, Ji ZL, and Chen YZ (2004) Enzyme family classification by support vector machines. *Proteins* **55:**66–76.

Cai YD, Liu XJ, Xu XB, and Chou KC (2003b) Support vector machines for prediction of protein domain structural class. *J Theor Biol* **221:**115–120.

Campfield LA, Smith FJ, and Burn P (1998) Strategies and potential molecular targets for obesity treatment. *Science (Wash DC)* **280:**1383–1387.

Chaix-Couturier C, Holtzer C, Phillips KA, Durand-Zaleski I, and Stansell J (2000) HIV-1 drug resistance genotyping. a review of clinical and economic issues. *Pharmacoeconomics* **18:**425–433.

Chantry D (2003) G protein-coupled receptors: from ligand identification to drug targets: 14–16 October 2002, San Diego, CA, USA. *Expert Opin Emerg Drugs* **8:**273–276.

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

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

Chong PH and Bachenheimer BS (2000) Current, new and future treatments in dyslipidaemia and atherosclerosis. *Drugs* **60:**55–93.

Clapham JC, Arch JR, and Tadayyon M (2001) Anti-obesity drugs: a critical review of current therapies and future opportunities. *Pharmacol Ther* **89:**81–121.

Collins JL (2004) Therapeutic opportunities for liver X receptor modulators. *Curr Opin Drug Discov Devel* **7:**692–702.

Cotsarelis G and Millar SE (2001) Towards a molecular understanding of hair loss and its treatment. *Trends Mol Med* **7:**293–301.

PHARM REV
PHARMACOLOGICAL REVIEWS
aspet
Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

Darnell JE Jr (2002) Transcription factors as targets for cancer therapy. *Nat Rev Cancer* **2:**740–749.
Debouck C and Metcalf B (2000) The impact of genomics on drug discovery. *Annu Rev Pharmacol Toxicol* **40:**193–207.
De Clercq E (2001) 2001 ASPET Otto Krayer Award Lecture: Molecular targets for antiviral agents. *J Pharmacol Exp Ther* **297:**1–10.
Desany B and Zhang Z (2004) Bioinformatics and cancer target discovery. *Drug Discov Today* **9:**795–802.
des Jardins M, Karp PD, Krummenacker M, Lee TJ, and Ouzounis CA (1997) Prediction of enzyme classification from protein sequence without the use of sequence similarity. *Proc Int Conf Intell Syst Mol Biol* **5:**92–99.
de Souza CJ and Burkey BF (2001) β3-adrenoceptor agonists as anti-diabetic and anti-obesity drugs in humans. *Curr Pharm Des* **7:**1433–1449.
Docherty AJ, O'Connell J, Crabbe T, Angal S, and Murphy G (1992) The matrix metalloproteinases and their natural inhibitors: prospects for treating degenerative tissue diseases. *Trends Biotechnol* **10:**200–207.
Dohrmann CE (2004) Target discovery in metabolic disease. *Drug Discov Today* **9:**785–794.
Dove A (1999) Proteomics: translating genomics into products? *Nat Biotechnol* **17:**233–236.
Drews J (1997a) Proceedings of the Roche Symposium "The Genetic Basis of Human Disease," in *Human Disease—From Genetic Causes to Biochemical Effects* (Drews J and Ryser S eds) pp 5–9, Blackwell, Berlin.
Drews J (1997b) Strategic choices facing the pharmaceutical industry: a case for innovation. *Drug Discov Today.* **2:**72–78.
Drews J (2000) Drug discovery: a historical perspective. *Science (Wash DC)* **287:**1960–1964.
Dubowchik GM and Walker MA (1999) Receptor-mediated and enzyme-dependent targeting of cytotoxic anticancer drugs. *Pharmacol Ther* **83:**67–123.
Ducruet AP, Vogt A, Wipf P, and Lazo JS (2005) Dual specificity protein phosphatases: therapeutic targets for cancer and Alzheimer's disease. *Annu Rev Pharmacol Toxicol* 45:725–750.
Eggert M, Kluter A, Zettl UK, and Neeck G (2004) Transcription factors in autoimmune diseases. *Curr Pharm Des* **10:**2787–2796.
Elsayed YA and Sausville EA (2001) Selected novel anticancer treatments targeting cell signaling proteins. *Oncologist* **6:**517–537.
Emilien G and Maloteaux JM (1998) Current therapeutic uses and potential of β-adrenoceptor agonists and antagonists. *Eur J Clin Pharmacol* **53:**389–404.
Evans WE and Johnson JA (2001) Pharmacogenomics: the inherited basis for interindividual differences in drug response. *Annu Rev Genomics Hum Genet* **2:**9–39.
Feldman M and Segal G (2004) A specific genomic location within the icm/dot pathogenesis region of different *Legionella* species encodes functionally similar but nonhomologous virulence proteins. *Infect Immun* **72:**4503–4511.
George RA and Heringa J (2002) Protein domain identification and improved sequence similarity searching using P51-BLAST. *Proteins* **48:**672–681.
Gerstein M (1998) Measurement of the effectiveness of transitive sequence comparison, through a third "intermediate" sequence. *Bioinformatics* **14:**707–714.
Gibbs JB, Pompliano DL, Mosser SD, Rands E, Lingham RB, Singh SB, Scolnick EM, Kohl NE, and Oliff A (1993) Selective inhibition of farnesyl-protein transferase blocks ras processing in vivo. *J Biol Chem* **268:**7617–7620.
Gong L, Grupe A, and Peltz GA (2002) inventors, Syntex LLC, assignee. 3-Indolyl-4-phenyl-1*H*-pyrrole-2,5-dione derivatives as inhibitors of glycogen synthase kinase-3 beta. U.S. patent 6,479,490. 2001 Jul 27.
Greene JM and Rosen CA (2001) inventors, Human Genome Sciences, Inc., assignee. Human tissue inhibitor of metalloproteinase-4. U.S. patent 6,544,761. 2001 Jul 11.
Greenfeder S and Anthes JC (2002) New asthma targets: recent clinical and preclinical advances. *Curr Opin Chem Biol* **6:**526–533.
Gronemeyer H, Gustafsson JA, and Laudet V (2004) Principles for modulation of the nuclear receptor superfamily. *Nat Rev Drug Discov* **3:**950–964.
Han LY, Cai CZ, Lo SL, Chung MC, and Chen YZ (2004) Prediction of RNA-binding proteins from primary sequence by a support vector machine approach. *RNA* **10:**355–368.
Hardy LW and Peet NP (2004) The multiple orthogonal tools approach to define molecular causation in the validation of druggable targets. *Drug Discov Today* **9:**117–126.
Helmuth L (2002) New therapies. New Alzheimer's treatments that may ease the mind. *Science (Wash DC)* **297:**1260–1262.
Hoffman EP and Dressman D (2001) Molecular pathophysiology and targeted therapeutics for muscular dystrophy. *Trends Pharmacol Sci* **22:**465–470.
Hopkins AL and Groom CR (2002) The druggable genome. *Nat Rev Drug Discov* **1:**727–730.
Hossain MA and Ghannoum MA (2000) New investigational antifungal agents for treating invasive fungal infections. *Expert Opin Investig Drugs* **9:**1797–1813.
Howe R, Rao BS, Holloway BR, and Stribling D (1992) Selective β3-adrenergic agonists of brown adipose tissue and thermogenesis. 1. [4-[2-[(2-Hydroxy-3-phenoxypropyl)amino]ethoxy]phenoxy]acetates. *J Med Chem* **35:**1751–1759.
Ilag LL, Ng JH, Beste G, and Henning SW (2002) Emerging high-throughput drug target validation technologies. *Drug Discov Today* **7:**S136–142.
Irizarry MC and Hyman BT (2001) Alzheimer disease therapeutics. *J Neuropathol Exp Neurol* **60:**923–928.
Jabbour E, Kantarjian H, and Cortes J (2004) Clinical activity of farnesyl transferase inhibitors in hematologic malignancies: possible mechanisms of action. *Leuk Lymphoma* **45:**2187–2195.
Jensen LJ, Gupta R, Blom N, Devos D, Tamames J, Kesmir C, Nielsen H, Staerfeldt HH, Rapacki K, Workman C, et al. (2002) Prediction of human protein function from post-translational modifications and localization features. *J Mol Biol* **319:**1257–1265.
Kanehisa M (2002) The KEGG database. *Novartis Found Symp* **247:**91–101; discussion 101–103, 119–128, 244–252.
Karchin R, Karplus K, and Haussler D (2002) Classifying G-protein coupled receptors with support vector machines. *Bioinformatics* **18:**147–159.
Karp JE and Lancet JE (2004) Farnesyltransferase inhibitors (FTIs) in myeloid malignancies. *Ann Hematol* **83 (Suppl 1):**S87–S88.
Kennedy BP and Ramachandran C (2000) Protein tyrosine phosphatase-1B in diabetes. *Biochem Pharmacol* **60:**877–883.
Kennedy T (1997) Managing the drug discovery/development interface. *Drug Discov Today* **2:**436–444.
Kobayashi H and Stringer MD (2003) Biliary atresia. *Semin Neonatol* **8:**383–391.
Koehl P and Levitt M (2002) Sequence variations within protein families are linearly related to structural variations. *J Mol Biol* **323:**551–562.
Kumar S, Blake SM, and Emery JG (2001) Intracellular signaling pathways as a target for the treatment of rheumatoid arthritis. *Curr Opin Pharmacol* **1:**307–313.
Lark MW and Morrison KE (2002) Musculoskeletal diseases: novel targets for therapeutic intervention. *Curr Opin Pharmacol* **2:**287–290.
Leurs R, Blandina P, Tedford C, and Timmerman H (1998) Therapeutic potential of histamine H3 receptor agonists and antagonists. *Trends Pharmacol Sci* **19:**177–183.
Lewis AJ and Manning AM (1999) New targets for anti-inflammatory drugs. *Curr Opin Chem Biol* **3:**489–494.
Lin PC, Bhatnagar KP, Nettleton GS, and Nakajima ST (2002) Female genital anomalies affecting reproduction. *Fertil Steril* **78:**899–915.
Lipinski CA, Lombardo F, Dominy BW, and Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* **46:**3–26.
Lizotte-Waniewski M, Tawe W, Guiliano DB, Lu W, Liu J, Williams SA, and Lustigman S (2000) Identification of potential vaccine and drug target candidates by expressed sequence tag analysis and immunoscreening of *Onchocerca volvulus* larval cDNA libraries. *Infect Immun* **68:**3491–3501.
Luchner A and Schunkert H (2004) Interactions between the sympathetic nervous system and the cardiac natriuretic peptide system. *Cardiovasc Res* **63:**443–449.
Lyon MA, Ducruet AP, Wipf P, and Lazo JS (2002) Dual-specificity phosphatases as targets for antineoplastic agents. *Nat Rev Drug Discov* **1:**961–976.
Macdonald IA (2000) Obesity: are we any closer to identifying causes and effective treatments? *Trends Pharmacol Sci* **21:**334–336.
Makriyannis A, Lin S, and Hill WA (2002) inventors, University of Connecticut, assignee. Anandamide amidase inhibitors as analgesic agents. U.S. patent 6,579,900. 2002 Feb 6.
Matter A (2001) Tumor angiogenesis as a therapeutic target. *Drug Discov Today* **6:**1005–1024.
Miller MD and Hazuda DJ (2001) New antiretroviral agents: looking beyond protease and reverse transcriptase. *Curr Opin Microbiol* **4:**535–539.
Nicholls H (2003) Improving drug response with pharmacogenomics. *Drug Discov Today* **8:**281–282.
Ohlstein EH, Ruffolo RR Jr, and Elliott JD (2000) Drug discovery in the next millennium. *Annu Rev Pharmacol Toxicol* **40:**177–191.
Olliaro PL and Yuthavong Y (1999) An overview of chemotherapeutic targets for antimalarial drug discovery. *Pharmacol Ther* **81:**91–110.
Peltonen L and McKusick VA (2001) Genomics and medicine. Dissecting human disease in the postgenomic era. *Science (Wash DC)* **291:**1224–1229.
Persidis A (1999) Cardiovascular disease drug discovery. *Nat Biotechnol* **17:**930–931.
Picard JA and Wilson MW (2002) inventors, Warner-Lambert Company, assignee. Benzo thiadiazine matrix metalloproteinase inhibitors. U.S. patent 6,656,932. 2002 Feb 13.
Poulos TL (1988) Cytochrome P450: molecular architecture, mechanism and prospects for rational inhibitor design. *Pharm Res (NY)* **5:**67–75.
Ramnath N and Creaven PJ (2004) Matrix metalloproteinase inhibitors. *Curr Oncol Rep* **6:**96–102.
Sali A (1998) 100,000 protein structures for the biologist. *Nat Struct Biol* **5:**1029–1032.
Sanseau P (2001) Impact of human genome sequencing for in silico target discovery. *Drug Discov Today* **6:**316–323.
Scheinfeld NS, Silverberg NB, Weinberg JM, and Nozad V (2004) The preauricular sinus: a review of its clinical presentation, treatment and associations. *Pediatr Dermatol* **21:**191–196.
Schwartz JC, Christiania R, Vargas F, Ganellin CR, Zhao L, Sanjeeda S, and Chen Y (2000) inventors, Institut National de la Sante et de la Recherche Medicale and Bioprojet, assignee. Tripeptidyl peptidase inhibitors. U.S. patent 6,335,360. 2000 Sep 18.
Scott MK, Lee DHS, Reitz AB, Ross TM, and Wang H-Y (2000) inventors, Ortho-McNeil Pharmaceutical, Inc., assignee. 1-4-dithiin and 1,4-dithiepin-1,1,4,4 tetroxide derivatives useful as antagonists of the human galanin receptor. U.S. patent 6,407,136. 2000 May 2.
Serrero G (2001) inventor, A & G Pharmaceutical, Inc., assignee. 88 kDa tumorigenic growth factor and antagonists. U.S. patent 6,670,183. 2001 Mar 21.
Smith C (2003) Drug target validation: hitting the target. *Nature (Lond)* **422:**341, 343, 345 passim.
Spina D (2003) Phosphodiesterase-4 inhibitors in the treatment of inflammatory lung disease. *Drugs* **63:**2575–2594.
Striessnig J, Grabner M, Mitterdorfer J, Hering S, Sinnegger MJ, and Glossmann H (1998) Structural basis of drug binding to L Ca^{2+} channels. *Trends Pharmacol Sci* **19:**108–115.
Sussman JL, Lin D, Jiang J, Manning NO, Prilusky J, Ritter O, and Abola EE (1998) Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. *Acta Crystallogr D Biol Crystallogr* **54:**1078–1084.
Teall MR (2001) inventor, Merck Sharp & Dohme Ltd., assignee. Gamma secretase inhibitors. U.S. patent 6,448,229. 2001 Jun 29.
Terstappen GC and Reggiani A (2001) In silico research in drug discovery. *Trends Pharmacol Sci* **22:**23–26.
Toda N (2003) Vasodilating β-adrenoceptor blockers as cardiovascular therapeutics. *Pharmacol Ther* **100:**215–234.

PHARM REV
PHARMACOLOGICAL REVIEWS
aspet
Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

Torphy TJ and Page C (2000) Phosphodiesterases: the journey towards therapeutics. *Trends Pharmacol Sci* **21:**157–159.
Turpin JA (2003) The next generation of HIV/AIDS drugs: novel and developmental antiHIV drugs and targets. *Expert Rev Anti Infect Ther* **1:**97–128.
van de Waterbeemd H and Gifford E (2003) ADMET in silico modelling: towards prediction paradise? *Nat Rev Drug Discov* **2:**192–204.
Vane JR, Bakhle YS, and Botting RM (1998) Cyclooxygenases 1 and 2. *Annu Rev Pharmacol Toxicol* **38:**97–120.
Wada CK (2004) The evolution of the matrix metalloproteinase inhibitor drug discovery program at Abbott Laboratories. *Curr Top Med Chem* **4:**1255–1267.
Wagman AS and Nuss JM (2001) Current therapies and emerging targets for the treatment of diabetes. *Curr Pharm Des* **7:**417–450.
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.
Wen YM, Lin X, and Ma ZM (2003) Exploiting new potential targets for anti-hepatitis B virus drugs. *Curr Drug Targets Infect Disord* **3:**241–246.
Wheeler DL, Church DM, Edgar R, Federhen S, Helmberg W, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, et al. (2004) Database resources of the National Center for Biotechnology Information: update. *Nucleic Acids Res* **32:**D35–D40.
Whythes MJ, Palmer MJ, Kemp MI, MacKenny MC, Maguire RJ, and Blake JJF (2000) inventors, Pfizer Inc., assignee. FKBP inhibitors, U.S. patent 6,495,549. 2000 Oct 30.
Widdicombe J and Lee LY (2001) Airway reflexes, autonomic function and cardiovascular responses. *Environ Health Perspect* **109 (Suppl 4):**579–584.
Windisch M, Hutter-Paier B, and Schreiner E (2002) Current drugs and future hopes in the treatment of Alzheimer's disease. *J Neural Transm Suppl* **62:**149–164.
Wise A, Gearing K, and Rees S (2002) Target validation of G-protein coupled receptors. *Drug Discov Today* **7:**235–246.
Wood TC and Pearson WR (1999) Evolution of protein sequences and structures. *J Mol Biol* **291:**977–995.
World Health Organization (1992) *International Statistical Classification of Diseases and Related Health Problems*, 3 p, World Health Organization, Geneva.
Yanase H, Sugino H, and Yagi T (2004) Genomic sequence and organization of the family of CNR/Pcdhα genes in rat. *Genomics* **83:**717–726.
Yu EW, McDermott G, Zgurskaya HI, Nikaido H, and Koshland DE Jr (2003) Structural basis of multiple drug-binding capacity of the AcrB multidrug efflux pump. *Science (Wash DC)* **300:**976–980.
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.
Zernov VV, Balakin KV, Ivaschenko AA, Savchuk NP, and Pletnev IV (2003) Drug discovery using support vector machines: the case studies of drug-likeness, agrochemical-likeness and enzyme inhibition predictions. *J Chem Inf Comput Sci* **43:**2048–2056.
Zhang Z, Burch PE, Cooney AJ, Lanz RB, Pereira FA, Wu J, Gibbs RA, Weinstock G, and Wheeler DA (2004) Genomic analysis of the nuclear receptor family: new insights into structure, regulation and evolution from the rat genome. *Genome Res* **14:**580–590.
Zhou GP and Assa-Munt N (2001) Some insights into protein structural class prediction. *Proteins* **44:**57–59.

PHARM REV
PHARMACOLOGICAL REVIEWS
aspet
Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012

PHARM REV

PHARMACOLOGICAL REVIEWS

Correction to "Therapeutic Targets: Progress of Their Exploration and Investigation of Their Characteristics"

During proofreading of the above article [Zheng CJ, Han LY, Yap CW, Ji ZL, Cao ZW, and Chen YZ (2006) *Pharmacol Rev* **58:**259–279], the authors entered the wrong text data for the percentages of targets and the numbers of similarity proteins, participating pathways, and affiliated tissues into the section "A. 'Rules' for Guiding the Search for Druggable Proteins." The corrected data, listed in bold, are reported below.

On page 275, under the third bullet point, "<6 human similarity proteins" should be replaced by "**<15** human similarity proteins"; "51% of the successful targets" should be replaced by "**78%** of the successful targets"; and "<6 human similarity proteins" should be replaced by "**<15** human similarity proteins."

In the first and second sentences under the fourth bullet point, "no more than two pathways" should be replaced by "no more than **three** pathways," and "(76% of the successful targets with pathway information are associated with no more than two pathways)" should be replaced by "(**87%** of the successful targets with pathway information are associated with no more than **three** pathways)," respectively.

In the first sentence under the fifth bullet point, "no more than two tissues in humans" should be replaced by "no more than **five** tissues in humans."

On pages 275 and 276, again under the fifth bullet point, "(53% of the successful targets with tissue distribution information are distributed in no more than two tissues)" should be replaced by "(**79%** of the successful targets with tissue distribution information are distributed in no more than **five** tissues)."

The online version of this article has been corrected in departure from print.

The authors regret these errors and apologize for any confusion they may have caused.

Downloaded from pharmrev.aspetjournals.org by guest on June 15, 2012