Knowledge-Based, Central Nervous System (CNS) Lead Selection and Lead Optimization for CNS Drug Discovery

Arup K. Ghose,* Torsten Herbertz, Robert L. Hudkins, Bruce D. Dorsey, and John P. Mallamo

Department of Chemistry, Discovery Research, Cephalon, Inc., 145 Brandywine Parkway, West Chester, Pennsylvania 19380, United States

S *Supporting Information*

ABSTRACT: The central nervous system (CNS) is the major area that is affected by aging. Alzheimer's disease (AD), Parkinson's disease (PD), brain cancer, and stroke are the CNS diseases that will cost trillions of dollars for their treatment. Achievement of appropriate blood–brain barrier (BBB) penetration is often considered a significant hurdle in the CNS drug discovery process. On the other hand, BBB penetration may be a liability for many of the non-CNS drug targets, and a clear understanding of the physicochemical and structural differences between CNS and non-CNS drugs may assist both research areas. Because of the numerous and challenging issues in CNS drug

discovery and the low success rates, pharmaceutical companies are beginning to deprioritize their drug discovery efforts in the CNS arena. Prompted by these challenges and to aid in the design of high-quality, efficacious CNS compounds, we analyzed the physicochemical property and the chemical structural profiles of 317 CNS and 626 non-CNS oral drugs. The conclusions derived provide an ideal property profile for lead selection and the property modification strategy during the lead optimization process. A list of substructural units that may be useful for CNS drug design was also provided here. A classification tree was also developed to differentiate between CNS drugs and non-CNS oral drugs. The combined analysis provided the following guidelines for designing high-quality CNS drugs: (i) topological molecular polar surface area of <76 Å^2 (25–60 Å^2), (ii) at least one (one or two, including one aliphatic amine) nitrogen, (iii) fewer than seven (two to four) linear chains outside of rings, (iv) fewer than three (zero or one) polar hydrogen atoms, (v) volume of 740–970 Å^3, (vi) solvent accessible surface area of 460–580 Å^2, and (vii) positive QikProp parameter CNS. The ranges within parentheses may be used during lead optimization. One violation to this proposed profile may be acceptable. The chemoinformatics approaches for graphically analyzing multiple properties efficiently are presented.

KEYWORDS: *Central nervous system, CNS, CNS drugs, non-CNS drugs, polar surface area, aliphatic amine, linear chains, polar hydrogen, molecular volume, solvent accessible surface area, chemical substructures, physicochemical property profile, chemoinformatics*
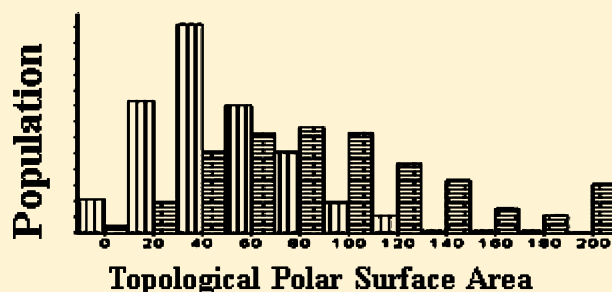
Statistics indicate that by the year 2020, the United States will have more than 20% of its population older than the age of 65.[1] It is estimated that central nervous system (CNS) diseases affected by aging, such as Alzheimer's disease (AD), Parkinson's disease (PD), brain cancer, and stroke, will cost in the trillions of dollars for their treatment. The health care expenditures will increase from ~15% of gross domestic product (GDP) today ($14.6 trillion) to ~29% in 2040[2] (http://www.nber.org/aginghealth/2008no4/w14361.html), and therefore, access to high-quality, effective CNS drugs remains an essential responsibility of the pharmaceutical industry. The objective of this study is to improve the understanding of differences between CNS and non-CNS oral drugs in terms of the large number of physicochemical properties that can be calculated using common molecular modeling packages. However, for such an objective, it is very important that the data set should be chosen carefully. Many non-CNS drugs cross the blood–brain barrier (BBB) at

sufficient concentrations to produce CNS-related side effects, such as drowsiness, headache, etc., and therefore were not included in this type of analysis. Establishing a so-called "CNS drug" may be challenging in some cases. The nervous system[3] is responsible for every activity taking place in all parts of the body of every bilateral organism. The nervous system (Figure 1) can be divided into two parts: (i) the central nervous system and (ii) the peripheral nervous system (PNS). The CNS is comprised of the brain and the spinal cord and is encased safe inside bone (skull and vertebrae). The PNS exists and extends outside the CNS. The PNS relays sensory information to the CNS and executes motor commands from the CNS. The CNS integrates and processes the sensory information from the sensory neurons of the PNS and sends back commands to the
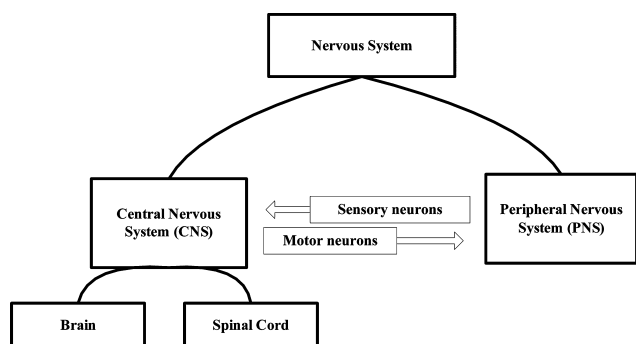
Figure 1. Schematic diagram of the nervous system (NS). The CNS is encased safe in bone (skull and vertebrae). The PNS exists and extends outside the CNS. The PNS relays sensory information to the CNS and executes motor commands from the CNS. The CNS integrates and processes the information from the sensory neurons and sends commands to the motor neurons.

motor neurons. Neurons send information using neurotransmitters that stimulate various receptors. Some of these receptors, like muscarinic acetylcholine receptors, are found both on autonomic effector cells of PNS and in high densities within the hippocampus, cortex, and thalamus of the CNS.[4] In such cases, and in diseases related to pain, allergy, muscle relaxation, etc. that are related to the PNS, without a detailed study of the mechanism of action, it may not always be easy to pinpoint with certainty whether the action is coming from the PNS or the CNS. In this article, such drugs were considered CNS drugs if there was evidence of BBB penetration.

Components of the BBB. The BBB is a cellular and multifunctional molecular barrier located at the capillaries in the brain that restricts the passage of many chemical substances and microscopic objects from the blood into the brain. The important components of this barrier, besides the cellular lipid barrier, are as follows.[5,6]

Endothelial Tight Junctions. The cerebral blood capillaries of the brain are made of brain microvascular endothelial cells (BMVEC). A major difference between BMVEC and normal endothelial cells is the tight junctions. The tight junctions prevent many water-soluble substances from freely passing through the cells and entering the fluid environment of the brain cells.

Enzymatic Barrier. Several enzymes on the lining of the blood capillaries of the brain actively destroy or inactivate undesirable peptides and other small molecules in the blood. For example, circulating adenosine enters the brain from the blood via BBB concentrative nucleoside transporter type 2 (CNT2) but does not have pharmacological effects in the brain, because of the rapid inactivation at the BBB by adenosine metabolizing enzyme. Conversely, the enzymatic BBB may serve to activate prodrugs. Circulating L-DOPA enters the brain via BBB large neutral amino acid transporter type 1 and is rapidly converted to pharmacologically active dopamine.

Active Efflux Barrier. Certain drugs may cross the endothelial barrier via free diffusion and undergo influx from the blood to the brain compartment. However, this influx can be immediately followed by active efflux from the brain back to the blood if the drug is a substrate for one of many different active efflux transporters (AETs) expressed within the brain microvasculature. P-Glycoprotein (P-gp) is by far the most prominent and studied AET. P-gp detoxifies cells by exporting hundreds of chemically unrelated toxins and foreign small

molecules. It is a member of the superfamily of ATP-binding cassette (ABC) transporters. Substrate promiscuity is a hallmark of P-gp activity. Its X-ray structures, both apo and drug-bound forms, have recently been determined.[7] These structures revealed an internal cavity of $\sim$6000 Å^3 with a 30 Å separation of the two nucleotide-binding domains. Apo and drug-bound P-gp structures have portals open to the cytoplasm and the inner leaflet of the lipid bilayer for drug entry. The inward-facing conformation represents an initial stage of the transport cycle that is competent for drug binding.

It is important to note that physicochemical property modulation alone during lead optimization cannot overcome all the BBB components; it will help mainly to optimize the passive diffusion and the P-gp activity. There are several articles covering the pharmacophores[8] and the quantitative definition[9] of P-gp substrates that may be useful for converting a P-gp substrate to a nonsubstrate.

Alternate Drug Delivery in the Brain. Essentially 100% of macromolecules and 98% of small molecule drugs[10] do not cross the BBB. Extensive research is underway for delivering such drugs[10,11] using transcranial delivery, transnasal delivery, BBB disruption, endogenous BBB transporters, etc. Though such methods may have a huge impact on CNS drug discovery in the future, any detailed discussion of those methods is outside the scope of this work. However, this work would not be complete unless the recent approaches in using P-gp inhibitors to improve the BBB penetration were mentioned.[12,13] The finding that the clinically applied Ca2C channel blocker verapamil inhibits drug efflux and restores drug sensitivity in multi-drug-resistant leukemia cell lines[14] gave rise to the idea that compounds that inhibit P-gp might reverse P-gp-mediated multidrug resistance in patients or improve the delivery of CNS drugs across the BBB, which otherwise failed because of efflux. However, such an approach may have to be a monitored therapy, because P-gp dysfunction can both induce the increased level of accumulation of toxins, as in Parkinson's disease, and weaken the ability of the brain to efflux proteins, as in Alzheimer's disease.

Physicochemical Properties of CNS Drugs, the Prior Art. A number of reports have attempted to establish selected physiochemical property differences in CNS drugs. More than 100 years ago, Overton[15] and Meyer[16] called attention to the relationship between the lipophilic character of simple neutral organic compounds and their narcotic action. Hansch et al.[17] discussed the hydrophobic–hydrophilic balance necessary to avoid CNS-related issues in antihistamines and the optimal logP (octanol–water partition coefficient) for drugs that penetrate or are precluded from penetrating the brain. This trend continued in many other publications, contributing to our understanding of BBB penetration and brain exposure. Abraham et al.[18−20] developed various physicochemical property-based equations to estimate log BB (log of the brain/blood ratio). One simple equation derived was log BB = 0.055 + 0.203 × logPoct − 0.507$\sum\alpha_2$H − 0.500$\sum\beta_2$H, where logPoct is the water–octanol partition coefficient, $\sum\alpha_2$H is the overall hydrogen bond acidity, and $\sum\beta_2$H is the overall hydrogen bond basicity. van de Waterbeemd et al.[21] studied the estimation of BBB penetration by plotting various combinations of two physicochemical properties, such as molecular size, shape, H-bonding descriptors, and polar surface area, and then formulated various guidelines for use by medicinal chemists, e.g., polar surface area (PSA) of ≤90 Å^2 and molecular mass of <450 Da. With respect to molecular shape, it was concluded

Table 1. List of Computed Physicochemical and a Few PK Properties and Their Qualifying and Preferred Ranges in CNS and Non-CNS Oral Drugs

property	description (source[a])	ranges[b] in non-CNS oral drugs				ranges in CNS drugs			
		QL	PL	PU	QU	QL	PL	PU	QU
Human Oral Absorption	human oral absorption (QP)	1	3	3	3	2	3	3	3
Percent Human Oral Absorption	percent of human oral absorption (QP)	10	77	100	100	61	95	100	100
#acid	no. of carboxylic acid groups (QP)	0	0	0	2	0	0	0	0
#amide	no. of amide groups (QP)	0	0	0	1	0	0	0	1
#amidine	no. of amidine groups (QP)	0	0	0	0	0	0	0	0
#amine	no. of basic amines (QP)	0	0	0	1	0	1	1	2
SASA	solvent accessible surface area (QP)	265	459	660	1023	348	487	620	798
Molecular_SurfaceArea	topology-based molecular surface area (AP)	95	230	365	646	144	236	320	426
AREA	solvent accessible surface area (TP)	269	419	613	1028	319	457	577	735
FISA	SASA on N, O, and H attached to heteroatoms (QP)	0	81	176	306	0	0	64	176
PSA	polar surface area (solvent accessible) (TP)	0	89	185	317	0	2	64	171
Molecular_Polar Surface Area	topology-based molecular polar surface area (AP)	20	36	96	220	3.2	23	56	97
PSA_Q	van der Waals surface area of polar nitrogen and oxygen atoms (QP)	0	61	120	194	3.8	12	54	109
FOSA	SASA on saturated carbon and attached hydrogen (QP)	0	69	304	667	16	178	314	464
PISA	π component of SASA (QP)	0	0	138	371	0	160	292	343
WPSA	weakly polar component of the SASA (halogens, P, and S) (QP)	0	0	0	144	0	0	0	126
C-Het-Ratio	ratio of C atom and non-C, non-H atoms (CP)	0	1.1	2.9	7.8	1.2	2.1	4.3	11
Carbon Atoms	no. of C atoms (CP)	2	10	20	36	6	16	21	25
Nonpolar H atoms	no. of H atoms attached to C (CP)	1	6	19	46	5	17	26	31
AtomCount	total no. of atoms, including H (TP)	9	26	49	92	20	29	44	63
PolarH-Atom	H atoms not attached to C (CP)	0	0	2	6	0	0	1	3
HeteroAtom	no. of non-C and non-H atoms (CP)	1	3	7	14	1	3	5	8
#NandO	no. of N and O atoms (QP)	0	3	6	13	1	2	4	7
#nonHatm	no. of non-H atoms (QP)	4	17	27	46	8	19	25	30
Num_Atoms	no. of non-H atoms (AP)	4	17	27	46	9	19	25	30
Br_Count	no. of Br atoms (AP)	0	0	0	0	0	0	0	0
C_Count	no. of C atoms (AP)	2	10	20	36	6	16	21	25
Cl_Count	no. of Cl atoms (AP)	0	0	0	2	0	0	0	1
F_Count	no. of F atoms (AP)	0	0	0	2	0	0	0	1
H_Count	no. of H atoms (AP)	4	8	21	50	7	17	26	31
I_Count	no. of I atoms (AP)	0	0	0	0	0	0	0	0
N_Count	no. of N atoms (AP)	0	0	2	6	0	1	2	4
O_Count	no. of O atoms (AP)	0	2	4	11	0	1	2	4
P_Count	no. of P atoms (AP)	0	0	0	0	0	0	0	0
S_Count	no. of S atoms (AP)	0	0	0	2	0	0	0	1
#in34	no. of atoms in three- or four-membered rings (QP)	0	0	0	4	0	0	0	0
#in56	no. of atoms in five- or six-membered rings (QP)	0	9	16	24	5	11	17	24
#noncon	no. of atoms not able to form conjugation in nonaromatic rings (QP)	0	0	2	14	0	0	4	10
QPlogBB	brain/blood partition coefficient (QP)	−3.1	−1.5	−0.36	0.78	−1.2	−0.06	0.75	1.2
VOLSURF_BBB	a qualitative blood−brain permeability parameter (TP)	−2.4	−0.81	0.3	1.5	−0.63	0.19	1.1	1.6
BondCount	total no. of bonds in a molecule (TP)	12	29	53	98	19	30	47	65
Num_AromaticBonds	no. of bonds in aromatic rings (AP)	0	0	6	18	0	10	15	16
Num_BridgeBonds	no. of bridge bonds, naphthalene does not have any (AP)	0	0	0	0	0	0	0	10
Num_Bonds	no. of bonds between non-H atoms (AP)	7	17	29	53	8	18	26	34
QPPCaco	apparent Caco-2 cell permeability (QP)	0	0	198	3975	0	0	810	3269
VOLSURF_Caco2	a qualitative Caco-2 cell permeability parameter (TP)	−1.7	−0.06	0.85	1.59	−0.04	0.34	1	1.7
QPPMDCK	predicted apparent MDCK cell permeability (nm/s) (QP)	0	0	133	5302	0	0	634	5899
CIQPlogS	log of conformation-independent solubility (QP)	−9.6	−5.1	−2.4	0.14	−6.3	−4.2	−2.3	−0.36
CNS	a qualitative CNS activity parameter, from −2 to 2 (QP)	−2	−2	−1	1	−2	0	1	2
Dipole	computed dipole moment (QP)	0.96	3.7	7.7	12	0.67	1.1	3.9	8.9

Table 1. continued

property	description (source[a])	ranges[b] in non-CNS oral drugs				ranges in CNS drugs			
		QL	PL	PU	QU	QL	PL	PU	QU
Glob	a globularity descriptor (1 for a sphere) (QP)	0.73	0.81	0.88	0.94	0.77	0.82	0.88	0.93
HB-Acceptor	H bond acceptors without protonating bases or ionizing acids, no lone pair count (CP)	0	2	5	12	1	1	3	6
accptHB	estimated no. of hydrogen bonds that would be accepted from the solvent water (QP)	0	4	8.2	16.1	1	2.8	5.2	8.3
Acceptor	no. of hydrogen bond acceptors in a molecule (TP)	0	3	5	11	0	0	2	5
Num_H_Acceptors	no. of hydrogen bond acceptors in a molecule (AP)	0	2	5	12	1	2	3	6
HB-Donor	no. of H bond donors without protonating bases or ionizing acids (CP)	0	0	2	6	0	0	1	3
donorHB	estimated no. of hydrogen bonds that would be donated to the solvent water (QP)	0	1	2.5	5	0	0	1	3
Donor	no. of hydrogen bond donors in a molecule (TP)	0	1	2	7	0	1	2	3
Num_H_Donors	no. of hydrogen bond donors according to the supplied structure without protonating bases or ionizing acids (AP)	0	1	2	5	0	0	1	2
logD2	octanol−water logP at pH 2	−4	0.49	3.7	6.5	−1.6	0.9	2.8	3.8
logD74	octanol−water logP at pH 7.4	−4.9	0.28	3.4	6.4	−0.55	1.2	3.1	5.5
ALogP_A	atom-based logP calculator (AP)	−2.5	1	4.2	7.8	−0.31	2.1	4.2	6.1
AlogP	atom-based logP calculator (CP)	−2.8	0.82	4	7.4	−0.31	2.5	4.6	5.9
QPlogPo/w	octanol−water logP (QP)	−2.6	0.76	4	7.3	−0.16	2.5	4.7	6.0
clogP	octanol−water partition coefficient, using Hansch and Leo's clogP method (TP)	−4.4	0.8	4.2	7.4	−0.66	2.1	4.4	6.1
MlogP	octanol−water partition coefficient, Moriguchi logP (TP)	−2.9	0.86	3.2	−5.8	0.39	1.8	3.5	5.3
Molecular_Solubility	solubility in log(moles/liter) (AP)	−9.1	−5.6	−2.7	0.93	−7.2	−5.9	−3.4	−1
QPlogS	solubility in log(moles/liter) (QP)	−9.4	−4.9	−2.3	0.47	−6.5	−4.6	−2.5	−0.42
VOLSURF_Soly	solubility in log(moles/liter) (TP)	−8.2	−4.9	−3.1	−0.47	−6.1	−5.2	−3.7	−1
ALogP_MR	molar refractivity (AP)	15	68	108	178	35	76	103	129
MolarRefrac	molar refractivity (CP)	14	66	109	178	33	76	104	127
CMR	molar refractivity (TP)	1.6	6.3	11	18	3.5	7.8	10.4	12.6
Qppolrz	predicted polarizability (Å^3) (QP)	10	25	41	71	14	28	38	49
MW	molecular weight (TP)	75	241	393	671	141	250	353	452
#stars	drug likeness penalty; the higher the value, the less druglike the molecule (QP)	0	0	0	8	0	0	0	3
pKa_BasicSite	pK_a of the most basic atom (AP)	−2	0	4.5	11	0	7.9	10.7	10.9
QPlogKhsa	prediction of binding to human serum albumin (QP)	−1.80	−0.67	0.24	1.42	−1	0.04	0.78	1.04
VOLSURF_ProtBinding	calculated plasma protein binding (%) (TP)	−26.3	51	86	129	24	73	98	108
Num_RingAssemblies	no. of ring assemblies, note that for naphthalene, anthracene it is 1 (AP)	0	1	2	4	0	1	2	3
RingCount	no. of rings in a molecule (TP)	0	2	3	5	1	2	3	5
Num_Rings	no. of rings in a molecule (AP)	0	2	3	5	1	2	3	5
Num_Rings3	no. of three-membered rings (AP)	0	0	0	0	0	0	0	0
Num_Rings4	no. of four-membered rings (AP)	0	0	0	1	0	0	0	0
Num_Rings5	no. of five-membered rings (AP)	0	0	0	2	0	0	0	2
Num_Rings6	no. of six-membered rings (AP)	0	1	2	4	0	1	2	4
Num_Rings7	no. of seven-membered rings (AP)	0	0	0	0	0	0	0	1
Num_Rings8	no. of eight-membered rings (AP)	0	0	0	0	0	0	0	0
Num_Rings9Plus	no. of nine- and higher-membered rings (AP)	0	0	0	0	0	0	0	0
Num_Aromatic Rings	no. of aromatic rings (AP)	0	0	1	3	0	1	2	3
Num_RingBonds	no. of bonds in rings (AP)	0	11	18	28	5	15	22	27
Num_Rotatable Bonds	no. of rotatable bonds (hydrogen suppressed) (AP)	0	1	5	12	0	1	4	8
Num_Chains	no. of unbranched chains to cover all the nonring bonds in a molecule (AP)	1	4	7	15	1	2	4	7
#rotor	no. of rotatable bonds (without CX3, alkene, amide, small ring) (QP)	0	2	6	17	0	3	6	8
RotBonds	no. of rotatable bonds (hydrogen suppressed) (TP)	0	3	7	14	0	1	4	8
RuleOfFive	no. of violations of Lipinski's rule of five (QP)	0	0	0	2	0	0	0	1
RuleOfThree	no. of violations of Jorgensen's rule of three (QP)	0	0	0	2	0	0	0	1
Num_StereoAtoms	no. of chiral atoms (AP)	0	0	0	0	0	0	1	4
Volume	solvent accessible volume (Å^3) (QP)	410	763	1178	2082	492	830	1104	1388

Table 1. continued

property	description (source[a])	ranges[b] in non-CNS oral drugs				ranges in CNS drugs			
		QL	PL	PU	QU	QL	PL	PU	QU
VOLUME_T	molecular volume (TP)	346	678	1047	1863	461	736	966	1248

[a]Abbreviations: QP, Schrodinger/QikProp; TP, Tripos; AP, Accelrys; CP, internal program. [b]Abbreviations: QL, qualifying lower limit; PL, preferred lower limit; QU, qualifying upper limit; PU, preferred upper limit.

that the principal length/width ratio should be <5. The authors also pointed out that poor brain uptake often may be related to P-glycoprotein efflux. Kelder et al.[22] derived the dynamic polar surface areas of 776 CNS active compounds that went to phase II clinical trials. Their prescribed value for brain penetration was 60−70 Å². They also showed that the dynamic and static polar surface areas are highly correlated properties. Ghose et al.[23] studied computed physicochemical properties [atom-based logP (AlogP), atom-based molar refractivity (AMR), molecular mass, and the total number of atoms] in different classes of clinical-trial CNS compounds (depressant, psychotic, and hypnotic agents). The preferred and qualifying property ranges for these CNS drugs were provided, where the "preferred" range was derived from the most densely populated range covering 50% of the reference compounds and the qualifying range covered 80% of the reference compounds. Clark[24−26] described the early development of rapid computational methods applicable to a large number of compounds for the prediction of log BB, including many historical "rules of thumb". Norinder and Haeberlein[27] provided an excellent review of several computational protocols for modeling the transport of drugs across the BBB, including quantum-mechanics-based approaches, molecular-mechanics-related techniques, and two-dimensional-graph procedures. Mahar Doan[28] studied the passive permeability and P-gp-mediated efflux and 18 physicochemical properties of 48 marketed CNS drugs and 45 non-CNS drugs. They concluded that the CNS drug set had fewer hydrogen bond donors, fewer positive charges, greater lipophilicity, smaller polar surface areas, and reduced flexibility. For CNS delivery, a drug should ideally have an in vitro passive permeability of >150 nm/s and should not be a P-gp substrate (B → A/A → B ratio of <2.5). Adenot et al.[9] developed several discriminating models between CNS and non-CNS compounds, including the P-gp substrate. They used 1336 BBB-crossing compounds, 259 non-BBB-crossing compounds, and 91 P-gp substrates from the WDI (World Drug Index, http://thomsonreuters.com/products_services/science/science_products/a-z/world_drug_index/). Pajouhesh et al.[29] reviewed the various works related to CNS drug properties and concluded that molecular mass, lipophilicity, and the number of hydrogen bond donors and acceptors for CNS drugs had smaller ranges compared to those of the general therapeutics. Rishton et al.[30] reviewed the computational approaches about the prediction of BBB permeability with a comprehensive analysis of the ACDLabs (http://www.acdlabs.com) log BB prediction module, using various literature inhibitors and drugs for CNS targets. Hitchcock and Pennington[31] reviewed brain exposure, focusing on a few physicochemical properties that influence brain partitioning [molecular mass, computed logP (clogP), PSA, and number of hydrogen bond donors (HDs)] of many investigational CNS agents and their analogues. They provided the mean values of these properties for the 25 top-selling drugs in 2004 and also gave suggested limits and preferred ranges for these properties. Manallack[32] studied the pK_a distribution in CNS and non-CNS drugs. He concluded

that for CNS drugs the acidic pK_a values rarely fell below 6 and the basic pK_a values were not >10.5. Wager et al.[33] recently conducted a thorough analysis of six physicochemical properties for 119 marketed CNS drugs and a set of 108 Pfizer CNS candidates. The following six physicochemical properties were examined: ClogP, ClogD, molecular mass, TPSA, HBD, and pK_a. The CNS drug space defined by these six physicochemical properties is quite broad, but this article points to optimal ranges for each of these properties. On the basis of the drug set, the following median values were found: ClogP = 2.8, ClogD = 1.7, molecular mass = 305.3 Da, TPSA = 44.8 Å², HBD = 1, and pK_a = 8.4. These authors also proposed a multiproperty optimization scheme based on six prescribed property range violations.[34,35] Such a scoring scheme will be especially useful in the virtual screening of compounds.

Many of these earlier studies suffered from a number of drawbacks. (1) The size of the data set was relatively small. (2) The analysis was typically conducted using investigational compounds and nonapproved drugs. (3) Non-CNS drugs were most often neglected in the analysis. (4) The computed property source was not well documented.

When we were building the learning set for this study, the main question was whether to include clinical candidates in the study or to use only the approved drugs. Using all compounds that entered into clinical trials offers the advantage that the conclusions would be statistically more robust. However, the disadvantage was that approximately 90% of the compounds in the analysis failed during clinical evaluation and their inclusion may significantly distort the conclusions.

In our analysis, only approved drugs were used. Both CNS and non-CNS drugs were used in a comparative fashion. Only well-documented properties were included in the analysis. The focus was on simpler physicochemical properties and not the complex ADME properties. "Rules" were derived from the difference in the shape or gradient of the distribution curves as well as from recursive partition (RP) trees. Some of the graphing or analysis techniques used for simultaneously keeping track of multiple properties were discussed. The substructural elements that made the current CNS drug set were also identified. The main objective of this analysis was to provide the direction of various physicochemical property changes that may help lead selection and lead optimization for CNS activity.

■ RESULTS AND DISCUSSION

The vendor-provided property names with a short description and the qualifying and preferred ranges for CNS and non-CNS oral drugs are listed in Table 1. This table will be useful for designing both CNS and non-CNS drugs. The same properties computed using different software programs are in the proximity of each other in Table 1 for a better comparison. Many properties directly related to the molecular structure, like the number of hydrogen bond donors and hydrogen bond acceptors and the number of rotatable bonds, might be calculated differently by different programs. A summary of
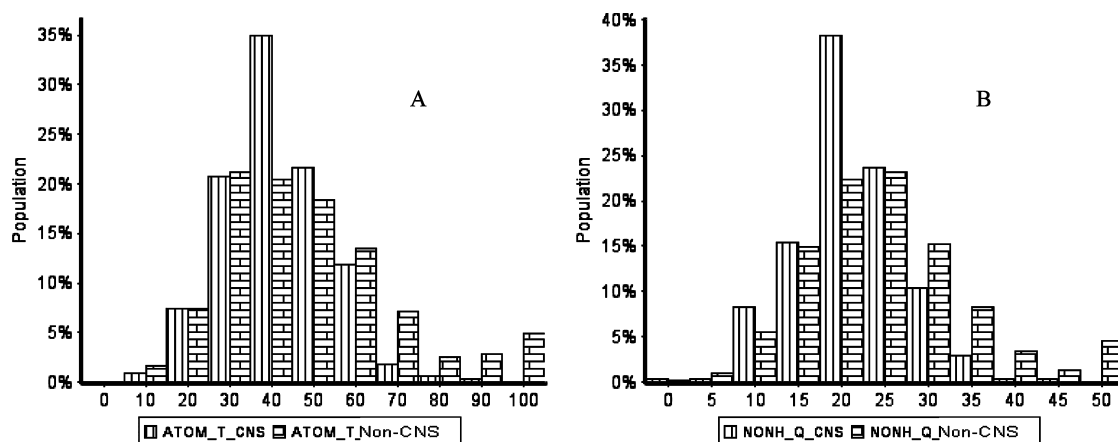
Figure 2. Distributions of the total number of atoms, including hydrogen (ATOM_T), and the number of non-hydrogen atoms (NONH_Q) in the non-CNS (brick columns) and CNS (vertically striped columns) drugs. The distribution indicates that 40−49 (A) total atoms and 20−24 non-hydrogen atoms (B) may be the best region for CNS drugs. All property distribution curves were generated using internally modified AP components, where each nonterminal column includes the value at which it was centered and anything less than the next bin. The left terminal column also includes all points below the bin, and the right terminal column includes all points higher than the last bin. In all figures, _A represents AP, _T represents TP, and _Q represents QP.
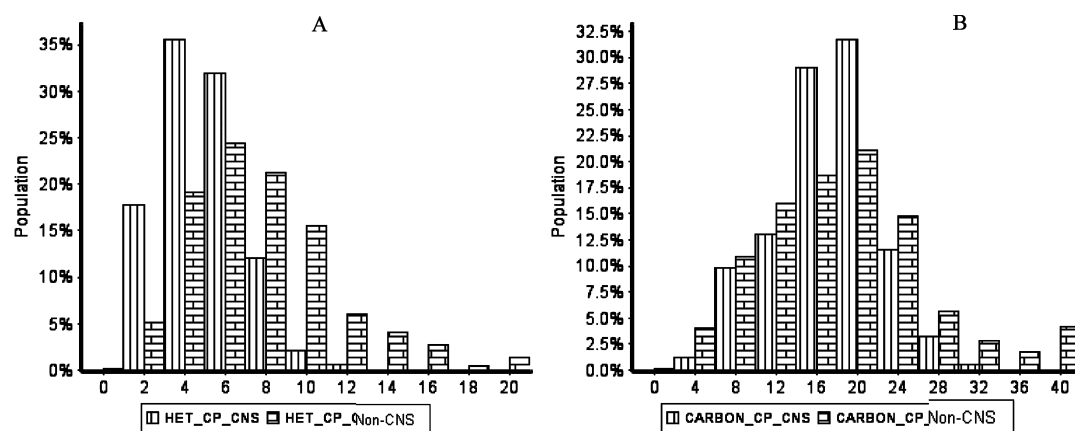
Figure 3. Distributions of the number of heteroatoms (non-C and non-H) (HET_CP) and the number of carbon atoms (CARBON_CP) in non-CNS and CNS drugs. The distributions indicate that four to seven heteroatoms (A) and 16−23 carbon atoms (B) may be the best region for CNS drugs.

computed values and their rules, which might themselves be debatable, and the distribution of various types of properties are described in the following sections.

Common Functional Groups. The parameter #amide computed in QP (QikProp, http://www.schrodinger.com/products/14/17/) is the number of nonconjugated amides. The amide group was in fact ubiquitous in the approved drugs. More than one-third of non-CNS oral drugs and a little less than one-third of CNS drugs had an amide group in their structure. This finding was not unexpected considering their ease of synthesis. The number of unsubstituted amides ($-CONH_2$) in both drug classes was low (6 in CNS and 18 in non-CNS oral). Monosubstituted amides were somewhat favored in the non-CNS oral drug class compared with disubstituted amides (154 vs 92). The opposite trend was observed in the CNS drugs: 37 monosubstituted amides versus 54 disubstituted amides. This might be a direct reflection of the fact that a smaller polar surface area is preferred for CNS drugs. The parameter #amine computed in QP was the number of nonconjugated amines. The basic aliphatic amines were not as common in the non-CNS oral drugs as they were in the CNS drugs. Approximately 70% of CNS drugs had an aliphatic amine

compared to 30% of non-CNS oral drugs (Figure S1 of the Supporting Information). The carboxylic acid group was not very common in the CNS drugs, because only 3−4% of CNS drugs had a carboxylic acid group, whereas 25% of non-CNS oral drugs contained a carboxylic acid group.

Atom Count. Atom counting parameters were omnipresent in most property calculators. Not only are such functionalities easy to implement, but they also can be used directly during the lead optimization process. When using the total atom count, one should ensure it includes the hydrogen atom in the calculation. Figure 2 illustrates the distribution of all atoms, including hydrogen and excluding hydrogen. The distributions suggested that both CNS and non-CNS drugs had uneven bell-shaped distributions. The major difference was in the slope of the curves and the locations of the maxima. The distribution clearly indicated that a total of 40−49 atoms and 20−24 non-hydrogen atoms were optimal for CNS drugs. Figure 3 illustrates the distribution of the number of heteroatoms (non-carbon, non-hydrogen) and the distribution of carbon atoms. The distribution indicated that four to five heteroatoms were noticeably preferred, but this number can easily reach as high as seven. It was established that 20−23 carbon atoms may
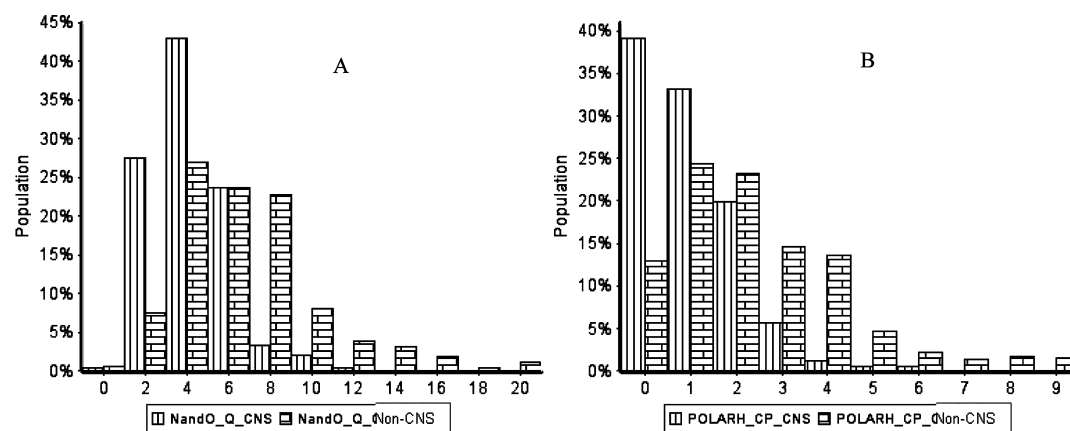
Figure 4. Distributions of the number of oxygen and nitrogen atoms (N and O_Q, respectively) (A) and the number of polar hydrogen atoms (POLARH_CP) (B) in non-CNS and CNS drugs. The distributions indicate that four or five oxygen and nitrogen (combined) atoms may be the best region, but it can be broadened to two to seven atoms.
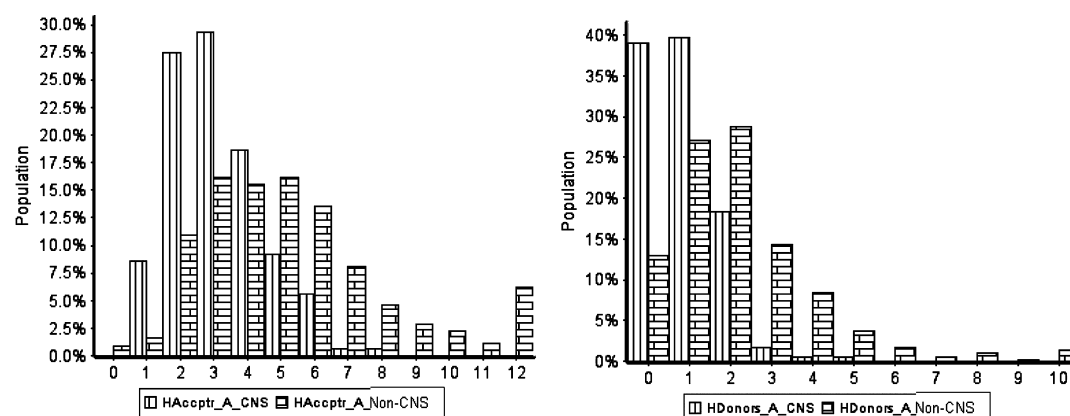

Figure 5. Distributions of the number of hydrogen bond acceptors (HAccptr_A) (A) and the number of hydrogen bond donors (HDonors_A) (B) in non-CNS and CNS drugs. The distribution indicate that two or three hydrogen bond acceptors and zero or one hydrogen bond donor may be the best ranges for the CNS drugs.

be preferred, but the range can safely be broadened to 16–23. Figure 4 shows the distribution of the number of oxygen and nitrogen atoms combined and the number of polar hydrogen atoms. Clearly, two to five nitrogen and oxygen atoms are the best range for CNS drugs. When the nitrogen and oxygen atoms are counted separately (Figure S2 of the Supporting Information), one or two nitrogen atoms and zero to two oxygen atoms may be optimal for successful CNS drugs. Similarly, zero or one polar hydrogen atom may be sufficient for CNS drugs.

H Bond Acceptors and Donors. Counting hydrogen bond acceptors and donors was not a straightforward process. A nitrogen atom may have multiple hydrogen atoms attached and therefore may form multiple hydrogen bonds. Similarly, an electronegative atom may have multiple lone pairs and may form more than one hydrogen bond in a suitable environment. A basic nitrogen atom may be protonated and become a hydrogen bond donor, rather than of a hydrogen bond acceptor. Many such issues complicate the counting and assignment of hydrogen bond acceptors and hydrogen bond donor groups. The best method for addressing these issues was to study the rules followed in the property calculator and benchmark a reference set with the calculator that was used. One recommendation is that one should never use the benchmark derived from one calculator to judge similar

properties using a different calculator. Figure 5 shows the distribution of hydrogen bond acceptors and donors as calculated by the AP (Accelrys Pipeline Pilot Component Collections, http://www.accelrys.com/products/pipeline-pilot/component-collections.html) molecular property calculator. According to the AP calculator, the number of HB-Acceptors is the number of heteroatoms (oxygen, nitrogen, sulfur, and phosphorus) with one or more lone pairs, excluding atoms with positive formal charges, amide and pyrrole-type nitrogens, and aromatic oxygen and sulfur atoms in heterocyclic rings. Similarly, the number of HB-Donors is the number of heteroatoms (oxygen, nitrogen, sulfur, and phosphorus) with one or more attached hydrogen atoms. Note that according to these definitions, water has one hydrogen bond acceptor and one hydrogen bond donor. These definitions are different from the Lipinski definitions where primary amines are counted as two hydrogen bond donors. Once these differences have been noted, it is important to know how much the distribution differed according to the different definitions. Figure 5 and Figure S3 of the Supporting Information show the distributions of hydrogen bond acceptors and donors in the CNS and non-CNS oral drugs. The distribution differed in terms of not only the percentage of occurrence for different values but also the locations of the maximum, which also varied significantly using different software. According to the AP calculator, the hydrogen
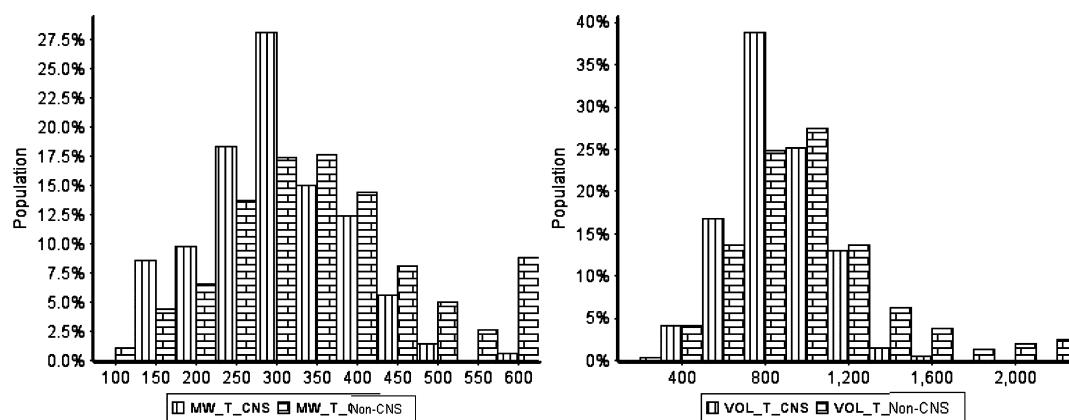
Figure 6. Distributions of molecular volume and molecular weight in the non-CNS and CNS drugs. The distributions indicate that the range of 300−350 was the best molecular weight region for CNS drugs even though the maximal population range for non-CNS oral drugs is the range of 300−400. The smaller size of CNS drugs was also reflected in the volume. The TP-computed molecular volume showed the most populated range was from 800 to 1000 for the CNS drugs. For the non-CNS oral drugs, the most populated range was from 1000 to 1200. In the case of CNS drugs, the peaks are considerably sharper than that of non-CNS oral drugs.
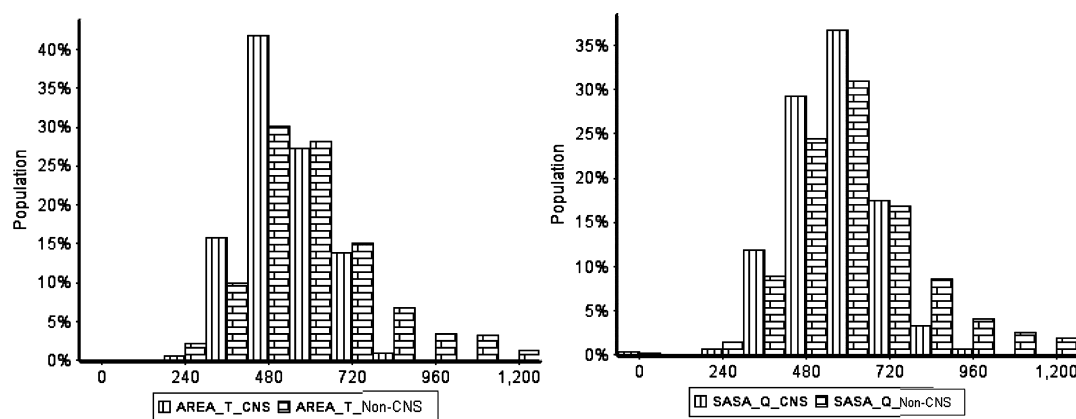
Figure 7. Distributions of solvent accessible surface areas in non-CNS and CNS drugs. The overall distributions of solvent accessible surface area differed noticeably not only in the location of the most populated bin but also in the relative population of the bins. The solvent accessible surface computed by Tripos showed a more noticeable difference between CNS and non-CNS drugs. According to this distribution, the CNS drugs have a considerably higher population in the range of 480−600.

bond acceptor peak was at three for the CNS drugs; according to TP (Tripos property calculator, http://www.tripos.com/index.php?family=modules,SimplePage,,,&page=sybyl_ligand_based_design), the value was two, and according to QP it was four. TP protonates several basic nitrogen atoms, and QP tries to calculate the number of hydrogen bonds with the solvent water, which is very close to counting the number of lone pairs available for hydrogen bond formation. For hydrogen bond donors, AP showed the maximum at one and zero being almost equally populated. TP showed the maximum at two, with one being a close competitor. QP showed the maximum at zero and one being the next populated area. When the distribution differences with the non-CNS oral drugs are considered, the best ranges to target for the hydrogen bond acceptor are two to three for AP, three to four for QP, and one to two for TP. For hydrogen bond donors, these ranges are zero to one, one to two, and zero to one, respectively.

Molecular Size-Related Parameters (molecular weight and volume). Molecular weight may be one of the few parameters that was consistently computed by all the software. On the other hand, the definition of volume can refer to either solvent accessible volume or molecular volume. The solvent accessible volume is larger than the molecular volume. Even

though TP and QP both claimed to compute the solvent accessible surface area, the volume computed by TP was considerably smaller than the volume computed by QP. The distribution of MW and the TP volume is shown in Figure 6. These distributions indicate that the range of 300−350 was the best molecular weight region for CNS drugs, even though the maximal population range for non-CNS oral drugs was the range of 300−400 (with considerable overlap). The smaller size of CNS drugs was also reflected in the volume parameter. The TP-computed molecular volume showed the most populated range was from 800 to 1000 $Å^3$ for the CNS drugs. For the non-CNS oral drugs, the most populated molecular volume range was from 1000 to 1200 $Å^3$. In the case of CNS drugs, the peaks were considerably sharper than those for non-CNS oral drugs. According to QP, the most populated volume range for CNS drugs was 1000−1200. The non-CNS oral drugs have a fairly flat distribution from 800 to 1400 (Figure S4 of the Supporting Information).

Molecular Surface Area (total and polar surface area). Comparison of the molecular surface area or polar surface area should be done with caution because there are three common types of molecular surfaces: (i) topological molecular surface area, which was derived from the atomic
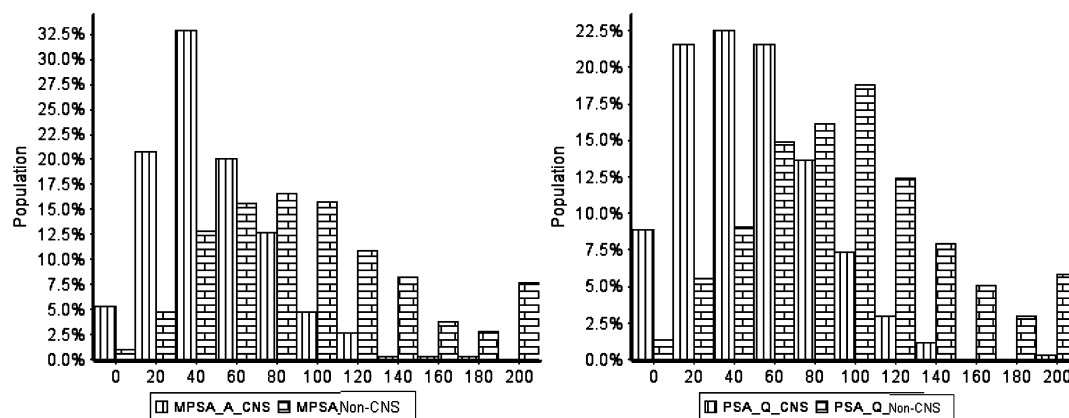
Figure 8. Distributions of topological polar surface areas in the non-CNS and CNS drugs. The topological polar surface areas were more discriminatory for CNS and non-CNS drugs over the solvent accessible polar surface area. These properties showed a noticeable difference between CNS drugs and non-CNS drugs. A small polar surface area (20−60 for MPSA_A and 20−80 for PSA_Q) was favored for CNS drugs, and a larger polar surface area (>80) was favored for the non-CNS drugs.
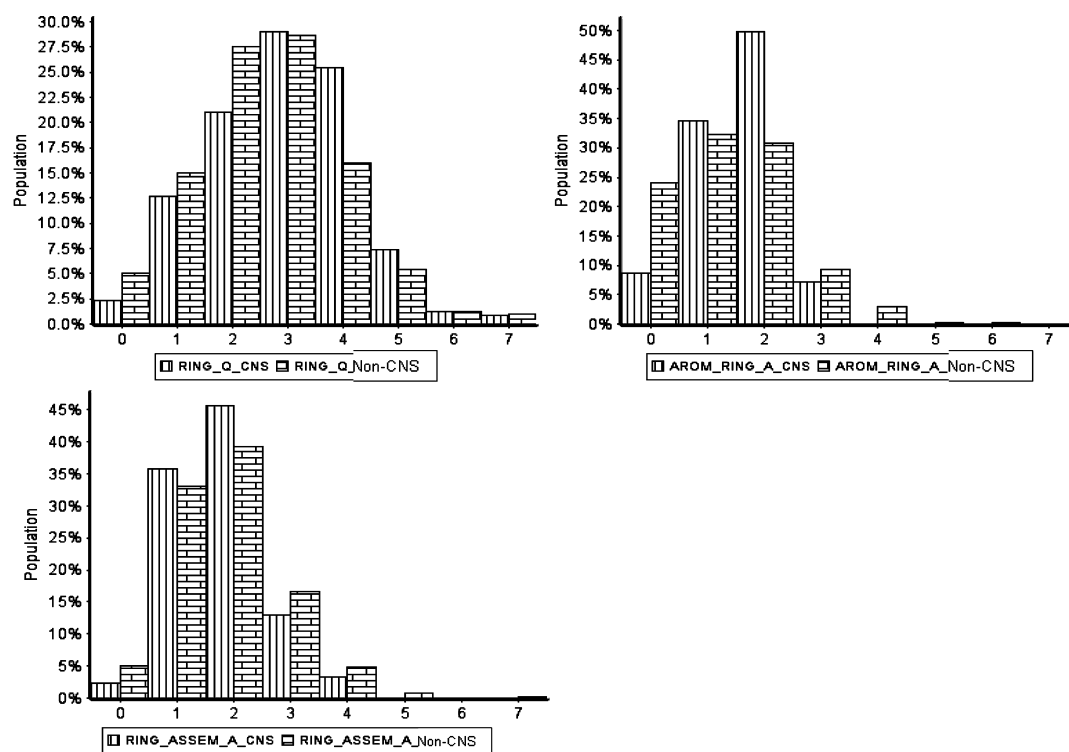

Figure 9. Number of rings and ring assemblies in CNS and non-CNS drugs. Even though the numbers of rings and ring assemblies were among the least discriminatory properties between CNS and non-CNS drugs, three or fewer rings and two or fewer aromatic rings and assemblies were the ideal regions for both CNS and non-CNS drugs.

connectivity and independent of the conformation used, (ii) molecular surface area derived from the van der Waals surface of a three-dimensional (3D) conformation, and (iii) solvent accessible surface area (SASA), where an approximate water dimension radius of 1.4 Å was enclosed over a 3D conformation to estimate the surface area. These computed surfaces vary considerably from one another. Even the values of SASA computed by TP or QP differed considerably in their distribution (Figure 7). The distribution of TP-computed SASA was slightly more discriminatory in differentiating CNS from non-CNS drugs, where the distribution of the CNS drugs had a considerably higher population in the range of 480−600. The polar surface area had three parallel counterparts as the SASA. The topological polar surface areas computed by AP or QP

were the most discriminatory properties in differentiating CNS drugs from non-CNS drugs (Figure 8). A small polar surface area of 20−60 for MPSA_A or 20−80 for PSA_Q was favored for the CNS drugs, while a larger polar surface area (>80) was preferential for non-CNS drugs.

Number of Rings and Ring Assemblies. All the software programs were in agreement in counting the total number of rings and ring assemblies. Further, the distribution was approximately identical in both the CNS and non-CNS drugs (Figure 9) and clearly showed that the total number of rings should be no more than three, and the number of aromatic rings and ring assemblies (e.g., naphthalene equals one) should be no more than two.

58

dx.doi.org/10.1021/cn200100h | *ACS Chem. Neurosci.* 2012, 3, 50−68
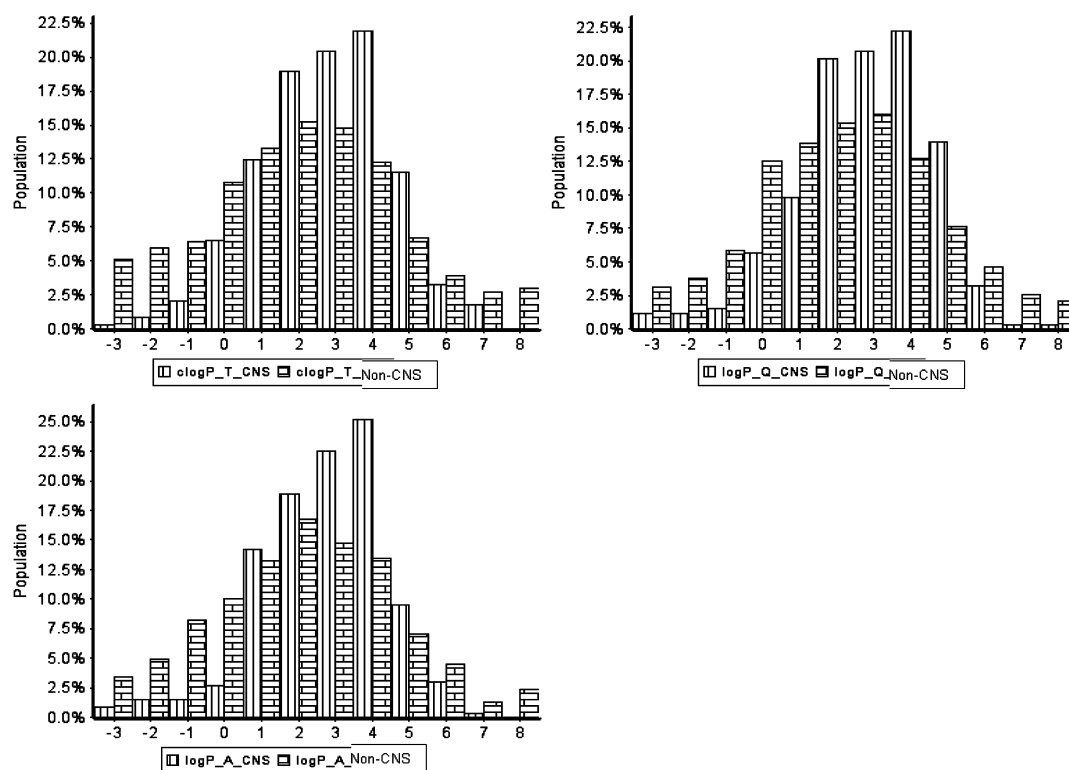
Figure 10. LogP values computed by different methods, in CNS and non-CNS drugs. The logP distributions determined by different methods were fairly consistent. These properties are among the least discriminatory properties for differentiating between CNS and non-CNS drugs.
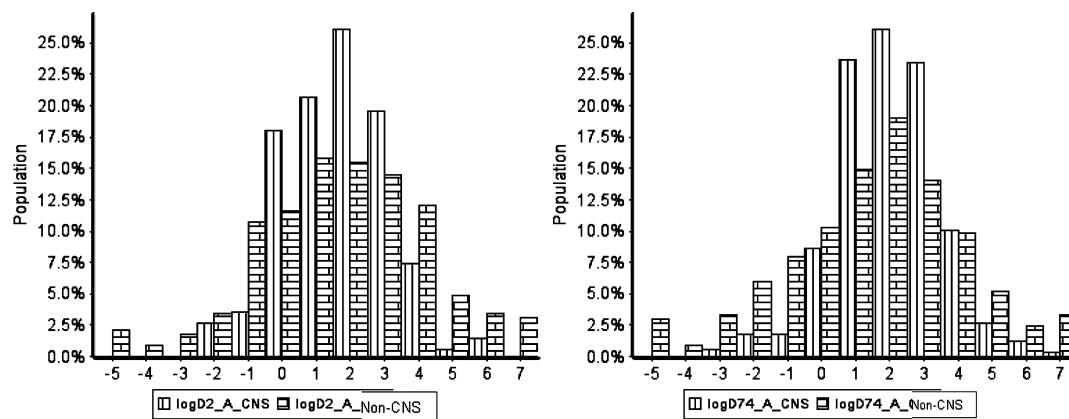
Figure 11. Computed logD values at pH 2 and 7.4 in CNS and non-CNS drugs. Even though the most populated range for logP in CNS drugs was between 4 and 5, logD values at both pH 2 and 7.4 were between 2 and 3.

Octanol−Water Partition Coefficient (logP and logD). The distributions of logP values (Figure 10) computed from Tripos (clogP), Accelrys (AlogP), and QikProp were fairly consistent and were among the least discriminatory properties for differentiating between CNS drugs and non-CNS drugs. It was unexpected that the largest population of CNS drugs was in the logP range of 4−5. An accurate calculation of logD is further complicated by the difficulty of accurately estimating the pK_a of a molecule.[36] However, logD is still a widely accepted parameter to use during the lead optimization process. LogD distributions at both pH 2 and 7.4 showed that the largest population for CNS drugs was from two to three (Figure 11). When compared with logP, the distribution suggested that CNS drugs were more hydrophobic compounds with a solublizing basic aliphatic amine group.

Chemical Makeup of the CNS Drugs. The critical substructural elements of the CNS drugs are listed in Table 2. These substructures occurred in ≥5% in CNS drugs. The percentage of occurrence alone may be useful for the design or optimization of a CNS drug. However, we also provided the percentage of occurrence of each fragment in the non-CNS drugs, which may be useful as a secondary parameter. Fragments that occurred frequently in the CNS drugs and less frequently in the non-CNS drugs may be a clear choice. Even though a benzene ring is a perfect choice for a CNS drug, it cannot be the only substructure, because it needs to satisfy other physicochemical requirements and more importantly it should have sufficient binding affinity with the biological target.

Graphical Representations of Multiple Properties Using a Radar Chart. The qualifying or the 95% property

Table 2. Important Substructures in the Chemical Makeup of CNS Drugs
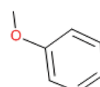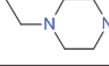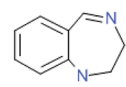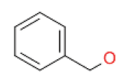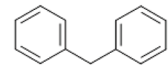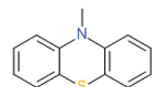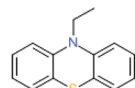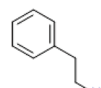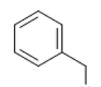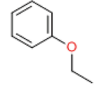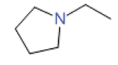
Table 2. The chemical make up of CNS drugs: important substructures.

Substructure ID	Substructure	Percent of CNS drugs having the substructure	Percent of non-CNS drugs having the substructure	Class†
1		86	66	C
2		58	51	E
3		31	41	N
4		30	42	N
5		24	37	N
6		23	7	C
7		18	12	C
8		15	16	E
9		13	8	C
10		13	4	C
11		11	17	N
12		11	5	C
13		11	24	N
14		10	4	C
15		9	37	N
16		9	18	N
17		9	18	N
18		9	0	C
19		9	3	C

Substructure ID	Substructure	Percent of CNS drugs having the substructure	Percent of non-CNS drugs having the substructure	Class†
20		9	0	C
21		8	0	C
22		8	0	C
23		8	11	N
24		7	4	C
25		7	0	C
26		7	0	C
27		7	11	N
28		6	11	N
29		6	8	N
30		6	6	E
31		5	3	C
32		5	8	N
33		5	2	C

†C stands for CNS preferred. N stands for non-CNS preferred. E stands for equally preferred.

Figure 12. Multiple-physicochemical property analysis using Radar charts, a tool for lead identification. The shaded area represents the property ranges of 95% of approved CNS drugs. The red line represents the properties of the test molecule. (A) Properties of biperiden, a muscarinic antagonist used for parkinsonism. (B) Properties of morphine, the principal alkaloid in opium.

ranges (Table 1) were important to identify if the test compound's property profile was within a reasonable range. The preferred or 50% property ranges (Table 1) on the other hand may help one to decide the direction of a change in property that might increase drugability during lead optimization. It might not be easy to follow multiple properties of a molecule without a graphical approach. The Radar chart provided a 360° snapshot view of the profile using as many as 10−15 properties. There are two types of Radar charts that may be useful. The simpler Radar charts use a single range for each property, and the objective is to determine if the test properties fall within the qualified ranges. Figure 12 illustrates the properties of two approved CNS drugs, biperiden and morphine, using the 95th percentile property ranges. Here all the properties were within the 95% percentile property range. Figure 13, on the other hand, illustrates the properties of bromocryptine (a CNS drug that originated from natural products), which showed multiple violations (A). Similarly, very small CNS drugs like acetazolamide may have a very large polar surface area (B) yet do not have any BBB permeability issues. Because 5% of approved drugs fall outside of this range, one can see that on average, one of 20 properties may be outside the qualifying range of approved CNS drugs.

Alternatively, if 10 properties in the Radar chart are routinely visualized, on average one of every two approved drugs should have a property violation. When a compound is selected as a possible CNS drug candidate, the best chance for success would be a compound that has properties that fall within the qualified ranges (Table 1). During a lead optimization process, it is often found that after satisfying various rule of five (ROF) type rules,[37,38] the goal still may not be achieved. A better strategy for selecting and advancing lead compounds may be to use the multicolor band Radar charting technology (Figure 14). In this approach, the qualifying range (Table 1) was divided into three color bands (QL to PL, blue; PL to PU, green; PU to QU, yellow). The properties in the yellow band or below should go outward during lead optimization until they reach the green area. Similarly, properties in the blue region or above should go inward during lead optimization until they reach the green area.

Classification Tree. Several simple recursive partition (RP) classification trees were developed to differentiate between CNS drugs and non-CNS oral drugs. The objective was to keep the tree simple enough to be physically meaningful. One reasonably simple tree is shown in Figure 15. The green box represents higher probability for the CNS drugs and the red box higher probability for the non-CNS oral drugs, with the
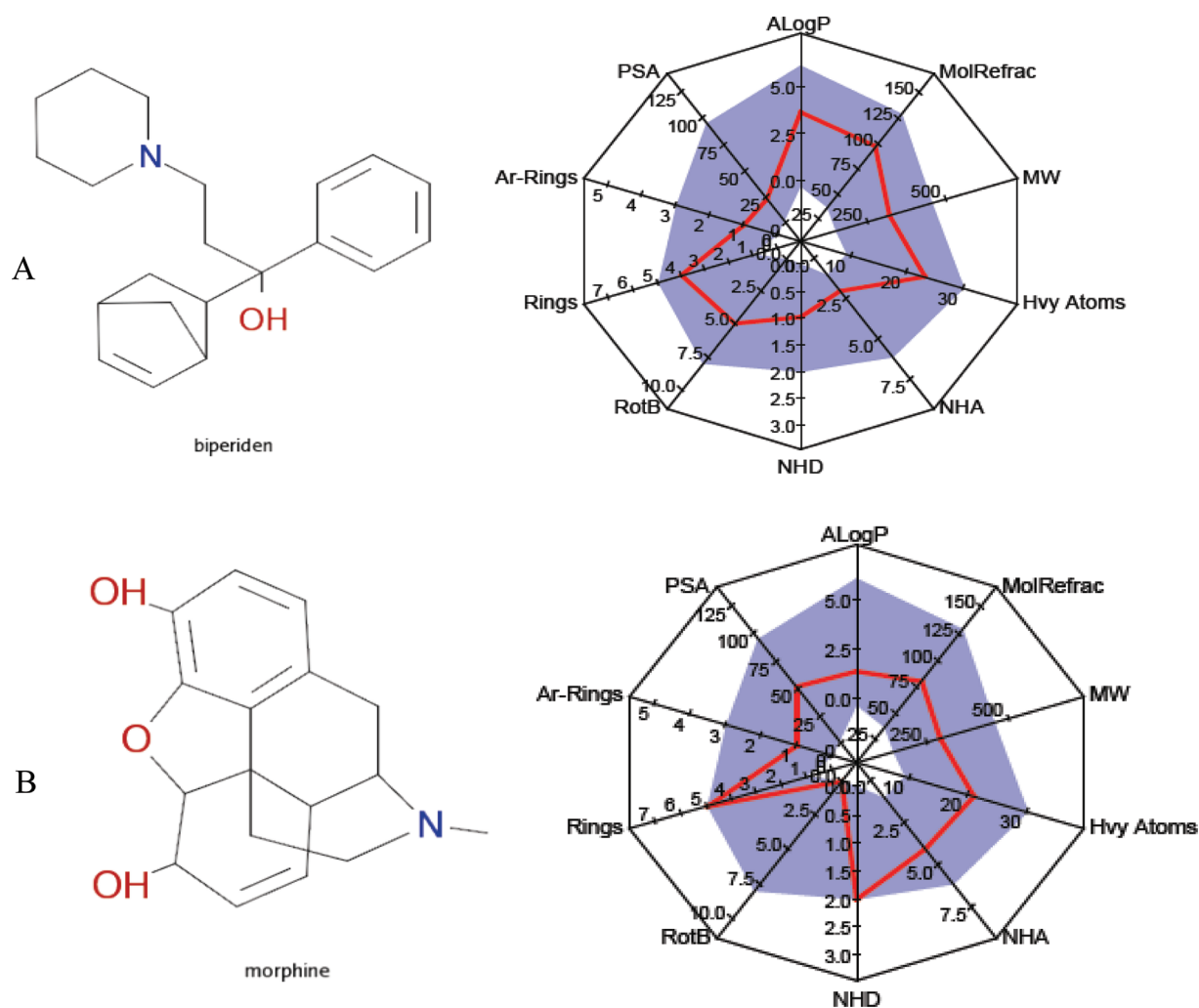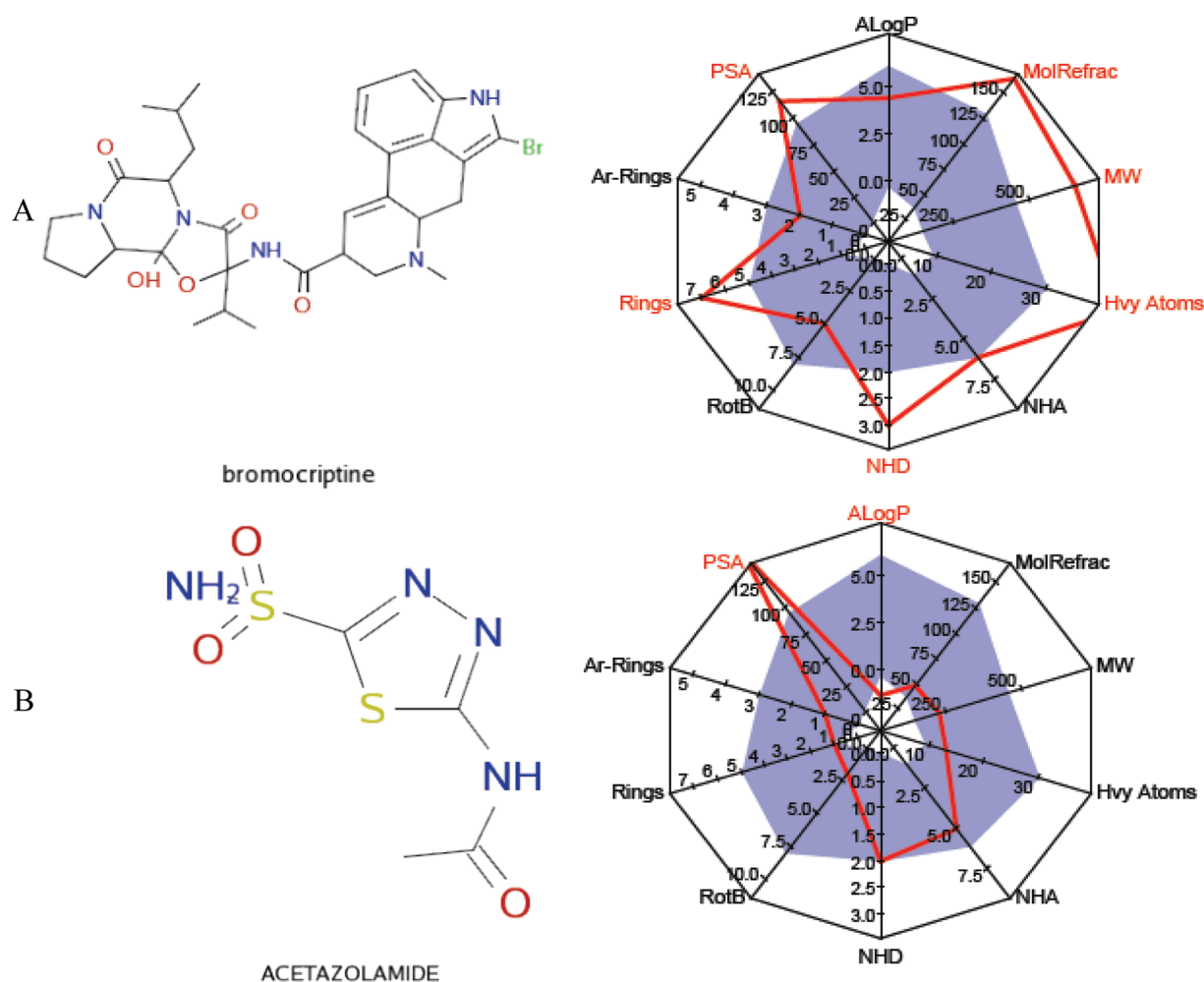
Figure 13. Multiple-physicochemical property analysis using Radar charts, a tool for lead identification. The shaded area represents the property ranges of 95% of approved CNS drugs. The red line represents the properties of the test molecule. Even though for many CNS drugs, all the properties were within our qualified range, this may not always be the case. Several CNS drugs, like bromocriptine (that originated from natural products), showed multiple violations (A). Similarly, very small CNS drugs like acetazolamide may have very a large polar surface area (B) yet do not have BBB permeability issues.

probabilities represented in parentheses. The green connecting arrows indicated that the condition was satisfied. When a decision tree is generated, it is important that it should not be overfitted (http://epub.ub.uni-muenchen.de/10589/1/partitioning.pdf). Cross validation, keeping sufficient members per nodes, using a minimum size tree helps to avoid the problem of overfitting. During the generation of the current RP tree, only physicochemical properties were used, the maximal tree depth was four, the minimal number of samples per node was 40, and the weighting method was by class. The split method was "Gini". The current RP model was validated (Table 3) on a test set of 80 CNS and non-CNS drugs reported in *Annual Reports in Medicinal Chemistry* Volumes 41–45, in the articles "To Market, To Market"-2005 to 2009. Except for tapentadol, the CNS versus non-CNS classification was accepted from those articles. The model, using only a few physicochemical properties, correctly classified 250 of the 317 (79%) CNS drugs used and 502 of the 626 (80%) non-CNS oral drugs used in the training set. It also correctly classified 90% of the test set (79% of the CNS drugs), making it a useful classification tree. During CNS drug optimization, if the lead belongs to node 2 and, because of the pharmacophoric restrictions, the PSA cannot be reduced, decreasing the volume

and number of atoms outside the rings may help the drug pass into the brain. Similarly, if a carboxylic acid occupies node 7, it may not be a good idea to optimize such a molecule as a CNS candidate unless there is active transportation. The guidelines for designing CNS drugs as learned from this classification tree as well as from the distribution curves are summarized below (the values within parentheses may be used as the direction of change during lead optimization): (i) topological molecular polar surface area (AP) of <76 Å^2 (25–60 Å^2), (ii) at least one nitrogen [one or two, including one aliphatic amine (note that one-third of CNS drugs did not have any basic amine)], (iii) fewer than seven (two to four) linear chains outside the rings (AP), (iv) fewer than three (zero or one) polar hydrogen atoms, (v) volume of 740–970 Å^3 (TP), (vi) solvent accessible surface area of 460–575 Å^2 (TP), (vii) positive QikProp parameter CNS. (viii) Other properties can be included from Table 1, preferably if the shape or gradient of the distribution curves differs between CNS and non-CNS drugs. The general strategy is CNS QL-QU for lead acceptance and PL-PU for the direction of property modulation during lead optimization. Effort will be to minimize the distance from the PL-PU range.

How many violations from the prescribed property profile may be acceptable? Even though we provide seven parameters,
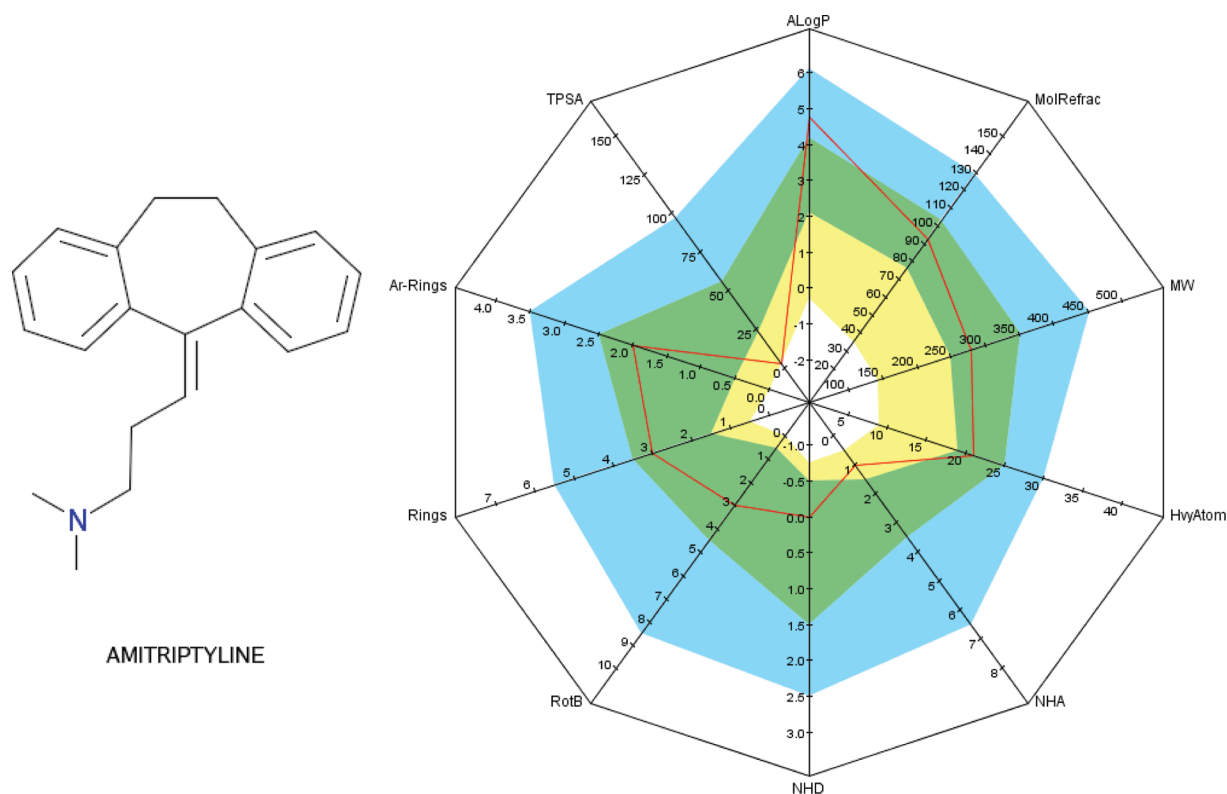
Figure 14. Multiple-color band Radar chart, a tool for lead optimization. Here the combined shaded area represents the property ranges of 95% of approved CNS drugs. The green band represents the most densely populated property range containing 50% of the approved CNS drugs. The red line represents the properties of the test molecule (amitriptyline, a tertiary amine tricyclic antidepressant, is structurally related to both the skeletal muscle relaxant cyclobenzaprine and the thioxanthene antipsychotics such as thiothixene). During lead optimization, it is often found that even after ROF type rules have been satisfied the lead optimization process does not achieve the goal. In such cases, optimizing the physicochemical properties toward the green region may have a better chance of success.

one can have many more combinations of rules using Table 1. Because QL to QU covered 95% of the approved CNS drugs, 10% property violations may be comfortably accepted. The ultimate acceptability is determined by the therapeutic success. When a large TPSA is unavoidable because of the nature of the protein target, decreasing the molecular size and side chains may help them pass into the brain. Carboxylic acids usually do not pass into the brain even though several amino acids can penetrate the BBB. The substructural elements that might be useful for the CNS drug design are provided in Table 2. We compiled Table 4 to compare the currently proposed profile with the profiles proposed by different authors.

CONCLUDING REMARKS

The pharmaceutical research industry has undergone a major change in recent years. Research facilities are slowly migrating to the more cost-effective areas of the globe. Project survival is becoming success-driven rather than need-based. However, the pharmaceutical research environment will stabilize, and CNS drug discovery research will accelerate soon. Both the CNS drug delivery research for BBB noncrossing entities and medicinal chemistry research to find CNS effective, BBB crossing oral drugs will continue. This work is primarily to help the medicinal chemistry research to select a better lead and to direct lead optimization more efficiently. The comparative analysis of physicochemical property profiles of 317 CNS and 626 non-CNS oral drugs helped to improve our understanding of the differences between CNS and non-CNS oral drugs in terms of the large number of physicochemical properties that

can be calculated using common molecular modeling packages. The ultimate objective was to derive guidelines that could be interpreted in terms of the chemical structure of the ligand to aid in reducing risk in the lead optimization and candidate selection phases of a discovery research project. The results derived from this study provided an ideal property profile in addition to a property modification strategy to utilize. A classification tree was also developed to differentiate between CNS drugs and non-CNS oral drugs. The property distribution study and the classification tree provided the following guidelines for designing high-quality CNS drugs: (i) topological molecular polar surface area of <76 Å^2 (25−60 Å^2), (ii) more than one nitrogen (one or two, including one aliphatic amine), (iii) fewer then seven (two to four) linear chains outside of rings, (iv) fewer than three (zero or one) polar hydrogen atoms, (v) volume of 740−970 Å^3, (vi) solvent accessible surface area of 460−580 Å^2, and (vii) a positive QikProp parameter CNS. (viii) Other properties can also be included from Table 1, preferably if the shape or gradient of the distribution curves differs between CNS and non-CNS drugs. The general strategies are as follows: (i) CNS QL-QU for lead acceptance and (ii) PL-PU for the direction of property modulation during lead optimization. The objective will be to minimize the distance from the PL-PU range. The chemo-informatics approaches to graphically analyzing multiple properties efficiently were also presented. Appropriate application of these results should be used in combination with the discovery of flow information (e.g., pharmacokinetics, PK/PD, selectivity, metabolic stability, P-gp substrate, etc.) to
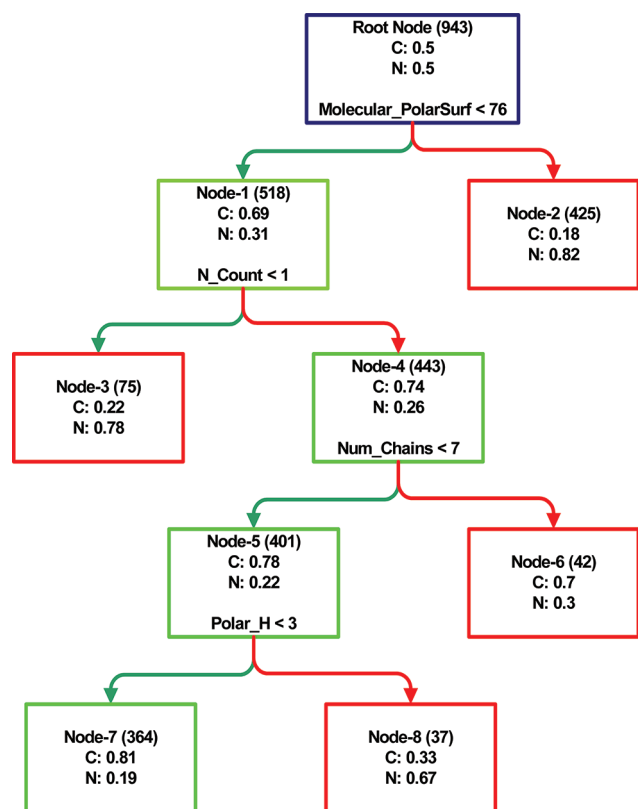
Figure 15. Recursive partition (RP) model generated by Accelrys cross-validated RP tree generator available under Pipeline Pilot components for differentiating between non-CNS oral drugs (red) and CNS drugs (green). The green connectors represent a satisfied condition. Green nodes favored CNS drugs. The numbers after C and N are the probabilities of CNS and non-CNS drugs, respectively, in the node. The number within parentheses in each node gives the total population of the drug candidates. During CNS drug optimization, if the lead belongs to node 2 and, because of the pharmacophoric restrictions, the PSA cannot be reduced, decreasing the volume may help the drug pass into the brain. Similarly, if a carboxylic acid occupies node 7, it may not be a good idea to optimize such a molecule as a CNS candidate unless there is active transportation.

help to expand our knowledge base, rather than used independently to eliminate out of range compounds.[39,40] The criteria proposed here should be used to design only a certain portion of compounds during lead selection and lead optimization. Definitely, we do not recommend 100% of compounds to follow the suggested profile that will hinder the expansion of knowledge.

■ METHODS

Preparation of Molecular Data Sets. Only the approved CNS and non-CNS drugs were analyzed to avoid skewed conclusions from "unsuccessful" data. The data set collection had two steps.

Primary Data Collection. Initial collection of the approved drugs was from several sources. The MDDR database from Symyx (http://www.symyx.com/products/databases/bioactivity/mddr/index.jsp) is an excellent source for collecting the structures, generic names, activity classes, and mechanisms of action. Searching the development phase field for "Launched" drugs identified 1844 approved drug structures. Leeson et al.[41] in their Supporting Information provided a comprehensive list of pre-1983 and 1983−2002 oral drugs. DrugBank (http://www.drugbank.ca/), maintained by the University of Alberta (Edmonton, AB), is also an excellent resource for obtaining approved drug names and structures with brief biological, biochemical, and

pharmaceutical properties. WOMBAT-PK from Sunset Molecular (http://www.sunsetmolecular.com) provides the measured PK and physicochemical properties of 1260 approved drugs. The chapter "To Market, To Market-Year"[42] in *Annual Reports in Medicinal Chemistry* is an excellent source for the drugs approved in the previous year.

Data Curation. Cleaning the primary data required a major effort and involved the following.

Merging Lists and Removing Duplication Using Accelrys/Pipeline Pilot (AP). Because it was possible to download an SD file containing all chemical structures of the approved drugs from a particular data source, it was very important not to enter the same compound multiple times in the final list.

Classification of the Preliminary List into CNS Drugs, Non-CNS Oral Drugs, and Miscellaneous Compounds. "Drug category", "indication", "mechanism of action", and "pharmacodynamics" in DrugBank were used for the initial identification of CNS and non-CNS drugs. The following step was the consistency checkup, in Wikipedia (http://en.wikipedia.org/wiki/Main_Page), Pharmacogenomics Knowledge Base (http://www.pharmgkb.org/), Drugs@FDA (http://www.accessdata.fda.gov/scripts/cder/drugsatfda/), and SciFinder (http://www.cas.org/products/scifindr/). Most often, the information provided in these resources was fairly consistent. The pain-, allergy-, and muscle relaxation-related drugs may work peripherally (PNS). Those compounds were accepted in the CNS list, only when there was clear evidence of BBB penetration. For example, most of the first-generation antihistamines went in the brain and were accepted here in the CNS list. The number of such compounds was a small fraction of the complete CNS list. The route of administration field "dosage forms" in DrugBank was used to identify the oral drugs.

Further Cleaning of the Non-CNS Drug List. Many non-CNS oral drugs have CNS side effects either at normal therapeutic doses or at large doses. The oral drugs with CNS side effects at therapeutic doses were removed from the non-CNS oral drug list. The final list contained 317 CNS drugs (Table S1 of the Supporting Information) and 626 non-CNS oral drugs (Table S2 of the Supporting Information). Table S1 contains structures, generic names, indications, and mechanisms of action.

Computation and Analysis of Physicochemical Properties. The property calculators from three widely used molecular modeling companies, Schrodinger/QikProp (QP) (http://www.schrodinger.com/products/14/17/), Tripos (TP) (http://www.tripos.com/index.php?family=modules,SimplePage,,,&page=sybyl_ligand_based_design), and Accelrys Pipeline Pilot (AP) Component Collections (http://www.accelrys.com/products/pipeline-pilot/component-collections.html), were used for calculating the various properties.

The SD files containing the two-dimensional (2D) structures of the two classes of drugs were subjected to the following steps.

Counterions were removed, and the base structure was neutralized where possible (except for quaternary ammonium salts) in AP.

2D structures were converted to 3D using Schrodinger/LigPrep. 3D structures were used for property calculations in the three molecular modeling packages listed above, even though some property calculators require only the topological information.

Customized property distribution curves were generated using internally modified AP components, where each nonterminal column includes the value where it was centered and anything less than the next bin. The left terminal column also includes all points below the bin, and the right terminal column includes all points higher than the last bin.

The "preferred" and "qualifying" property ranges were determined using internally developed programs. The preferred ranges were the shortest (most populated) property ranges covering 50% of the drugs. The qualifying ranges were the shortest (most populated) property ranges covering 95% of the drugs. The preferred range is a shorter range within the qualifying range. Because the population density of approved drugs is considerably higher in the preferred range than the population density in the qualifying range, compounds satisfying the preferred range are expected to have a better survival rate. The population density usually increases if the percentage of drug coverage

Table 3. Test Set for the RP Classification Tree

generic name	RP class	node[a] occupied	indication	ARMC[b]	drug class[c]
Aliskiren	non-CNS	2	antihypertensive	43	
Alvimopan	non-CNS	2	postoperative ileus	44	
Ambrisentan	non-CNS	2	pulmonary arterial hypertension	43	
Armodafinil	non-CNS	2	sleep disorder treatment	45	CNS
Asenapine	CNS	7	antipsychotic	45	CNS
Besifloxacin	non-CNS	2	antibacterial	45	
Blonanserin	CNS	7	antipsychotic	44	CNS
Ciclesonide	non-CNS	2	asthma, COPD	41	
Clevudine	non-CNS	2	hepatitis B	43	
Clofarabine	non-CNS	2	anticancer	41	
Conivaptan	non-CNS	2	antidiuretic	42	
Dapoxetine	CNS	7	premature ejaculation	45	
Darifenacin	CNS	7	urinary incontinence	41	
Darunavir	non-CNS	2	antiviral HIV	42	
Dasatinib	non-CNS	2	anticancer	42	
Decitabine	non-CNS	2	anticancer	42	
Deferasirox	non-CNS	2	chronic iron overload	41	
Desvenlafaxine	CNS	7	antidepressant	44	CNS
Dexlansoprazole	non-CNS	2	gastroesophogeal reflux disease	45	
Doripenem	non-CNS	2	antibiotic	41	
Dronedarone	non-CNS	2	antiarrhythmic	45	
Eberconazole	CNS	7	antifungal	41	
Eltrombopag	non-CNS	2	antithrombocytopenic	45	
Entecavir	non-CNS	2	antiviral	41	
Eslicarbazepine-acetate	CNS	7	antiepileptic	45	CNS
Eszopiclone	non-CNS	2	hypnotic	41	CNS
Etravirine	non-CNS	2	antiviral	44	
Febuxostat	non-CNS	2	antihyperuricemic	45	
Fesoterodine	non-CNS	6	overactive bladder	44	
Garenoxacin	non-CNS	2	anti-infective	43	
Imidafenacin	CNS	7	overactive bladder	43	
Indacaterol	non-CNS	2	chronic obstructive pulmonary disease	45	
Ivabradine	non-CNS	6	angina pectoris	42	
Ixabepilone	non-CNS	2	anticancer	43	
Lacosamide	CNS	7	anticonvulsant	44	CNS
Lapatinib	non-CNS	2	anticancer	43	
Lenalidomide	non-CNS	2	immunomodulator	42	
Lisdexamfetamine	non-CNS	2	ADHD	43	CNS
Lubiprostone	non-CNS	2	chronic idiopathic constipation	42	
Luliconazole	non-CNS	2	antifungal	41	
Lumiracoxib	CNS	7	anti-inflammatory	41	
Maraviroc	non-CNS	6	anti-infective	43	
MinodronicAcid	non-CNS	2	osteoporosis	45	
Mozavaptan	non-CNS	6	hyponatremia (low blood sodium level)	42	
Nalfurafine	non-CNS	2	pruritus	45	
Nelarabine	non-CNS	2	anticancer	42	
Nepafenac	non-CNS	2	anti-inflammatory	41	
Nilotinib	non-CNS	2	anticancer	43	
Pirfenidone	CNS	7	idiopathic pulmonary fibrosis	44	
Posaconazole	non-CNS	2	antifungal	42	
Pralatrexate	non-CNS	2	anticancer	45	
Prasugrel	CNS	7	antiplerelet therapy	45	
Raltegravir	non-CNS	2	anti-infective HIV	43	
Ramelteon	CNS	7	insomonia	41	CNS
Ranolazine	non-CNS	6	antiangina	42	
Rasagiline	CNS	7	Parkinson's disease	41	CNS
Retapamulin	non-CNS	2	anti-infective	43	
Rimonabant	non-CNS	6	antiobesity	42	
Rivaroxaban	non-CNS	2	anticoagulant	44	
Rotigotine	CNS	7	anti-Parkinson	42	CNS
Rufinamide	CNS	7	anticonvulsant	43	CNS

Table 3. continued

generic name	RP class	node[a] occupied	indication	ARMC[b]	drug class[c]
Saxagliptin	non-CNS	2	antidiabetic	45	
Silodosin	non-CNS	2	dysuria (painful urination)	42	
Sitafloxacin	non-CNS	2	antibacterial	44	
Sitagliptin	non-CNS	2	antidiabetic	42	
Sitaxsentan	non-CNS	2	pulmonary hypertension	42	
Sorafenib	non-CNS	2	anticancer	41	
Sunitinib	non-CNS	2	anticancer	42	
Tafluprost	non-CNS	3	antiglaucoma	44	
Tamibarotene	non-CNS	6	anticancer	41	
Tapentadol	CNS	7	analgesic	45	CNS
Telbivudine	non-CNS	2	hepatitis B	42	
Tigecycline	non-CNS	2	antibiotics	41	
Tipranavir	non-CNS	2	HIV	41	
Tolvaptan	non-CNS	6	hyponatremia, antidiuretic	45	
Udenafil	non-CNS	2	erectile dysfunction	41	
Ulipristalacetate	non-CNS	6	contraceptive	45	
Varenicline	CNS	7	nicotine dependence	42	CNS
Vildagliptin	non-CNS	2	antidiabetic	43	
Vorinostat	non-CNS	2	anticancer	42	

[a]Check Figure 15 for node description. [b]*Annual Reports in Medicinal Chemistry* volume. [c]The drug class was accepted from ARMC with only one change, tapentadol, which was reported to be centrally acting.

Table 4. Comparison to the Currently Proposed Physicochemical Property Profile with the Prior Art

property name	current suggestion[a]	prior art	conclusion
TPSA	<76 Å^2 (25−60 Å^2)	<90 Å^{2,22} 60−70 Å^{2,29} 40−90 Å^{234}	TPSA was a better differentiator for CNS and non-CNS drugs than the 3D structure-based PSA
no. of N	≥1 (1−2, including one aliphatic amine)	not available	was combined with oxygen count as shown below
molecular flexibility	<7 (2−4) linear chains outside rings; 0−8 (1−4) rotatable bonds	≤5 rotatable bonds[41]	the current may be qualitatively comparable to that of Iyer et al.,[43] who concluded that some amount of flexibility enhances log BB but too much flexibility will diminish log BB
no. of polar hydrogen atoms per H bond donor	<3 (0−1)	<3[29]	
volume	460−1250 Å^3 (740−970 Å^3)	not available	it is related to molecular size, and most workers provided guidance for MW (see below)
solvent accessible surface area	320−735 Å^2 (455−575 Å^2)	not available	it is related to molecular size, and most workers provided guidance for MW (see below)
QikProp parameter CNS	>0		
no. of carboxylic acids	0 (unless an amino acid)	avoid acid[29]	
dealing polar compounds	decrease size and side chains		
logD at pH 7.4	−0.55 to 5.5 (1.2−3.1)[b]	1.4−2.6,[17] 1−4[21]	in the case of computed logD, validation of the data with a few experimental logD values is advised
molecular weight	140−450 (250−355)[b]	≤450,[21] ≤360[34]	
N + O	1−7 (2−4)[b]	≤5[27]	
ClogP − (N + O)	not available	>0[27]	should be used along with the previous rule
other properties	derived from Table 1, QL-QU (PL-PU)	not available	inclusion of other properties is advised only if distributions of CNS and non-CNS drugs differ in shape or gradient

[a]The ranges within the parentheses are preferred and should be used to derive the direction for property modulation during lead optimization. [b]Derived from Table 1 using the general guideline.

is lowered. However, decreasing the percentage of drug coverage increases the chance of neglecting good compounds. Preferred and qualifying ranges described here were a compromise between these two opposing conditions. Should all approved drugs be considered to derive the acceptable property ranges? Omitting 5% of the drugs would help to neglect the outliers resulting from prodrugs or compounds with active transport mechanisms, which are rare, but the properties of such compounds are often well outside the more common range. The 95% ranges were fairly wide and would accommodate most compounds that could be optimized within an

acceptable time period. The preferred range was a highly populated narrow band covering 50% of the approved drugs, which would be used as a directional guide for future compound design by moving the property closer to this range.

Qualitative rules for CNS and non-CNS drugs were derived from the properties whose distribution curves have different shapes or gradients. The shape of a distribution curve depended on the location and size of the bin. When distribution curves are generated, the data range is rounded and the rounded range is divided into approximately 10 bins.

Several classification trees were derived to differentiate between CNS and non-CNS oral drugs. The most promising one was presented here.

Evaluation and Analysis of Substructural Elements. SARvision (http://www.chemapps.com) was used to evaluate the relevant substructural units in the CNS drug set. Only the substructures that were present in 5% of the CNS data set were used to check their percentage of occurrence in the non-CNS drug set. The comparative occurrence was used to classify the substructure's class (C for CNS, N for non-CNS, and E for equally favored).

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

List of CNS and non-CNS drugs in the form of SMILES notation with their generic names, several property distribution curves, and the means and standard deviations of various physicochemical properties. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Department of Chemistry, Discovery Research, Cephalon, Inc., 145 Brandywine Parkway, West Chester, PA 19380. Phone: (610) 738-6882. Fax: (610) 738-6643. E-mail: aghose@cephalon.com.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Chen, L. C. (1996) in *World Population and Health. 2020 Vision: Health in the 21st Century*, The National Academies Press, Washington, DC.

(2) Fogel, R. W. (2009) Forecasting the cost of U.S. Health Care in 2040. *Journal of Policy Modeling 31*, 482−488.

(3) Evans-Martin, F. F. (2010) *The Human Body How It Works: The Nervous System*, Infobase Publishing, New York.

(4) Brunton, L. L., Lazo, J. S., and Parker, K. L., Eds. (2006) *Goodman and Gilman's The Pharmacological Basis of Therapeutics*, McGraw-Hill, New York.

(5) Pardridge, W. M. (2002) Drug and Gene Delivery to the Brain: Minireview The Vascular Route. *Neuron 36*, 555−558.

(6) Cardoso, F. L., Brites, D., and Brito, M. A. (2010) Looking at the blood−brain barrier: Molecular anatomy and possible investigation approaches. *Brain Res. Rev. 64*, 328−363.

(7) Aller, S. G., Yu, J., Ward, A., Weng, Y., Chittaboina, S., Zhuo, R., Harrell, P. M., Trinh, Y. T., Zhang, Q., Urbatsch, I. L., and Chang, G. (2009) Structure of P-Glycoprotein Reveals a Molecular Basis for Poly-Specific Drug Binding. *Science 323*, 1718−1722.

(8) Stouch, T. R., and Gudmundsson, O. (2002) Progress in Understanding the Structure-Activity Relationships of P-Glycoprotein. *Adv. Drug Delivery Rev. 54*, 315−328.

(9) Adenot, M., and Lahana, R. (2004) Blood-Brain Barrier permeation Models: Discriminating between potential CNS and non-CNS drugs including P-glycoprotein substrates. *J. Chem. Inf. Comput. Sci. 44*, 239−248.

(10) Pardridge, W. M. (2006) Blood-brain barrier delivery. *Drug Discovery Today 12*, 54−61.

(11) Misra, A., Ganesh, S., Shahiwala, A., and Shah, S. P. (2003) Drug delivery to the central nervous system: A review. *J. Pharm. Pharm. Sci. 6*, 252−273.

(12) Breedveld, P., Beijnen, J. H., and Schellens, J. H. M. (2006) Use of P-glycoprotein and BCRP inhibitors to improve oral bioavailability

and CNS penetration of anticancer drugs. *Trends Pharmacol. Sci. 27*, 17−24.

(13) Colabufo, N. A., Berardi, F., Cantore, M., Contino, M., Inglese, C., Niso, M., and Perrone, R. (2010) Perspectives of P-Glycoprotein Modulating Agents in Oncology and Neurodegenerative Diseases: Pharmaceutical, Biological, and Diagnostic Potentials. *J. Med. Chem. 53*, 1883−1897.

(14) Tsuruo, T., Iida, H., Tsukagoshi, S., and Sakurai, Y. (1981) Overcoming of Vincristine Resistance in P388 Leukemia in Vivo and in Vitro through Enhanced Cytotoxicity of Vincristine and Vinblastine by Verapamil. *Cancer Res. 41*, 1967−1972.

(15) Overton, E. (1897) Osmotic properties of cells in the bearing on toxicology and pharmacology. *Z. Phys. Chem. 22*, 189−209.

(16) Meyer, H. (1899) On the theory of alcohol narcosis I. Which property of anesthetics gives them their narcotic activity? *Arch. Exp. Pathol. Pharmakol. 42*, 109−118.

(17) Hansch, C., Björkroth, J. P., and Leo, A. (1987) Hydrophobicity and central nervous system agents: On the principle of minimal hydrophobicity in drug design. *J. Pharm. Sci. 76*, 663−687.

(18) Abraham, M. H., Chadha, H. S., and Mitchell, R. C. (1994) Hydrogen bonding. 33. Factors that influence the distribution of solutes between blood and brain. *J. Pharm. Sci. 83*, 1257−1268.

(19) Abraham, M. H., Chadha, H. S., and Mitchell, R. C. (1995) Hydrogen bonding. Part 36. Determination of blood brain distribution using octanol-water partition coefficients. *Drug Des. Discovery 13*, 123−131.

(20) Abraham, M. H., Takács-Novák, K., and Mitchell, R. C. (1997) On the partition of ampholytes: Application to blood-brain distribution. *J. Pharm. Sci. 86*, 310−315.

(21) van de Waterbeemd, H., Camenisch, G., Folkers, G., Chretien, J. R., and Raevsky, O. A. (1998) Estimation of blood-brain barrier crossing of drugs using molecular size and shape and H-bonding descriptors. *J. Drug Targeting 6*, 151−165.

(22) Kelder, J., Grootenhuis, P. D. J., Bayada, D. M., Delbressine, L. P. C., and Ploemen, J.-P. (1999) Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs. *Pharm. Res., 16*.

(23) Ghose, A. K., Viswanadhan, V. N., and Wendoloski, J. J. (1999) A knowledge-based approach in designing combinatorial and medicinal chemistry libraries of drug discovery. 1. A qualitative and quantitative characterization of known drugs databases. *J. Comb. Chem. 1*, 55−68.

(24) Clark, D. E. (1999) Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena: 2. Prediction of blood-brain barrier penetration. *J. Pharm. Sci. 88*, 815−821.

(25) Clark, D. E. (2003) In silico prediction of blood−brain barrier permeation. *Drug Discovery Today 8*, 927−933.

(26) Clark, D. E. (2005) Computational prediction of blood-brain barrier permeation. *Annu. Rep. Med. Chem. 40*, 403−415.

(27) Norinder, U., and Haeberlein, M. (2002) Computational approaches to the prediction of blood-brain distribution. *Adv. Drug Delivery Rev. 54*, 291−313.

(28) Mahar Doan, K. M., Humphreys, J. E., Webster, L. O., Wring, S. A., Shampine, L. J., Serabjit-Singh, C. J., Adkison, K. K., and Polli, J. W. (2002) Passive permeability and P-glycoprotein-mediated efflux differentiate central nervous system (CNS) and non-CNS marketed drugs. *J. Pharmacol. Exp. Ther. 303*, 1029−1037.

(29) Pajouhesh, H., and Lenz, G. R. (2005) Medicinal chemical properties of successful central nervous system drugs. *NeuroRx 2*, 541−553.

(30) Rishton, G. M., LaBonte, K., Williams, A. J., Kassam, K., and Kolovanov, E. (2006) Computational approaches to the prediction of blood-brain barrier permeability: A comparative analysis of central nervous system drugs versus secretase inhibitors for Alzheimer's disease. *Curr. Opin. Drug Discovery Dev. 9*, 303−313.

(31) Hitchcock, S. A., and Pennington, L. D. (2006) Structure-Brain Exposure Relationships. *J. Med. Chem. 49*, 7559−7583.

(32) Manallack, D. T. (2007) The pKa distribution of drugs: Application to drug discovery. *Perspect. Med. Chem. 1*, 25−38.

(33) Wager, T. T., Chandrasekaran, R. Y., Hou, X., Troutman, M. D., Verhoest, P. R., Villalobos, A., and Will, Y. (2010) Defining desirable central nervous system drug space through the alignment of molecular properties, in vitro ADME, and safety attributes. *ACS Chem. Neurosci. 1*, 420−434.

(34) Wager, T. T., Hou, X., Verhoest, P. R., and Villalobos, A. (2010) Moving beyond Rules: The Development of a Central Nervous System Multiparameter Optimization (CNS MPO) Approach To Enable Alignment of Druglike Properties. *ACS Chem. Neurosci. 1*, 435−449.

(35) Wager, T. T., Villalobos, A., Verhoest, P. R., Hou, X., and Shaffer, C. L. (2011) Strategies to optimize the brain availability of central nervous system drug candidates. *Expert Opin. Drug Discovery 6*, 371−381.

(36) Lee, A. C., and Crippen, G. M. (2009) Predicting pKa. *J. Chem. Inf. Model. 49*, 2013−2033.

(37) Lipinski, C. A., Lombardo, F., Dominy, B. W., and Freeney, P. J. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev. 23*, 3−25.

(38) Ursu, O., and Oprea, T. I. (2010) Model-free drug-likeness from fragments. *J. Chem. Inf. Model. 50*, 1387−1394.

(39) Ghose, A. K., Herbertz, T., Pippin, D. A., Salvino, J. M., and Mallamo, J. P. (2008) Knowledge based prediction of ligand binding modes and rational inhibitor design for kinase drug discovery. *J. Med. Chem. 51*, 5149−5171.

(40) Ghose, A. K., Herbertz, T., Salvino, J. M., and Mallamo, J. P. (2006) Knowledge-based chemoinformatic approaches to drug discovery. *Drug Discovery Today 11*, 1107−1114.

(41) Leeson, P. D., and Davis, A. M. (2004) Time-related differences in the physical property profiles of oral drugs. *J. Med. Chem. 47*, 6338−6348.

(42) Hegde, S., and Schmidt, M. (2010) To market, to market-2009. *Annu. Rep. Med. Chem. 45*, 467−537.

(43) Iyer, M., Mishra, R., Han, Y., and Hopfinger, A. J. (2002) Predicting Blood−Brain Barrier Partitioning of Organic Molecules Using Membrane-Interaction QSAR Analysis. *Pharm. Res. 19*, 1611−1621.