and/or screen against a specific target protein, to a manageable number of compounds that exhibit the highest chance to lead to a drug candidate (10, 19). The major sources of information to guide virtual screening for a particular target are derived from the following questions:

1. What does a drug look like in general?
2. What is known about compounds that interact with the receptor?
3. What is known about the structure of the target protein and the protein-ligand interactions?

In the following subsections we address these three points, outlining concepts of assessing the overall druglikenss of molecules, the concentration of subsets of molecules in focused libraries, and the identification of specific leads through structure-based virtual screening techniques.

2.1 Druglikeness Screening

Many drug candidates fail in clinical trials because of reasons unrelated to potency against the intended drug target. Pharmacokinetics and toxicity issues are blamed for more than half of all failures in clinical trials. Therefore, the first part of virtual screening evaluates the *druglikeness* of small molecules, mostly independent of their intended drug target (there are specific drug classes such as those acting in the central nervous system that require specific drug profiles). Druglike molecules exhibit favorable absorption, distribution, metabolism, excretion, and toxicological (ADMET) parameters (20–24). They are synthetically feasible and possess pharmacophore features that offer the chance of specific interactions with the intended protein target. Druglikeness is currently assessed using the following types of methods: simple counting methods, functional group filters, topological filters, and pharmacophore filters. Computational techniques used to identify druglikeness include neural networks (25–27), recursive partitioning approaches (25, 28), and genetic algorithms (29). These methods are further discussed below.

Table 6.1 Typical Ranges for Parameters Related to Druglikeness[a]

Parameter	Minimum	Maximum
Log *P*	−2	5
Molecular weight	200	500
Hydrogen bond acceptors	0	10
Hydrogen bond donors	0	5
Molar refractivity	40	130
Rotatable bonds	0	8
Heavy atoms	20	70
Polar surface area [Å^2]	0	120
Net charge	−2	+2

[a]Data taken from ref. 21.

2.1.1 Counting Schemes. Database collections of known drugs [e.g., CMC (30), WDI (31), or MDDR (32)] are typically used to extract knowledge about structure and properties of potential drug molecules. Key physicochemical properties such as molecular weight, charge, and lipophilicity (33, 34) of drug collections are profiled to extract simple counting rules for relevant descriptors of ADMET-related parameters. Examples include Lipinski's "rule-of-five" (33), which limits the range for molecular weight (MW ≤ 500), computed octanol-water partition coefficient (Clog P ≤ 5), and hydrogen-bond donors and acceptors (OHs + NHs ≤ 5; Ns + Os ≤ 10). Other authors limit the number of rotatable bonds (RB ≤ 8) or rings in a molecule (number of rings ≤ 4) (34). Table 6.1 shows a list of typical boundaries of counting parameters. Figure 6.1 illustrates the profiling procedure for these counting parameters using polar surface area (PSA) (35) as a descriptor. Collections of 776 orally administered CNS drugs and 1590 orally administered non-CNS drugs that reached phase II efficacy studies were analyzed for their PSA. It was found that 90% of the non-CNS compounds have a PSA below 120 Å^2; 90% of CNS drugs have a PSA below 80 Å^2. Although it is possible that drugs have higher PSA values and are still orally bioavailable or penetrate the blood-brain barrier (as the result of active transport or other reasons), the profile suggests that it is much less likely. It is therefore a reasonable assumption in a virtual screening approach to discriminate against compounds outside the most populated descriptor space (in this case, PSA