may not equal conventional synthetic approaches. The beginnings of combinatorial chemistry are attributed to Furka[67] with applications in peptide synthesis by Geysen and coworkers[68] and Houghten.[69] These initial efforts in peptide library synthesis were followed by synthesis of peptoids by Zuckermann and coworkers[70] and small molecule nonpeptide libraries by Ellman and coworkers[71] and Terrett and coworkers.[72]

b. Split Synthesis: Peptide Libraries

The initial approach, known as a *split synthesis* (also called *mix and split, split and pool*, or the *divide, couple, recombine* method), is the most common general lead discovery approach for making large libraries (10^4–10^6 compounds) that are assayed as library mixtures.[73] The result of a split synthesis is a collection of polymer beads, each containing one library member, i.e., one bead, one compound. The library contains every possible combination of every building block. The serious limitations are that it is applicable only to the synthesis of sequenceable oligomers and each bead carries only about 100–500 pmol of product, which makes structure determination difficult or impossible. For simple compounds mass spectrometric methods may be used,[74] but this is not applicable if the library contains many thousands or millions of members that may not be pure or are isomeric with other library members. In that case, encoding methods (see next section) need to be utilized.

Below is an example of how the split synthesis approach would be applied to a small (27-member) library of all possible tripeptides of three amino acids. This method can be extrapolated to any size library. A homogeneous mixture of all of the tripeptides of His, Val, and Ser ($3^3 = 27$) could be synthesized on a Merrifield resin as shown in Scheme 2.2. Note that a Merrifield synthesis starts at the C terminus and builds to the N terminus. The homogenization step is very important to ensure that each tube contains the same mixture of resin-bound compounds.

What if you want to determine the most active peptide for binding to a particular receptor? Houghten and coworkers[75] prepared a combinatorial library of more than 52 million L-[76] and D-hexapeptides[77] to identify the best hexapeptide antagonist of the μ opioid receptor (the receptor to which morphine and endorphins bind). The process shown in Scheme 2.2 was carried out, but starting with 20 separate tubes containing methylbenzhydrylamine (MBHA) polystyrene as the resin. (This resin produces peptide amides when peptides are cleaved from it.) A combinatorial library of pentapeptides containing the 20 standard amino acids was constructed on the MBHA resin, homogenized, then separated into 20 tubes. To each of the 20 different tubes was added a different N-acetylamino acid, so that in each tube there was a combinatorial library of all possible resin-linked N-acetylhexapeptides having the same N terminus; each tube contained all of the N-acetylhexapeptides starting with a different N-terminal amino acid (Figure 2.5A). For example, the first tube may contain all resin-linked N-acetylhexapeptides that have N-acetylalanine at the N terminus (this resin synthesis, as with the Merrifield synthesis, builds the peptide from the C terminus back to the N terminus), the second tube could contain all resin-linked N-acetylhexapeptides that start with N-acetylcysteine, and so forth. All of the N-acetylpeptides can be cleaved from the resin to give N-acetylhexapeptide amides. An aliquot from each of the 20 tubes is removed and assayed. The most potent aliquot indicates which amino acid is best at the N terminus (in this case N-acetylArg was found to be best). Then this process is repeated, except in the next iteration a combinatorial library of MBHA-bound tetrapeptides is made, is split into 20 tubes, a different amino acid is coupled in each tube at the next-to-N-terminal position, then each tube is N-terminal capped with N-acetylArg, because that was shown in the previous assay
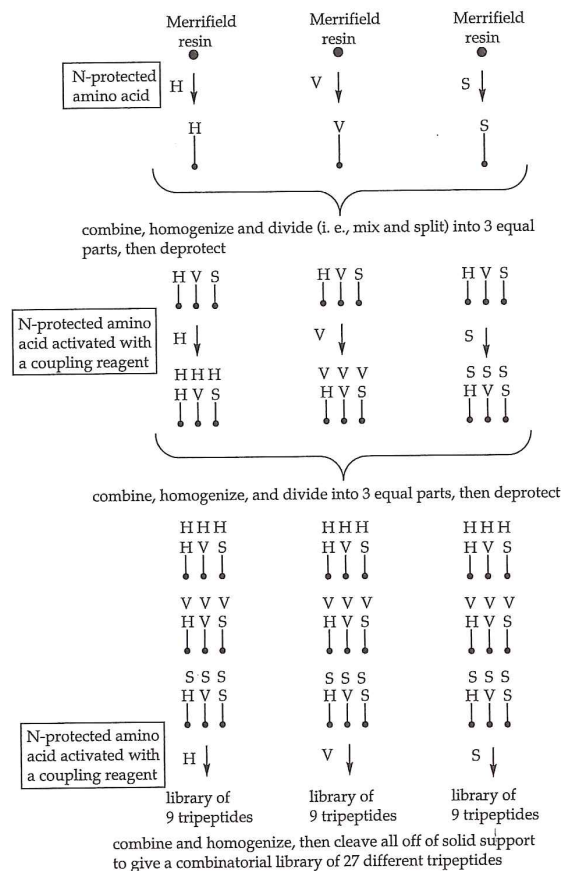
Scheme 2.2 ▶ Solid-phase synthesis of a combinatorial library of all possible tripeptides of histidine, valine, and serine

to be best (Figure 2.5B). Again, an aliquot from each tube, after cleavage from the resin, is assayed, and the best amino acid at the penultimate position is determined. This process is repeated until there are 20 tubes, each containing only one N-acetylhexapeptide, and the best of the 20 is determined by assay. Following this procedure, the most potent antagonist for the μ opioid receptor found up to that time was identified (Ac-Arg-Phe-Met-Trp-Met-Thr-NH$_2$). The entire process took about 2 months! Can you imagine how long this would take to carry out one compound at a time?

This methodology sounds foolproof, but it is not. Sometimes, the most potent aliquot is less potent than what was observed in the assays from the previous iteration. How is that possible, because the previous iteration had to have contained all of the peptides in the next iteration (and then some)? This is a common phenomenon with assaying multiple compounds simultaneously, particularly peptides. One explanation is peptide–peptide interactions. In the next iteration there are fewer peptides in the tube than in the previous assay; the necessary peptide–peptide interaction may be lost in the later assay. Possibly the active component is really two or more interacting peptides and one is removed in the next iteration. Or maybe
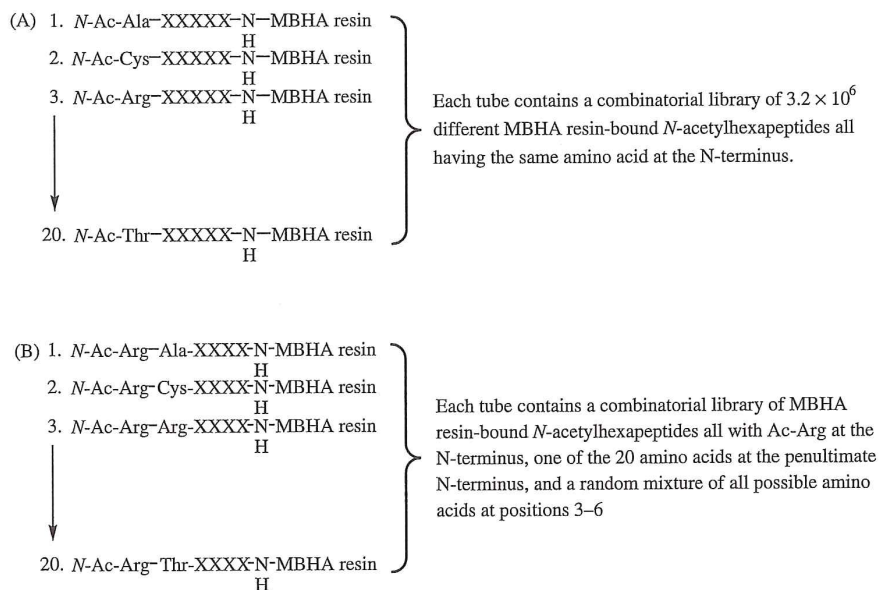
(A) 1. N-Ac-Ala—XXXXX—N—MBHA resin
                                  H
     2. N-Ac-Cys—XXXXX—N—MBHA resin
                                  H
     3. N-Ac-Arg—XXXXX—N—MBHA resin
                                  H

                  ↓

     20. N-Ac-Thr—XXXXX—N—MBHA resin
                                  H

Each tube contains a combinatorial library of 3.2×10^6 different MBHA resin-bound N-acetylhexapeptides all having the same amino acid at the N-terminus.

(B) 1. N-Ac-Arg—Ala-XXXX-N-MBHA resin
                                  H
     2. N-Ac-Arg-Cys-XXXX-N-MBHA resin
                                  H
     3. N-Ac-Arg-Arg-XXXX-N-MBHA resin
                                  H

                  ↓

     20. N-Ac-Arg-Thr-XXXX-N-MBHA resin
                                  H

Each tube contains a combinatorial library of MBHA resin-bound N-acetylhexapeptides all with Ac-Arg at the N-terminus, one of the 20 amino acids at the penultimate N-terminus, and a random mixture of all possible amino acids at positions 3–6

Figure 2.5 ▶ Combinatorial synthesis of all of the N-acetylhexapeptides of the 20 commonly encoded amino acids

there is a conformational difference in the active peptide as the number of peptides diminishes. Peptides generally do not make useful drugs, as discussed in Section 2.2.E.7, p. 47.

### c. Encoding Combinatorial Libraries

Before turning our attention to the more important nonpeptide libraries, let's consider an alternative approach to the identification of the most potent analog in a combinatorial library other than repeated iterations until one compound remains. A more rapid approach would be to test the entire library at once and identify the active component of the library directly. As mentioned above, with large libraries of complex molecules it is not readily possible to determine the structure of the active component. In that case, encoding methods are needed.[78] This is similar to the way in which proteins are often sequenced in biology; the protein is not sequenced, but the gene that encodes the protein is.[79] Although the structure of the actual compound may not be directly elucidated, certain tag molecules that encode the structure may be determined.[80] One important approach that involves the attachment of unique arrays of readily analyzable, chemically inert, small molecule tags to each bead in a split synthesis was reported by Still and coworkers.[81] Ideal encoding tags must survive organic synthesis conditions, not interfere with screening assays, be readily decoded without ambiguity, encode large numbers of compounds, and the test compound and the encoding tag must be able to be packed into a very small volume. In the Still method, groups of tags are attached to a bead at each combinatorial step in a split synthesis. The tags create a record of the building blocks used in that step. At the end of the synthesis, the tags are removed and analyzed, which decodes the structure of the compound attached to that bead. As depicted in Scheme 2.3, one or more readily cleavable tag molecules (TagsX) are attached to about 1% of the polymer bead sites (about 1 pmol/bead), and these encode building block 1 (BB1). Then
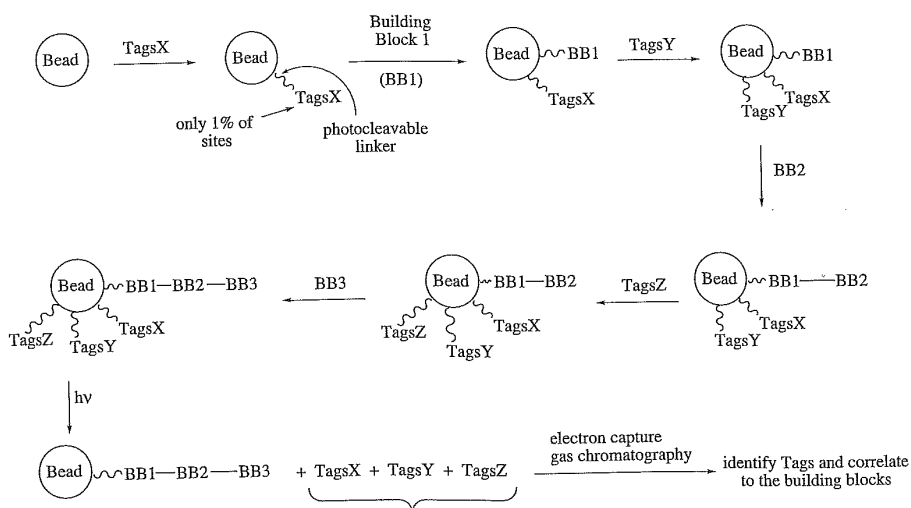
Scheme 2.3 ► Still methodology for encoding combinatorial peptide libraries on a polymer bead

one or more other cleavable tags (TagsY) are attached to encode building block 2 (BB2), followed by TagsZ to encode BB3, and so forth. Although a different tag could be used for each building block, it is more efficient to use mixtures of tags because mixtures of N different tags can represent 2^N different syntheses; with just 10 tags, 1024 (2^{10}) syntheses can be performed.

The tags need to be chemically inert and reliably analyzed on the femtomolar scale. Two examples of tags are **2.44** and **2.45**. The thirty **2.44** tags are photocleavable, and the forty **2.45** tags are oxidatively (ceric ammonium nitrate) cleaved. The released tags are analyzed by capillary gas chromatography using electron capture detection; all of the tags have different retention times. Other sensitive detection methods are fluorescence-based HPLC[82] and GC mass spectrometry.[83]
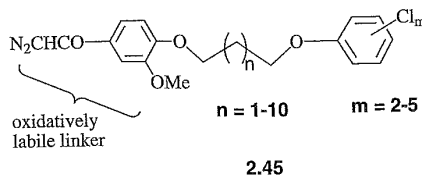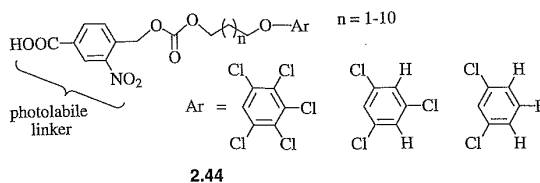
If you are using the split synthesis method, how will you know which of the large number of polymer beads has an active compound? One approach[84] is to chemically attach a commercial dye to the target receptor. The assay is run with the library still attached to each polymer bead. If a compound binds to the dye-labeled receptor, then the bead to which the compound is attached will take on the color of the dye, and the colored beads can be removed manually and decoded. The intensity of the color in the bead is an indication of the tightness of binding of the compound to the receptor.

Let's go through an example, which may make this process clearer (or maybe not). A combinatorial library of all of the hexapeptides of serine, leucine, lysine, isoleucine, and glutamate (5^6- or 15,625-member library) is prepared with the appropriate tag molecules at each iteration. To do this you need 6×3 (3-bit binary code) or 18 different tag molecules. Tags 1–3 are only used to define building block 1, tags 4–6 are for the second building block, tags 7–9 for the third, tags 10–12 for the fourth, tags 13–15 for the fifth, and tags 16–18 for the sixth building block. Arbitrarily assign a 3-bit binary code to each amino acid building block, for example, 001 = Ser, 010 = Leu, 011 = Lys, 100 = Ile, and 110 = Glu. If a tag is used, it represents binary bit 1; if no tag is used, it means binary bit 0. Because peptide syntheses are typically done on a resin to which the C-terminal amino acid is attached, and then coupling occurs back to the N terminus, tags 16–18 encode the N-terminal amino acid, and tags 1–3 encode the C-terminal amino acid. Let's say an active bead is identified, and you want to know what the hexapeptide structure is. The tags are removed from the bead (by photolysis or oxidation, depending on which linker was used), the carboxylic acid produced by cleavage is trimethylsilylated to make it more volatile, and the electron capture GC is run. In this example, let's say that tags 1, 2, 5, 6, 8, 9, 12, 14, and 16 were detected. These tags can be decoded to identify the hexapeptide as shown in Figure 2.6.

Another encoding process segregates the test compound from the coding tag molecule by attaching the test compound to the exterior of the bead and the coding tag molecule to the interior of the bead.[85] This prevents the tag molecules from interfering with binding of the test compound to the target receptor. A polymer bead system was developed in which only the surface of the bead is exposed to an organic solvent that contains the organic-soluble derivatizing reagent; this allows a nonpeptidic test compound to be constructed while the interior of the bead remains in water without derivatizing reagent. The coding molecule is a
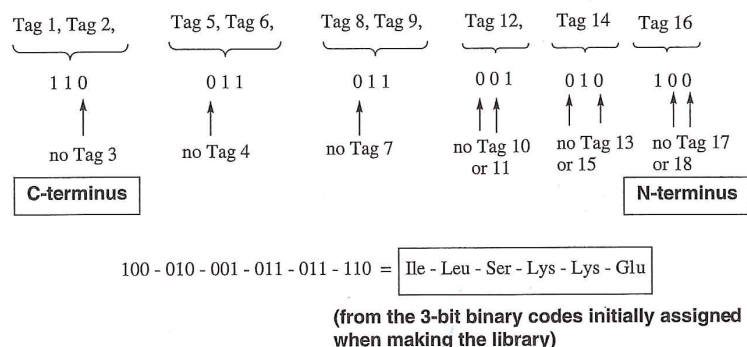

**Figure 2.6** ▶ An example of encoding the structure of a peptide bound to a polymer bead by the Still methodology

peptide, which can be synthesized in the aqueous medium of the interior of the bead. Once the active bead is identified by a colorimetric assay,[86] the tag peptide molecule in the interior of that bead is sequenced by Edman degradation or by mass spectrometry,[87] which encodes the structure of the test compound attached to the exterior of the bead.

Other interesting alternatives to molecular tag encoding are encoding by radio-frequency tags[88] and encoding with a polymeric matrix having unusual shapes.[89] Encoding methods may some day be displaced by mass spectral analyses, such as imaging time-of-flight secondary ion mass spectrometry (TOF-SIMS).[90] Many hundreds of beads can be assayed in a single measurement at the rate of about 10 beads/sec, although there is a problem with fragmentation of compounds, leading to a complicated analysis. Mass accuracy is about ± 0.01 amu, so in the 3.2×10^6-member library of pentapeptides from 20 amino acids, all peptides, except those containing leucine and isoleucine (same molecular weight), can be separated. By incorporation of a ^{15}N label into either leucine or isoleucine, even those peptides can be differentiated.

d. Nonpeptide Libraries

As discussed later in Section 2.2.E.7, p. 47, peptides do not make very useful drugs, especially if an orally active drug is sought. The same techniques described above for the synthesis of peptide libraries could be utilized to prepare nonpeptide libraries; however, there is an important difference between the chemistry with peptides versus nonpeptides, namely, reactivity. In a typical peptide coupling reaction the carbodiimide-activated N-protected amino acids are all about the same in reactivity with the different amino acids in the growing peptide chain. Because of that, the split synthesis method works well. However, with nonamino acid reagents, such as different acid chlorides, the structure of the acid chloride will affect the rates of reaction with different nucleophiles. That could lead to mixtures in which some of the components have reacted and others have not. For each reaction the conditions have to be worked out to be sure complete reaction has occurred.

Over the years it has been recognized that when large numbers of nonpeptide analogs are screened simultaneously, many *false negatives* (an active compound that does not produce a *hit*, i.e., a compound that shows a predetermined level of activity in the assay) and *false positives* (an inactive compound that gives a hit) are observed. A false positive may arise from an impurity in the sample tested or as a result of a complex between more than one compound. False positives are a waste of time, but false negatives mean that potential drugs (or at least lead compounds) are being overlooked. It is typical for pharmaceutical companies to carry out single entity screens to avoid these problems. Because of this, individual compounds, rather than mixtures, are synthesized. Nonetheless, synthesis on a solid support allows the synthesis of large numbers of individual compounds rapidly and robotically. The reactions are carried out individually in separate microtubes containing the polymeric support. This method is referred to as *parallel synthesis* rather than combinatorial synthesis because the library of compounds (in the range of $50-10^4$ compounds in amounts of 1–50 mg) is synthesized in parallel without combining any of the tubes. One strategy that can be used for potentially more effective libraries is to select privileged structures as the scaffold. Another strategy is to design a scaffold based on an important molecular recognition motif in the target receptor. The libraries should incorporate different sets of (commercially available) building blocks to provide a large number of diverse structures, and they should contain as much functionality as possible as recognition elements. Molecular diversity, however, is difficult to determine; Dixon and Villar have found that a protein can bind a set of structurally diverse molecules with
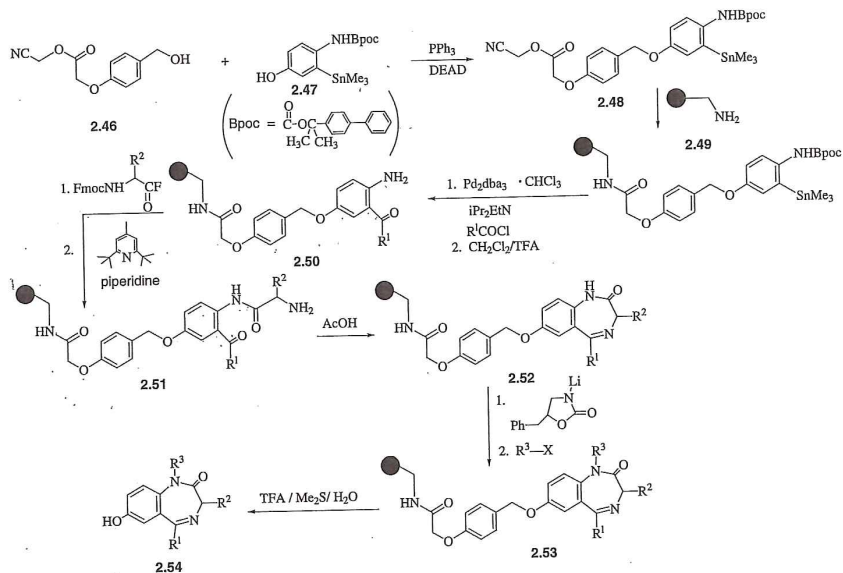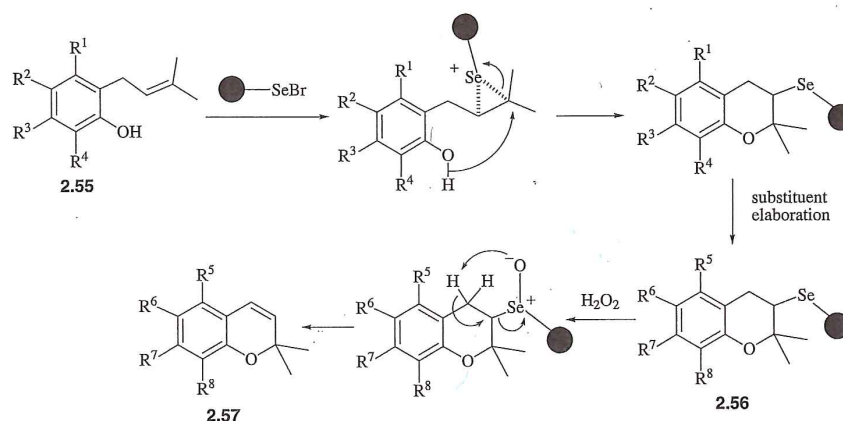
Scheme 2.4 ▶   Solid-phase synthesis of a nonpeptide library of privileged structures

similar potent binding affinities, but analogs closely related to these compounds can exhibit very weak binding.[91] Parallel synthesis can generate many more compounds than can be synthesized traditionally, and the cost per compound is much lower.[92]

An example of a nonpeptide library of a privileged structure (benzodiazepines) is shown in Scheme 2.4.[93] Note that the first piece of the benzodiazepine (**2.47**) is not attached directly to the polymer (aminomethylpolystyrene, **2.49**), but is attached to **2.46** instead to give **2.48** in a Mitsunobu coupling (PPh$_3$/diethylazodicarboxylate or DEAD). If **2.47** were attached directly to the polymer, then steric hindrance by the polymer to the first chemical reaction may result, i.e., the Stille coupling of an acid chloride to the aryl stannane in the presence of palladium to give **2.50**. To avoid that problem, a spacer group is typically attached to the polymer which moves the first reactant away from the polymer so that steric hindrance is not a problem. Compound **2.46** serves as the spacer, which will be removed at the end of the synthesis. In this solid-phase synthesis, three *diversity elements* can be varied: R^1, R^2, and R^3. The Stille coupling can be carried out with as many acid chlorides as are available (to vary R^1). Each of those products (**2.50**) can be coupled to the same Fmoc-protected amino acid fluoride in the next step or each of the **2.50** products can be treated with different Fmoc-protected amino acid fluorides, so that a wide variety of R^2 groups can be incorporated. Acetic acid causes autocyclization of **2.51** to **2.52**. Again, each **2.52** can be treated with base and the same alkyl halide (R^{3}X) or each of **2.52** can be treated with a different alkyl halide to give a library of **2.53**; cleavage from the spacer and polymer resin gives the library of benzodiazepines (**2.54**). If 25 different acid chlorides were used in the Still coupling, and each of those products was treated with 25 different Fmoc-protected amino acid fluorides, then each of those compounds alkylated with 25 different alkyl halides, there would be a library of $25 \times 25 \times 25$ or 15,625 different benzodiazepines using only 75 different building blocks.

Despite the potential of combinatorial library (or parallel) synthesis, natural products seem to provide greater structural diversity than standard combinatorial chemistry. About 40% of the chemical scaffolds in the Dictionary of Natural Products and the Bioactive Natural Products Database (a total of more than 100,000 compounds) have not been synthesized.[94] Furthermore, natural products that are biologically active in assays generally have drug-like properties, i.e., are capable of being absorbed and metabolized.[95a] Isolation of compounds from natural sources and from combinatorial approaches, however, should be complementary. Generally the excessive time taken to isolate and characterize bioactive compounds from natural product extracts is a disadvantage of the method, but the reward is greater molecular diversity,[95b] which gives the greatest opportunity to identify a variety of scaffolds for screening. To attain a wide diversity of chemical structures for screening purposes, computational chemists often reject compounds that are similar in structure, believing that similar compounds would have similar biological activities. In general, structurally similar compounds have similar biological activity. However, the biological similarity may not be very strong; it has been shown that only 30% of compounds considered to be 85% structurally similar to an active compound will themselves have the same activity.[96] Adding just one methylene group to a 4-hydroxypiperidine analog changed it from a poor binder of the chemokine receptor CCR1 into a potent binder.[97] This may be because similar compounds do not necessarily bind to the target receptor the same way.

Construction of chemical libraries based on natural product hits is a sensible compromise approach. For example, Nicolaou and coworkers developed a solid-phase method for the preparation of large natural product-like combinatorial libraries based on the privileged structure 2,2-dimethylbenzopyran,[98] a scaffold found in many natural products. The general methodology is shown in Scheme 2.5. In this example the 2,2-dimethylbenzopyran scaffold (2.57) is generated by reaction of the starting o-prenyl phenol (2.55) with a polystyrene-based selenenyl bromide resin. A variety of reactions are possible to elaborate the side chains to further enhance the library (2.56). The final products can be released from the polymeric support by oxidation/elimination to 2.57. A library of more than 10,000 analogs was readily prepared by this approach.

Scheme 2.5 ▶ Solid-phase synthesis of a natural product-like combinatorial library

E.6   SAR by NMR/SAR by MS

Fesik and coworkers at Abbott Laboratories developed a NMR-based approach to screen libraries of small organic molecules and to identify and optimize high-affinity *ligands* (compounds that bind to receptors) for proteins.[99] This approach, termed *SAR by NMR*, was initially used to discover compounds with nanomolar affinities (highly potent; see Chapter 3, Sections 3.2.A and 3.2.C) for the immunosuppressant FK506 binding protein by tethering two molecules with micromolar affinities (low potency). The first step of the process (Figure 2.7) involves screening a library of small compounds, 10 at a time, by observation of the amide ^{15}N-chemical shift in the heteronuclear single quantum coherence (HSQC) NMR spectrum. Once a lead is identified, a library of analogs is screened to identify compounds with optimal binding at that site. Then a second library of compounds is screened to find a compound that binds at a nearby site, and again this compound is optimized by screening a library of related compounds. Based on the NMR spectrum of the ternary complex of the protein and the two bound ligands, the location and orientation of these ligands are determined, and compounds are synthesized in which the two ligands are covalently attached. Although each individual ligand may be a relatively weak binder, when the two are attached, the binding affinity increases dramatically. This is because the free energy of binding becomes the sum of three free energies: the two ligands and the linker; the binding affinity is the multiplier of the three binding affinities. There is a gain of about a factor of 100 in binding affinity by freezing out one bond rotation. Therefore, it is not necessary to optimize the lead much, because ligands with micromolar or even millimolar affinities can attain nanomolar affinities when linked.

An example of this is the identification of the first potent inhibitor of the enzyme stromelysin, a *matrix metalloprotease* (a family of zinc-containing hydrolytic enzymes responsible for degradation of extracellular matrix components such as collagen and proteoglycans in normal tissue remodeling and in many disease states such as arthritis, osteoporosis, and cancer),[100] as a potential antitumor agent.[101] Matrix metalloproteases are generally inhibited by compounds that contain a hydroxamate moiety to bind to the zinc ion. A library of hydroxamates was screened, and acetohydroxamic acid (**2.58**) was identified with a K_d of 17 mM (very poor binding affinity). A focused screen of hydrophobic compounds was carried out in the presence of saturating (excess) amounts of acetohydroxamic acid, and biphenyl analogs were identified; optimization led to **2.59** with a K_d of 20 μM. From the NMR spectrum, the best site for a linker was expected to be between the methyl of acetohydroxamic acid and the hydroxyl group of **2.59**. Consequently, alkyl linkers of varying chain length were
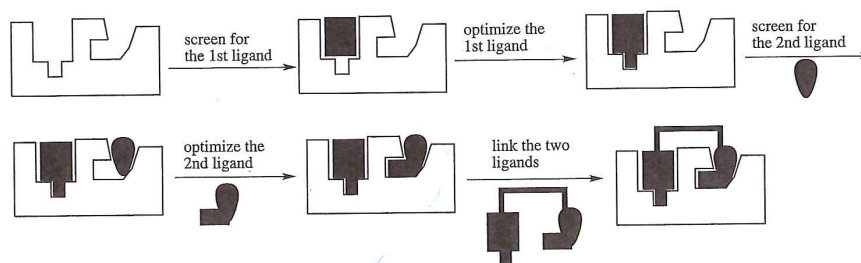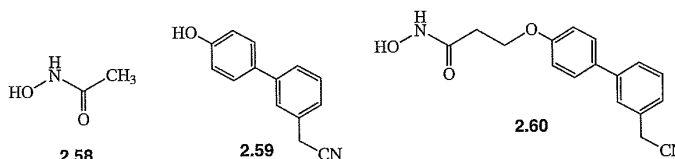

Figure 2.7 ▶ SAR by NMR methodology

tried, and the best was a one-carbon linker, giving **2.60** having a K_d of 15 nM! The ΔG for **2.59** is -2.4 kcal/mol, for **2.60** is -4.8 kcal/mol, and for the linker is -2.6 kcal/mol; the total, therefore, is -9.8 kcal/mol. It took about 6 months to identify this inhibitor; prior to this study 115,000 compounds had been screened with no leads.

Sounds simple, doesn't it? But let's think about what is involved. The method requires screening compounds and observing a specific ^{15}N-amide chemical shift for binding. Where did the ^{15}N come from? This had to be incorporated into the protein because natural abundance ^{15}N is not sufficiently high in concentration for detection. To incorporate ^{15}N, it is necessary to be able to express the protein in a microorganism, then grow the microorganism on [15]NH₄Cl as its sole nitrogen source; that gives the protein with all ^{15}N-containing amino acids. To perform the NMR experiments, large amounts of soluble ($>100\,\mu$M) protein (>200 mg per spectrum) are needed; therefore, an efficient overexpression system for the protein is needed. Then the protein has to be purified, and its complete structure determined by three- and four-dimensional NMR techniques, so that the position of every amino acid residue in the protein is known (which is needed to determine when the two ligands have bound in nearby sites). This means that the protein target should have a mass less than about 40 kDa (the current limit for rapid protein NMR spectra, although spectra of larger proteins is possible).[102] Although it appears that this is a highly specialized technique, it is used widely because molecular biology and protein chemistry techniques are well developed, making overexpression of proteins in microorganisms and their purification routine.[103] Newer NMR instrumentation and methods also have made structure determination plausible. If the structure can be determined, SAR by NMR provides a technique to screen by automation about 1000 compounds a day and identify, relatively rapidly, potent protein binders.[104] Even covalent binders can be identified by high-throughput NMR-based screens.[105] In drug discovery programs, it is often not too difficult to find compounds that bind to proteins in the micromolar range; what becomes time consuming is increasing the potency of the lead into the low nanomolar range. SAR by NMR may shorten that time.

Ellman and coworkers have developed a combinatorial lead optimization approach using the basic principles of SAR by NMR, except without the use of NMR and without needing any structural or mechanistic information about the target protein![106] First, a diverse library of compounds is synthesized in which each molecule incorporates a common chemical linkage group (Figure 2.8). Next, the library is screened to identify any members that show even weak binding to the target. Third, a new library is constructed containing all combinations of any two of the active compounds linked to each other by the common chemical linkage group through a set of flexible linkers. Then this combinatorial library is screened to identify the most potent analog. The method depends on two analogs binding in nearby sites (although it is not known which two will bind or where the sites are) and finding the appropriate linker size combinatorially so the linked active compounds take advantage of the additive free-energy gain of the three elements, the two compounds and the linker. This approach was used to identify a potent ($IC_{50} = 64$ nM) and selective inhibitor of one subtype of tyrosine kinase.
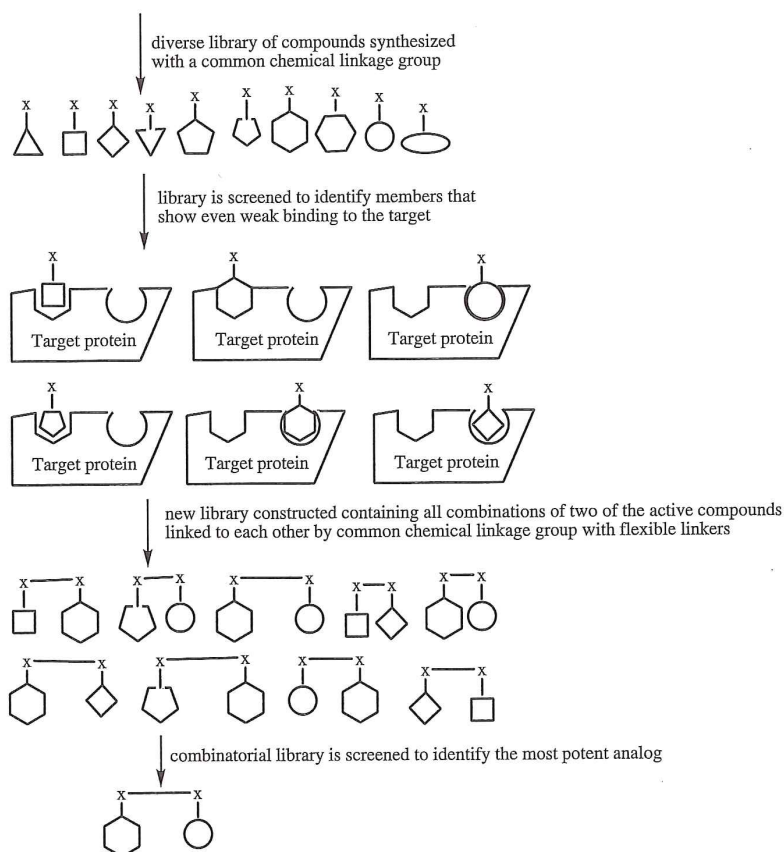
Figure 2.8 ▶ Ellman combinatorial methodology for lead generation with an unknown or impure receptor or enzyme

A complementary method to SAR by NMR is *SAR by MS*.[107] This is a high-throughput mass spectrometry-based screen that quantifies the binding affinity, stoichiometry, and specificity over a wide range of ligand binding energies. A set of diverse compounds is screened by mass spectrometry to identify those that bind to the receptor. Competition experiments are used to identify the ones that bind to the same site and those that do not. If two compounds bind at different binding sites, then a ternary complex of the two molecules plus the receptor is detected in the mass spectrum. If the two compounds bind at the same site, the tighter binding molecule displaces the other from the binding site, and only a binary complex is detected. By varying the substituent size on various classes of compounds and rescreening, it is possible to identify those molecules that bind at nearby sites as the ones that become competitive once a larger substituent is appended to one of the molecules. Once adjacent binding sites are realized, then the same methodology can be employed as in SAR by NMR, namely, attaching the two molecules to each other with linkers. This approach was applied to the discovery of lead compounds that bind to an RNA target for which no leads could be identified by conventional high-throughput screening methods.[108]

E.7 Peptidomimetics

In Section 2.2.E.5 on combinatorial chemistry, methods for the synthesis of libraries of peptides were described. Peptides are very important endogenous molecules that bind to a variety of receptors in their action as neurotransmitters, hormones, and neuromodulators,[109] and numerous enzymes are involved in the biosynthesis and catabolism of these peptides. Plants and animals,[110] including human skin,[111] contain a variety of antibiotic peptides. Endogenous peptides also function as analgesics,[112] antihypertensive agents,[113] and antitumor agents.[114] However, peptides do not make good drug candidates because they are rapidly proteolyzed in the GI tract and serum, and they are poorly bioavailable, rapidly excreted, and can bind to multiple receptors. What is needed is a compound that mimics or blocks the biological effect of a peptide by interacting with its receptor or enzyme, but does not have the undesirable characteristics of peptides; these are *peptidomimetics*.

Earlier in this chapter (Section 2.2.A, p. 17) morphine and morphine analogs (**2.21**) were discussed as potent binders to the μ opioid receptor. In the mid-1970s Hughes and coworkers[115] showed that the endogenous peptides, the enkephalins and β-endorphin, also bound to the same site on the opioid receptor as did morphine. A remarkable resemblance was demonstrated between the N-terminal tyrosine structure of these opioid peptides and the morphine phenol ring system, which suggested why they all interacted with these receptors in a similar way.[116] Farmer then proposed that this may be a general phenomenon and that other nonpeptide structures may mimic natural peptide effectors.[117] His postulate was that peptide mimetics (which later became "peptidomimetics") could be designed that would replace peptide backbones while retaining the appropriate topography for binding to a receptor; this initiated the field of peptidomimetics.

The design of peptidomimetics can be a lead optimization approach, which uses the desired peptide as the lead compound and modifies it to minimize (or preferably, eliminate) the undesirable pharmokinetic properties. The generation of peptidomimetics is based on the conformational, topochemical, and electronic properties of the lead peptide when bound to its target receptor or enzyme.[118] The goal is to replace as much of the peptide backbone as possible with nonpeptide fragments while still maintaining the pharmacophoric groups (usually the amino acid side chains) of the peptide. This makes the compound more lipophilic, which increases its bioavailability. Replacement of the amide bond with alternative groups prevents proteolysis and promotes metabolic stability. Initially, conformational flexibility has to be retained to allow the pharmacophoric groups a better opportunity to find their binding sites, but further lead refinement should favor the formation of more conformationally restricted analogs that hold appropriate pharmacophoric groups in the bioactive conformation for binding to the target receptor.[119]

Increased lipophilicity and conformational modification of amino acids can be designed into the peptidomimetic. These groups may not be recognized by peptidases. For example, conformationally restricted analogs of phenylalanine shown in Figure 2.9[120] can be incorporated into peptidomimetic receptor ligands. Likewise, conformational restriction and lipophilicity can be incorporated into peptides (Figure 2.10).[121]

Another approach involves the design of conformationally restricted analogs that mimic characteristics of the receptor-bound conformation of the endogenous peptide,[122] such as β-turns (**2.61**,[123] **2.62**,[124] Figure 2.11), α-helices (**2.63**),[125] Ω-loops (**2.64**),[126] and β-strands (**2.65**).[127] This idea can be extended to *scaffold peptidomimetics* in which important pharmacophoric residues are held in the appropriate orientation by a rigid template. Compounds that block the binding of fibrinogen to its receptor (glycoprotein IIb/IIIa) can prevent platelet aggregation and are of potential value in the treatment of strokes and heart attacks.[128]
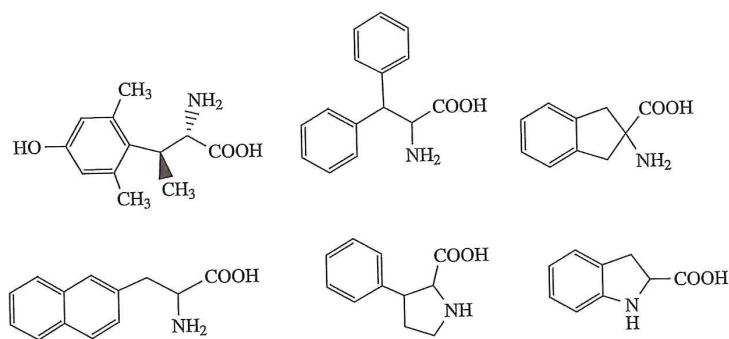
Figure 2.9 ▶ Conformationally restricted phenylalanine analogs
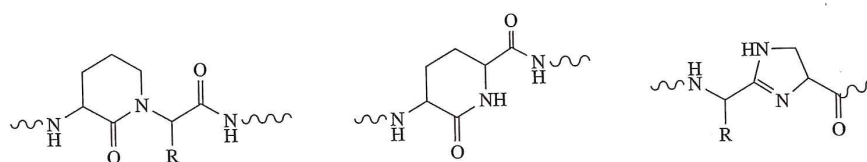
Figure 2.10 ▶ Conformationally restricted peptide analogs

2.61

2.62

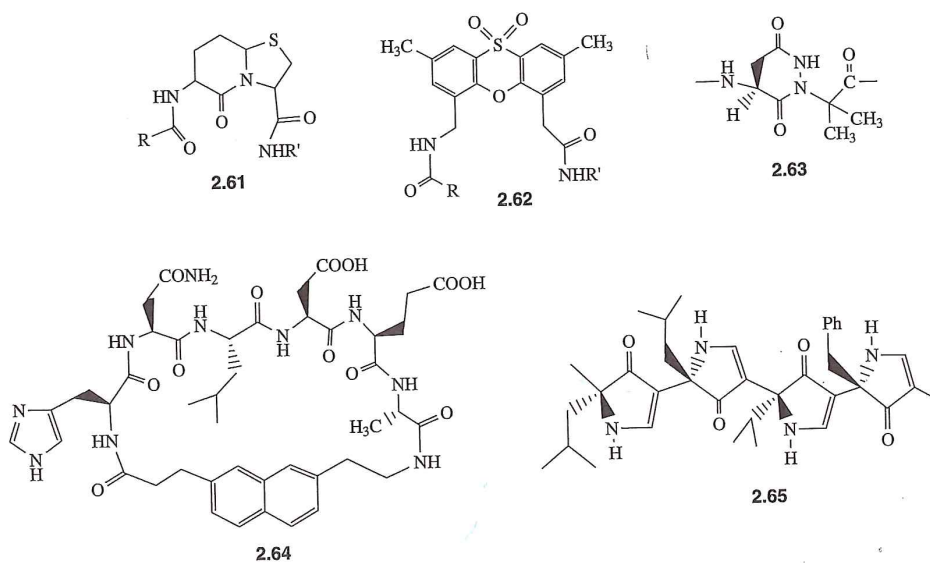**2.63**

2.64

2.65

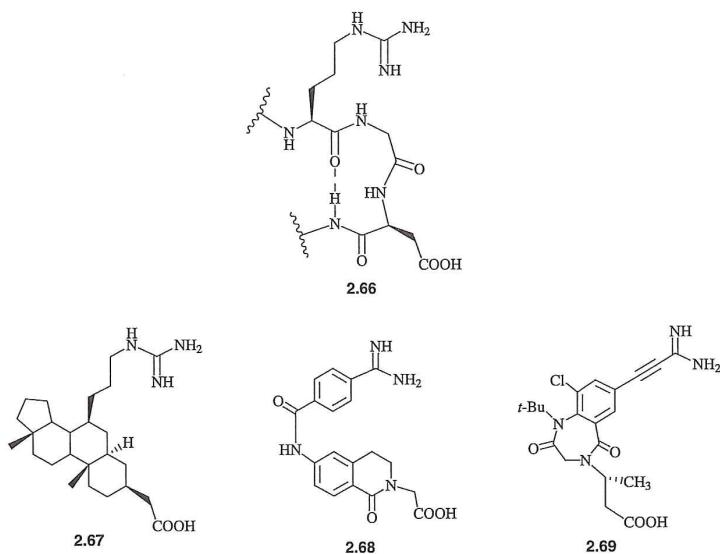Figure 2.11 ▶ Conformationally restricted secondary structure peptidomimetics

Figure 2.12 ▶ RGD scaffold peptidomimetics

A common β-turn motif that has been found to bind to GPIIbIIIa is arginine-glycine-aspartic acid (or in the one-letter amino acid code, RGD, **2.66**, Figure 2.12). Consequently, a variety of scaffold peptidomimetics for RGD have been designed based on the hypothesis that the glycine residue only represents a spacer between the two important recognition residues, arginine and aspartate. Several potent binders to this receptor have been found by replacement of the glycine with more rigid mimics, such as steroid (**2.67**),[129] isoquinolone (**2.68**),[130] and benzodiazepinedione (**2.69**)[131] spacers. A β-D-glucose-based nonpeptide scaffold (**2.71**) was designed as a mimic (note the darkened groups) of the potent somatostatin agonist (see Chapter 3, Section 3.2.C) **2.70**.[132] A target peptide for the treatment of cognitive disorders, such as Alzheimer's disease, is thyrotropin-releasing hormone (TRH, pyroGlu-His-ProNH$_2$, **2.72**);[133] a scaffold peptidomimetic for this hormone is **2.73**.[134]
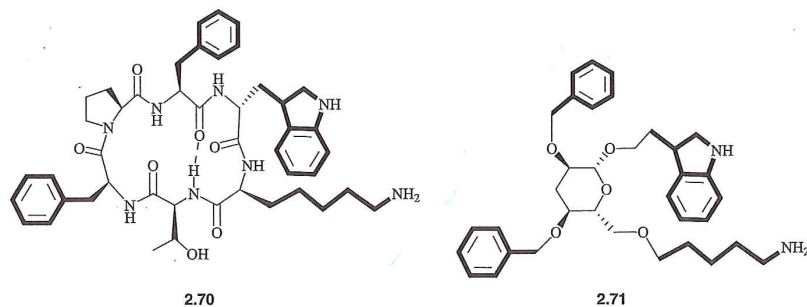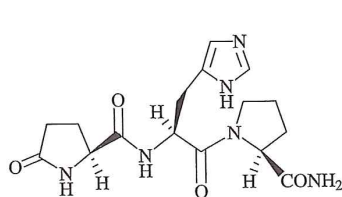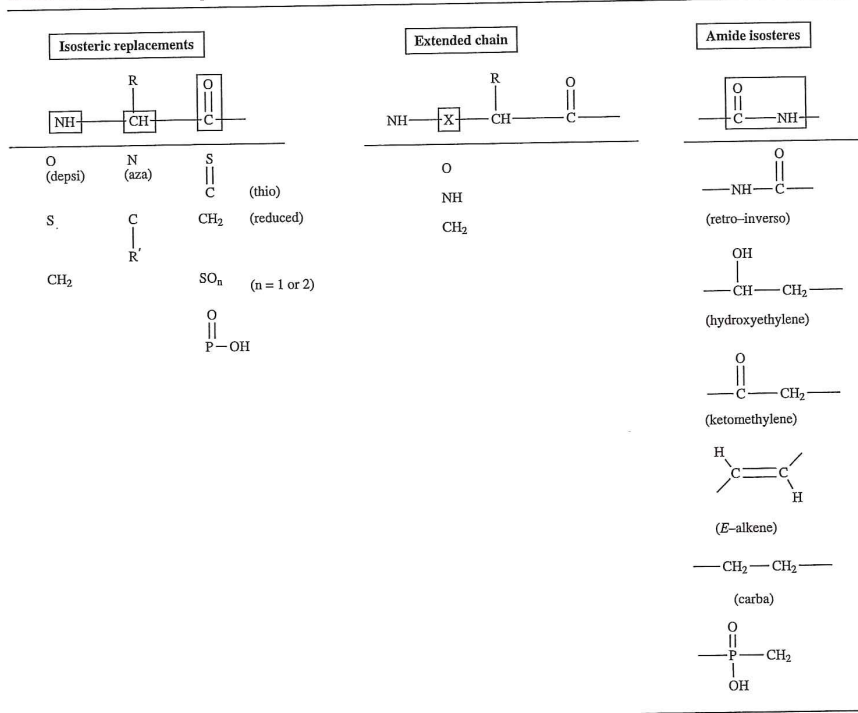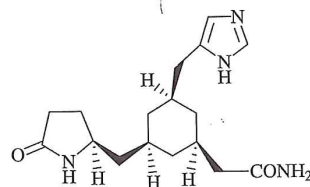
TABLE 2.5 ▶ Peptide Backbone Isosteres for Peptidomimetics

A common and important approach for the conversion of a peptide lead into a peptidomimetic is the use of peptide backbone isosteres (Table 2.5). Peptides in which the amide bonds are replaced with alternative groups are known as *pseudopeptides*.[135] These isosteric replacements remove the peptide linkage (thereby stabilizing the peptidomimetics to metabolism) and/or make them less polar and more lipophilic. The hydroxymethylene (also called statine)[136] isostere is one of the early mimetics used in the design of inhibitors of proteases, particularly of HIV protease.[137] Other variants of azapeptides (**2.74**, in which one or more of the α-carbons are replaced by N)[138] include azatides (**2.75**, azapeptides in which *all* of the α-carbons are replaced by N)[139] and peptoids (**2.76**, in which the α-CHR groups are replaced by NR units and the NH groups are replaced by CH_2 units, i.e., an amino acid sequence that is opposite that in peptides).[140]

2.74 **2.75** **2.76**
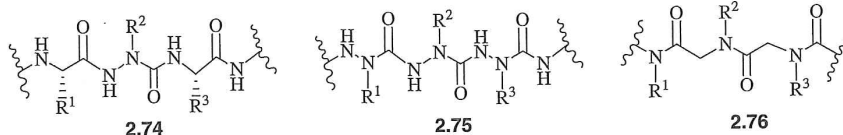
2.2.F Structure Modifications to Increase Oral Bioavailability

About three-quarters of drug candidates do not make it to clinical trials because of problems with pharmacokinetics in animals.[141] Less than 10% of drug candidates entering clinical trials become marketed products. About 40% of the molecules that fail in clinical trials do so because of pharmacokinetic problems, such as poor oral bioavailability or short plasma half-lives.[142] Because of the huge waste of time and resources by having a drug candidate fail late in the drug discovery process, a more recent trend is to examine pharmacokinetic aspects of molecules as early as possible in this process.[143] The use of mass spectrometry for this purpose is discussed in Chapter 7 (Section 7.3.C).

Low water solubility of a compound (high lipophilicity) can be a limiting factor in oral bioavailability,[144] and highly lipophilic compounds also are easily metabolized (see Chapter 7) or bind to plasma proteins. However, low lipophilicity is typically more of a problem, because that leads to poor permeability through membranes. Membrane permeability for a number of drugs is known.[145] In this section I try to assess how to incorporate better pharmacokinetic properties into lead modification design.

Several of the lead modification approaches discussed earlier were directed at improving both pharmacodynamics as well as pharmacokinetics, such as homologation, chain branching, ring-chain transformations, and bioisosterism. Increases in potency *in vivo* using these approaches could be explained either by pharmacodynamics (enhanced binding to a receptor) or by pharmacokinetics (increased lipophilicity, leading to improved absorption and distribution).

Because of the importance of lipophilicity in drug design,[146] it is essential to understand not only how to determine lipophilicities of compounds but also how to determine lipophilicities of substituents so that the correct substituent can be selected in lead modification approaches. The basis for the determination of the lipophilicities of substituents, as presented by Corwin Hansch and coworkers,[147] is derived from the earlier postulate by L. P. Hammett on how the electronic effects of substituents affect the reactivity of organic molecules, known as the Hammett equation. Those of you who know how to derive this equation can skip the next section.

F.1 Electronic Effects: The Hammett Equation

Hammett's postulate was that the electronic effects (both the inductive and resonance effects) of a set of substituents should be similar for different organic reactions. Therefore, if values could be assigned to substituents in a standard organic reaction, these same values could be used to estimate rates in a new organic reaction. This was the first approach that allowed the prediction of reaction rates. Hammett chose benzoic acids as the standard system.

Consider the reaction shown in Scheme 2.6. Intuitively, it seems reasonable that as X becomes electron withdrawing (relative to H), the equilibrium constant (K_a) should increase (the reaction should be favored to the right) because X is inductively pulling electron density

$$X \!\!-\!\!\text{CO}_2\text{H} + \text{H}_2\text{O} \; \underset{}{\overset{K_a}{\rightleftharpoons}} \; X \!\!-\!\!\text{CO}_2^- + \text{H}_3\text{O}^+$$
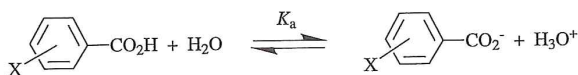
Scheme 2.6 ▶ Ionization of substituted benzoates

$$X \!\!-\!\!\text{CO}_2\text{Et} + \text{HO}^- \; \underset{}{\overset{k}{\rightleftharpoons}} \; X \!\!-\!\!\text{CO}_2^- + \text{EtOH}$$

Scheme 2.7 ▶ Saponification of substituted ethyl benzoates
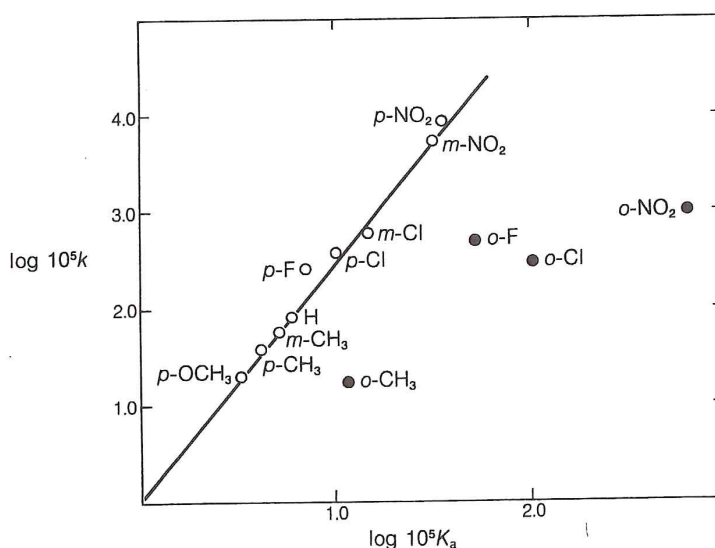
Figure 2.13 ▶ Linear free-energy relationship for the dissociation of substituted benzoic acids in water at 25°C (K_a) against the rates of alkaline hydrolysis of substituted ethyl benzoates in 85% ethanol-water at 30°C (k). [Reprinted with permission from Roberts, J. D. and Caserio, M. C. (1977). *Basic Principles of Organic Chemistry*, 2nd ed., p. 1331. W. A. Benjamin, Menlo Park, CA. Copyright ©1977 Benjamin/Cummings Publishing Company.]

from the carboxylic acid group, making it more acidic (reactant argument); it also is stabilizing the negative charge on the carboxylate group of the product (product argument). Conversely, when X is electron donating, the equilibrium constant should decrease. A similar relationship should exist for a rate constant (k); an electron-withdrawing substituent would stabilize a negative change in the transition state, thereby lowering the activation energy, and increasing the rate, and an electron-donating group would destabilize the transition state, decreasing the rate. Hammett chose the reaction shown in Scheme 2.7 as the standard system to determine electronic effects of substituents on the rate constant of a reaction.

If K_a is measured from Scheme 2.6 and k from Scheme 2.7 for a series of substituents X, and the data are expressed in a double logarithm plot (Figure 2.13), then a straight line can be drawn through most of the data points. This is known as a *linear free-energy relationship*. When X is a meta- or para-substituent, then virtually all of the points fall on the straight line;

the ortho-substituent points are badly scattered. The initial Hammett relationship does not hold for ortho-substituents because of steric interactions and polar effects. The linear correlation for the meta- and para-substituents is observed for rate and equilibrium constants for a wide variety of organic reactions. The straight line can be expressed by Equation 2.3,

$$\log k = \rho \log K + C \tag{2.3}$$

where the two variables are $\log k$ and $\log K$. The slope of the line is ρ, and the intercept is C. When there is no substituent, i.e., when X = H, then Equation 2.4 holds:

$$\log k_0 = \rho \log K_0 + C \tag{2.4}$$

Subtraction of Equation 2.4 from Equation 2.3 gives Equation 2.5,

$$\log k/k_0 = \rho \log K/K_0 \tag{2.5}$$

where k and K are the rate and equilibrium constants, respectively, for compounds with a substituent X, and k_0 and K_0 are the rate and equilibrium constants, respectively, for the parent compound (X = H). If $\log K/K_0$ is defined as σ, then Equation 2.5 reduces to Equation 2.6, the *Hammett equation*:

$$\log k/k_0 = \rho \sigma \tag{2.6}$$

The *electronic parameter*, σ, depends on the electronic properties and position of the substituent on the ring and, therefore, is also called the *substituent constant*. The more electron withdrawing a substituent, the more positive its σ value (relative to H, which is set at 0.0); conversely, the more electron donating, the more negative its σ value. The meta σ constants result from inductive effects, but the para σ constants correspond to the net inductive and resonance effects. Therefore, σ_{meta} and σ_{para} for the same substituent, generally, are not the same.

The ρ values (the slope) depend on the particular type of reaction and the reaction conditions (e.g., temperature and solvent) and, therefore, are called *reaction constants*. The importance of ρ is that it is a measure of the sensitivity of the reaction to the electronic effects of the meta- and para-substituents. A large ρ, either positive or negative, indicates great sensitivity to substituent effects. Reactions that are favored by electron donation in the transition state (such as reactions that proceed via carbocation intermediates) have negative ρ values (i.e., the linear free-energy relationship has a negative slope); reactions that are aided by electron withdrawal (such as reactions that proceed via carbanion intermediates) have positive ρ values.

F.2 Lipophilicity Effects

a. Importance of Lipophilicity

Hansch believed that, just as the Hammett equation relates the electronic effects of substituents to reaction rates, there should be a linear free-energy relationship between lipophilicity and biological activity. Hansch proposed that the first step in the overall drug process was a random walk, a diffusion process, in which the drug made its way from a dilute solution outside of the cell to a particular site in the cell. This was visualized as being a relatively slow process, the rate of which is highly dependent on the molecular structure of the drug. For the drug to reach the site of action, it must be able to interact with two different environments, lipophilic (e.g., membranes) and aqueous (the exobiophase, such as the cytoplasm). The cytoplasm of a cell is

essentially a dilute solution of salts in water; all living cells are surrounded by a nonaqueous phase, the membrane. The functions of membranes are to protect the cell from water-soluble substances, to form a surface to which enzymes and other proteins can attach to produce a localization and structural organization, and to separate solutions of different electrochemical potentials (e.g., in nerve conduction). One of the most important membranes is known as the *blood–brain barrier*, a membrane that surrounds the capillaries of the circulatory system in the brain and protects it from passive diffusion of undesirable polar chemicals from the bloodstream. This is an important prophylactic boundary, but it also can block the delivery of central nervous system drugs to their site of action.

Although the structure of membranes has not been resolved, the most widely accepted model is the fluid mosaic model (Figure 2.14).[148] In this depiction integral proteins are embedded in a lipid bilayer; peripheral proteins are associated with only one membrane surface. The structure of the membrane is primarily determined by the structure of the lipids of which it is comprised. The principal classes of lipids found in membranes are neutral cholesterol (**2.77**) and the ionic phospholipids, e.g., phosphatidylcholine (**2.78**, $R = CH_2CH_2N(CH_3)_3^+$, phosphatidylethanolamine (**2.78**, $R = CH_2CH_2NH_3^+$), phosphatidylserine (**2.78**, $R = CH_2CH(NH_3^+)COO^-$), phosphatidylinositol (**2.78**, R = inositol), and sphingomyelin (**2.79**, $R = {}^-OPO_2CH_2CH_2N(CH_3)_3^+$; R′CO and R″CO in **2.78** and **2.79**
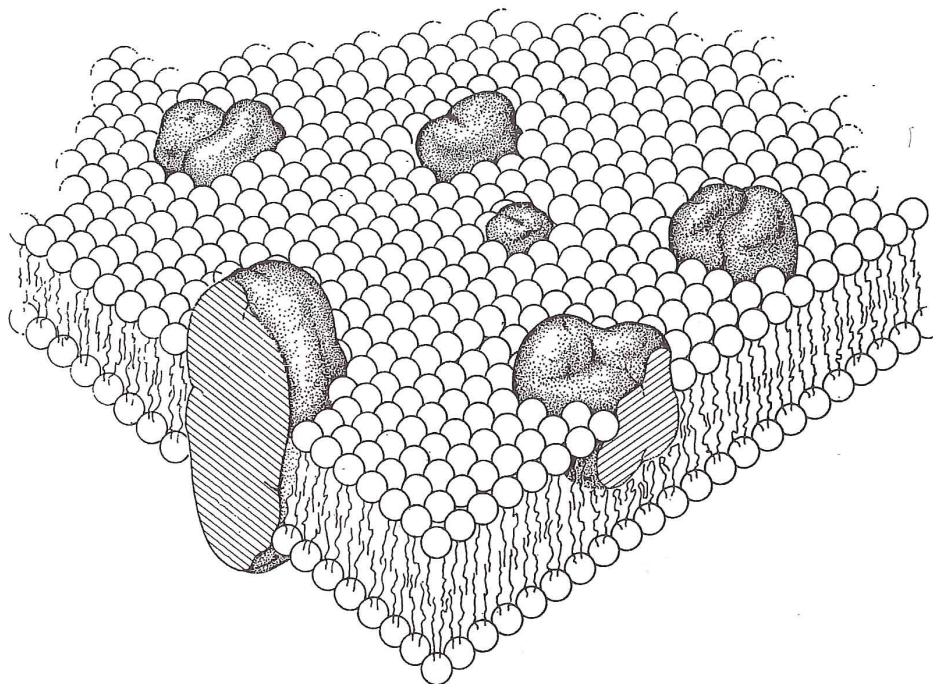

Figure 2.14 ▶ Fluid mosaic model of a membrane. The balls represent polar end groups, and the wavy lines are the hydrocarbon chains of the lipids. The masses embedded in the lipid bilayer are proteins. [Reprinted with permission from Singer, S. J. and Nicolson, G. L. (1972). *Science* 175, 720. Copyright ©1972 by American Association for the Advancement of Science.]

are derived from fatty acids. Glycolipids (**2.79**, R = sugar) also are important membrane constituents.

2.77   2.78   2.79
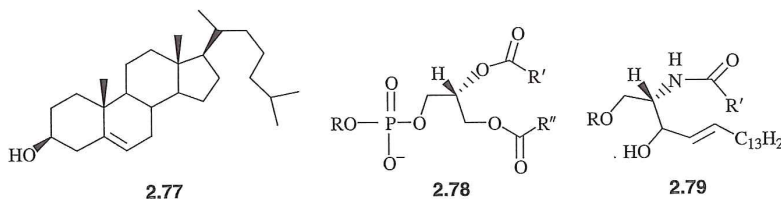
All of these lipids are amphipathic, which means that one end of the molecule is hydrophilic (water soluble) and the other is hydrophobic or, if you wish, lipophilic (water insoluble; soluble in organic solvents). Thus, the hydroxyl group in cholesterol, the ammonium groups in the phospholipids, and the sugar residue in the glycolipids are the polar, hydrophilic ends, and the steroid and hydrocarbon moieties are the lipophilic ends. The hydrocarbon part (R' and R'') actually can be a mixture of chains from 14 to 24 carbon atoms long; approximately 50% of the chains contain a double bond. The polar groups of the lipid bilayer are in contact with the aqueous phase; the hydrocarbon chains project toward each other in the interior with a space between the layers. The stability of the membrane arises from the stabilization of the ionic charges by ion–dipole interactions (see Chapter 3, Section 3.2.B.3, p. 125) with the water and from association of the nonpolar groups. The hydrocarbon chains are relatively free to move; therefore, the core is similar to a liquid hydrocarbon.

b.   Measurement of Lipophilicities

It occurred to Hansch that the fluidity of the hydrocarbon region of the membrane may explain the correlation noted by Richet,[149] Overton,[150] and Meyer[151] between lipid solubility and biological activity. He first set out to measure the lipophilicities of various compounds and then to determine the lipophilicities of substituents. But how should the lipophilicities be measured? The most relevant approach would be to determine their solubility in membranes or vesicles. However, as an organic chemist, Hansch probably realized that if he set a scale of lipophilicities based on membrane solubility, which required the researcher to prepare membranes or vesicles, there was no way organic chemists, especially in the 1960s, would ever bother to use this method, and it would become very limited. So he decided to propose a model for a membrane and determine lipophilicities by a simple methodology that organic chemists would not hesitate to employ. The model for the first step in drug action (transport to the site of action) would be the solubility of the compound in 1-octanol, which simulates a lipid membrane, relative to that in water (or aqueous buffer, the model for the cytoplasm). 1-Octanol has a long saturated alkyl chain, a hydroxyl group for hydrogen bonding, and it dissolves water to the extent of 1.7 M (saturation). This combination of lipophilic chains, hydrophilic (head) groups, and water molecules gives 1-octanol properties very close to those of natural membranes and macromolecules.

As a measure of lipophilicity, Hansch proposed the *partition coefficient*, P, a measure of the solubility in 1-octanol versus water,[152,153] and P was determined by Equation 2.7,

$$P = \frac{[\text{compound}]_{\text{oct}}}{[\text{compound}]_{\text{aq}}(1 - \alpha)} \qquad (2.7)$$

where α is the degree of dissociation of the compound in water calculated from ionization constants. (Ionization makes the compound more soluble in water than the structure appears, so that must be taken into account.) The partition coefficient is derived experimentally by placing a compound in a shaking device (like a separatory funnel) with varying volumes of 1-octanol and water, determining the concentration of the compound in each layer after mixing (by gas chromatography or HPLC), and employing Equation 2.7 to calculate P.[154] The value of P varies slightly with temperature and concentration of the solute, but with neutral molecules in dilute solutions (<0.01 M) and small temperature changes ($\pm 5°$C), variations in P are minor.

Collander[155] had shown previously that the rate of movement of a variety of organic compounds through cellular material was approximately proportional to the logarithm of their partition coefficients between an organic solvent and water. Therefore, as a model for a drug traversing through the body to its site of action, the relative potency of the drug, expressed as $\log 1/C$, where C is the concentration of the drug that produces some standard biological effect, was related by Hansch *et al.*[156] to its lipophilicity by the parabolic expression shown in Equation 2.8:

$$\log 1/C = -k(\log P)^2 + k'(\log P) + k''   \tag{2.8}$$

On the basis of Equation 2.7, it is apparent that if a compound is more soluble in water than in 1-octanol, $P < 1$, and, therefore, $\log P$ is negative. Conversely, a molecule more soluble in 1-octanol has a $P > 1$, and the $\log P$ is positive. Therefore, the more positive the $\log P$, the more lipophilic it is. The larger the value of P, the more there will be an interaction of the drug with the lipid phase (i.e., membranes). As P approaches infinity, micelles will form and/or the drug interaction will become so great that the drug will not be able to cross the aqueous phase, and it will localize in the first lipophilic phase with which it comes into contact. As P approaches zero, the drug will be so water soluble that it will not be capable of crossing the lipid phase and will localize in the aqueous phase. Somewhere between $P = 0$ and $P = \infty$, there will be a value of P such that drugs having this value will be least hindered in their journey through macromolecules to their site of action. This value is called $\log P_0$, the logarithm of the optimum partition coefficient for biological activity. This random walk analysis supports the parabolic relationship (Equation 2.8) between potency ($\log 1/C$) and $\log P$ (Figure 2.15). Note the correlation of Figure 2.15 with the generalization regarding homologous series of compounds (Section 2.2.E.1, p. 26; Figure 2.3, p. 26). An increase in the alkyl chain length increases the lipophilicity of the molecule; the $\log P_0$ generally occurs in the range of 5–9 carbon atoms. Hansch *et al.*[157] found that a number of series of
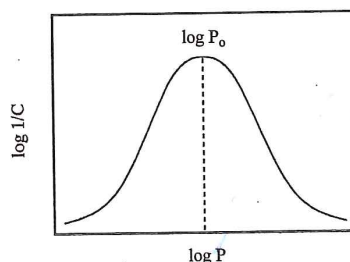

Figure 2.15 ▶   Effect of $\log P$ on biological response. P is the partition coefficient, and C is the concentration of the compound required to produce a standard biological effect

nonspecific hypnotics had similar log P_0 values, approximately 2, and they suggested that this is the value of log P_0 needed for penetration into the central nervous system (CNS), i.e., for crossing the blood–brain barrier. If a hypnotic agent has a log P considerably different from 2, then its activity probably is derived from mechanisms other than just lipid transport. If a lead compound has modest CNS activity and has a log P value of 0, it would be reasonable to synthesize an analog with a higher log P.

Because of the problem associated with ionization of compounds, which leads to greater water solubility than predicted from the neutral structure, often the term log D (the log of the distribution coefficient, generally between 1-octanol and aqueous buffer) is used to describe the lipophilicity of an ionizable compound.[158] Because ionization is a function of the pK_a of the compound and the pH of the solution in which the compound is dissolved, log D describes the log P of an ionizable compound at a particular pH. For example, log $D_{4.5}$ is the log P of an ionizable compound at pH 4.5. The log D value will change for ionizable compounds as a function of pH, whereas log P of nonionizable compounds will be independent of pH. For example, Table 2.6 shows how the log D for the antihypertensive drug metoprolol (**2.80**, Toprol-XL) changes as a function of pH. Note that at low pH values, the amine is protonated, lowering the log D value. As the pH is increased, the equilibrium starts to favor the neutral free base form, which is more lipophilic.

Although it is valuable to be able to determine the lipophilicity of a molecule, for lead modification purposes you need to be able to predict, prior to synthesis of the compound,

TABLE 2.6 ▶ Change in log D as a Function of pH for Metoprolol (2.80)[a]

2.80
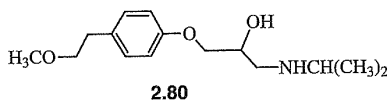
pH	log D
2.0	−1.31
3.0	−1.31
4.0	−1.31
5.0	−1.28
5.5	−1.21
6.0	−1.05
6.5	−0.75
7.0	−0.34
7.5	0.12
8.0	0.59
8.5	1.03
9.0	1.39
10.0	1.73

[a] The author is grateful to Karolina Nilsson and Ola Fjellström (AstraZeneca) for providing the log D values as a function of pH using ACD software.

what the lipophilicity of an unknown molecule will be. To do that it is necessary to know the lipophilicities of substituents and atoms. In the same way that substituent constants were derived by Hammett for the electronic effects of atoms and groups (σ constants), Hansch and coworkers[159] derived substituent constants for the contribution of individual atoms and groups to the partition coefficient. The lipophilicity substituent constant, π, is defined by Equation 2.9,

$$\pi = \log P_X - \log P_H = \frac{\log P_X}{\log P_H} \tag{2.9}$$

which has the same derivation as the Hammett equation. The term P_X is the partition coefficient for the compound with substituent X, and P_H is the partition coefficient for the parent molecule (X = H). As in the case of the Hammett substituent constant σ, π is additive and constitutive. *Additive* means that multiple substituents exert an influence equal to the sum of the individual substituents. *Constitutive* indicates that the effect of a substituent may differ depending on the molecule to which it is attached or on its environment. Alkyl groups are some of the least constitutive groups. For example, methyl groups attached at the meta- or para-positions of 15 different benzene derivatives had π_{CH_3} values with a mean and standard derivation of 0.50 ± 0.04. Because of the additive nature of π values, π_{CH_2} can be determined as shown in Equation 2.10, where the log P values are obtained from standard tables:[160]

$$\pi_{CH_2} = \log P_{\text{nitroethane}} - \log P_{\text{nitromethane}}$$
$$= 0.18 - (-0.33) = 0.51 \tag{2.10}$$

Because, by definition, $\pi_H = 0$, then $\pi_{CH_2} = \pi_{CH_3}$. However, be aware that π_{CH_2OH} does not equal π_{CH_3O} because there is a difference in hydrogen bonding for substituents with and without hydroxyl groups, and, therefore, a difference in water solubility between these two substituents. Note that π represents the lipophilicity of a substituent and log P is the lipophilicity of a compound.

As was alluded to in Section 2.2.E.2 on molecular modification, branching in an alkyl chain lowers the log P or π as a result of larger molar volumes and shapes of branched compounds. As a rule of thumb, the log P or π is lowered by 0.2 unit per branch. For example, the π_{i-Pr} in 3-isopropylphenoxyacetic acid is 1.30; π_{n-Pr} is $3(0.5) = 1.50$, or 0.2 greater than π_{i-Pr}.

Another case where π values are fairly constant is conjugated systems, as exemplified by $\pi_{CH=CHCH=CH}$ in Table 2.7.
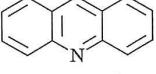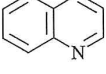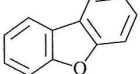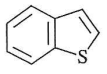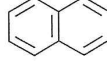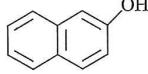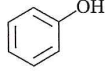
Inductive effects are quite important to lipophilicity.[161] In general, electron-withdrawing groups increase π when a hydrogen-bonding group is involved. For example, π_{CH_2OH} varies as a function of the proximity of an electron-withdrawing phenyl group (Eq. 2.11), and π_{NO_2} varies as a function of the inductive effect of the nitro group on the hydroxyl group (Eq. 2.12). The electron-withdrawing inductive effects of the phenyl group (Eq. 2.11) and the nitro group (Eq. 2.12)

$$\pi_{CH_2OH} = \log P_{Ph(CH_2)_2OH} - \log P_{PhCH_3} = -1.33$$
$$\pi_{CH_2OH} = \log P_{PhCH_2OH} - \log P_{PhH} = -1.03 \tag{2.11}[162]$$

$$\pi_{NO_2} = \log P_{PhNO_2} - \log P_{PhH} = -0.28$$
$$\pi_{NO_2} = \log P_{4-NO_2PhCH_2OH} - \log P_{PhCH_2OH} = 0.11 \tag{2.12}[163]$$

make the nonbonded electrons on the OH group less available for hydrogen bonding, thereby reducing the affinity of this functional group for the aqueous phase. This, then, increases

TABLE 2.7 ▶ Constancy of π for $-CH=CH-CH=CH-$ [a]

$\pi_{CH=CHCH=CH}$

log P	(indole)	$-$log P (pyrrole)	$= 2.14 - 0.75 = 1.39$
log P	(quinoline)	$-$log P (pyridine)	$= 2.03 - 0.65 = 1.38$
log P	(acridine)	$-$log P (quinoline)	$= 3.40 - 2.03 = 1.37$
log P	(dibenzofuran)	$-$log P (benzofuran)	$= 4.12 - 2.67 = 1.45$
log P	(dibenzothiophene)	$-$log P (thiophene)	$= 3.12 - 1.81 = 1.31$
log P	(naphthalene)	$-$log P (benzene)	$= 3.45 - 2.13 = 1.32$
2/3 log P	(benzene)		$= 2/3(2.13) = 1.42$
log P	(naphthol)	$-$log P (phenol)	$= 2.84 - 1.46 = 1.38$
		ave.	1.38 ± 0.046

[a] Hansch, C.; Steward, A. R.; Anderson, S. M.; Bentley, D. *J. Med. Chem.* **1968**, *11*, 1.

the log P or π. Also note in Equations 2.11 and 2.12 that, because $\pi_H = 0$ by definition, log $P_{benzene} = \pi_{Ph}$. These examples enforce the notion of the constitutiveness of π values.

Resonance effects also are important to the lipophilicity, much the same way as are inductive effects. Delocalization of nonbonded electrons into aromatic systems decreases their availability for hydrogen bonding with the aqueous phase and, therefore, increases the π. This is supported by the general trend that aromatic π_X values are greater than aliphatic π_X values, again emphasizing the constitutive nature of π and log P.

Steric effects are variable. If a group sterically shields nonbonded electrons, then aqueous interactions will decrease, and the π value will increase. However, crowding of functional groups involved in hydrophobic interactions (Chapter 3, Section 3.2.B.6, p. 129) will have the opposite effect.

Conformational effects also can affect the π value. The π_X values for Ph(CH$_2$)$_3$X are consistently lower (more water soluble) than π_X values for CH$_3$(CH$_2$)$_3$X (Table 2.8). This phenomenon is believed to be the result of folding of the side chain onto the phenyl ring (**2.81**), which means a smaller apolar surface for organic solvation. The folding may be caused by the interaction of the CH$_2$–X dipole with the phenyl π-electrons and by intramolecular hydrophobic interactions.

TABLE 2.8 ▶ Effect of Folding of Alkyl Chains on π

X	π_X (aromatic)[a]	π_X (aliphatic)[b]	$\Delta\pi_X$
OH	−1.80	−1.16	0.64
F	−0.73	−0.17	0.56
Cl	−0.13	0.39	0.52
Br	0.04	0.60	0.56
I	0.22	1.00	0.78
COOH	−1.26	−0.67	0.59
CO$_2$CH$_3$	−0.91	−0.27	0.64
COCH$_3$	−1.26	−0.71	0.55
NH$_2$	−1.85	−1.19	0.66
CN	−1.47	−0.84	0.63
OCH$_3$	−0.98	−0.47	0.51
CONH$_2$	−2.28	−1.71	0.57
		Average	0.60 ± 0.05

[a] Log $P_{Ph(CH_2)_{3X}}$ − log $P_{Ph(CH_2)_3H}$
[b] Log $P_{CH_3(CH_2)_{3X}}$ − log $P_{CH_3(CH_2)3H}$
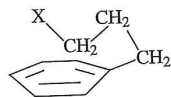
2.81

Two examples follow to show the additivity of π constants in predicting log P values. A calculation of the log P for the anticancer drug diethylstilbestrol (**2.82**, DES) is shown in Equation 2.13.

$$\text{Calc. log } P = 2\pi_{CH_3} + 2\pi_{CH_2} + \pi_{CH=CH} + 2\log P_{PhOH} - 0.40$$
$$= 2(0.50) + 2(0.50) + 0.69 + 2(1.46) - 0.40 \tag{2.13}$$
$$= 5.21$$

In Equation 2.13, $\pi_{CH=CH} = 1/2(\pi_{CH=CHCH=CH})$, which was shown in Table 2.7 to be 1.38; −0.40 is added into the equation to account for two branching points (each end of the alkene). The calculated log P value of 5.21 is quite remarkable considering that the experimental log P value is 5.07.

**diethylstilbestrol
2.82**

A calculation of the log P for the antihistamine diphenhydramine (**2.83**, Benedryl), is shown in Equation 2.14.

$$\begin{aligned}
\text{Calc.}\log P &= 2\pi_{\text{Ph}} + \pi_{\text{CH}} + \pi_{\text{OCH}_2} + \pi_{\text{CH}_2} + \pi_{\text{NMe}_2} - 0.2 \\
&= 2(2.13) + 0.50 - 0.73 + 0.50 - 0.95 - 0.2 \\
&= 3.38
\end{aligned} \tag{2.14}$$

In this equation, 2.13 is log P for benzene, which is the same as π_{Ph}; 0.50 is π_{CH} (same as π_{CH_3}); −0.73 was obtained by subtracting 1.50 ($2\pi_{\text{CH}_3} + \pi_{\text{CH}_2}$) from log $P_{\text{CH}_3\text{CH}_2\text{OCH}_2\text{CH}_3}$ (= 0.77); −0.95 is the value for π_{NMe_2} obtained by subtracting $\pi_{\text{Ph(CH}_2)_3}(2.13 + 3(0.5) = 3.63)$ from log $P_{\text{Ph(CH}_2)_3\text{NMe}_2}(2.68)$; −0.2 is for branching at the CH. [Note that there is no branching at the $\text{N(CH}_3)_2$ because we used that whole substituent to obtain π.] The experimental log P value is 3.27.

diphenhydramine
2.83
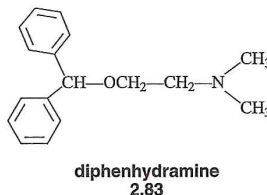
A more rapid approach than the standard shake-flask method[164] for determination of log P values that was described above has been reported[165] for neutral compounds. The reversed-phase HPLC method takes about 20 minutes per compound with a wide range of lipophilicities (6 log P units) with good accuracy and excellent reproducibility. The value obtained by this method is referred to as the E log P_{oct}. A reversed-phase HPLC method for determination of log D values (E log D_{oct}) also was devised by the same Pfizer group.[166]

c. Computerization of Log P Values

The determination of log P values has become less of a chore as a result of computerization of the method.[167] A nonlinear regression model for the estimation of partition coefficients was developed by Bodor *et al.*[168] using the following molecular descriptors: molecular surface, volume, weight, and charge densities. It was shown to have excellent predictive power for the estimation of the log P for complex molecules. A semiquantitative method for calculating log P values (M log P) was developed by Moriguchi *et al.*[169] using a multiple regression analysis of 1230 organic molecules having a wide variety of structures; excellent correlation was observed between the observed log P and the calculated log P.

Probably the simplest way to get log P values for unknown compounds is with the use of one of the numerous software packages that are now commercially available, such as those from Daylight (C log P; the one developed by the medicinal chemistry group at Pomona College), Advanced Chemistry Development (ACD/log P DB), CTIS (AUTOLOG$^{\text{TM}}$), Scivision (Sci log P), and Bio-Rad (PredictIt$^{\text{TM}}$ log P and log D). The problem with the software packages, however, is that the results can differ widely (2 or more log P units) and differ from the experimental value.[170] The reason is related to the fact that, as mentioned above, π values are constitutive; depending on the structure of the compound, the π value can differ. Also, ionization of groups varies with concentration and counter ions. So, no software package can account for π values for substituents on *every* scaffold. Most software packages will always

give an answer, but $C \log P$ will not calculate a $\log P$ when it does not have sufficient data (e.g., if an atom is poorly parameterized). Of course, it is frustrating when a computer tells you it cannot compute (which may be detrimental to the software company sales), but it may be the most honest approach. One way to obtain the most accurate predictive results for your particular family of compounds is to determine experimentally the actual $\log P$ value for one of the members of the family, then ask a variety of software packages to predict the $\log P$ value, and use the program that comes closest to the experimental value for that family of compounds.

d. Membrane Lipophilicity

Although the $\log P$ values determined from 1-octanol/water partitioning are excellent models for *in vivo* lipophilicity, it has been found that for a variety of aromatic compounds whose $\log P$ values are greater than 5.5 (very lipophilic) or whose molar volumes are greater than $230\,\text{cm}^3/\text{mol}$, there is a breakdown in the correlation of these values with those determined from partitioning between L-α-phosphatidylcholine dimyristoyl membrane vesicles and water.[171] Above $\log P$ of 5.5, the solvent solubility for these molecules is greater than their membrane solubility. As the compound increases in size, more energy per unit volume is required to form a cavity in the structured membrane phase. This is consistent with observations that branched molecules have lower $\log P$ values than their straight-chain counterparts, and that this effect is even greater in membranes than in organic solvents.

Note that although $\log P$ values are most commonly determined with 1-octanol/water mixtures, this is not universal because hydrophilicity resulting from acceptance of a hydrogen bond is not reflected well by partitioning in 1-octanol, which can accept hydrogen bonds almost as well as does water.[172] Consequently, this gives an apparently higher lipophilicity value than is reflected in membrane partitioning. Other nonhydroxylic solvents, such as cyclohexane, can provide insights into these processes. Because of this, Seiler[173] introduced a new additive constitutive substituent constant for solvents other than 1-octanol. Therefore, when using $\log P$ values, it is important to be aware of the solvent used to obtain the $\log P$ data.

F.3 Effects of Ionization on Lipophilicity and Oral Bioavailability

Receptors are typically proteins, comprised of amino acids with varying ionization states depending on the pH of the environment. For example, anionic groups in proteins include carboxylic acids (aspartic and glutamic acids, pK_a of 4–4.5), phenols (tyrosine, pK_a of 9.5–10), sulfhydryls (cysteine, pK_a of 8.5–9), and hydroxyls (serine and threonine, pK_a of 13.5–14). Cationic groups in proteins include imidazole (histidine, pK_a of 6–6.5), amino (lysine, pK_a of 10–10.5), and guanidino (arginine, pK_a of 12–13) groups. At physiological pH (pH 7.4), even the mildly acidic groups, such as carboxylic acid groups, will essentially be completely in the carboxylate anionic form; phenolic hydroxyl groups may be partially ionized. Likewise, basic groups, such as amines, will be partially or completely protonated to give the cationic form. The same is true for a drug; the ionization state of a drug will depend on the pH of the medium with which it has to interact and the pK_a values of the ionizable groups. Ionization will have a profound effect not only on its interaction with a receptor, but also on its lipophilicity. Consequently, it is important to appreciate the effects of ionization in lead modification approaches.

What if the drug you are attempting to discover binds at an ionized site in the receptor, so ionization of your drug favors binding to the receptor (see Chapter 3, Section 3.2.B.2),

$$RNH_2 + H^+ \rightleftharpoons RNH_3^+$$

$$RCOOH \rightleftharpoons RCOO^- + H^+$$
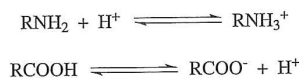
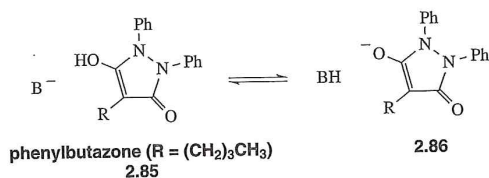**Scheme 2.8** ▶ Ionization equilibrium for an amine base and a carboxylic acid

but ionization of the drug also blocks its ability to cross various membranes prior to reaching the receptor? How is it possible to design a compound that is neutral when it needs to cross membranes, but ionized when it finally reaches the target receptor? This is possible because an equilibrium is established between the neutral and ionized form of a molecule or group that depends on the pH of the medium and the pK_a of the ionizable group (Scheme 2.8). When the pH of the medium equals the pK_a of the molecule, half of the molecules are in the neutral form and half in the ionized form. The ones that are neutral may be able to cross membranes, but once on the other side, the equilibrium with the ionized form is reestablished (the equilibrium mixture will again depend on the pH on the other side of the membrane), so there are now ionized molecules on the other side of the membrane that can interact with the target receptor. The ionized molecules that did not cross the membrane also reestablish an equilibrium and become a mixture of ionized and neutral molecules, so more neutral molecules can get across the membrane. If the equilibria could be reestablished indefinitely, eventually all of the molecules would cross the membranes and bind to the target receptor. However, drugs get metabolized and excreted (see Chapter 7), so they may never get across the membrane before they are excreted. To adjust the ionization equilibrium of the lead compound, you need to add electron-withdrawing or electron-donating groups to vary the pK_a of the molecule. Electron-withdrawing groups will lower the pK_a, making acids more ionizable and bases less ionizable; the opposite holds for electron-donating groups.

The importance of ionization was recognized in 1924 when Stearn and Stearn[174] suggested that the antibacterial activity of stabilized triphenylmethane cationic dyes was related to an interaction of the cation with some anionic group in the bacterium. Increasing the pH of the medium also increased the antibacterial effect, presumably by increasing the ionization of the receptors in the bacterium. Albert and coworkers[175] made the first rigorous proof that a correlation between ionization and biological activity existed. A series of 101 aminoacridines, including the antibacterial drug, 9-aminoacridine or aminacrine (**2.84**, Monacrin), all having a variety of pK_a values, was tested against 22 species of bacteria.
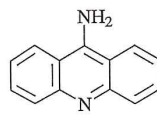
2.84

A direct correlation was observed between ionization (formation of the cation) of the aminoacridines and antibacterial activity. However, at lower pH values, protons can compete with these cations for the receptor, and antibacterial activity is diminished. When this was realized, Albert[176] quips, the Australian Army during World War II was advised to pretreat wounds with sodium bicarbonate to neutralize any acidity prior to treatment with aminacrine. This, apparently, was quite effective in increasing the potency of the drug. The mechanism of action of aminoacridines is discussed in Chapter 6 (Section 6.3.A.3.a, p. 349).

phenylbutazone (R = (CH₂)₃CH₃)
2.85

2.86

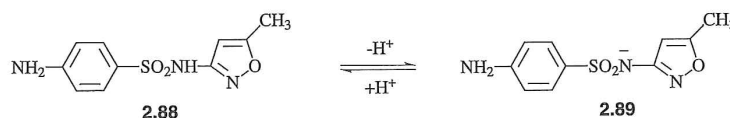Scheme 2.9 ▶ Ionization equilibrium for phenylbutazone

Antihistamines and antidepressants tend to have pK_a values of about 9. The great majority of alkaloids that act as neuroleptics, local anesthetics, and barbiturates have pK_a values between 6 and 8; consequently, both neutral and cationic forms are present at physiological pH. This may allow them to penetrate membranes in the neutral form, and exert their biological action in the ionic form.

The uricosuric drug (increases urinary excretion of uric acid) phenylbutazone (**2.85**, Scheme 2.9; Butazolidine), $R = (CH_2)_3CH_3$ has a pK_a of 4.5 and is active as the anion (**2.86**). However, because the pH of urine is 4.8, suboptimal concentrations of the anion were found in the urinary system. Sulfinpyrazone (**2.85**, $R = CH_2CH_2SOPh$; Anturane) has a lower pK_a (2.8) and is about 20 times more potent than phenylbutazone; the anionic form blocks reabsorption of uric acid by renal tubule cells.[177]

The antimalarial drug pyrimethamine (**2.87**, Daraprim) has a pK_a of 7.2 and is best absorbed from solutions of sufficient alkalinity that it has a high proportion of molecules in the neutral form (to cross membranes). Its mode of action, the inhibition of the parasitic enzyme dihydrofolate reductase, however, requires that it be in the protonated cationic form.

pyrimethamine
2.87

The effect of ionization can be rationalized either from a pharmacokinetic or pharmacodynamic perspective. For example, if changing the pK_a increases its potency, it could be because the neutral form becomes more prevalent and, therefore, crossing membranes becomes favored (pharmacokinetic argument), or it could be because there is a hydrophobic pocket in the receptor that the neutral form prefers to bind into (pharmacodynamic argument). How can the relative importance of these two properties be determined? If the drugs act on microbial systems, one way is to compare results of assaying the test compounds in a cell-free system (in which there are no membranes to cross) and in an intact cell system (in which it is necessary to cross a membrane to get to the receptor). For example, the pharmacokinetics of the antibacterial agent sulfamethoxazole (**2.88**, Scheme 2.10; Bactrim) depend on their nonionized form (**2.88**), but the pharmacodynamics depend on the anionic form (**2.89**). In a cell-free system the antibacterial activity of **2.88** and other sulfonamides is directly proportional to the degree of ionization, supporting the importance of ionization on pharmacodynamics, but in intact cells, where the drug must cross a membrane to get to the site of action, the antibacterial activity also is dependent on the neutral form,[178] supporting the notion that the neutral form is not important to pharmacodynamics, only to pharmacokinetics.

Scheme 2.10 ▶ Ionization equilibrium for sulfamethoxazole

The structure and function of a receptor and of a drug can be strongly dependent on the pH of the medium, especially if an *in vitro* assay is being used. However, you must be careful when trying to assess pK_a values of groups within a binding site of a receptor, because these values can be quite variable, and will depend on the microenvironment. On the basis of molecular dynamics simulations of several proteins in water, the interiors of these proteins were calculated to have dielectric constants of about 2–3,[179] which is comparable to the dielectric constant of nonpolar solvents such as benzene ($\varepsilon = 2.28$) or *p*-dioxane ($\varepsilon = 2.21$). This is quite different from the dielectric constant of water ($\varepsilon = 78.5$), which is a result of the strong dipole moment of the O–H bonds. If a carboxyl group is in a nonpolar region, its pK_a will rise because the anionic form will be destabilized. Glutamate-35 in the lysozyme-glycolchitin complex has a pK_a of 8.2[180] the pK_a of glutamate in water is 4–4.5. The pK_a of Asp-99 in a nonpolar region of 3-oxo-Δ^5-steroid isomerase is a remarkable 9.5![181] That is a change in equilibrium of a factor of 10^5 (remember, pK_a is a logarithm) in favor of the neutral form! If the carboxylate forms a salt bridge, it will be stabilized, and its pK_a will be lowered. If a carboxylic acid group is near an essential active site carboxylic acid, the anionic form will be destabilized, and its pK_a will be raised.[182a] Likewise, an amino group buried in a nonpolar microenvironment will have a lower pK_a because protonation will be disfavored (to avoid the polar cationic character). If the ammonium group of lysine forms a salt bridge, it will be stabilized, deprotonation will be inhibited, and the pK_a will rise. If basic residues are adjacent, the pK_a will drop to avoid two neighboring cations; the ε-amino group of the active site lysine residue in the enzyme acetoacetate decarboxylase has a pK_a of 5.9,[182b] whereas in water, it is about 10–10.5. Given this large change in pK_a values in different microenvironments, it is worthwhile to make large changes in pK_a values of compounds in a lead modification library to see how the potency changes in both *in vitro* and *in vivo* assays. Once it is established whether the potency of a molecule is favored in the neutral or ionized form, then pK_a considerations can be employed in further lead modification approaches.

F.4 Other Properties that Influence Oral Bioavailability and Ability to Cross the Blood–Brain Barrier

Pharmacokinetics is as important to drug discovery as pharmacodynamics, so Lipinski[183] proposed "the rule of five" as a guide to improve oral bioavailability during lead modification. Based on a large database of known drugs, the *rule of five* states that it is highly likely (>90% probability) that compounds with two or more of the following characteristics will have **poor** oral absorption and/or distribution properties:

▶ The molecular weight is >500.

▶ The log *P* is >5.

▶ There are more than 5 H-bond donors (expressed as the sum of OH and NH groups).

▶ There are more than 10 H-bond acceptors (expressed as the sum of N and O atoms).

Antibiotics, antifungals, vitamins, and cardiac glycosides are the exception because they often have active transporters to carry them across membranes, so lipophilicity is not relevant.

Therefore, when low potency of compounds in an *in vivo* assay is observed, the rule of five should be applied to determine if low potency is the result of a pharmacokinetic problem. To get a drug across the blood–brain barrier, the upper limits really should be 3 H-bond donors and 6 H-bond acceptors.[184]

In contrast to the rule of five, Veber and coworkers[185] measured the oral bioavailability of 1100 drug candidates and found that reduced molecular flexibility, as determined by the number of rotatable bonds (10 or fewer), and low polar surface area ($\leq 140\,\text{Å}^2$) or total hydrogen bond count (less than or equal to a total of 12 donors and acceptors) are important predictors of good oral bioavailability, *independent of molecular weight*. Both the number of rotatable bonds and hydrogen bond count tend to increase with molecular weight, which may explain Lipinski's first rule. Reduced polar surface area was found to correlate better with an increased membrane permeation rate than did lipophilicity. Nonetheless, molecular weight and lipophilicity were shown to have the greatest influence on getting a drug to the market.[186a]

Ajay and coworkers[186b] carried out computations to determine what drug properties were important for crossing the blood–brain barrier and for CNS activity. CNS-active and -inactive compounds were selected from the Comprehensive Medicinal Chemistry (CMC) and the MDDR (a database from MDL Inc.; MDL Drug Data Report) databases. Each molecule was described by seven 1-D descriptors (e.g., molecular weight, number of hydrogen bond donors, and number of hydrogen bond acceptors) and 166 2-D descriptors. Using all of these descriptors, 83% of the CNS-active compounds and 79% of the CNS-inactive compounds in these databases were correctly predicted. In general, they concluded that if the molecular weight, the degree of branching, the number of rotatable bonds, or the number of hydrogen bond acceptors is increased, the compound will be *less* likely to be CNS active. If the aromatic density, number of hydrogen bond donors, or log P is increased, the compound is *more* likely to be CNS active.

Absorption, distribution, metabolism, and excretion (ADME) characteristics of compounds are very important because a large percentage of drug candidates that reaches clinical trials are discontinued as a result of ADME and toxicity problems. If these properties could be predicted, much time and expense would be saved in designing, synthesizing, and testing compounds. Numerous computational methods have been devised that deal with these properties,[187] but reliable predictive capabilities are still lacking. A graphical model for estimating high, medium, or low oral bioavailability of drugs in humans, rats, dogs, and guinea pigs, based on both their permeability through human intestinal epithelial (Caco-2) cells and their *in vitro* liver enzyme metabolic stability rates, gave excellent results.[188]

Up to this point we have been discussing more or less random molecular modifications to make qualitative differences in a lead compound. In 1868 Crum-Brown and Fraser[189] predicted that some day a mathematical relationship between structure and activity would be expressed. It was not for almost 100 years that this prediction began to be realized, and a new era in drug design was born. In 1962 Corwin Hansch attempted to quantify the effects of particular substituent modifications, and from his studies the area of quantitative structure–activity relationships developed.[190]

2.2.G   Quantitative Structure–Activity Relationships

G.1   Historical

The concept of quantitative drug design is based on the fact that the biological properties of a compound are a function of its physicochemical parameters, that is, physical properties, such as solubility, lipophilicity, electronic effects, ionization, stereochemistry, and so forth,

that have a profound influence on the chemistry of the compounds. The first attempt to relate a physicochemical parameter to a pharmacological effect was reported in 1893 by Richet.[191] He observed that the narcotic action of a group of organic compounds was inversely related to their water solubility (*Richet's rule*). Overton[192] and Meyer[193] related tadpole narcosis induced by a series of nonionized compounds added to the water in which the tadpoles were swimming to the ability of the compounds to partition between oil and water. These early observations regarding the depressant action of structurally nonspecific drugs were rationalized by Ferguson.[194] He reasoned that when in a state of equilibrium, simple thermodynamic principles could be applied to drug activities, and that the important parameter for correlation of narcotic activities was the relative saturation (termed *thermodynamic activity* by Ferguson) of the drug in the external phase or extracellular fluids. This is known as *Ferguson's principle*, which is useful for the classification of the general mode of action of a drug and for predicting the degree of its biological effect. The numerical range of the thermodynamic activity for structurally nonspecific drugs is 0.01 to 1.0, indicating that they are active only at relatively high concentrations. Structurally specific drugs have thermodynamic activities considerably less than 0.01 and normally below 0.001.

In 1951 Hansch *et al.*[195] noted a correlation between the plant growth activity of phenoxyacetic acid derivatives and the electron density at the *ortho* position (lower electron density gave increased activity). They made an attempt to quantify this relationship by the application of the Hammett σ functions (see Section 2.2.F.1, p. 51), but this was unsuccessful.

The crucial breakthrough in QSAR came when Hansch and coworkers[196] conceptualized the action of a drug as depending on two processes. The first process is the journey of the drug from its point of entry into the body to the site of action (pharmacokinetics), and the second process is the interaction of the drug with the specific site (pharmacodynamics). Because of the importance of pharmacokinetics to the success of a drug, he developed the octanol-water scale for lipophilicity (see Section 2.2.F.2.b, p. 55) as a measurable physicochemical parameter to consider in addition to electronic effects that were developed by Hammett (see Section 2.2.F.1, p. 51). Another physicochemical parameter that Hansch thought should be important in lead discovery/lead optimization was steric effects, particularly for receptor binding.

G.2 Steric Effects: The Taft Equation and Other Equations

Because interaction of a drug with a receptor involves the mutual approach of two molecules, another important parameter for QSAR is the steric effect. In much the same way that Hammett derived quantitative electronic effects (see Section 2.2.F.1, p. 51), Taft[197] defined the steric parameter E_s as shown in Equation 2.15:

$$E_s = \log k_{XCO_2Me} - \log k_{CH_3CO_2Me} = \log k_X/k_0 \qquad (2.15)$$

Taft used for the reference reaction the relative rates of the acid-catalyzed hydrolysis of α-substituted acetates (XCH_2CO_2Me). This parameter is normally standardized to the methyl group ($XCH_2 = CH_3$) so that $E_s(CH_3) = 0.0$; it is possible to standardize it to hydrogen by adding 1.24 to every methyl-based E_s value.[198] Hancock *et al.*[199] claimed that this model reaction was under the influence of hyperconjugative effects and, therefore, developed corrected E_s values for the hyperconjugation of α-hydrogen atoms (Equation 2.16):

$$E_s^c = E_s + 0.306(n - 3) \qquad (2.16)$$

where E_s^c is the corrected E_s value and n is the number of α-hydrogen atoms.