ever, not all aspects of molecular structure may be important for understanding structure-activity relationships (SARs) for a particular receptor. This led Pearlman and Smith (11d) to introduce the concept of a receptor-relevant subspace (RRSS) of a full chemistry space. For example, starting with a chemistry space of six dimensions, defined to best represent the diversity of all druglike compounds in the MDDR (MDL Drug Data Report) database (13a), they showed how to perceive the three-dimensional subspace that conveys information that is particularly relevant for affinity to the ACE (angiotension converting enzyme) receptor. ACE inhibitors of diverse structure were tightly clustered with respect to the receptor-relevant metrics, thereby providing an obvious near-neighbor strategy for lead follow-up. They (11d) also emphasized the importance of not considering metrics that are not "receptor-relevant" when computing distances for such near-neighbor-based discovery efforts. This also enables diversity in these other dimensions to be explored (e.g., with combinatorial libraries), to obtain compounds with a modified profile for other properties such as bioavailability.

Work on the design and diversity analysis of large combinatorial libraries at Pharmacopeia using BCUT metrics and DiverseSolutions was reported by Schnur (32). A cell-based analysis of synthon-derived libraries was performed, using full product libraries, including library comparisons. Active molecules in these libraries, which involved multiple scaffolds, were found to cluster in various three-dimensional subspaces of the diversity spaces. The utility of a simple property-based reactant/synthon selection tool was also described, targeted at the synthetic chemists, with reactants binned according to patterns based on the ranges of a set of user-selected properties that form a diversity hypothesis.

Chemistry space metrics have been used at Rhône-Poulenc Rorer for diversity analysis, library design, and compound selection (59, 80) using DiverseSolutions to generate a "universal" chemistry space for use as a standard for profiling structural sets of interest. The complementarity of three different diversity measures for comparing and profiling compound collections (a corporate database, combinatorial libraries, and the MDDR drugs database) was also shown. The methods used were a 2D structural characterization (Daylight fingerprints), DiverseSolutions, and 3D pharmacophore fingerprints. A combinatorial library of 100,000 structures appeared structurally different from the other databases by the Daylight fingerprint clustering, yet the bulk of its compounds overlapped with druglike compounds (MDDR) in DiverseSolutions BCUT chemistry space and 3D pharmacophore space ("cells" in fingerprints). It was shown and "quantified" that new diversity relative to the company database was explored, with much of this new diversity in desirable areas occupied by MDDR compounds. The nonuniform binning scheme was developed to enable the use of chemistry spaces scaled to include all structures within a set, while maintaining a reasonable distribution of compounds within cells. The method was used to select a subset for initial screening of a large set of combinatorial libraries designed for 7-TM GPCR targets.

2.2.1.2 Pharmacophore Fingerprints. Pharmacophore fingerprints can also be considered as a high dimensional partitioning of the compound space (35). Underrepresented pharmacophores within a population can be identified and act as a possible focus for library design or compound acquisition. Using six feature types (hydrogen bond acceptor, donor, acid, base, hydrophobe, and aromatic ring centroid) with four-point pharmacophores and 7–10 binned distance ranges, it is possible to resolve about 2–10 million different phannacophoric shapes. Different databases can be compared using this fingerprint, and differences identified. For example, by comparing a corporate screening file (100,000 structures) with the MDDR database (62,000 structures) of biologically active compounds (as discussed above for DiverseSolutions, Refs. 62, 80) "holes" could be identified, in terms of about 1 million 3D pharmacophores exhibited only by MDDR compounds (about 2.7 million were in common and 0.2 million unique to the corporate set). This provides a design space for which combinatorial libraries were designed and synthesized. A total of 100,000 combinatorial library compounds were able to match about 40% (0.4 million) of the pharmacophore "holes" (i.e.,