Software from MDL includes the ISIS scientific information system (ISIS/Draw, ISIS/Base, and ISIS/Direct), Cheshire for chemical structure manipulation, and Chime and Chemscape for Web access. Combinatorial and high-throughput chemistry programs include Afferent, Central Library, Project Library, Reagent Selector, and Elan. Biological data management programs include Apex and Assay Explorer; literature access through LitLink; reaction access through Reaction Browser/Web; and finally, molecular modeling through Sculpt (101).

Tripos, Inc. Originally the major provider of molecular modeling software, Tripos now offers chemical information content in the form of databases and the tools to manage them. These include the following:

- Several Chapman and Hall databases including ones for organic structures (180,000 structures), inorganic and organometallic structures (40,000 structures), natural products (105,000 structures), and pharmacological agents (22,000 structures)
- The National Cancer Institute structures in a Tripos-compatible format
- The Derwent World Drug Index (60,000 structures)

Chemical information software offered by Tripos now also extends beyond just molecular modeling. Their programs include the following:

- The Unity 3D database system, which features rapid flexible 3D pharmacophore searching
- Concord and Stereoplex—for generating 3D models of database structures including multiple stereochemical isomers
- ChemEnlighten for chemical data mining
- The AUSPYX structure data cartridge for Oracle
- A suite of programs for combinatorial chemistry—Legion to build and store virtual libraries, CombiLibMaker to enumerate structures, Selector to define diversity measures to select diverse subsets of structures, and DiverseSolutions to apply chemical diversity techniques to chemical populations to characterize and populate chemical space (102)

3.6 Sequence and 3D Structure Databases

Sequence databases of biological macromolecules are useful when defining new therapeutic targets. Databases for DNA, RNA, and proteins are available from such sources as the National Center for Biotechnology Information (NCBI) (103) and the European Bioinformatics Institute (104). Numerous online programs and tools are available to researchers to search and align sequences, generate phylogenetic analyses (chemical evolutionary trees), map genes, and predict secondary structure (105). The Protein Data Bank stores the largest collection of crystallographic, NMR, and molecular-modeling derived protein and nucleic acid 3D models (106). The Cambridge Crystallographic Data Center is the primary source for crystal structure data on small molecules, with more than 250,000 entries. The Cambridge Database can be searched using the programs ConQuest for searching, Mercury for structure visualization, and Vista for numerical display and statistical analysis (107).

3.7 In-House Proprietary and Academic Database Systems

Larger chemical and pharmaceutical firms have, over the years, developed in-house systems with capabilities that are specific to the chemist's needs. Today, the costs of developing from scratch and maintaining an in-house system are prohibitive, especially because commercial chemical information systems are highly efficient and customizable. Personal chemical information software is still being developed and reported in the literature. Examples include a relational database patterned after the Upjohn Cousin system (108), and CheD, which is a SQL-based system with a Web client (109).

Commercial personal database systems are available from several vendors, as described above. These products extend the productivity of an individual chemist or a small workgroup, but are not designed for corporate or enterprise applications. Other personal chemical