Computer-Assisted Analysis and Perception of Stereochemical Features in Organic Molecules Using the CHIRON Program

STEPHEN HANESSIAN,* JONATHAN FRANCO, GILBERT GAGNON, DOMINIC LARAMÉE, and BENOIT LAROUCHE

Department of Chemistry, Université de Montréal, Montréal, Québec H3C 3J7, Canada

Received July 15, 1990

A program for stereochemical analysis, chiral substructure searching, and synthesis planning based on the heuristic matching of frameworks and functions with appropriate precursors has been developed.

As practitioners of organic chemistry on a daily basis, whether in the laboratory or in the classroom, we inevitably have to deal with structural formulas. By tradition, if not for lack of other means, chemical structures are drawn on paper or on the blackboard. Although we occasionally rely on molecular models, the large part of our mode of communication and thinking relies on a translation of dynamic three-dimensional molecular structures into flat 2-D representations on paper. In recent years, the computer has begun to play an increasingly important role in our planning of projects dealing with organic, bioorganic, and medicinal chemistry, in interpreting results, and in the rapid retrieval of information.[1] Computer-assisted organic synthesis,[2,3] for example, has evolved to great levels of sophistication and application since the inception of pioneering concepts along these lines by Corey and co-workers[4] some 20 years ago. In a recent review, Hrib[2] has divided computer programs used by chemists into "passive" and "active" categories depending on their function. An impressive number of programs are now available that address different yet important issues dealing with organic synthesis[5] and molecular modeling.[6] In spite of the availability of these innovative programs, few if any specifically address a fundamentally critical problem, namely, the ability to analyze and perceive stereochemical features in the relative as well as the absolute sense.[7] Such information is crucial when planning a synthesis strategy of a target molecule. Thus, knowledge of the sense of chirality of individual carbon atoms, the relationships between segments bearing identical functional groups, and the overall stereochemical and topological features relating to various projections in space, is imperative for the synthesis of optically active target molecules. Molecular models provide a tactile and visual device that link out 2- and 3-D worlds of molecules. Consider, however, the time involved in the building of such molecules as rifamycin, erythronolide A, or amphotericin B with CPK space-filling or Dreiding models. To obtain the required stereochemical information from such models may be equally tedious and at times frustrating. Being cognizant of these problems, both in teaching and in research, we reasoned that part of the "missing" information from the third dimension dealing with hidden functional and stereochemical features in molecules could well be probed by the computer and presented to us in a visually clear and useful format on screen. By using appropriate algorithms, such an analysis could be free of the biases often imposed upon the user by his or her own powers of perception or predisposed attitude toward the specific problem.

THE CHIRON PROGRAM

CHIRON is an interactive program for the analysis and perception of stereochemical and functional features in organic molecules.[8] It is also a powerful tool for heuristic synthesis planning with appropriate starting materials based on pattern recognition of certain structural features. The program consists of four subprograms that deal with drawing structures: CARS-2D (Computed-Assisted Reaction Schemes and Drawing), CARS-3D (Simulation of 3-D Perception by Simultaneous Projection on Two Planes and Stereoscopic Viewing in Real-Time Mode), CASA (Computer-Assisted Stereochemical Analysis), and CAPS (Computer-Assisted Precursor Selection). In this paper we outline salient features concerned with all these subprograms except for CAPS, which is discussed only briefly.[9]

Primary consideration in any computer graphics program is the facility of drawing structures and the efficiency of interaction with the user, without prior knowledge of programming skills. CHIRON was developed based on such a philosophy, with the practicing organic chemist, teacher, and student in mind.

Graphics. Structure input in CARS-2D or -3D is done through primary menus by using a mouse or other device. Plotting or printing the information on the screen is possible through various devices, including laser printing in PostScript mode. Although many graphics programs have common features for drawing structures, CHIRON offers a variety of options with minimum use of the keyboard. Of interest is the Grid option which allows chemical structures to be drawn uniformly by connecting corners of a hexagonal grid on the screen (Figure 1). The Build option has a set of predrawn shapes (templates) of common organic molecules which can be accessed, moved, duplicated, etc. Common atoms as well as others from the periodic table can be incorporated into the drawings. Up to 250 atoms (excluding hydrogens) are accepted by the program. Bonds are drawn as straight lines or using the Alpha or Beta commands for stereochemical depiction. Coordinate information is automatically relayed to the computer as the structure is drawn, and a connection table is created. Many other options in the CARS-2D menu and submenus such as Move, Scale, Rotate, Text, Arrow, Delete, Undo, etc. are available. It is possible to draw reaction schemes with reagents that are suitable for publications, reports, etc.

3-D Graphics. CARS-3D offers the user a unique way to draw and simulate three-dimensional shapes.[10] Structures can be drawn by using many templates with the 3-D Build option, and bonds or appendages can be introduced with the 3-D Draw command. Alternatively, molecules can be drawn in the 2-D mode and automatically converted to 3-D structures. Molecules are thus seen in a "box", with reflections of the structure, including bonds appearing on horizontal (floor) and vertical (wall) planes simultaneously (Figure 2). A semiautomatic rotation in three different planes along X, Y, and Z coordinates produces the desired perspectives, which are rotated by increments of designated angles. The *trans*-decalin structure in Figure 2 was constructed with templates found in the 3-D Build option. The bonds were introduced using the 3-D Draw option, where the computer suggests directions for bonds once

0095-2338/90/1630-0413$02.50/0

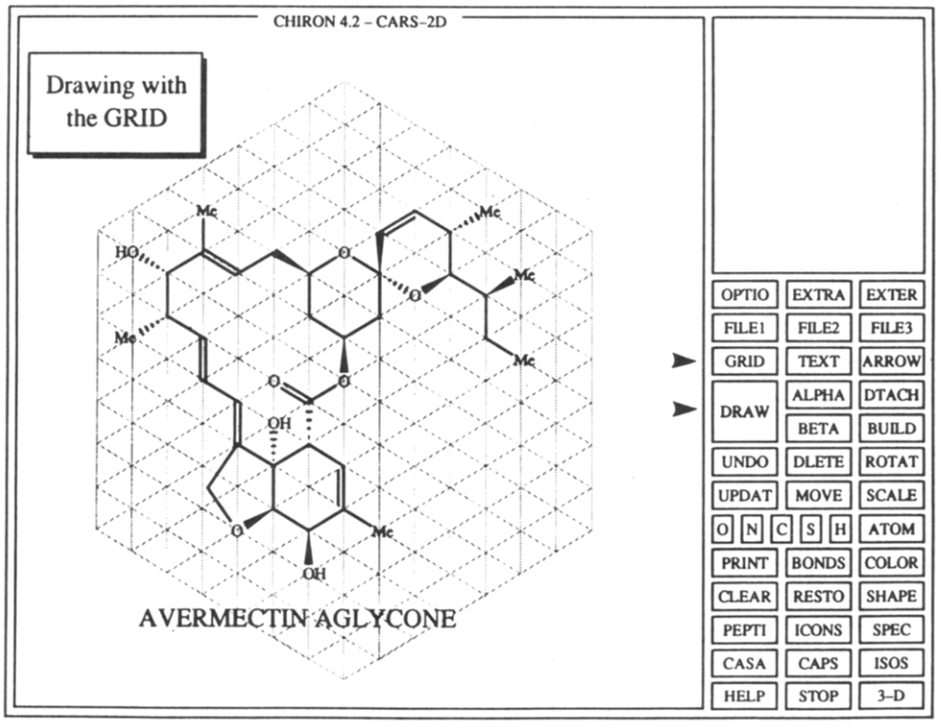


Figure 1. Drawing molecules using the Grid and Draw commands in the CARS-2D menu.

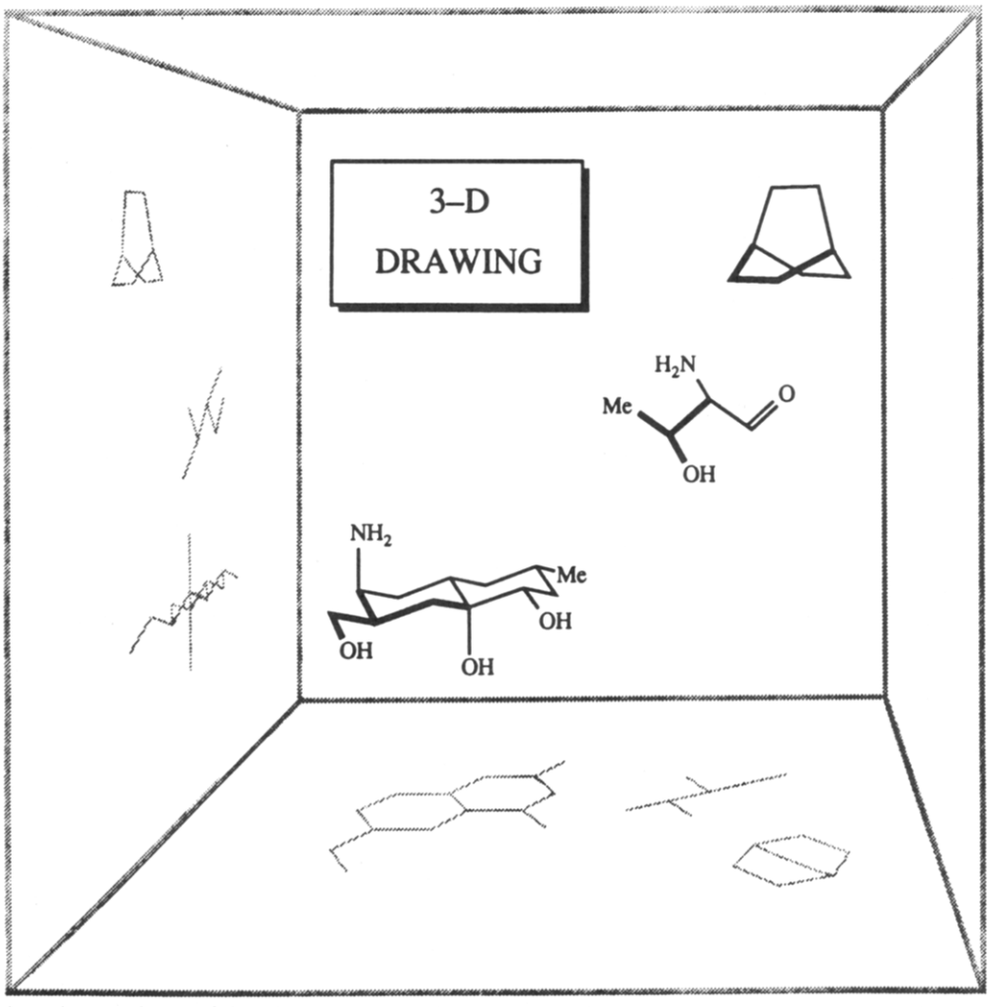


Figure 2. 3-D drawing using templates and the 3-D Draw command. Bottom and left-hand reflections represent "top" and "side" views, respectively. Bonds can be added to the reflected structures.

an atom is indicated. By observing the orientation of a bond in one or both perspective reflections, and following simple drawing commands, a simulated three-dimensional structure results. Figure 3 shows the automatic 2-D → 3-D conversion

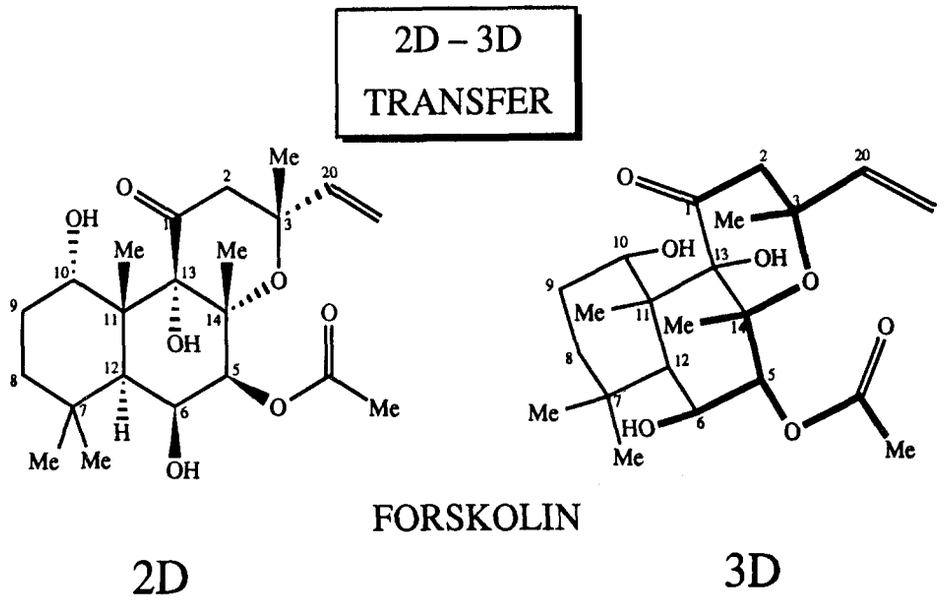


Figure 3. Automatic 2-D → 3-D conversion of the structure of forskolin in the CHIRON 3-D option. Highlighting of the "front" end of the molecule and rotating around the *Y* axis are done by activating appropriate commands from the 3-D menu.

of the structure of forskolin, which is rotated in the *Y* axis before the highlighting option is activated. The 2-D structure was drawn using the Grid or Build options.

Although the cyclic templates in the 3-D Build option represent energy-minimized structures, the resulting molecules (e.g., forskolin) are not subjected to further minimization routines in CHIRON. This could be done by interfacing with one of several other programs (see below). Since the structures can be made to rotate in increments of predesignated angles along the *X*, *Y*, or *Z* axes in the "box", a preferred perspective can be "frozen" and analyzed. Clearly, this mode of perception of molecules has direct applications in teaching stereochemistry and conformational analysis.

Real-Time Rotation and Stereoscopic Viewing. The 3-D option in the CHIRON program features stereoscopic viewing with color-coding on the Silicon Graphics 4-D series workstations. Molecules can be viewed in a number of modes (line bonds, dot matrix, space-filling, ball-and-stick as well as atomic radii). They can be made to rotate, translate, zoom, and tumble along the *X*, *Y*, and *Z* axes, while being viewed in stereoscopic mode with a stereoviewer. Molecules can be juxtaposed along a designated bond, thus simulating an automatic docking routine.

Computer-Assisted Stereochemical Analysis (CASA). As previously mentioned, one of the important prerequisites in the planning of a synthesis of a complex natural product lies in an accurate analysis of stereochemical features. The CASA subprogram with its separate menu offers a full complement of stereochemical information on structures drawn using the 2-D or 3-D modes. These include *R*/*S* absolute configurational assignment, *E*/*Z* geometry, Fischer, extended (zig-zag), and mirror images, chiral substructure analysis and interrelations, and a reshape option. Newman projections can be obtained in the CARS-3D option. Several of these features are unique to the CHIRON program, and they constitute a superb aid to synthesis planning, particularly when the target molecules are polyfunctional. Only a selected number of the above options will be discussed and illustrated below.

***R*/*S* Configurational Assignment.** The Cahn–Ingold–Prelog rules for configurational notation[11] of stereogenic carbon atoms have been implemented in the program. Simply touching the *R*/*S* command box will result in the appearance of the appropriate symbol next to a given stereogenic carbon atom almost instantly, and without the necessity for including the hydrogen atoms. Configurational assignment of simple chiral sulfoxides is also possible. Figure 4 illustrates a collection of molecules containing one or multiple stereogenic carbon atoms with different degrees of difficulty (e.g., fredericamycin) and the response given by the program. The speed of assignment and the exceedingly high percentage of accuracy is impressive. Problems and misassignments may arise due to the human error of perception when drawing a structure, or occasionally because of program-related reasons. These isolated instances, mostly the result of an incorrect perspective in drawing in 2-D, can be remedied in the corresponding 3-D structures. The configuration of asymmetric sulfoxides can be assigned by including a lone pair as a ligand and representing the sulfoxide as a sulfenic acid.

Excluded from consideration are stereogenic carbon atoms carrying equal weights on either side as in some inositols and other highly symmetrical molecules. In such cases the middle carbon will remain unassigned, and an asterisk will appear next to it. Asymmetry by virtue of hindered rotation of biaryls, for example, and compounds with special axes of symmetry are for the time being not considered.

Another feature of the CHIRON program is the automatic configurational assignment of a stereogenic center when it is represented by a line bond, i.e., neither α nor β. Simply pointing to the carbon atom and designating *R* or *S* from the menu will redraw the structure with a hashed or wedged bond. This feature can greatly accelerate the registering of structures originally drawn without stereochemical perspective (α or β), but with a knowledge of *R* or *S* stereochemistry for each center.

The *R*/*S* Priority Tree. A pedagogically useful feature in the CASA subprogram is the possibility to generate a hierarchical priority for the *R*/*S* sequence rule by pointing to a specific carbon atom. Figure 5 illustrates two such examples where the hierarchy is depicted from left to right with the "weights" of the chains in decreasing order. For example, CHIRON assigned an *S* configuration to C_8 in gibberellic acid because the priority rule establishes the order of carbon atoms as 6 > 9 > 15. Color-coded branches help visualize the priorities even better on the screen.

Fischer and Extended Projections. The rapid generation of Fischer projections for molecules or segments thereof in graphic format can be an invaluable asset in the stereochemical decoding of configurational relationships. Such projections have been the basis for all stereochemical interrelationships since they ingeniously depict the chiral backbone of a sub-

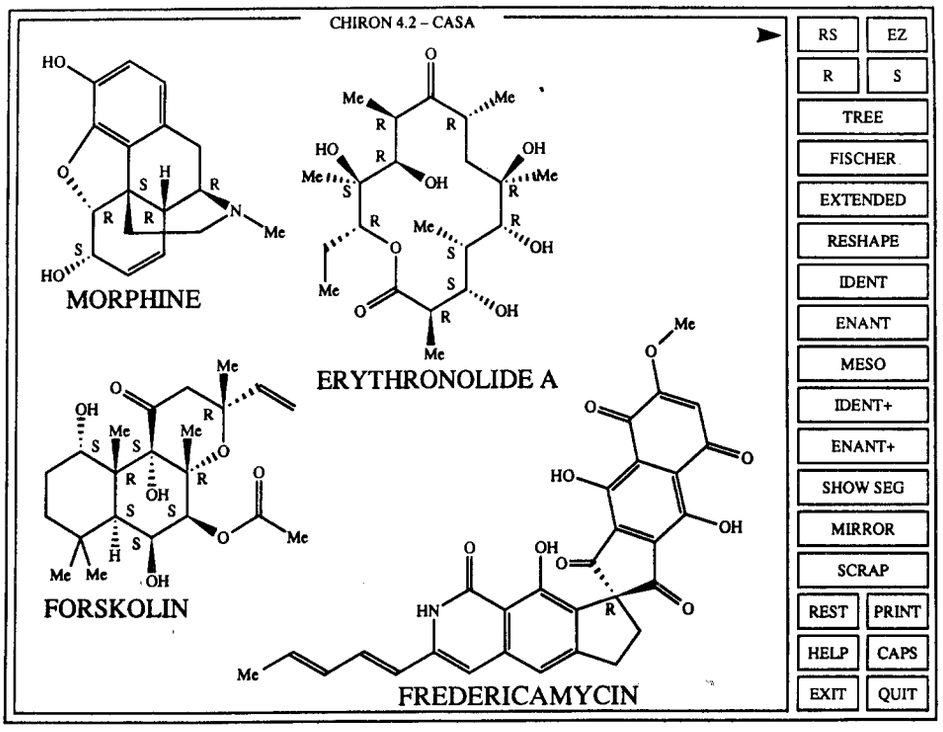


Figure 4. Automatic assignment of *R,S* configurational notation using the RS command in the CASA menu.

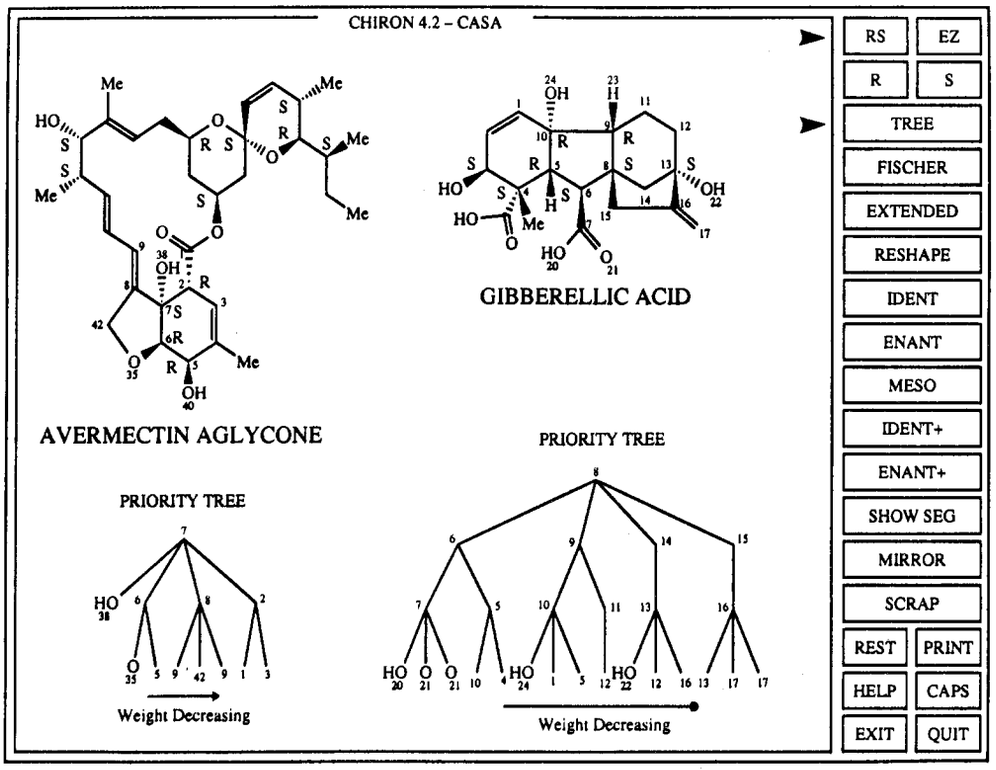


Figure 5. *R,S* configurational assignments with a hierarchical "priority tree" showing the priority sequence at a designated carbon atom (e.g., C_7 in avermectin aglycon and C_8 in gibberellic acid).

stituted carbon chain in a linear format. In effect only one stereogenic carbon atom is considered to be in one plane at any time, with the vertical substituents (usually the carbon chain) drawn back and the two other substituents wedging outward. The resulting loop is "stretched" and seen head-on in the linear perspective of a Fischer projection.[12] Clearly this is not an easy exercise in visual perception for a given structure, particularly with a number of stereogenic centers and rings are involved. More than often, molecular models come to the rescue. The same arguments are true for extended (zig-zag) projections. Using the CASA menu in CHIRON, it is now possible to display one or more segment(s) of a given molecule as Fischer projections simply by touching the first and last carbon atoms of the segment(s). Following clearly displayed instructions on the screen, one indicates two points on a vertical plane and the desired projection will appear almost instantly. The same procedure is followed for extended projections, except that a horizontal perspective is indicated. Consider for example a polycyclic molecule such as monensin. How long would it take for the average organic chemist to draw the

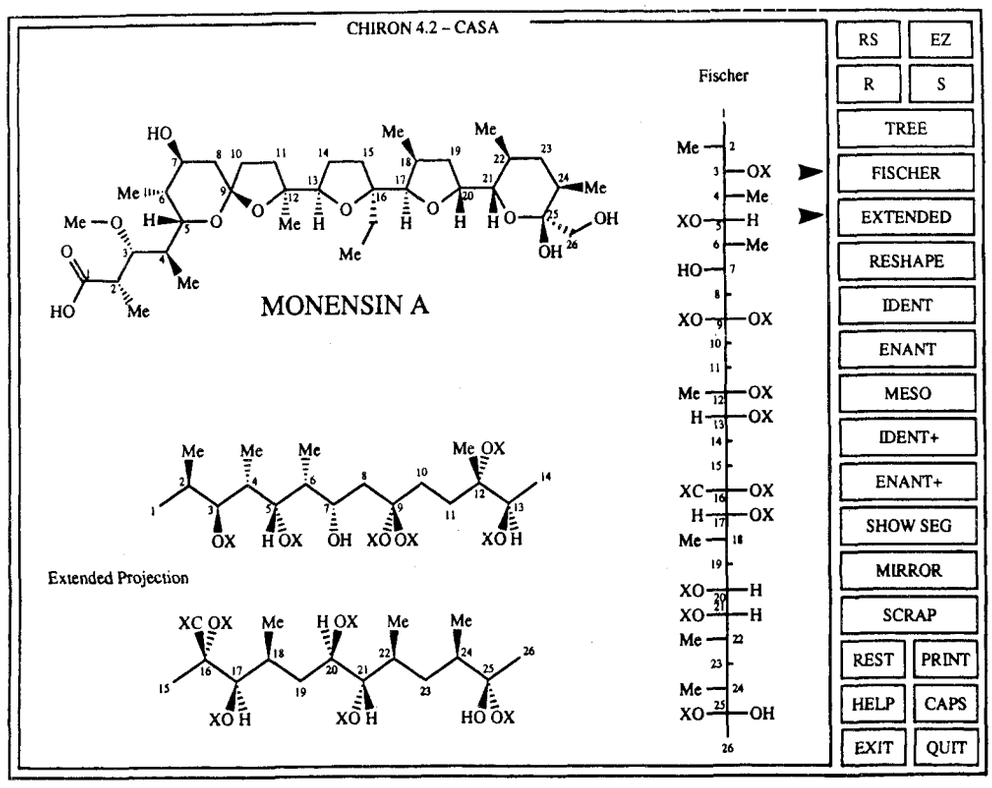


Figure 6. Depiction of Fischer and extended (zig-zag) projection for monensin using the CASA menu. Substituents are indicated as X.

molecule as Fischer projections? Figure 6 illustrates the perceptive and analytical powers of CHIRON, which is capable of displaying such relationships almost instantly. Note that the reverse projection (i.e., C_{26}–C_1) may unravel an altogether different stereochemical situation, hence an alternative synthetic strategy based on a different chiral progenitor. A subtle yet pedagogically useful feature emerges from the visualization of the Fischer and extended projections. The relationship between the C_2 and C_3 substituents in the extended projection is syn which in actuality corresponds to a threo absolute configuration in the Fischer projection.

A most useful feature of such projections made possible by the CHIRON program is the comparison of substructures within two or more molecules simultaneously. Molecules with similar biosynthetic pathways such as the polyether antibiotics[13] will most likely have coincident segments that can be easily visualized in the projections rather than the structures themselves. The same comparison can also be done for two or more molecules that have different biogenetic origins, hence even more remote in their structural resemblance. The emergence of common segments may be helpful in planning syntheses using a common precursor.

Newman Projections. The possibility to simulate depth and perspective drawing in CHIRON-3D has led to the development of the automatic transformation of two adjacent atoms into a Newman projection (Figure 7). The ability to view the projection from "front to back" and vice versa, coupled with the possibility to rotate along three planes, offers rapid and unique modes of visualization along a given axis of a bond. As in the previous options for stereochemical analysis, this option offers another valuable tool in perception which can find utility in teaching and in research.

Chiral Substructure Searching. It is often very useful to obtain information on the functional and stereochemical relationships of substructures or segments within a molecule or between two or more molecules. For example, in the case of a molecule that contains a repeating set of functional groups, one may wish to know whether various segments harboring such functionalities are identical (superimposable) or enantiomeric (non-superimposable mirror images). A meso relationship may also be desirable to uncover in more intricate structures. Finding identical or quasi-identical relationships between two structurally unrelated molecules may lead to the discovery of a common starting material or precursor for synthesis. By activating the Ident command, CHIRON will find the identical (superimposable) segments consisting of four carbon atoms or more within a given structure, or between two different structures, and it will display them on the screen in a visually clear perspective. Terminals with color graphics depict the identical segments with a different color, compared to the original molecule. On monochrome terminals, the pertinent substructures will appear in bold lines. Since the property in question is related to a particular orientation in space, an arrowhead at the end of the redrawn substructure will indicate the direction in which the substructures are superimposable. Let us examine the results of such an analysis for monensin, avermectin aglycon, and erythronolide A (Figure 8). The discovery by CHIRON that the C_3–C_7 and C_9–C_{13} segments in erythronolide A are identical augurs well for a common precursor strategy consisting of a five-carbon chiron[14] (chiral synthon) harboring two stereogenic centers with differentiable terminal functionality.[15] There are two sets of identical four-carbon substructures each bearing a vicinal C-methyl and hydroxyl group. A text option allows the user to read the result of such substructure searches by displaying them on the screen. CHIRON rapidly uncovers a five-carbon "polypropionate"-derived identical substructure with three contiguous stereogenic centers corresponding to $C_4 \rightarrow C_8$ of monensin and $C_{11} \rightarrow C_{16}$ of the immunosuppressive agent FK506[16] (Figure 9). Can the reader "see" a six-carbon common substructure in monensin and erythronolide A? CHIRON sees C_1–C_6 and C_3–C_8 as being identical! Consider now the structures of gibberellic acid and morphine shown in Figure 10. Is the average reader capable of uncovering four, five- (or higher) carbon subunits with patterns of substitution that are related in an identical way (with allowances made for functional convergence)? In the Ident search mode CHIRON was able to find at least 20 such "identities" which are listed

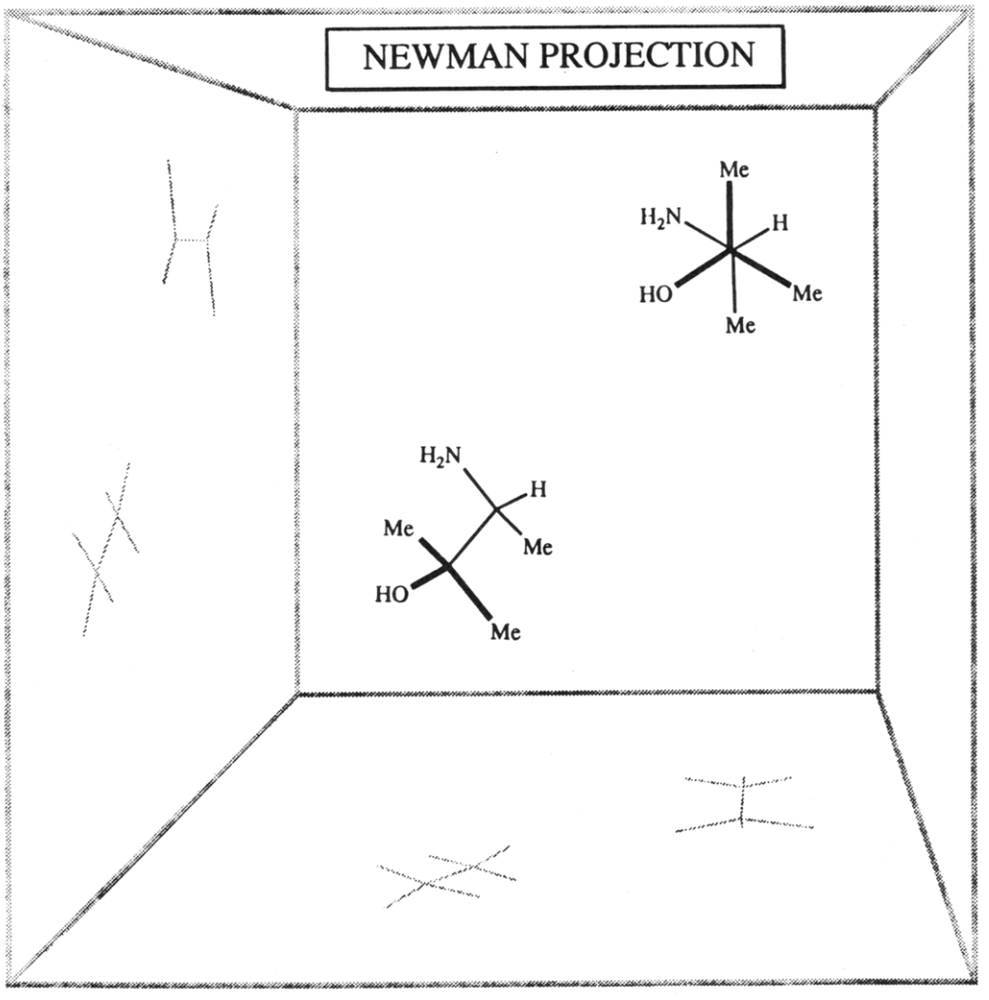


Figure 7. Visualization of a Newman projection (top right) by pointing to the "front" (highlighted) carbon atom first and the "back" carbon atom next of the staggered structure (bottom left). Note the simultaneous projection of reflected "top" and "side" view perspectives.

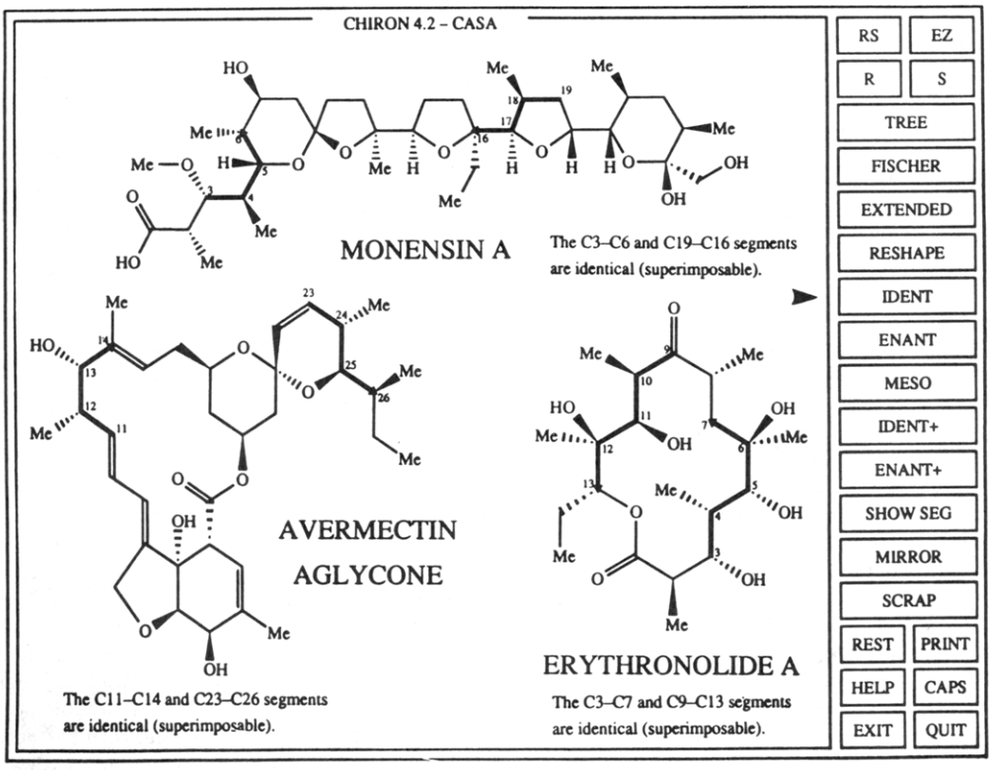


Figure 8. Chiral substructure search for identical (superimposable) segments in monensin, avermectin aglycon, and erythronolide A. Note the direction of the arrowhead in each case.

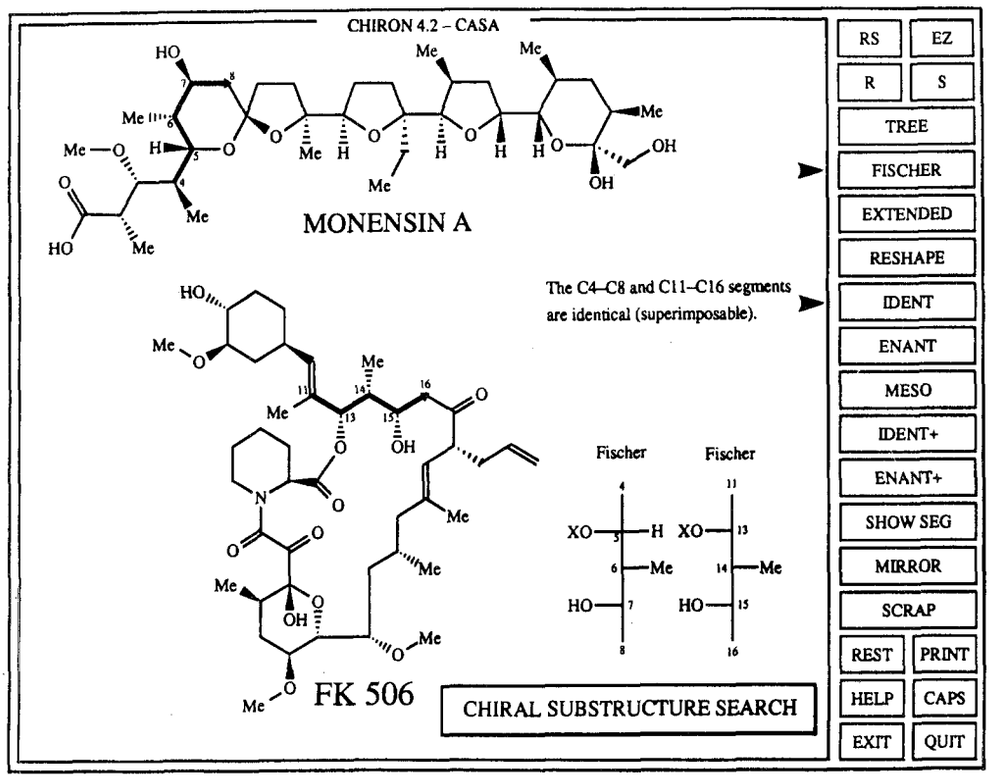


Figure 9. Chiral substructure search between monensin A and FK-506 showing "identity" of the C_4–C_8 and C_{11}–C_{16} segments, respectively. This relationship is also evident from the Fischer projections of the same segments.

in Figure 10. Take, for example, the segments in entry 1 corresponding to C_{12}–C_{14} of morphine and C_2–C_6 of gibberellic acid. Although as entire structural entities there are no evident superpositions of rings with the same orientation of substituents, once hypothetical cleavages are allowed, the program tries to superpose carbon atoms in such a way that identical orientations of functional group (or their equivalents) are obtained. Thus the C_{11} and C_3 hydroxy groups and C_9 and C_5 angular hydrogen atoms in morphine and gibberellic acid, respectively, are "coincident" and superimposable once the substructures are "carved" out of their original forms. The quaternary carbon atoms are also considered although the C-methyl group in gibberellic acid has no exact equivalent in morphine.

CHIRON can also uncover four carbon subunits or higher within one molecule or more that have enantiomeric or quasi-enantiomeric relationships (Enant command). Again, taking gibberellic acid and morphine as examples, there are at least 20 such mirror-image relationships, particularly when substituents at the secondary and tertiary stereogenic carbon atom are considered (Figure 10).

The discovery of stereochemical relationships based on common biogenetic pathways between structurally diverse molecules is thus directly accessible by using the Ident and Enant options. It is obvious that this type of visual perception is exceedingly difficult without the capability of a program such as CHIRON or a related one. This feature can be very useful not only in synthesis planning with a common precursor in mind, but in the search for similar pharmacophores in structurally unrelated molecules.

Diastereomer Searches and Duplicate Structures. CHIRON has the capability to recognize diastereomeric compounds that have the same constitutional structures but differ in stereochemistry. For example, all the 2-amino-2-deoxyhexoses in the D-series that are part of the CHIRON database of precursors will be displayed on the screen when one of the isomers is drawn. CHIRON will also automatically scan its files in search for duplicate structures with the same molecular formula.

Peptide Builder Option. The Pepti option in the CARS-2D menu generates a submenu that allows the user a rapid mode for the drawing of peptide chains. By activating the Build command, one can type the desired sequence of the appropriate amino acid units as one or three letter symbols. The peptide sequence will then be drawn on the screen from left (amino terminal) to right, ending with the carboxyl group. A transoid amide linkage is maintained for all acyclic amino acids, and the initial sequence can be manipulated and modified in several ways. For example, one can easily Insert, Replace, or Extend within the sequence by a new amino acid by following simple commands. These operations are shown in Figure 11, starting with the sequence Ala-Pro-Lys-Thr. The peptide builder option also offers the possibility to work with D-amino acids as well as to draw reverse peptidic sequences. Provisions are made for drawing cyclic peptides and for "twisting" cisoid to transoid amide linkages (e.g., proline). It should be noted that these structures can be transformed in 3-D and viewed using the "box" and Real-Time modes, but they do not represent energy-minimized conformations.

Computer-Assisted Precursor Selection (CAPS). The main feature of the CHIRON program is concerned with a heuristic analysis of target molecules and searching for appropriate starting materials from an existing data base for synthesis.[17] The search is done by a process of pattern recognition involving carbon skeleton, functional, and stereochemical convergences, with allowances made for chemical manipulation. CHIRON will favor the best functional and stereochemical overlap between a precursor and a target (The Chiron Approach).[14] The better the overlap the higher the "score" that it assigns to the particular precursor. Presently a database with 2000 precursor molecules is available with the program. Although the subject matter of this article is not concerned with precursor selection,[9] it may be informative to briefly outline the procedure that one follows in the CAPS menu.

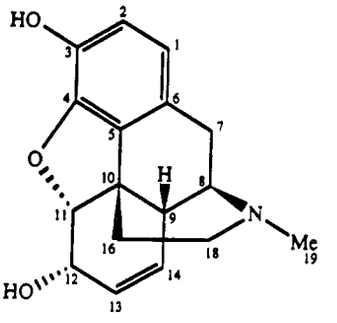
Once a target is drawn, it takes a minute or two to define the search parameters from the menu. Several options are available to the user:

CHIRAL SUBSTRUCTURE SEARCH

IDENT. SEARCH

MORPHINE GIBBERELLIC ACID

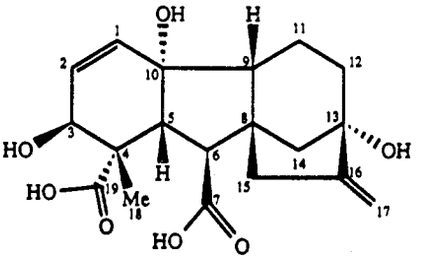
1 : 12, 11, 10, 9, 14 and : 2, 3, 4, 5, 6
2 : 12, 11, 10, 9, 8 and : 2, 3, 4, 5, 10
3 : 12, 11, 10, 5 and : 2, 3, 4, 5
4 : 12, 11, 10, 16 and : 2, 3, 4, 5
5 : 12, 11, 10, 5 and : 2, 3, 4, 18
6 : 12, 11, 10, 9 and : 2, 3, 4, 18
7 : 12, 11, 10, 16 and : 2, 3, 4, 18
8 : 12, 11, 10, 5 and : 2, 3, 4, 19
9 : 12, 11, 10, 9 and : 2, 3, 4, 19
10 : 12, 11, 10, 16 and : 2, 3, 4, 19
11 : 5, 10, 9, 14 and : 3, 4, 5, 6
12 : 16, 10, 9, 14 and : 3, 4, 5, 6
13 : 5, 10, 9, 8 and : 3, 4, 5, 10
14 : 16, 10, 9, 8 and : 3, 4, 5, 10
15 : 8, 9, 10, 5 and : 5, 6, 8, 9
16 : 8, 9, 10, 11 and : 5, 6, 8, 9
17 : 8, 9, 10, 16 and : 5, 6, 8, 9
18 : 8, 9, 10, 5 and : 5, 6, 8, 14
19 : 8, 9, 10, 11 and : 5, 6, 8, 14
20 : 8, 9, 10, 16 and : 5, 6, 8, 14



MORPHINE

ENANT. SEARCH

1 : 5, 10, 9, 8 and : 3, 4, 5, 6
2 : 11, 10, 9, 8 and : 3, 4, 5, 6
3 : 16, 10, 9, 8 and : 3, 4, 5, 6
4 : 5, 10, 9, 14 and : 3, 4, 5, 10
5 : 11, 10, 9, 14 and : 3, 4, 5, 10
6 : 16, 10, 9, 14 and : 3, 4, 5, 10
7 : 14, 9, 10, 5 and : 5, 6, 8, 9
8 : 14, 9, 10, 11 and : 5, 6, 8, 9
9 : 14, 9, 10, 16 and : 5, 6, 8, 9
10 : 14, 9, 10, 5 and : 5, 6, 8, 14
11 : 14, 9, 10, 11 and : 5, 6, 8, 14
12 : 14, 9, 10, 16 and : 5, 6, 8, 14
13 : 14, 9, 10, 5 and : 5, 6, 8, 15
14 : 14, 9, 10, 11 and : 5, 6, 8, 15
15 : 14, 9, 10, 16 and : 5, 6, 8, 15
16 : 8, 9, 10, 5 and : 6, 5, 4, 18
17 : 8, 9, 10, 11 and : 6, 5, 4, 18
18 : 8, 9, 10, 16 and : 6, 5, 4, 18
19 : 8, 9, 10, 5 and : 6, 5, 4, 19
20 : 8, 9, 10, 11 and : 6, 5, 4, 19



GIBBERELLIC ACID

Figure 10. (Top left): Chiral substructure search for "identical" segments (Ident or Enant search) in morphine and gibberellic acid, showing a pair of five-carbon segments and 19 pairs of four-carbon segments. (Bottom left): Chiral substructure search for "enantiomeric" relationships at two stereogenic centers (Enant search).

a. A "convergence" score is chosen by typing in a minimum number. By default the score is set at 65% convergence.
b. General search in the database without considering the parameters c–g.
c. Searching by nature of precursor carbon skeleton (acyclic, carboxylic, heterocyclic, etc.).
d. Searching by functional groups (e.g., OH, terminal olefin, COOH, etc.).
e. Searching for precursors to a specific substructure in the target.
f. Searching by number of stereogenic centers in the precursor.
g. Searching by number of carbon atoms in the precursor.
h. Searching for chiral, achiral, or racemic precursors from a database.
i. Searching for cleaved and reshaped precursors (olefins, diols, ketones, etc.).

The search takes a few minutes or less, with the number of hits appearing on the screen as the search progresses. There are four ways of viewing the hits: 1, Rsp (rapid scanning of precursors), where the highest scoring one appears first; 2, Place, where the precursor structure and its parameters (name, reference, or source, overlap designations) appear on the screen; 3, Transformations, same as Place but with keywords indicating reaction types; 4, List, where a listing of the names, references, etc. of the precursors appears on the screen and can be printed.

Precursor searching can also be done in the 3-D mode, but for the purposes of synthesis planning this may not be necessary. The precursors are color-coded in a way so as to match the corresponding substructure in the target, thus making the entire recognition process more aesthetically pleasing and

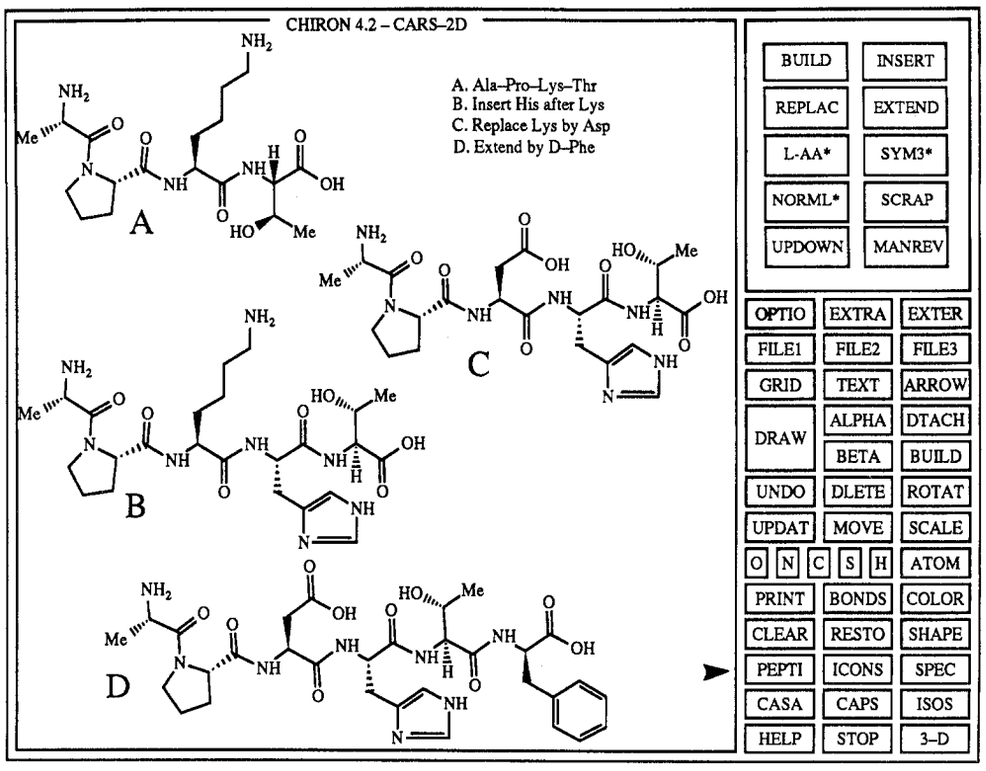


Figure 11. Automatic peptide sequence building using the Build command in the Pepti submenu. The original Ala-Pro-Lys-Thr sequence (A) was modified by inserting His after Lys (B), then by replacing Lys by Asp (C), and finally by extending the carboxyl terminal end by D-Phe.

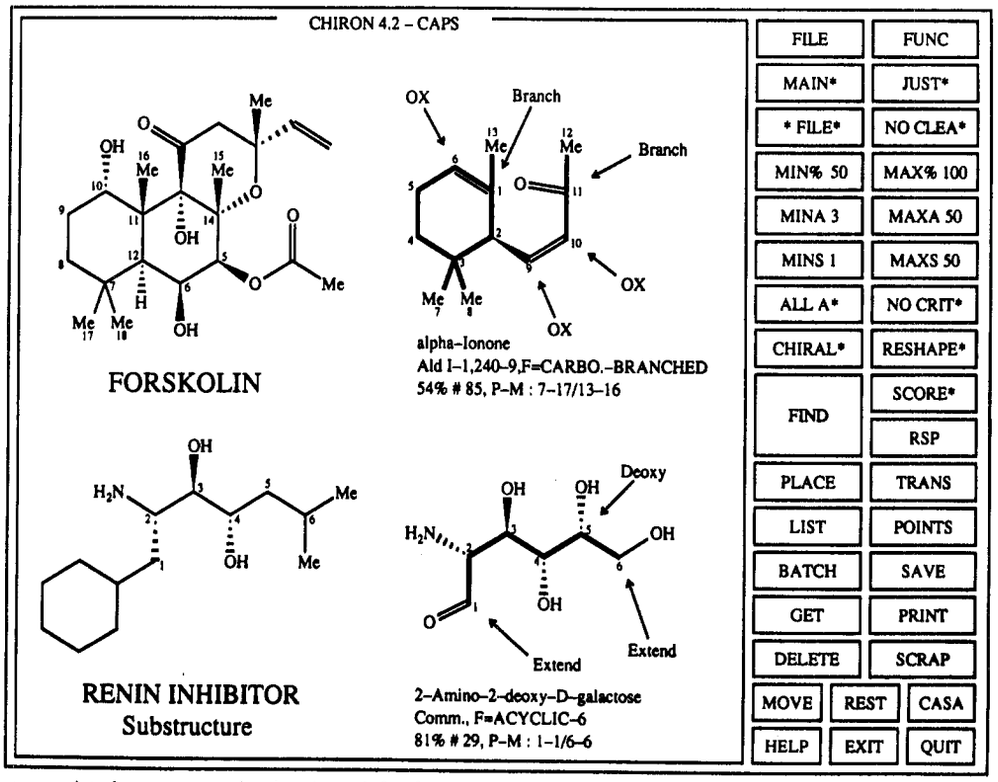


Figure 12. Computer-assisted precursor selection (CAPS) for forskolin and a renin inhibitor substructure using the Transformation option in the CAPS menu.

visually facile. The searches can be done in a Batch mode and the results viewed at another time using the Get command.

Figures 12 and 13 illustrate the results from such a heuristic analysis. For example, α-ionone was found as one of many possible precursors to the decalin segment of forskolin[18] (Figure 12). Note that the carbon skeleton is "reshaped" to adopt the contours of the target molecule as it was originally drawn, and that the convergent portion in the precursor is highlighted with bold lines (color-coded).

The coincident atoms (generally extremeties) are indicated below the precursor (P = precursor, M = molecule). The name, source, file name (F), are convergence score (54%) is also given. The Transformation option generates key words that are pointed toward a particular group that must be ma-

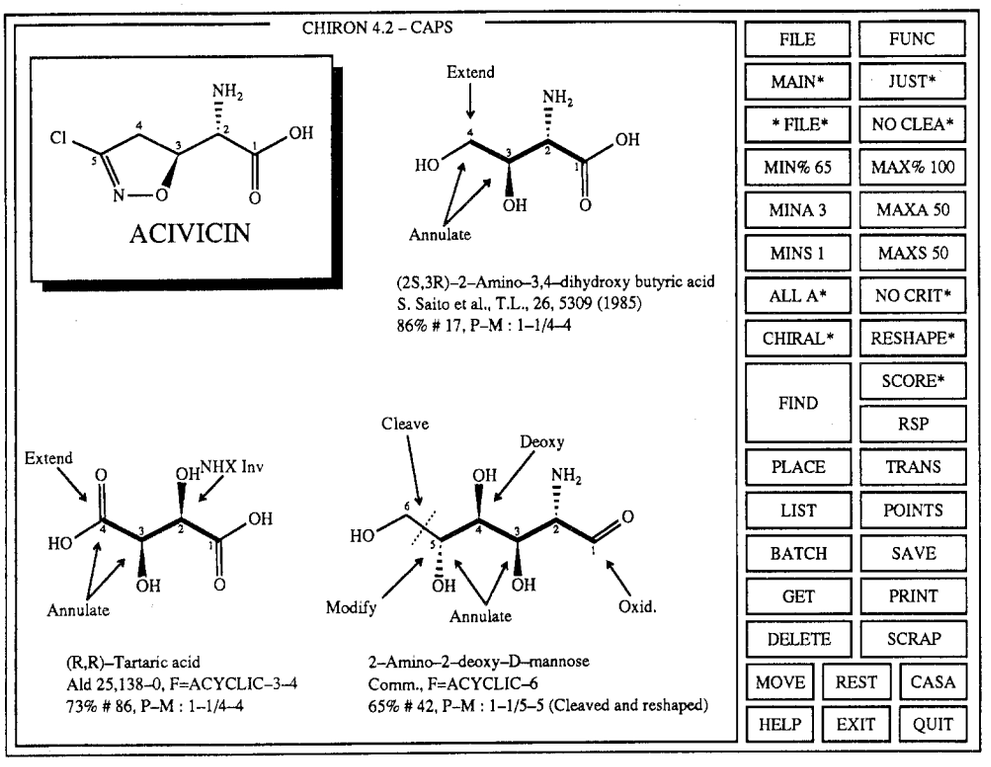


Figure 13. CAPS analysis for acivicin where acyclic 4,5- and 6-C precursors were searched with the Cleave and Reshape options.

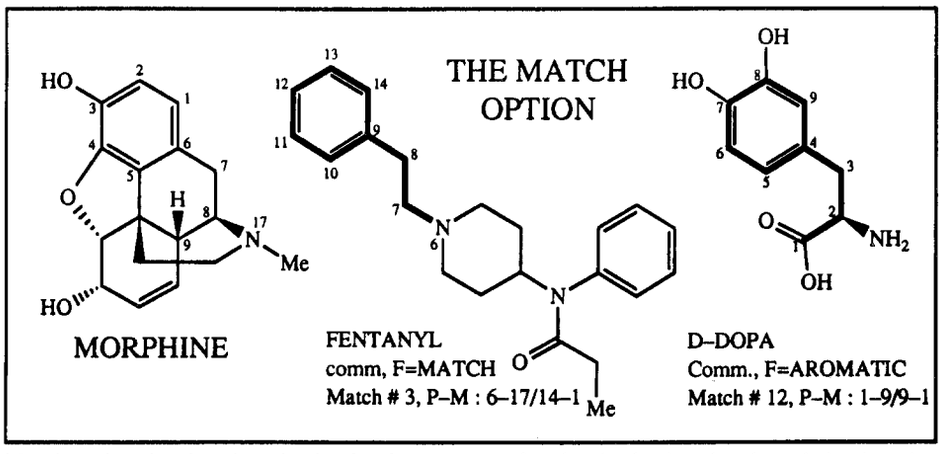


Figure 14. The 2-D Match option showing the selection by the program of molecules having the phenethylamine pharmacophore found in morphine. Fentanyl and D-dopa were entered in a file with the phenethylamino substructure as "essential".

nipulated for best convergence (functional and stereochemical). These key words provide the user with a sequence of "chemical thoughts" and activate a sense of timing which can serve to stimulate our reflexes and power of association.

In Figure 13, several precursors to acivicin are shown with the appropriate key words. Note that the amino sugar has undergone a "cleavage" which the program indicates as a dotted line. A higher score (86%) is assigned to (2*S*,3*R*)-2-amino-3,4-dihydroxy butyric acid because of the obvious functional and stereochemical convergence with C_1–C_4 of the target molecule.

Other aspects of the CAPS subprogram, including the Cleave and Reshape options which represent a new dimension in heuristic analysis and synthesis planning will be discussed elsewhere.[9]

The Match Option. This option will search and match the carbon framework of a molecule with that of a target structure, without consideration of stereochemical or functional overlap. In the Match All option, all atoms are interchangeable in a matching operation, whereas in Match Exact, carbon atoms are matched with the like and hetero atoms may be interchangeable. Because of the large number of overlap possibilities in this mode, it is best to specify a substructure to be matched in the target molecule. This option may be a method for the rapid scanning of possible pharmacophore substructures, even though the search is conducted in a simple 2-D overlapping way. Figure 14 illustrates the discovery of fentanyl and D-dopa as possible pharmacophores that contain the phenethylamine moiety found in morphine. Since energy considerations are ignored in this option, the matching operation may generate unusual or distorted structures which should be dismissed. Allowances have been made for "stretching" certain bonds which are chemically cleavable (e.g., lactones). Note, for example, the matching of erythronolide A aglycon molecule onto monensin as seen in Figure 15, where a C_1–C_{15} overlap is seen. It is of interest that the spiroacetal carbon atom at C_9 in monensin is convergent with C_9 of the macrolide. Although some stereochemical divergence can be found, a good deal of functional overlap exists. Many modes of convergence are suggested by the program, proceeding from left to right along the monensin framework, then working backward. It is possible that in one of these matches a good

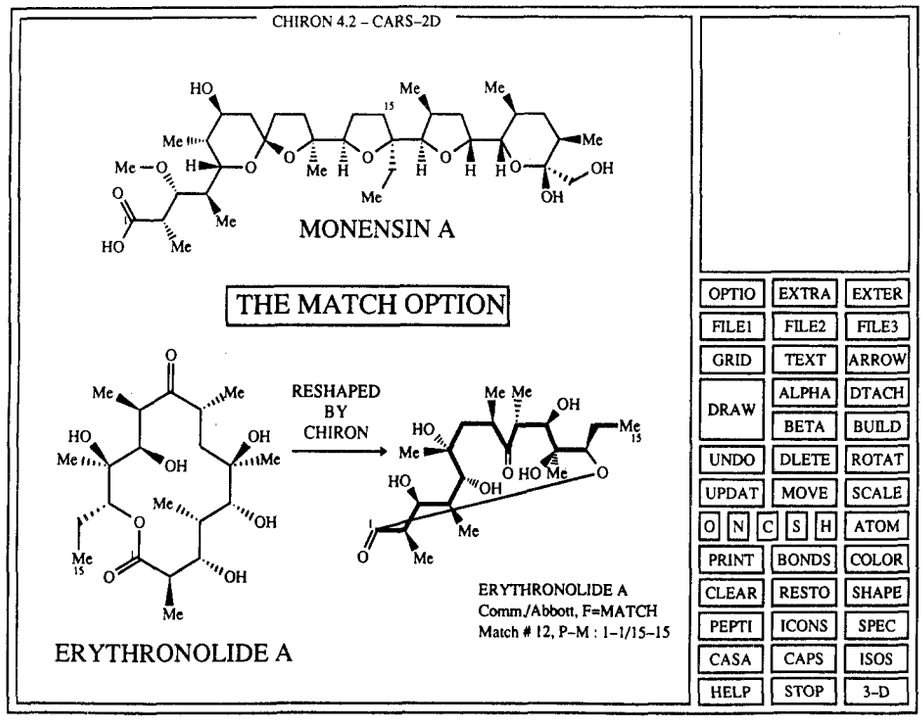


Figure 15. The 2-D Match option using erythronolide A aglycon as a "precursor" to monensin A, showing a C_1–C_{15} overlap with automatic reshaping of the lactone to match the corresponding substructure in the target. Note the exact convergence of several stereogenic centers and the coincidence of a number of functional groups and substituents.

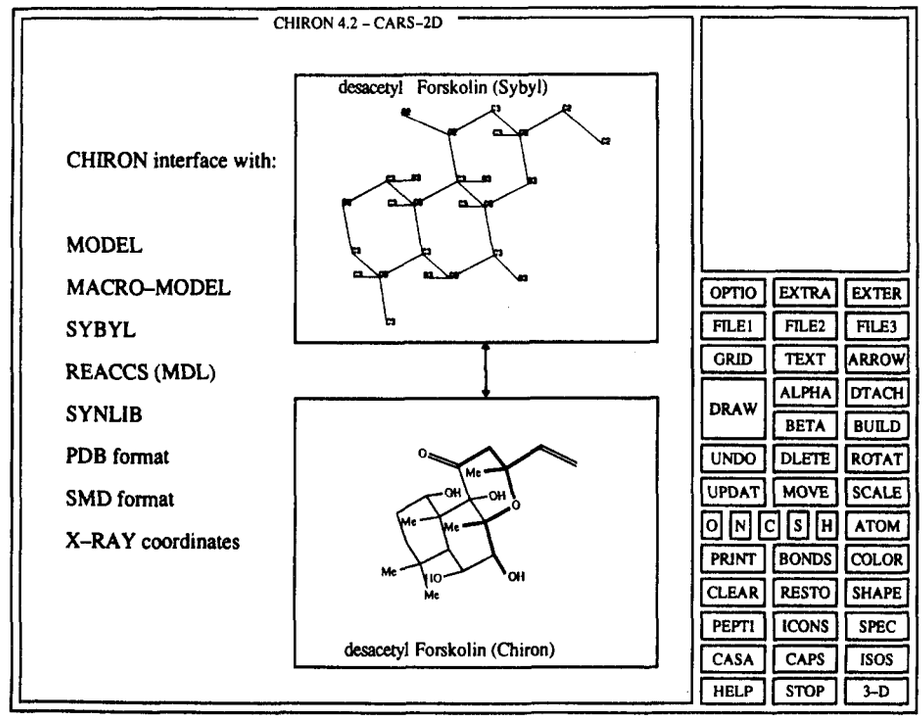


Figure 16. Two-way interface between CHIRON 3-D (bottom) and SYBYL (top) for *des*-acetyl forskolin. Other interface capabilities are shown.

level of functional *and* stereochemical overlap can be found, hence the prospect of using erythronolide A for the synthesis of monensin-like molecules. The Match option will select precursors with the same or lower number of essential carbon atoms compared to the target structure, since it is easier to extend a carbon chain rather than to cleave it. Using this option, a program involving chemical modification by adjustments of existing functionality in certain molecules can be initiated, and new uses for old compounds can be uncovered once important target molecules have been chosen.

Interface with Other Software. CHIRON can interface with a number of other programs, and molecules can be transferred to and from different files. Figure 16 shows CHIRON's interfaces and illustrates an example of a transfer of the forskolin structure from a CHIRON format to a SYBYL input format. Thus, after interacting with the CHIRON program, the user can transfer target molecules to one or more files dealing with other programs and continue with different computer-assisted methods (molecular modeling, database searching, etc.).

CHIRON 3-D capability will enable the user to transfer structures from X-ray coordinates and to display them in various modes on the screen. Figure 17 shows the structure of calcium ionomycin[19] in the stereoviewing mode using the Ball and Stick option. The Rotate command in the Real-Time

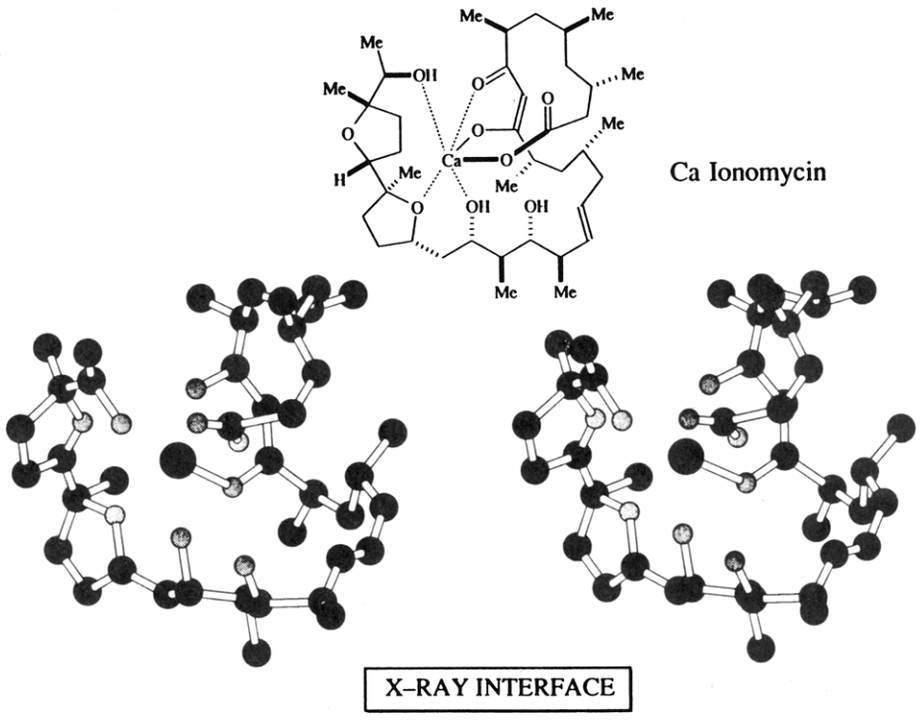
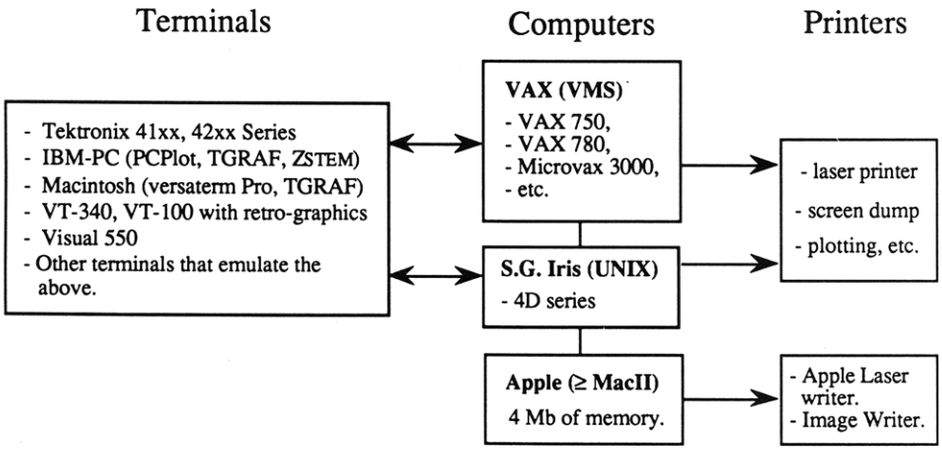


Figure 17. Stereoscopic view of calcium ionomycin shown in the Ball and Stick model format as generated from X-ray coordinates. For clarity the corresponding 2-D chemical structure is also shown.

Chart I. Hardware Diagram for CHIRON



menu enables the user to visualize the molecule in different perspectives along the *X*, *Y*, or *Z* axes.

Other Amenities. CHIRON provides the user with a number of features that can be helpful on a day-to-day basis. In addition to the possibility for drawing reaction schemes and incorporating diagrams in reports or publications, CHIRON will calculate molecular composition, molecular weight, and percent elemental composition for a given structure. Drawing can be done freehand with automatic adjustment of distorted rings using the Reshape command. Charges and valences are considered by the program.

Duplicate entries of the same molecule in a file will be recognized by the program using the Morgan algorithm,[20] and indicated on the screen. Text, arrows of different forms, icons, atoms, and commonly used functional groups (e.g., OMe, SPh, NHCbz, etc.) are included in the program in various submenus or can be typed in for inclusion in the drawing of reaction schemes. Molecules, text, etc. can be moved, rotated, scaled-up and down in size, and saved in a file. Molecules are numbered automatically as they are drawn, and their coordinates, including absolute configuration, are recorded in the connection table. The numbered atoms can be renumbered or their order changed.

Appendix. CHIRON version 4.2 is a multiuser program written in Pascal (68 000 lines). A hardware diagram for CHIRON is shown in Chart I.

ACKNOWLEDGMENT

We thank the Université de Montréal and a number of contributors to the CHIRON program for financial assistance. At various stages, the program has benefitted from the contributions of the following past co-workers to whom we are grateful: S. Meynard, A. Glamyan, E. Lauzon, L. Trépanier, B. Leboeuf, S. Léger, D. Forest, F. Major, and D. Léveillé. We are particularly grateful to Luc Forest for his valuable contributions to the programming of CHIRON and for the management of the group.

REFERENCES AND NOTES

(1) Haggin, J. *Chem. Eng. News* **1985**, *63* (34), 13.
(2) Hrib, N. J. *Annu. Rep. Med. Chem.* **1986**, *21*, 303. Larsen, D. L. *Annu. Rep. Med. Chem.* **1981**, *16*, 281. Zurer, P. S. *Chem. Eng. News* **1985**,

63 (33), 21. Bersohn, M.; Esack, A. *Chem. Rev.* **1976**, *25*, 455. Wipke, W. T.; Howe, W. J., Eds. *Computer-Assisted Organic Synthesis*; American Chemical Society Symposium Series 61; American Chemical Society: Washington, DC, 1977.

(3) Corey, E. J.; Long, A. K.; Rubenstein, S. D. *Science* **1985**, *228*, 408.

(4) Corey, E. J. *Q. Rev. Chem. Soc.* **1971**, *25*, 455. For a selection of seminal references, see: Corey, E. J.; Wipke, W. T. *Science* **1969**, *166*, 178. Corey, E. J.; Howe, W. J.; Pensak, D. D. *J. Am. Chem. Soc.* **1974**, *96*, 7724. Corey, E. J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1976**, *98*, 189. Corey, E. J., et al. *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 221.

(5) For some examples, see for LHASA: Long, A. K.; Rubenstein, S. D.; Joncas, L. J. *Chem. Eng. News* **1983**, *61* (19), 22. For SECS: Wipke, W. T.; Ouchi, G. I.; Krishnan, S. *Artif. Intell.* **1978**, *11*, 173. For CASP: Sieber, W. *Abstracts of Papers*, 191st ACS National Meeting, New York, April 14, 1986; ORGN 85. For CAMEO: Salatin, T. D.; Jorgensen, W. L. *J. Org. Chem.* **1980**, *45*, 2043. For EROS: Gasteiger, J.; Jochum, C. *Top. Curr. Chem.* **1978**, 93. For IGOR: Bauer, J.; Herges, R.; Fontain, E.; Ugi, I.*Chimia* **1985**, *39*, 43. For SYNGEN: Hendrickson, J. B.; Grier, D. L.; Toczko, A. G. *J. Am. Chem. Soc.* **1985**, *107*, 5228. REACSS, SYNLIB, ORAC, CAS ONLINE, etc. are programs available commercially for reaction retrieval, see ref 2.

(6) For recent reviews, see: Cohen, N. C.; Blaney, J. M.; Humblet, C.; Gund, P.; Barry, D. C. *J. Med. Chem.* **1990**, *33*, 883. Frühbeis, H.; Klein, R.; Wallmeier, H. *Angew. Chem., Int. Ed. Engl.* **1987**, *26*, 403. Various modeling programs are available commercially or for academic use: SYBYL, PC-MODEL, CHEM-X, BIOGRAF, QUANTA, CHARMM, etc. See also: Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440 for MACRO-MODEL.

(7) See, for example: Wipke, W. T.; Dyott, T. M. *J. Am. Chem. Soc.* **1974**, *96*, 4825, 4834.

(8) See the CHIRON manual, 1987. See also: Hanessian, S.; Forest, L.; Trêpanier, L.; Lêger, S.; Major, F.; Glamyan, A. Abstracts of *Papers*, 191st ACS National Meeting, New York, April 14, 1986; ORGN 87. Hanessian, S.; Sakito, Y.; Dhanoa, D.; Baptistella, L. *Tetrahedron* **1989**, *45*, 6623. Hanessian, S.; Faucher, A.-M.; Lêger, S. *Tetrahedron* **1990**, *46*, 231.

(9) For a more detailed discussion, see: Hanessian, S.; Franco, J.; Larouche, B. *Pure Appl. Chem.* **1990**, *62*, 1887.

(10) See, for example: Hahn, M.; Wipke, W. T. *Tetrahedron Comput. Methodol.* **1988**, *1*, 81 and references cited therein.

(11) Cahn, R. W.; Ingold, C. K.; Prelog, V. *Angew. Chem. Int. Ed. Engl.* **1966**, *5*, 385.

(12) See, for example: Klyne, W.; Buckingham, J. In *Atlas of Stereochemistry*; Chapman and Hall: London, 1974.

(13) Cane, D. E.; Celmer, W. D.; Westley, J. W. *J. Am. Chem. Soc.* **1983**, *105*, 3594.

(14) Hanessian, S. In *The Total Synthesis of Natural Products: The Chiron Approach*; Baldwin, J. E., Ed.; Pergamon Press: Oxford, 1983.

(15) For a synthetic approach to erythronolide A using a common precursor approach, see: Hanessian, S.; Rancourt, G. *Pure Appl. Chem.* **1977**, *49*, 1201, *Can. J. Chem.* **1977**, *55*, 1111.

(16) Isolation and structure: Tanaka, H.; Kuroda, A.; Marusawa, H.; Hatanaka, H.; Kino, T.; Goto, T.; Hashimoto, M. *J. Am. Chem. Soc.* **1987**, *169*, 5031. Synthesis: Jones, T. K.; Mills, S. G.; Reamer, R. A.; Askin, D.; Desmond, R.; Volante, R. P.; Shinkai, I. *J. Am. Chem. Soc.* **1989**, *111*, 1157.

(17) Wipke, W. T.; Rogers, D. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 71.

(18) The synthesis of forskolin from α-ionone has been recently described: Corey, E. J.; Jardine, P. D.; Rohloff, J. C. *J. Am. Chem. Soc.* **1988**, *110*, 3672. See also: Koft, E. R.; Kotnis, A. S.; Broadbent, T. A. *Tetrahedron Lett.* **1987**, *28*, 2799. Nicolaou, K. C.; Li, W. S. *Chem. Soc., Chem. Commun.* **1985**, 421.

(19) Toeplitz, B. K.; Cohen, A. I.; Funke, P. T.; Parker, W. L.; Gougoutas, J. Z. *J. Am. Chem. Soc.* **1979**, *101*, 3344.

(20) Morgan, H. L. *J. Chem. Doc.* **1965**, *5*, 107.