(c) Visser, T.; Van Der Maas, J. H. *Anal. Chim. Acta* **1981**, *133*, 451.

(10) Debska, B.; Duliban, J.; Guzowska, B.; Hippe, Z. *Anal. Chim. Acta* **1981**, *133*, 303.

(11) (a) Szalontai, G.; Simon, Z.; Csappo, Z.; Farkas, M.; Pleifer, G. *Anal. Chim. Acta* **1981**, *133*, 31. (b) Farkas, M.; Markos, J.; Szepesvary, P.; Bartha, I.; Szalontai, G.; Simon, Z. *Anal. Chim. Acta* **1981**, *133*, 19.

(12) Passlack, M.; Bremser, W. In *Computer-Supported Spectroscopic Databases*; Zupan, J., Ed.; Ellis Horwood: Chichester, 1986, p 92.

(13) Carhart, R. E.; Smith, D. H.; Brown, H.; Djerassi, C. *J. Am. Chem. Soc.* **1975**, *97*, 5755.

(14) Carhart, R. E.; Smith, D. H.; Gray, N. A. B.; Nourse, J. G.; Djerassi, C. *J. Org. Chem.* **1981**, *46*, 1708.

(15) Carabedian, M.; Dagane, I.; Dubois, J. E. *Anal. Chem.* **1988**, *60*, 2186.

(16) Cabrol, D.; Rabine, J. P.; Rouillard, M.; Ricard, D.; Forrest, T. P. (University of Nice and Dalhousie University of Halifax) EXP'AIR PROGRAM V.2, 1989.

(17) Schrader, B.; Bougeard, D.; Niggeman, W. Computer Evaluation of IR and Raman Spectra. Comput. Methods Chem. (*Proc. Int. Symp.*) **1977**, *80*, 37 (44BBAM).

(18) Elyashberg, M. E.; Gribov, L. A. *J. Mol. Comput. Sci.* **1981**, *21*, 48.

(19) Klopman, G.; McGonigal, M. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 48.

(20) GAUSSIAN-80: AB-INITIO MO PROGRAM QCPE 446; Quantum Chemistry Program Exchange: Bloomington, IN, 1980.

(21) Small, G. W. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 232.

(22) Panaye, A.; Doucet, J. P.; Peguet, P.; Dubois, J. E. Proceedings of the 10th International CODATA Conference. *CODATA Bull.* **1986**, *64*, 32.

(23) Peguet, P. Thèse de Docteur Ingénieur, Université Paris 7, 1985.

(24) Socrates, G. *Infrared Characteristic Group Frequencies*; Wiley-Interscience Publications: New York, 1980.

(25) Nakanishi, K. *Infrared Absorption Spectroscopy*; Practical Holden-Day Inc.: San Francisco, 1962.

(26) Bellamy, L. J. *The Infrared Spectra of Complex Molecules*, 2nd ed.; Methuen: London, 1958.

(27) Colthup, N. B.; Daly, L. H.; Wiberley, S. E. *Introduction to IR, Raman Spectroscopy*, Academic Press: New York, 1964.

(28) Bellamy, L. J. *Advances in Infrared Group Frequencies*; Methuen: London, 1968.

(29) Kirrman, A.; Janot, M. M.; Ourisson, G. *Structures et Propriêtês Molêculaires*: *VII Fonctions Monovalentes. VIII Fonctions Divalentes. IX Fonctions Trivalentes. Monographies de Chimie Organique*; Masson et Cie: Paris, 1970.

(30) (a) Dubois, J. E.; Viellard, H. *Bull. Soc. Chim. France* **1968**, 900. (b) Dubois, J. E.; Viellard, H. *Bull. Soc. Chim. France* **1968**, 905. (c) Dubois, J. E.; Viellard, H. *Bull. Soc. Chim. France* **1968**, 913.

(31) Dubois, J. E. In *The Chemical Applications of Graph Theory*; Balaban, A. T., Ed.; Academic Press: New York, 1976; p 330.

(32) Dubois, J. E.; Laurent, D.; Viellard, H. *C. R. Acad. Sci., Ser. C* **1966**, *236C*, 764, 1245.

(33) Dubois, J. E.; Laurent, D.; Viellard, H. *C. R. Acad. Sci., Ser. C* **1967**, *264C*, 348.

(34) Attias, R. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 102.

(35) Dubois, J. E.; Carabêdian, M.; Ancian, B. *C. R. Acad. Sci. Ser. C* **1980**, *290C*, 369, 383.

(36) (a) Dubois, J. E.; Carabêdian, M.; Dagane, I. *Anal. Chim. Acta* **1984**, *158*, 217. (b) Carabêdian, M.; Dagane, I.; Dubois, J. E. Systêmes Experts et leur Applications. *INA* **1985**, *1*, 401.

(37) Lioutas, A. Thèse de Doctorat, Universitê Paris 7, 1986.

(38) Laurent, D.; Aranda, A. *J. Phys. Chim. Biol.* **1973**, *70*, 1068.

(39) Dubois, J. E.; Mercier, C.; Panaye, A. *Acta Pharm. Jugosl.* **1986**, *36*, 135.

(40) Dubois, J. E.; Panaye, A.; Attias, R. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 74.

(41) Panaye, A. ASC: Adaptation Structurale Conventionnelle. Thêse de Doctorat, Universitê Paris 7, 1976.

(42) Zupan, J.; Hadži, D.; Penca, M. *Comput. Chem.* **1976**, *1*, 71.

(43) Delaney, M. F.; Warren, F. V.; Hallowell, J. R. *Anal. Chem.* **1983**, *55*, 1925.

(44) Lowry, S. R.; Huppler, D. A. *Anal. Chem.* **1983**, *55*, 1288.

(45) Razinger, M.; Zupan, J.; Penca, M.; Barlic, B. *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 158.

Computational Perception and Recognition of Digitized Molecular Structures[†]

M. LEONOR CONTRERAS,* CARLOS ALLENDES, L. TOMAS ALVAREZ, and ROBERTO ROZAS

University of Santiago de Chile, Department of Chemistry, Casilla 5659, Santiago 2, Chile

Received March 7, 1990

Molecular structures containing both common and special alphanumeric characters are efficiently recognized by a program written in C. The program was designed to process type- and hand-printed structures. A scanner digitizes the corresponding images. Treatment of the binary information obtained in this way includes molecular graph perception and character recognition. Known and new image processing methods for molecular graph perception and an intelligent pattern-recognition principle for character processing were used. A graphic interface allows one to display and manipulate the recognized molecular images. Applications of the software to different areas such as molecular design, automatic input of structures to databases like ARIUSA, and others are also presented.

INTRODUCTION

Representation of molecular structures allows one to describe and study sophisticated molecules such as vitamins, alkaloids, antibiotics, pheromones, organometallic complexes, etc., all of which may contain a 2-D stereochemical representation (dot and wedge convention) and delocalized bonds as in donor–acceptor complexes. Thus, the natural way of knowledge communication and management of information in chemistry is done using these structures. This is true in databases,[1,2] in CAMD,[3] in structure–activity relationships,[4] in synthesis design,[5] etc.

The structures themselves consist of two basic components: (a) a graph[6] or skeleton of the structure and (b) common and special alphanumeric characters (symbols, parenthesis, charges). A program that works with molecular structures must handle both components. That is what most interfaces do for manual input of structures to computer systems.[1-6] The internal representation of that information through a connectivity table is known as recognition of the molecular structure by the system. This recognition of chemical structures is necessary for selective retrieval of information.[1,2] However the input of the structures, especially when they have more than 20 atoms and stereochemical specifications, is a time-consuming process normally requiring specialized people.[1]

In this paper we present a system which supports capture, perception, and recognition of type- and hand-printed molecular structures. In addition, as a part of this system, a graphic interface for the display and manipulation of the recognized structures is also presented.

DESCRIPTION OF THE SYSTEM

The process basically consists of four steps: (a) scanning of molecular structures, (b) graph recognition, (c) character recognition, and (d) display.

[†] Presented in part at the XII Workshop of Systems Engineering, Santiago, Chile, July 1989.

0095-2338/90/1630-0302$02.50/0 © 1990 American Chemical Society

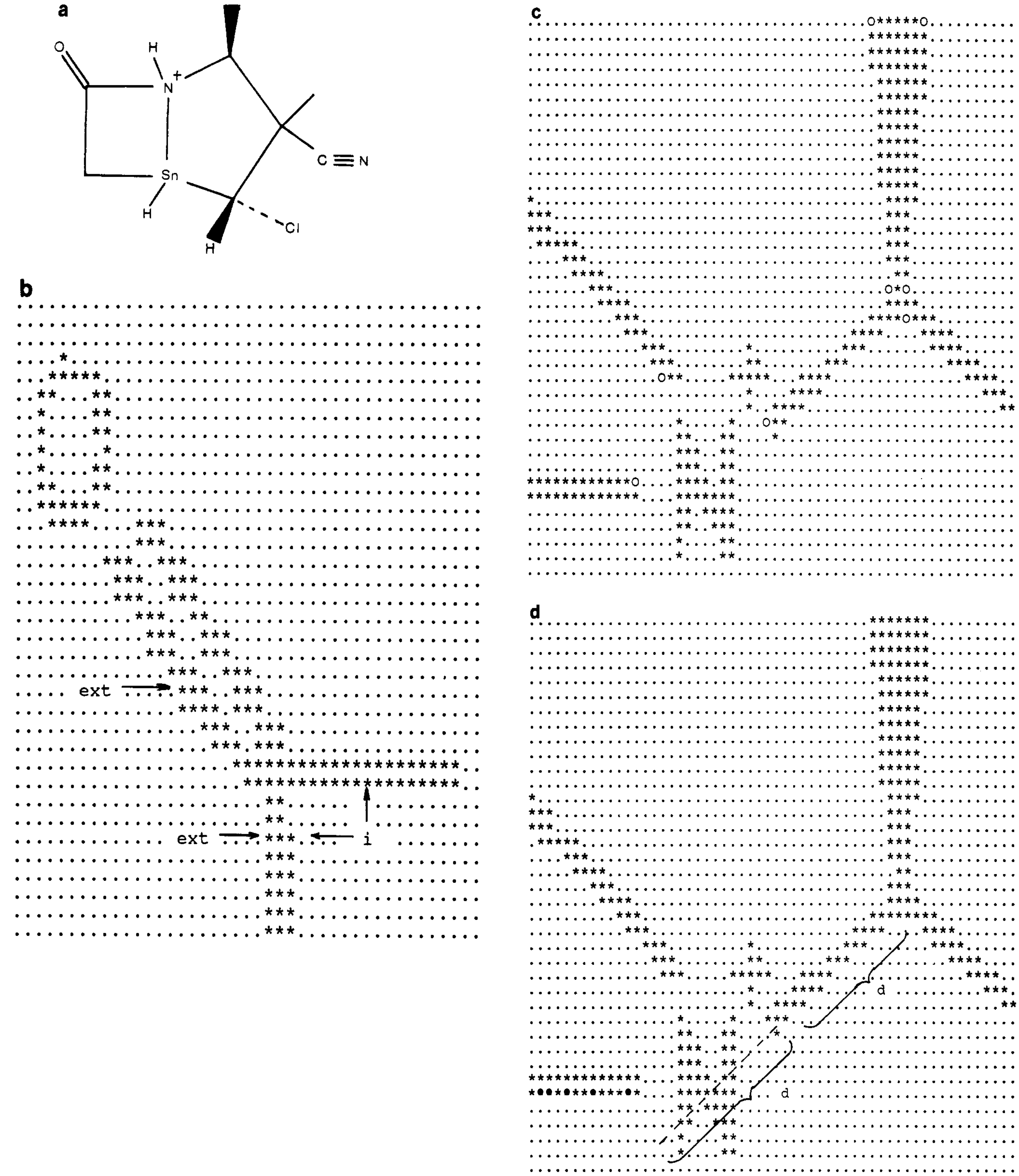

Figure 1. (Panel a) Type-printed molecular structure to be digitized and recognized. (Panel b) Partial view of the digitized structure showing the external (ext) and internal (i) border of a bond. (Panel c) Digitized partial view of the graph showing vertices position at the O points. (Panel d) Perception of an atom by linear projection and search of pixels *on* along that path.

The scanning or capture step is done with a PC-AT microcomputer, under MS-DOS, to which the scanner HP-Scanjet is linked. The scanner digitizes the printed structures with variable resolution (between 75 × 75 and 300 × 300 dpi), according to the size of the molecular structure. The digitized image thus obtained is kept as a matrix for further processing in a microVax II under an Ultrix operating system.

The graph recognition step is in charge of processing the graph of the molecular structure, including its stereochemistry when specified. Figure 1a shows an example of a molecule to be digitized. Here, it is convenient to have in mind that the binary description of the graph is represented by several points or pixels in state *on*. A single bond between two atoms is described by a string of pixels that has a length and a width determined by the original size of the image and the resolution of the scanning. Formally a digitized single bond looks like a thick line that has both an external and an internal margin or border (see Figure 1b). First, for the perception of the graphs, the program does a left-to-right horizontal sweeping. It starts from the left-superior part of the image until it finds

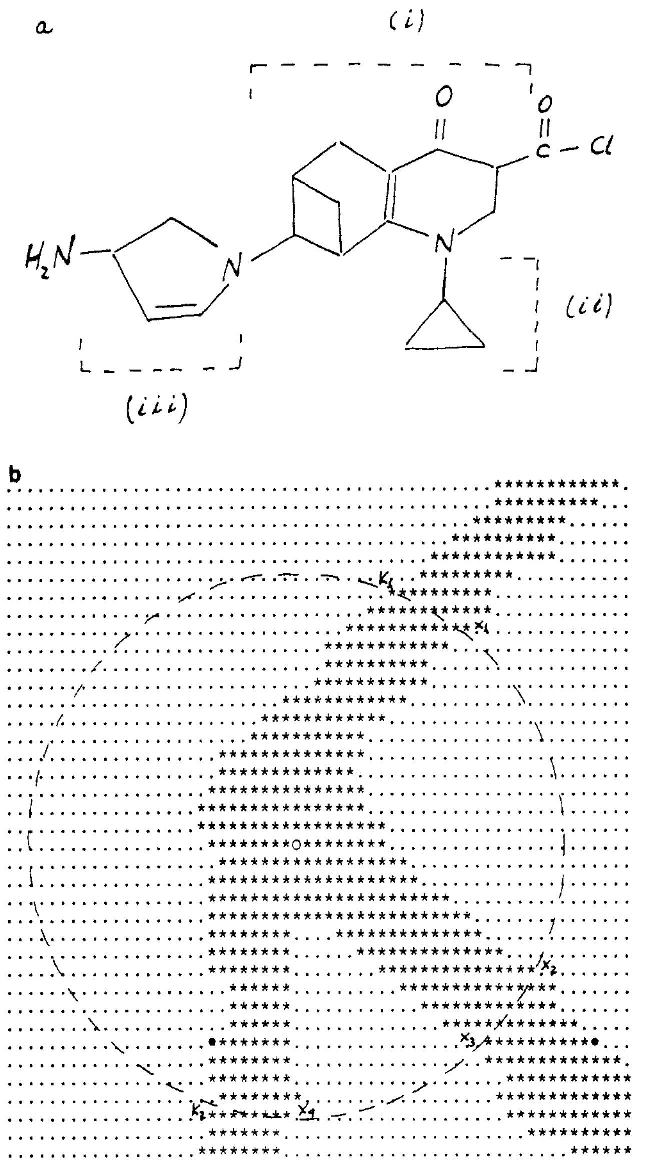

Figure 2. (Panel a) Hand-printed molecular structure having three subgraphs: i, ii, and iii; (panel b) Circular sweeping where k_i and x_i represent known and unknown points, respectively.

the first pixel *on* of the digitized image. Then a counterclockwise contour search algorithm is applied until arriving back at the first pixel. In this way the coordinates of every pixel of the contour are kept in a 2 × *n* matrix, where *n* represents the number of pixels of the contour. This value is dependent on the original image and the scanning resolution as was mentioned before. When *n* is bigger than 300 (for a typical printed structure scanned at 300 dpi), this is interpreted as the graph contour. Otherwise it is considered as a chemical symbol and it is treated as shown later. In the first case, deflection of the linear trajectory of any external or internal border indicates the existence of a vertex.[7] Two or more vertices within a defined small space indicate the point of the graph where an atom should be located (see Figure 1c).

Atoms are numbered, and a neighborhood relationship among them is kept. Perception of terminal atoms—having a single neighbor after the first sweeping—is done by making a linear projection of its previous bond up to a distance similar to the length of that bond (see Figure 1d). If no pixels *on* are encountered, along that path, the atom itself is considered as a carbon atom. Otherwise, a contour determination is done over the new found pixels *on*. In this way a window that contains chemical symbols is detected and submitted to the character-recognition module.

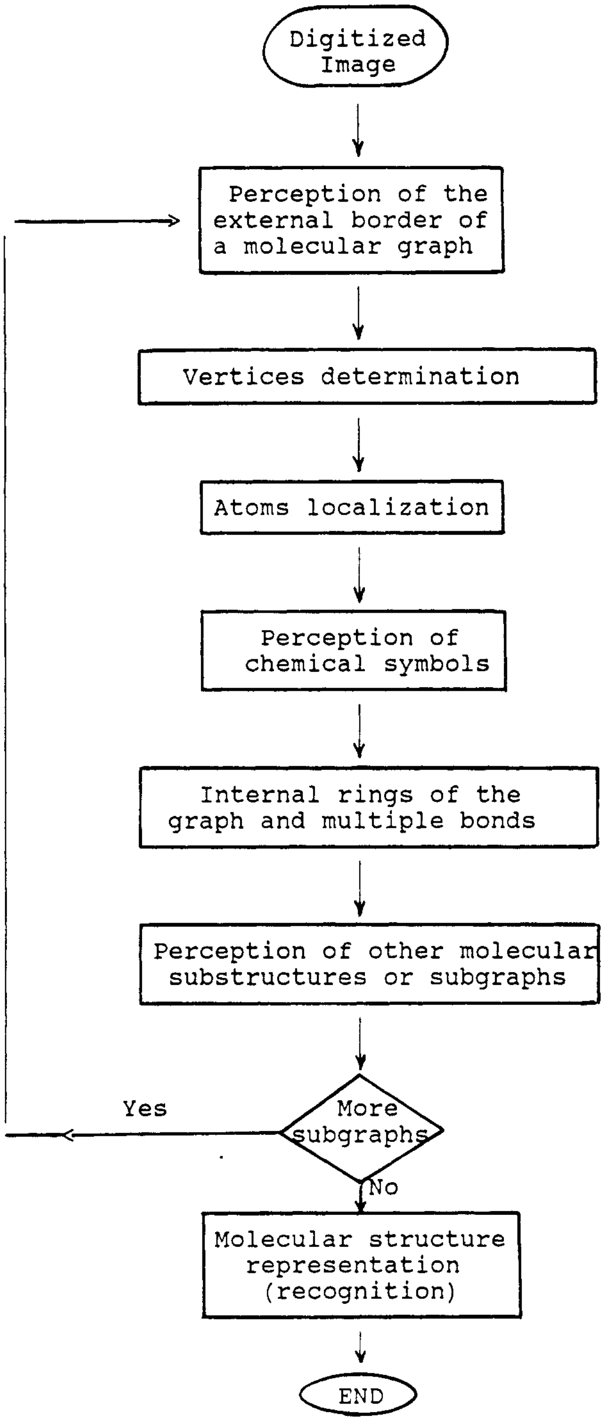
Digitized Image
Perception of the external border of a molecular graph
Vertices determination
Atoms localization
Perception of chemical symbols
Internal rings of the graph and multiple bonds
Perception of other molecular substructures or subgraphs
More subgraphs
Yes
No
Molecular structure representation (recognition)
END

Figure 3. Scheme of the graph-recognition process.

Perception of multiple bonds and internal rings as well as perception of other molecular substructures or subgraphs (see Figure 2a) is done through a circular inspection method. This is applied to every detected atom. For that, a circle of inspection centered on the middle of the space assigned to each atom is considered (see Figure 2b). The radius of this circle is chosen as equal to 0.3 times the value corresponding to a single bond length. Unknown border pixels found in this way are kept (x_i points on Figure 2b). They are used as the initial point for both a new counterclockwise contour search and a perception of new vertices and probable new atoms as it was described before. Figure 3 shows a flow sheet of the graph-recognition process. The source programs of this module occupy 50 kb, and the compiled version (main) occupies 89 kb.

The character-recognition module consists of two principal parts: one that separates each character into a matrix and

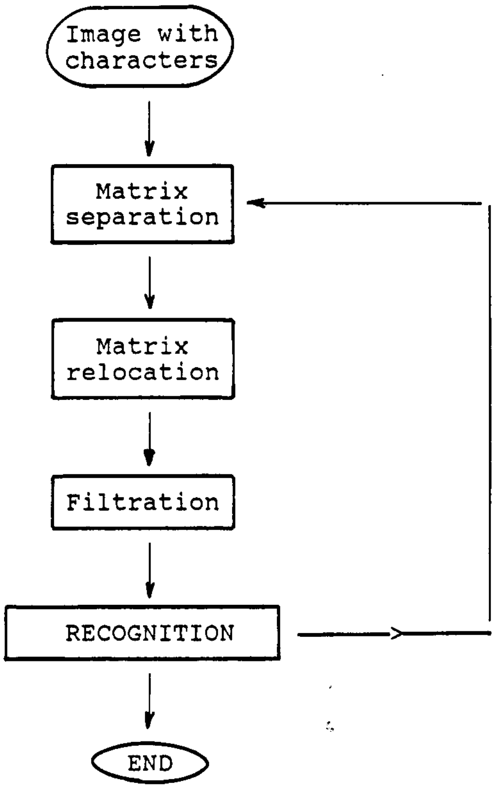

Figure 4. Block diagram for character separation.

the other one that makes the recognition of every matrix.

Character separation is done according to the block diagram shown in Figure 4. The image received from the graph-recognition module is submitted to separation; this is done because characters are often overlapped, and therefore recognition is difficult or unreliable. Once the system separates the characters, it copies each matrix in the left upper corner of a new matrix, to save time in processing and for standard comparison. Then this separated and relocated matrix is submitted to noise filtration to eliminate isolated pixels *on* that were introduced during the scanning process. This last refined matrix is sent to the recognition process. If a character cannot be recognized, the matrix is again submitted to separation until it is recognized.

The recognition process (see Figure 5) starts with a perception of every isolated matrix, known in pattern-recognition as feature extraction. Matrices are analyzed, and a set of parameters is defined for each one on the basis of an intelligent pattern-recognition principle[8] and taking into consideration a certain threshold of pixels *off* (see matrices i–iv of Figure 6). The threshold is found in the direction indicated by the arrows. This digital perception gives the following parametrization to each matrix:

matrix i gives the semibyte | 1 | 0 | 0 | 0 |
matrix ii gives the semibyte | 0 | 0 | 0 | 1 |
matrix iii gives the semibyte | 0 | 0 | 0 | 0 |
matrix iv gives the semibyte | 0 | 1 | 1 | 0 |

Then, classification of the characters is done by assigning to each matrix an identification number (ID), in hexadecimal, in such a way that:

matrix i has an ID equal to 8
matrix ii has an ID equal to 1
matrix iii has an ID equal to 0
matrix iv has an ID equal to 6

Several characters may have the same ID. For instance, characters 'b' and 'h' have the same semibyte as matrix i above. To achieve a unique recognition, other parameters determined in this step and based on the pixels gradient (e.g., concavities, vertical and horizontal lines, and other characteristics of the digitized characters) allow the system to make the final ASCII assignment.

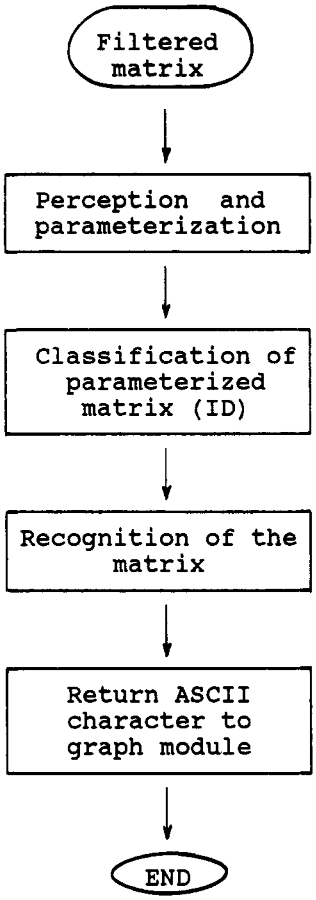
Filtered matrix
Perception and parameterization
Classification of parameterized matrix (ID)
Recognition of the matrix
Return ASCII character to graph module
END

Figure 5. Block diagram for character recognition.

```
          ***.........
  4       ***.........      1
 --->     ***.........    <---
          ***.........
          ***.........
          ***.........
          ***.........
          ********....
          **********..
          *****.****..
          ****...*****
          ****...*****
          ***.....****
          ***.....****
          ***.....****
  3       ****...****.      2
 --->     ****...***..    <---
          **********..
          **.*****....

               (i)
```

```
          .......****.
  4       .......****.      1
 --->     ........***.    <---
          ........***.
          ........***.
          ........***.
          ........***.
          ....*******.
          ..*********.
          .**********.
          ****...****.
          ****....***.
          ****....***.
          ****....***.
          .***....***.
  3       .***....***.      2
 --->     .****..*****    <---
          ..**********
          ...*********

               (ii)
```

```
          ***...****.
  4       ***...****.      1
 --->     ***....***.    <---
          ***....***.
          ***....***.
          ***....***.
          ***.....**.
          ***....***.
  3       ****...***.      2
 --->     .**********    <---
          .**********
          ...***..***

              (iii)
```

```
          .**.....***
  4       ****....***      1
 --->     ****...****    <---
          ****...****
          ****...**..
          .***...**..
          .****.**...
          ..******...
  3       ..******...      2
 --->     ...****....    <---
          ...****....
          ....**.....

              (iv)
```

Figure 6. Examples of matrix parametrization and classification.

Display and manipulation of the molecular graph and its characters, once they have been recognized, are done through

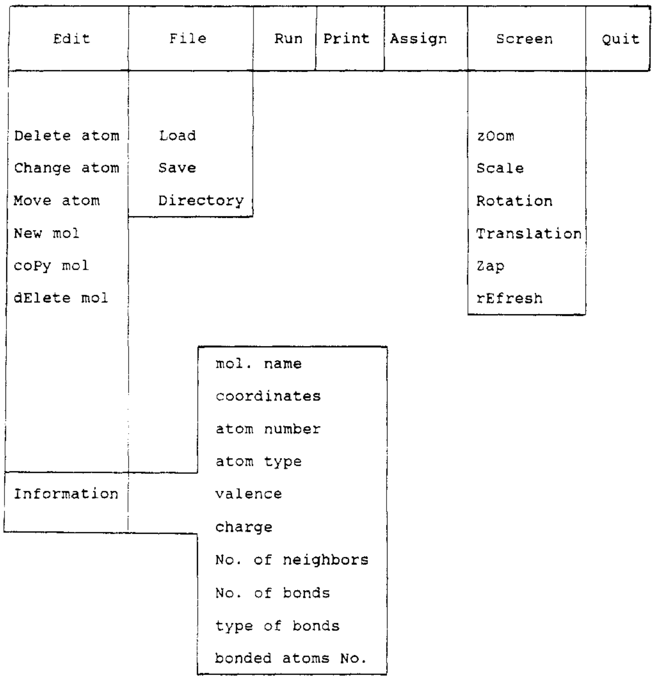

Figure 7. Content of some menus of the graphic interface.

a graphic interface provided by the system.[9] The interface has been implemented in a modular way and can easily be used with a mouse. Actually, it has 28 menus, including different options and many archives. The option active at a particular moment is highlighted. Figure 7 presents a brief scheme showing the contents of some of the menus. The principal menu has the following options: Edit, File, Run, Print, Assign, Screen, and Quit. *Edit* allows one to modify, create, and manipulate molecules. The *File* option allows one to load a particular file, save a file that has been modified, or scan a directory of files to choose the one to be loaded. Files created by the recognition process can be chosen. This facility represents an option to change a part of the scanned molecule or eventually to correct computer errors in character recognition. *Run* and *Assign* are used in the area of expert systems. The *Screen* option allows one to select one molecule from a set, scale it up or down, translate or rotate it within a defined area, erase at once all the molecules from that set, or refresh the screen.

The menu that manipulates molecules has the options: coPy mol, Save mol, mOve mol, Center mol, dElete mol, and Information. The *mOve mol* option has a displacement menu that allows one to choose the desired displacement of the structure on the screen. The *Information* option displays for any of the atoms selected with the mouse the following information: name of the molecule which it belongs to, coordinates, number it has at the connectivity table, type or identity, valence, charge, number of neighbor atoms different from hydrogen, number and type of bonds in which it participates, and number of the attached atoms.

The molecular structure already recognized may be stored as such or sent to a database,[2] to a CAD system,[9] or to an expert system[10] which could profit from the automatic input of this type of graphic information.

DISCUSSION AND CONCLUSIONS

Capture. The capture of the molecular structures to be recognized is actually done on a PC-AT and the information processed on a microVax/II with 8 Mb of real memory and a 159-Mb hard disc, running under Ultrix-32 (version 2.0). Transmission of the digitized structures is done through a Kermit protocol or through an Ethernet network with a PC-NFS board and its software. However the whole process can be done on the PC microcomputer. For continuous route work, software for manipulating the images has to be developed to avoid the highly interactive process provided by the scanner software package. In this way the image capture will be accelerated.

The image has now been converted to a matrix format. The required storage memory is dependent on the scanning resolution; e.g., for an average size of a structure—3 × 5 in.—and for a resolution of 75 × 75 dpi, the memory required is 12 kb; with a resolution of 300 × 300 dpi, the memory required is 175 kb. The processing time (capture plus recognition) is dependent on the amount of pixels to be considered and is 0.7 and 1.0 min, respectively. The necessary memory occupied by a recognized molecular structure depends only on the number of atoms and bonds. For instance, the *N*-butyl,*N*-pentyl-2-aminotetraline molecule, $C_{19}H_{31}N$, needs 4.1 kb of storage. If the graph of the same molecule is stored (skeleton only, without hydrogen atoms), only 1.6 kb are needed independent of the original size of the scanned image.

Graph Recognition. The implemented graph algorithms for any molecular graph are based on vertex determination and on different sweeping procedures. Vertices are found during perception, when a deflection angle of the linear trajectory higher than 18° is detected. This empirical value is suitable for hand- and type-printed graphs with bond lengths corresponding to about 130–150 pixels. Both the deflection angle and the number of pixels associated to bonds are parameters which can externally be defined according to the work.

Trajectory is determined by the pixels neighborhood connectivity method of image processing, where the direction of displacement is described by a number from 0 to 7. This is in function of the relative position of any of the eight neighbors of a central point. The number chosen for the neighbor that is bonded to the central point describes a line coincident with the direction of the displacement. External and internal borders of a graph arc permit detection of the type of bond associated with a molecular structure (wedge, dot, single, and multiple) as a function of the thickness of the bond and the number of lines joining the atoms.

Branching detection is done according to a specifically created algorithm based on a circular sweeping principle. Each time a subgraph is recognized, the matrix that contains the coordinates of all the points of the subgraph is changed to a window. That window keeps just the first and the last points of a rectangle containing the matrix information. This is done to optimize memory and to avoid sweeping of this part of the image again. These basic concepts are used for the perception of molecular skeletons of any complexity as was shown in Figure 2a. After all of the regions of the image are swept and recognized, the system integrates the several subgraphs into the corresponding structure.

Once the topological characterization of the graph is finished, each attribute is represented as a typical Prolog data structure (see Table I). From this representation a connectivity table is constructed. These processes constitute the final recognition of the graph as a unique molecular structure. The procedure is equally valid for type- and hand-printed molecular structures. In the last case, however, the recognition process needs that drawn structures be straight enough in order to keep deflection angles similar to those of the typed ones. A reliability of 94% was found either with type- (200 structures) or hand-printed (50 structures) molecular graphs.

Character Recognition. For a general automatic molecular structure recognition, this is a necessary step. As in the previous module, first a topological feature extraction is done. The implemented method, is based on a known principle of pattern-recognition.[8] The implementation is facilitated by the

Table I. Data Generated by Automatic Recognition of the Molecule in Figure 1a

attribute	name	no.	X-coord.	Y-coord.	charge	type	natch[a]
atom	mol58	1	272	227	neutral	C	3
		2	269	302	neutral	C	2
		3	354	305	neutral	Sn	4
		4	356	226	positive	N	4
		5	428	177	neutral	C	3
		6	482	261	neutral	C	4
		7	423	337	neutral	C	4
		8	328	369	neutral	H	1
		9	202	182	neutral	O	1
		10	434	104	neutral	C	1
		11	534	213	neutral	C	1
		12	534	303	neutral	C	2
		13	596	302	neutral	N	1
		14	500	366	neutral	Cl	1
		15	395	401	neutral	H	1
		16	328	155	neutral	H	1

		no.	atom-1	atom-2	type
bond	mol58	1	1	2	single
		2	2	3	single
		3	3	4	single
		4	1	4	single
		5	4	5	single
		6	5	6	single
		7	6	7	single
		8	3	7	single
		9	3	8	single
		10	1	9	double
		11	5	10	wedge
		12	6	11	single
		13	6	12	single
		14	12	13	triple
		15	7	14	dot
		16	7	15	wedge
		17	16	4	single

[a] Number of drawn attached atoms.

reduced number of characters. Although characters of different size (sub- and superscript versus normal ones) work against the generality of the module, the implemented algorithms make this character-recognition method size independent.

Human recognition of characters makes use of syntactic and semantic analysis of whole words. The system however works with isolated characters, and due to degradation some of them are quite similar to each other, such as 's' and '8' or '2' and 'z'. In spite of these limitations the system recognizes isolated characters with over 99% reliability.

With hand-printed characters, especially if they are small and written in cursive style, only recognition that is human dependent and therefore limited can be done.

Recognition of any molecular structure creates a Prolog data structure and a connectivity table that describes all of its atoms and bonds. Automatic generation of this information can be used by any program or system that works with molecular structure representation. In this way it can profit from the automatic input provided. In our case we used this facility to feed a personal database[2] that works with molecular structures. Also we used the system to feed a retrosynthetic program,[5] a direct synthetic program,[10] and other modules for CAMD.[9] These proved applications can easily be extended because the important contribution done is avoiding the manual molecular drawing.

Manual input of the structures involves human perception and assignment of every atom and bond. So any computational system just receives the information and constructs the corresponding connectivity table (recognition process). The system here developed instead is in charge of doing the whole perception, as a topological characterization process, and also the recognition in the standard way.

Display. In the graphic interface, created at our laboratory, the molecular attributes are stored and treated as typical Prolog data structures. This interface utilizes the ReGIS graphic routines provided by the Digital VT 340 graphic terminal under Ultrix. Also it can work with the graphic facilities of VMS Workstation Software and with GKS. Its dynamic data structure allows one to work with several molecules with different number of atoms and bonds limited only by the available memory. In addition to the display and manipulation of the recognized structures the interface allows one for the printing of them.

Summing up, the system described here, entirely developed in C (Vax C/Ultrix-32, version 1.0), can perceive and recognize complex type-printed molecular structures and allows one for the automatic input of this graphic information to a computer system. For hand-printed molecular structures however, there are some limitations for character recognition but not for graph (skeleton) recognition.

The system was proved with entry of structures to a database, to molecular design, and to organic synthesis programs and can be extended to other applications such as reaction representations, intelligent desktop publishing (rotation, scaling), etc. The system is actually about 3–5 times faster than a qualified person when making the input of a molecule of average size (20 atoms). Furthermore it is free of common human mistakes when introducing type-printed molecular structures.[11]

Beyond the immediate practical application of the system, it also provides a means to understand some basic cognitive processes utilized by chemists to learn and represent chemical concepts. For instance, coordinate generation of a condensed representation into a structure, or assignment of a general R substituent used in structures, requires a detailed description before they become implemented into a computer. These problems represent further developments we are working on.

ACKNOWLEDGMENT

We thank the Andes Foundation (Grant C10390/88), Fondecyt (Grant 774/89), and the University of Santiago for financial support.

REFERENCES AND NOTES

(1) Milne, G. W. A.; Feldman, A.; Miller, J. A.; Daly, G. P.; Hammel, M. J. The NCI Drug Information System. 2. DIS Pre-Registry. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 159–167.

(2) Contreras, M. L.; Deliz, M.; Rozas, R. Personal Microcomputer Based System of Chemical Information with Topological Structure Data Elaboration. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 163–167.

(3) Fruhbeis, H.; Klein, R.; Wallmeis, H. Computer Assisted Molecular Design. *Angew. Chem. Int. Ed. Engl.* **1987**, *26*, 403–418.

(4) Jurs, P. C.; Stouch, T. R.; Czerwinski, M.; Narvaez, J. N. Computer-Assisted Studies of Molecular Structure–Biological Activity Relationships. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 296–308.

(5) Chanon, M.; Barone, R.; Contreras, M. L. Microcomputer and Organic Synthesis Application of an Interactive Program to the Photochemical Synthesis of Pheromones. *Nouv. J. Chim.* **1984**, *8*, 311–315.

(6) Wilkins, C. L.; Randić, M. A Graph Theoretical Approach to Structure–Property and Structure–Activity Correlations. *Theor. Chim. Acta* **1980**, *58*, 45–68.

(7) Anderson, I. M.; Bezdek, J. C. Curvature and Tangential Deflection of Discretes Arcs. *IEEE Pattern Anal. Machine Intelligence* **1984**, *PAMI-6*, 27–40.

(8) Selfridge, O. C.; Neisser, U. Pattern Recognition by Machine. In *Computers and Thought*; Feigenbaum, E. A., Feldman, J., Eds.; McGraw Hill: New York, 1963; pp 237–250.

(9) Contreras, M. L.; Cabezas, A.; Pinto, A.; Rozas, R. Molecular Modeling. Structure-Property Relationships via Topological Indices. *Proceedings of the XVIII International Latino-american Congress of Theoretical Chemists*; La Plata, Argentina, Sept 1989; p 44.

(10) Alvarez, C.; Deliz, M.; Rannou, F.; Rozas, R. Computer Prediction of Free Radicals Products in Organic Synthesis. *Proceedings of the XVIII Latino-american Congress of Chemistry*; Santiago, Chile, Jan 1988; pp 785–786.

(11) The graph recognition program is available on request from MLC.