if they wish to do so. Changes in hardware and software—especially hardware—hold out the prospect for faster machines with remarkable storage capacity. We have already seen the introduction of microcomputers that have gone from 8- to 16- and even 32-bit words, thus blurring the distinctions between and among micros, minis, and mainframes. Among other things, this means that we can already have personal computers for home, school, and business that represent most of what we need for stand-alone jobs, and which can also be used as intelligent terminals for interfacing with other computers through existing networks. Many of the inconveniences associated with complicated protocols and sign-on procedures can be eliminated by using these micros intelligently. The gradual replacement of copper wire with fiber optics will allow us to transmit vastly greater quantities of data while also providing greater security for the transmitted messages. And laser disks hold out the promise of extremely dense storage for both pictorial and textual material, perhaps permitting us to download entire sections of on-line encyclopedias or other information sources.

In addition, there is hope emanating from the field of artificial intelligence, which has already introduced so-called "expert systems" into a number of areas—including medicine, where routine preliminary diagnoses can now be initiated by computers. And there is also progress in voice recognition, which would release us from the constraints imposed by keyboards—still the most efficient and cost-effective method for communicating on-line with computer-based systems.

These and other exciting possibilities are being explored by a variety of people in a number of settings, sometimes in industry, but often in our universities, where those teaching information science at the graduate level are also engaging in research.

REFERENCES AND NOTES

(1) Shera, J. H. "The Foundations of Education for Librarianship"; Wiley: New York, 1972.
(2) Skolnik, H. "Chemical Information Science in Academe". *J. Chem. Inf. Comput. Sci.* **1980**, *20* (2), 2A.
(3) Wiggins, G. "The Indiana University Chemical Information Specialist Program: Training the Library User and the Librarian". *Sci. Technol. Libr.* **1981**, *1* (3), 5-11.
(4) Malinowski, H. R.; Richardson, J. M. "Science and Engineering Literature: A Guide to Reference Sources", 3rd ed.; Libraries Unlimited: Littleton, CO, 1980.
(5) Price, D. J. de S. "Little Science, Big Science"; Columbia University Press: New York, 1963; pp 90-91.
(6) Davis, C. H.; Rush, J. E. "Information Retrieval and Documentation in Chemistry"; Greenwood Press: Westport, CT, 1974; pp 9-10.
(7) Davis, C. H.; Rush, J. E. "Guide to Information Science". Greenwood Press: Westport, CT, 1979; pp 7-14.
(8) Nair, J. H. "Chemical Safety Literature in Data Banks". *Chem. Eng. News* **1985**, *63* (5), 4.

DARC System for Documentation and Artificial Intelligence in Chemistry

JACQUES-EMILE DUBOIS* and YVES SOBEL

Association pour la Recherche et le Développement en Information Chimique, 75005 Paris, France, and Institut de Topologie et de Dynamique des Systèmes, associé au CNRS, Université Paris 7, 75005 Paris, France

Received June 4, 1985

The DARC System deals with structural information both for documentation and for artificial intelligence (AI) endeavors in chemistry. Its topological concepts are briefly reviewed in conjunction with the creative data needs in knowledge information processing systems (KIPS). Knowledge base, inference engine, and user interface are discussed with reference to the DARC potential in the field of AI and expert systems. AI methodology and its impact on knowledge production are reviewed. New chemical computer-aided design (CAD) tools to develop more creative and innovative research in synthesis planning, structure elucidation, and prediction in drug design are no longer pure prospective challenges.

INTRODUCTION

Chemistry is an excellent testing ground for serious development in the years ahead, both in cognitive science and in artificial intelligence. It has a long association with graph and combinatorial theories. Chemical facts, well structured and diverse, can be managed easily by a class of structural languages adapted to various needs ranging from bibliographic data to the most advanced correlations and to modelization connecting morphological facts to properties, reactivities, and synthesis. Moreover, it is, by tradition, very user friendly, and its ideograph presentation is naturally adapted to graphics display.

This paper was conceived as a follow-up of "French National Policy for Chemical Information and the DARC System as a Potential Tool of This Policy". Born about 20 years ago, the DARC System was presented in the *Journal of Chemical Documentation* in 1973 for its documentation potential, but its second facet, Automatic Research of Correlations (ARC), was, at that time, left in the dark!! Indeed, concepts and tools of artificial intelligence were still in their infancy. Artificial intelligence was then less than 17 years old and struggling for recognition. The concept of "decision support systems" that we have used in chemistry has also been developed only very recently.

DARC Concepts. The *Journal of Chemical Information and Computer Science* (JCICS) Silver Anniversary Issue is an excellent occasion on which to sum up a field whose recent progress has been significant, with the development of much basic formalization. DARC concepts and tools for artificial intelligence are based on more structured knowledge than is used by either yesterday's or today's expert systems.

The same concepts and tools reflect the choice of an associative approach, one that relates factual documentation to achievements of artificial intelligence. In many fields, the passage from one to the other is still difficult, and communication between documentation data and inference-creating information usually remains a problem. In chemistry, factual data banks (FDB) must be linked to knowledge banks, since

0095-2338/85/1625-0326$01.50/0 © 1985 American Chemical Society

Jacques-Emile Dubois is Professor of Physical Organic Chemistry and Chemical Informatics at the University of Paris VII and Director of the Institute of Topology and Dynamics of Systems. In 1947 he received his Ph.D. from the University of Grenoble in infrared studies of hydrogen bonding and regioselectivity of aldolization. Besides his work on various kinetics, such as fast proton transfer in prototropy, control of aldolization stereoselectivity, reactivity of structural constraints, and the physical chemistry of congested structures and thin films, he developed, as of 1957, a strong interest in chemical information. As early as 1963, this led to his pioneer work on the DARC Topological System and the subsequent development of EURECAS, the DARC on-line service utilizing the CAS file. He chaired the IUPAC Interdivisional Committee on Machine Documentation in the Chemical Field (1969–1977) and is presently Vice-President of CODATA. He founded several chemical information agencies, such as the Centre National de l'Information Chimique and the Association pour la Recherche et le Développement en Information Chimique.

Yves Sobel, born in Budapest, Hungary, in 1947, graduated from the Ecole Nationale Supérieure des Industries Chimiques in Nancy, France, with a chemical engineering degree in 1972 and obtained a Master's degree in physical organic chemistry at Paris VII University in 1973. Under the guidance of Professor J. E. Dubois, he prepared a doctoral work on "Populations of Structures and Hyperstructures" within the framework of the DARC System at the Institute of Topology and Dynamics of Systems. Its inclusion in the DARC/GEANT has played an important role as a structural space generation tool useful for different DARC applications. His main research interest is in the topology and computerization of complex chemical data, which he teaches in the "Computer Science in Chemistry" and "Organic Spectroscopy" courses at Paris VII University.

---

much unexplicit and uncorrelated knowledge lies dormant in the former. The structural side of these FDB seems most favorable to communication between knowledge and data. It is generally admitted that the chemical corpus as expressed by formulas of entities—expressed for us by their chromatic graphs—could be used more efficiently and be more fully consulted and exploited with some structural functions dealing with local and global ordering of specific and generic substructures (SS) and structures (S). These elements, SS and S, are topologically localized in population graphs, which, through their ordering as hyperstructures of structures (HS), are instrumental in providing neighboring distances as well as indications concerning some underlying properties shown in these special HS spaces. Few sciences provide the possibility that chemistry does, that of handling analogies by computer consultation of ordered chromatic tables. In few other sciences can one back up elucidations and correlations resulting from consultation of various databases with the aid of syntheses of as yet nonexistent key compounds associated with the essential nodes for prediction.

The originality of chemical informatics in France, from its inception within the framework of the DARC System, was to constitute various factual databases with a double objective. The first, quite naturally, was to provide the user with data stored in a more or less interactive mode. The second was to dispose of a corpus whose specific organization made it possible to generate *rule bases*, so useful in present expert systems. The methodology used in managing this kind of problem is based on a thorough formalization of factual bases and on the creation of rule generators. Part of this methodology is found in the "expert systems" of the second generation of "metaexpert systems", developed in order to get around the limitations of earlier systems. Present realizations are still quite modest and can only be credited with some degree of artificial intelligence. They are mainly based, in fact, on sequences of limited statements made by experts.

Knowledge Information Processing Systems. We feel it is a pity to have some potentially very important information in some databases yet to be unable to exploit it fully because of the limitation of scientific laws and of systems of pragmatic expert advice being used. Our views on the conception of knowledge information processing systems (KIPS) of the future is that these systems, at different levels of creation and consultation, should dispose of intelligent procedures to mobilize the strategic sets of existing data needed to answer given queries. In this respect, KIPS should also have the built-in capacity to evaluate their own potential as well as their weakness with regard to a given target problem.

To satisfy all these requirements, we are currently developing adequate mechanisms to (1) *W*ake *U*p *S*tored *D*ata in order to produce *creative data* and linked metarules (WUSD software), (2) *S*pot *M*issing *K*ey *D*ata (SMKD software), and (3) instigate *E*xternal *S*earch for production of certain *K*ey *D*ata (ESKD software). This is part of our original attempt to create a general *Knowledge Autoproduction Generator* (KAG), optimized for various structural and associated data sets.

Operational Realizations. There are, at present, three types of DARC realizations oriented to artificial intelligence (AI) applications: (1) several operational factual data banks; (2) some second generation expert systems (elucidation, combinatorics, correlation, and synthesis); (3) a set of informatic tools for AI handling of the information stored in these factual databases. These tools (files and inference engine) have been improved for greater efficiency in dealing with the complexity of real situations, difficult enough to lie beyond common sense dealings.

Prior to this updated presentation of our AI activities, it is worth while recalling that our EURECAS [Chemical Abstracts Service (CAS) file] Management System is based on an AI tool called DARC/TSS (Topological Screen System). This generates sets of hierarchical generic screens starting from a specific database of defined structural entities. The inference engine uses TSS rules according to the DARC/TDB (Topo-

Table I. Theoretical Graph Structures and Chemical Modeling

chemical object	graphlike representation tool	features
compound	chromatic graph	precise description at different structural levels
reaction network	oriented chromatic hypergraph	general tool for *n*-ary symmetrical relationships, and/or graphs, articulation networks
reaction invariant	loose chromatic graph	quantitative evaluation of variation sites
reaction variations	chromatic opegraph	minimizing redundancy with invariant
reaction transformation	chromatic transfograph	compact and suggestive dynamic representation
population	chromatic pterograph	complete analytical representation

logical Data Base) organization. It also uses a very efficient heuristic strategy to avoid any silence in the query–answer cycle of the chemist's dialogue via a special graphic input–output subsystem. This last tool, called GRAPHEDIT, is an expert system handling topological and stereo structure representations. These "expert systems" were used to establish the first on-line services based on topological querying of large bibliographic files (CBAC, 1978–1979; EURECAS, 1979–1980). Although TSS, TDB, and GRAPHEDIT are very important in factual data banks and bibliographic components cannot be ignored, the specific computer handling of physical or other associated data and their estimation or prediction present essential challenges. Factual expert systems are very complex, for they include many other tools besides the modern TSS and TDB.

The design and realization of some AI expert systems or KIPS within the framework of DARC concepts are presented here through the usual components of expert systems: first the knowledge database, then the functions of inference engines, and, finally, interfaces with the users.

KNOWLEDGE DATABASE

Concept of the Chromatic Graph and Its Use. In reductionist theories, *analogy and causality* correlations are linked to studied morphology fragments. Most work is done with tables containing sufficient data for a database management system (DBMS) using relational or other management systems. These procedures in chemistry tend toward a priori fragmentation of structures in modules of groups or of functions.

The original approach of the DARC System consists in a global description with organized local space, a generative grammar, and an approach to fragmentation only a posteriori and within some coherent generation processes. Structure representation is, in fact, a compromise between a very fine description and excessive mutilation. The guiding principle in DARC coding is to keep the description open for eventual additions and, moreover, to maintain constant ties with the neighboring sites information. To this end, a homogeneous formalism was devised, thanks to which solutions can be transposed from one task to another as communication can take place between specific and generic procedures. Coherence is assured by maintenance of a central concept, that of the chromatic graph. This underlies all our work on the structural formula and the spatial structure of compounds, global and local handling of reactions, interconnection between compounds by reactions in synthesis trees or by elementary processes in reaction mechanisms, and analogy relations between compounds or reactions. Any fundamental description is progressive; local and global visions of entities are handled by proceeding gradually from a generic perception at the starting point to a desired specific clarification of the final and global vision.

Either by acting on the canonical code grammar of the chromatic graph or by broadening the concept of graph description, all problems of structural designation can be handled by a single generative logic. Partial indeterminations can be managed either as proper Markush generic structures or as specific populations.

Transformation problems, on the other hand, are a challenge on the structural level, especially if one wishes to go further than simple description of initial and final products toward true management of a transformation with appropriate operators. The chromatic graph concept alone is not powerful enough to handle transformations. In our recent work, we propose new data structures by an *extension of the chromatic graph concept*. Where a situation goes beyond the framework of this concept in the handling of isolated structures, it has generally been brought back into this framework by Conventional Structure Adaptation (CSA). This is not always the case in handling populations of structures, of reactions, and more generally of structured sets of structures, which we call hyperstructures (HS). However, the new abstract kinds of data that we have introduced in graph theory (Table I) are part of the same family as chromatic graphs. In general, a representative chromatic graph can be associated with them.

Nonstructural Chemical Information Associated with Structures. Information associated with chemical structures is often simpler than that of structural data. However, a spectrum contains no isolated scalar information but a *set of such information* linked together. Of course, this type of data is found in spectrometry measurements but also in other fields where several different measurements participate in constituting a more complex information: chromatographic retention times on several columns of different characteristics, biological activity measured in various doses and in different conditions, and sets of experimental conditions in which a reaction is carried out.

Dealing with this "complexity of information associated" with chemical structures is still a very young technique and is not yet as elaborately formalized as the structural formula of organic compounds. We have also approached this problem by defining appropriate new Abstract Data Types.

"Topology–Information" Relationships. We must note that structural precision is exploited easily and fully only when an element of associated information is made to *correspond directly* with a specific structural unit: assigning peaks to atoms in NMR, UV absorption bands to chromophores, and biochemical or reactional activities to active sites. The rules for behavior prediction or for structural elucidation can thus take the shape of *generic data banks* of structure–information. These procedures are in fact more involved when one takes into account the real complexity due to environment influences on specific structural units. Large deviations can then be accounted for by some metarules created by Knowledge Autoproduction Generators (KAG).

Factual Data Banks (FDB): SS Search and HS Spaces. The operational achievement of DARC tools of Description–Acquisition–Restitution presented in 1973 as a "potential tool of French national policy for chemical information" enabled us to set up the querying of the CAS bibliographic file (more than six million entities) within EURECAS and of the Institute for Scientific Information (ISI) file (some three million entities). At the ARDIC (Association pour la Recherche et le Développement en Information Chimique), we have also constituted reliable factual data banks in the fields of spectroscopy, crystallography, biology, and reactivity. DARC-Regular (DARC-R) as a documentation tool using mainly

Generic (G) and Specific (S) substructures at the DBMS level was extended to take care of Markush patent problems and to combine generic and combinatorics in DARC-G. These tools were transferred within a homogeneous framework to a whole variety of chemical objects on the level of their properties and of their transformations, specific or generic. These elaborate factual data banks, associated with KAG items, gave rise to knowledge bases (factual bases and rule bases) with behavior prediction, simulation of synthesis pathways, and structural elucidation.

The strength and the all-encompassing nature of the DARC System as illustrated by these realizations stem from its principle of structure generation set forth as one of its bases in 1966: "We propose to define a structure by following its construction by logical, unequivocal steps; the intermediate entities generated during this process all belong to one and the same family." The organization part of our systems is always based on fluid and controlled interrelations between SS, S, and HS graph representations.

This link between the target object and the family of generated intermediates was stated more precisely in the form of "Structure–Hyperstructure" (S–HS) and "Structure–Hyperstructure–Information" (S–HS–I) *synchronous generation laws*.

From a *static viewpoint*, these laws lead to linear, structural descriptions, unique and unambiguous: major or canonical DARC code of defined or fuzzy structures and reactions; finalized code made up of sequences of operator adjunctions or of topological vectors capable of comparing structures of the same family by localizing them in a hyperstructure space. Different descriptive forms (matrices, tables, lists, bi- or tridimensional drawings) are derived for specific purposes.

From a *dynamic viewpoint*, these generation laws are the heart of the DARC System inference engines.

INFERENCE ENGINES

Search space, constraint handling, rule translation, rule base access, and traversing method are the essential aspects of most inference engines. DARC inference engines strive to work beyond the "common sense" level, with metarules. They also have the tools for creating ordered structural spaces (HS) as well as localization power for situating data within these spaces. In liaison with a knowledge autoproduction generator (KAG), these inference engines adapt hyperstructuration modes and hyperstructure traversing to performing important tasks of design in structural chemistry: structure–information correlation (behavior prediction, spectra simulation, therapeutic activity optimization, ...), elucidation of structure or reaction mechanism elucidation, search for synthesis pathways, ... The search for structural families within large databases cannot be deemed a purely documentary task. It calls on true expert systems, endowed with KAG tools for achieving acceptable efficiency.

Graphs, Space, and Localization. The structural subsystem provides the HS space wherein inference engine functions are often best organized. A structural space of states is created, like the procedures used in puzzle game theories, where the conversion graph of the game and the puzzle's different states are considered as nodes or states. This has the advantage of proposing a large number of pathways leading from the origin focus to the target structural entity. A good algorithm can supply many more alternatives than the traditional filiations that chemists envisage when handling chemical problems. Compounds thus compared are much more numerous than the usual chemical ones, since they may differ by their local chromatism (nature of sites) and their generic aspect; they also differ by their internal ordering in the various HS conversion graphs. According to how we look at topological representation, an infinity of spaces can regroup totally generic items, fully described entities, or entities that are partly generic and specific at their representation sites.

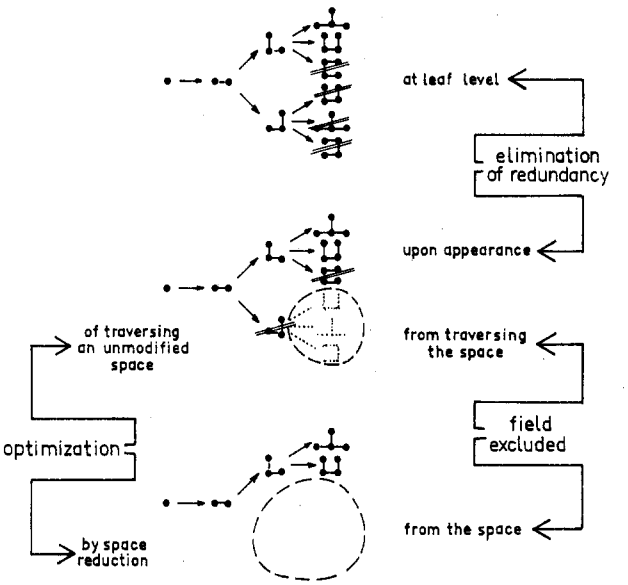

Figure 1. Generation of all isomorphically distinct trees with four vertices. Levels of heuristic optimization.

The ordering functions used to generate and handle structures impose certain types of ordering, not only on structural sites but also within the filiations or pathways. These aspects control important developments in the similarity correlations where the topological neighbors depend on those ordering functions. The originality of the DARC generation resides in the production of a space of states in a synchronous manner that monitors a progressive elaboration of the target entity to be described. The choice of a set of canonical rules generates a corresponding arbitrary hyperstructure, useful for certain problems of pattern recognition and other AI procedures. This is, in fact, a reference problem based on customary methods in this field.

As the creation of ordered search spaces is a crucial item of most DARC inference engines, we have developed a special subexpert system called DARC/GEANT (GEneration by ANTeriology), which has a broad potential for generating different spaces for a given target and can thus lead to different organizations defined by precise constraints.

Hyperstructure Building with DARC/GEANT. When a chromatic graph is generated through an organized construction, the resulting desired graph is endowed with an imposed order. This type of *heuristic* generation leads to a unique final graph as shown in Figure 1, where several types of generation of four-vertex graphs are compared. One of these avoids redundant creations and, later, tree pruning.

The fundamental aspect of synchronous building of HS and S by the DARC/GEANT system stems from the same necessity for heuristic production. All the graphs created during the generation of the chromatic target graph constitute a defined family whose members are well placed in relation to each other (see Figure 2). The whole of their respective relationships form a network in which the family members are the nodes. In the DARC System there is a *close synchronous* relationship between the progressive construction of the structure (S) and that of the associated hyperstructure (HS). In fact, while the structure is being generated, the generation is expressed by the ordered adjunction of a new site into a structure (S). One thus obtains a new structure called S′. In this way, for each ordered adjunction of a site into a structure, there is an ordered adjunction of a site (edge S–S′ and node S′) into a hyperstructure (HS).

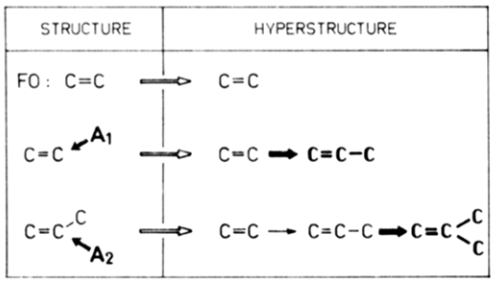

Figure 2. Synchronism between structure and hyperstructure generations. When generating the target compound—isobutene—from the source compound—ethylene—an intermediate compound is created formally during the introduction of each new site. The set of these intermediate compounds constitutes a generation series called linear hyperstructure.

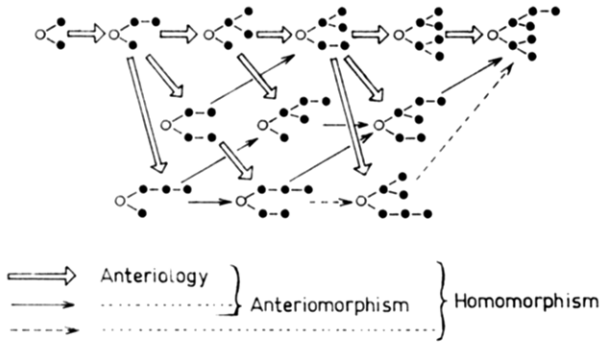

Figure 3. Example of formal hyperstructure HS(P,R). P is the population of anteriors of a target structure. Here all the anteriors of a target structure are generated and ordered according to three formal anteriority relationships R.

Insofar as the *adjunction and ablation* operations are the most fundamental ones, and as every structural transformation (whether formal or reactional) boils down to carrying out these generations, one may, through synchronous generation, foresee some very diversified situations.

A target compound and the hyperstructure of the focus-target space are linked together here by anteriority relationships (see Figure 3). Three relationships that we consider essential express a more powerful order. The homomorphism of two compounds lies on an adjunction-measured (ad) distance. In the anteriomorphism, the new ordered graph stems from the old one through ordered adjunctions. In anteriology, the order is that of the ordered generation sequence law. The HS's thus derived fit into each other. The anteriority notion accounts for the classical homology and allows for governing other "homologies".

The traversing step by step backward or forward enables us to enumerate either the precursors of a target or its isomers. With our DARC/GEANT, we have for instance enumerated 1011 bicyclic *isomers* of quinuclidine but also 14 791 *anteriomorphs* of cholestanol. These examples stress the importance and power of the anteriority and HS concepts for all enumeration problems.

DARC/GEANT takes care of varied relations (R) that can be used to organize a population (P) in hyperstructures HS-(P,R). This methodology makes it possible to explicitly construct not only the R relations but also the P populations of a formal hyperstructure HS(P,R). The explicit generating of such hyperstructure thus solves by subproduct *the enumeration problem* and subsequently that of counting all those structures *satisfying the given formal constraints* marking the boundaries of P.

The DARC/GEANT program is able to generate the structures required in a population. Such achievements can be attained by strategic traversing of a controlled HS under strict grammatical ordering rules (Table II). The present

Table II. GEANT System: "Adapt" Option Characteristics

goal	trace: entire hyperstructure network and population covered
supplementary constraints	cycles, unsaturations, structure valency, space depth
rules	trace: elementary adjunction type production rules; linearly ordered rule base
space	totally ordered rooted tree: anteriology HC
constraint handling	heuristic pruning using hierarchical constraint implementation
rule translation	direct interpretation
rule base access	heuristic GEneration by ANTeriology sequence rule, using intelligent ordering of the rule base and leading to exact solution
traversing	breadth-first search adaptation
input	parametric applicability conditions included in rule base

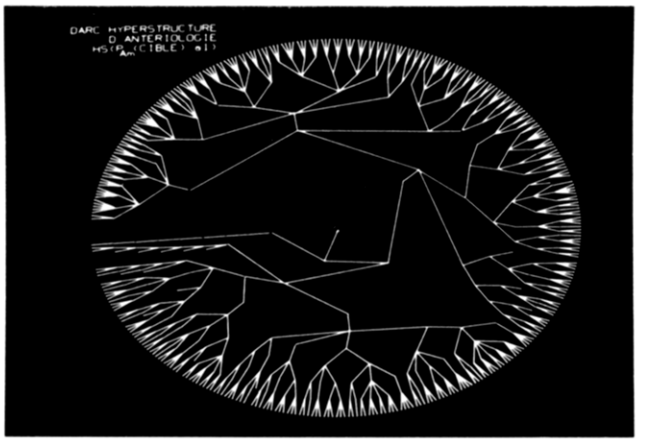

Figure 4. Hyperstructure ordering of all hydrocarbons C_nH_{2n+2} from C_1 to C_{12}.

implementation of DARC/GEANT by interactive communication accepts the following: the general constraints of chemistry, where the nature of the atoms limits the number and nature of bonds; independent constraints as to the number and nature of nodes, of edges, of cycles, and of certain substructures (these include isomeric constraints); anteriomorphism constraints. In other words, the targets are extremely diversified and may be a compound, a substructure, or a population of isomers of homologues, all of them defined with precise constraints. The HS obtained are visualized, and interactive actions can be conducted with them. Their global presentation is sometimes difficult, and DARC/GEANT can present trees in different ways. As an example, the anteriority hyperstructure of aliphatic alkanes from 1 to 12 carbon atoms regroups 644 structures (see Figure 4); 355 of them contain 12 carbon atoms and are displayed at the periphery of the graph presentation. Such an HS has proved extremely valuable in identifying key atoms in the carbon-13 NMR predictions.

Inference Engines and Choice of Search Spaces. Even in a globalizing view of structural properties and of their representation by topology and the DARC System, it is unreasonable to unify all study procedures. The nature of properties studied affects the role and functioning of the inference engine, whether in a strict expert system or in an interactive KIPS system. Two main types of properties can be considered: those where formal or real structure transformations involve atom transfers (mass, synthesis, combinatorics, ...) and those where basic or heavy graphs are unaltered. A few examples will show the influence of the extreme diversity of chemical problems accessible through Computer Aided Systems. We shall also see how our "decision support system" tools have helped us build some KIPS interactive subsystems, bearing in mind that the diversity of chemical problems implies certain intelligent

Table III. Examples of Systems for Handling Chemical Transformations[a]

transformation type	problem	T-DARC-based system	other systems
chemical reactions	computer-aided synthesis	SYNOPSYS	Corey–Wipke, Hendrickson, Ugi
reaction mechanisms	simulation of reaction mechanisms	MECOPSYS	Jorgensen
ionization, fragmentation and rearrangements in mass spectrometry	elucidation of mass spectra	MASOPSYS	Dendral
formal stages	generation of isomers	GISOPSYS	Dendral, DARC/GEANT

[a] *SYN*thesis; *MEC*hanisms; *MAS*s spectroscopy; *G*eneration of *IS*omers; *OP*timization *SYS*tem.

adaptations. While maintaining the chromatic graph and the synchronous generation of a search space as the heart of our systems, we must not forget the necessity of linking structural topology (T) and other property information (I). Whether the changes are static or go through structural transformations, this strong association with characteristic information ⟨T/I⟩ must be considered when conceiving the representation and when organizing the system and the nature of the inference engines. Several examples will help clarify these ideas.

In all problems where one refers to fairly precise structure handling and where knowledge, topology, and associated topography are essential, DARC/GEANT allows one to deal with S, SS, and, above all, HS spaces with little or no tree-pruning problems. For example, structure–activity relations and those of statistical calculation of data call for DARC/GEANT.

In graph transformations, problems concerning reactions, for example, are handled by a dynamic and semantic approach to the reaction, called IGLOO (*Invariant Graph and Localized Ordered Operators*). A special general computer assistance program called PARIS (*Pertinent Analysis of Reactions by IGLOO Simulation*) activates the structural operators of IGLOO. The complexity in the field is such that one has to prune the answers to avoid redundant information from emerging at different levels of the solution generation. Thus, PARIS acts as the inference engine for many applications.

Last but not least, DARC/EPIOS (*Elucidation by Progressive Intersection of Ordered Substructures*) is an information processing system developed for predicting structures from carbon-13 NMR spectral information. It is linked to a carbon-13 NMR factual data bank and disposes of an original knowledge autoproduction generator (KAG) that deals with the complexity introduced by the topochromatic environments of an entity's molecular sites. Instead of combining general algorithms, dealing in parallel with different types of data, this represents a case where an early combination of spectral and structural data at the level of substructures (SS) is very successful. This system can, moreover, conduct a complex strategy of assembling substructures, as in a puzzle, in order to build, in an appropriate space, prospective structures as candidates to match the spectra.

USER INTERFACE

This part deals with input, interaction, and output mechanisms that regulate man/machine relations or, more simply, the dialogue between chemist and computer. From the inception of the DARC System, we planned for this dialogue to take place in natural chemical language, i.e., through chemical formulas, global entities, or substructures. The DBMS must contain structures and substructures, and we had to decide whether to store all formulas and fragments at the input editing point or to invent an input cycle that would encode the image and reconstitute it each time on request.

The GRAPHEDIT system adopted reconstitutes the image via the stored code and provides a legible display based on atom connectivity. In its inference engine, this system of building rules takes into account certain habits of cycle presentation, such as respect for specific symmetries. A knowledge base and learning possibilities provide for adaptation to special situations (i.e., crystallographic FDB).

Displaying a large amount of data requires a particular set of rules, as, for example, when plotting hyperstructures (often delicate) where one must preserve the possibility of local intervention with possible extraction of ordered subgraphs and also be able to "zoom" in on nodes so as to see the structures. This merger of powerful graph possibilities with our data handling systems is essential and is coherent with our choosing to give priority to the structure drawing over the systematic name, which is only associated with the structure. This aspect led the DARC System, at an early stage, to implement interactive systems on different graphic material (black or color MPS of Evans & Sutherland or diverse Tektronix displays).

Graphic assistance to the input of structural data has been achieved by data blocks with adequate software or by step by step input with our home-made "topocodeur", which provides fast input thanks to a special chemical keyboard and matrix network serving as an electronic drawing board. Flexibility and compatibility of access to different visual terminals and printing devices have been attained. In conclusion, the user has access to the DARC substructures, structures, fuzzy or generic entities, or sets of entities. Stereopresentations are made in these types of codes: Cahn–Ingold–Prelog (CIP), DARC, and an intermediate DARC-CIP compromise code, which is easier to code both manually and for computer handling of stereochemistry.

COMPUTER-AIDED DESIGN

Our computer aided design applications are based on our conception of creating strong KIPS from accurate data banks. We have heretofore described the kind of software tools used for expert systems. We also felt the need to package some of these tools; PARIS is packaged as an inference engine with different knowledge database subsystems and user interfaces. This package, called T-DARC, oriented to solve problems where real and formal transformations are described, has been used to manage several expert systems. Table III shows how, through a unitary methodology, it handles different systems (synthesis, mechanisms, mass elucidation, enumeration) usually dealt with elsewhere by several different, isolated systems.

We have developed specific reaction data banks for SYNOPSYS along the Entity-Relationship approach, with a physical scheme providing for access optimization. Interactivity with the user is such that he can express his queries in a near natural language for the textual part and with normal formulas for the graphic part. These banks can be used notably to (1) assess the scope of application of a reaction and (2) facilitate the extraction of rules for production systems in chemistry (SYNOPSYS and other products).

DARC/CALPHI and Correlation/CAD (CORCAD). In the field of CAD (computer-aided design), a system has been extensively developed to handle structure–activity and structure–property relationships with a topological methodology. It is called PELCO (*P*erturbation of an *E*nvironment which is *L*imited, *C*oncentric, and *O*rdered). PELCO is a similarity method used for analogy search among various compounds of a family and, eventually, for sorting out certain causality functions. CALPHI (*C*omputer *A*ided *L*aw *P*rediction by *H*yperstructure *I*nvestigation) is our general system for implementing the PELCO methodology in a computer-assistance

framework. It employs two subsystems, one for structure display and one for 2D and 3D hyperstructure display, an HS space-builder program, DARC/GEANT, and, finally, some management modules for structure and property databases. DARC/CALPHI is most useful in elucidating data prediction rules or laws in general chemistry and in drug design. This CAD system searches the best structural representation space for the problem under study and makes it possible to work in several spaces, to handle complex structural focuses and to use, during the solution process, external data methods such as cluster analysis, factorial analysis, and molecular and quantum mechanics. A scanning search of the HS working space is conducted to identify regularities and singularities of the HS-associated information surface in order to glean potential correlations.

Developed to implement the PELCO topological method, it is becoming a more general tool, gradually encompassing most of the topological and group methods based on structure–information relationships. The strength of such a complex structural-solving CAD system and its flexibility are best used to complement a structural KIPS, including specific pharmaceutical data banks and files with access to three types of structural knowledge, i.e., charge densities (carbon-13 NMR), bond breaking (mass spectroscopy), and structure topography (crystallography).

This paper has tried to point out the constant need for well-evaluated data as a source for creative information. It is clear that convergence of bibliographic data, factual data, and knowledge information processing systems will influence the future development of knowledge at all levels.

Advances in knowledge and in engineering will rely on and benefit from research in the cognitive sciences and in artificial intelligence. Such research can only be valid when based on serious testing in concrete areas. Chemistry is an excellent field, perhaps the best, for this purpose. We are thus convinced that the reasons for innovating in chemistry with the DARC are as strong now and as promising for the future as they were in the Sixties.

DARC SYSTEM BIBLIOGRAPHY

GENERAL INFORMATION

Dubois, J. E. "Principles of the DARC Topological System. Applications Pointing to Structural Influence on Oxidation of Hydrocarbons". *Entropie* **1969**, *25*, 5–13.

Dubois, J. E. "French National Policy for Chemical Information and the DARC System as a Potential Tool of This Policy". *J. Chem. Doc.* **1973**, *13*, 8–13.

Dubois, J. E. "Ordered Chromatic Graph and Limited Environment Concept". In "Chemical Applications of Graph Theory"; Balaban, A. T., Ed.; Academic Press: London, 1976; Chapter 11, pp 333–370.

Dubois, J. E. "Prévisions d'activité pharmacodynamique à l'aide du système DARC". *Man Comput.* **1972**, 309.

Dubois, J. E. "DARC System in Chemistry". In "Computer Representation and Manipulation of Chemical Information"; Wipke, W. T.; Heller, S. R.; Feldmann, R. J.; Hyde, E., Eds.; Wiley: New York, 1974; Chapter 10, pp 239–264.

Dubois, J. E. "Analysis of Data Bank Implementation and Computer Aided Design in Chemistry". In "Proceedings of the Fifth Biennial International CODATA Conference", Boulder, CO, June 28–July 1, 1976; Dreyfus, B., Ed.; Pergamon Press: Oxford, 1977; pp 501–514.

Dubois, J. E. "Structural Organic Thinking and Computer Assistance in Synthesis and Correlation". *Isr. J. Chem.* **1975**, *14*, 17–32.

Dubois, J. E. "DARC Creative Data to Meet the Challenge of the Fifth Generation in Chemistry". In "Proceedings of the International Conference on Information and Knowledge", Tokyo, Japan, May 3–10, 1984; in press.

DARC CODE

Dubois, J. E.; Laurent, D.; Viellard, H. "Système de documentation et d'automatisation des recherches de corrélations (DARC). Principes généraux". *C. R. Seances Acad. Sci., Ser. C* **1966**, *263*, 764–767.

Dubois, J. E.; Viellard, H. "Système DARC. VII. Théorie de génération-description. I. Principes généraux". *Bull. Soc. Chim. Fr.* **1968**, *No. 3*, 900–904.

Dubois, J. E.; Viellard, H. "Système DARC. XIII. Théorie de génération-description. IV. Représentation des composés cycliques". *Bull. Soc. Chim. Fr.* **1971**, *No. 3*, 839–848.

Dubois, J. E.; Viellard, H.; Panaye, A. "Système DARC. XIV. Théorie de génération-description. V. Description d'entités nécessitant un chromatisme secondaire: ions, radicaux, isotopes, sels, hydrates". *Bull. Soc. Chim. Fr.* **1973**, *No. 6*, 1988–1996.

Razinger, M.; Chrétien, J. R.; Dubois, J. E. "Graphes chimiques: génération automatique des descripteurs topologiques". *Vestn. Slov. Kem. Drus.* **1984**, *31*, 211–227.

Panaye, A.; Dubois, J. E. "Representation and Coding (2D/3D) of 3D Chemical Objects". In "Proceedings of the Ninth International CODATA Conference", Jerusalem, Israel, June 24–28, 1984; Glaeser, P. S., Ed.; North-Holland: Amsterdam, 1985; in press.

DOCUMENTATION AND INFORMATION

Dubois, J. E.; Hennequin, F.; Boussu, M. "Utilisation du système topologique DARC à des fins documentaires. Méthodes de préparation des cétones aliphatiques saturées". *Bull. Soc. Chim. Fr.* **1969**, *No. 10*, 3615–3623.

Dubois, J. E.; Couesnon, T.; Laurent, D.; Azema, C.; Saillard, J. C. "Application des traitements graphiques en documentation automatique et en conception assistée en chimie". *Automatisme* **1974**, *No. 4*, 227.

Attias, R. "DARC Substructure Search System: A New Approach to Chemical Information". *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 102–108.

Picchiottino, R.; Georgoulis, G.; Sicouri, G; Panaye, A.; Dubois, J. E. "DARC-SYNOPSYS. Designing Specific Reaction Data Banks: Application to KETO-REACT". *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 241–249.

Sobel, Y.; Dagane, I.; Carabedian, M.; Dubois, J. E. "Specific Features of Scientific Data Banks". In "Proceedings of the Ninth International CODATA Conference", Jerusalem, Israel, June 24–28, 1984; Glaeser, P. S., Ed.; North-Holland: Amsterdam, 1985; in press.

COMPUTER-AIDED DESIGN

Hyperstructure Concept

Dubois, J. E.; Laurent, D. "Système DARC. Théorie de population-corrélation. Organisation et description d'une population". *C. R. Seances Acad. Sci., Ser. C* **1968**, *266*, 943–945.

Dubois, J. E.; Anselmini, J. P.; Chastrette, M.; Hennequin, F. "Système DARC. XI. Théorie de population-corrélation. I. Organisation d'une population chimique". *Bull. Soc. Chem. Fr.* **1969**, *No. 7*, 2439–2448.

Dubois, J. E.; Laurent, D.; Panaye, A.; Sobel, Y. "Système DARC. Concept d'hyperstructure formelle". *C. R. Seances Acad. Sci., Ser. C* **1975**, *280*, 851–854.

Dubois, J. E.; Laurent, D.; Panaye, A.; Sobel, Y. "Système DARC. Hyperstructures formelles d'antériorité". *C. R. Seances Acad. Sci., Ser. C* **1975**, *281*, 687–690.

Dubois, J. E.; Picchiottino, R.; Sicouri, G.; Sobel, Y. "Système DARC: Relations et hyperstructures par blocs". *C. R. Seances Acad. Sci., Ser. I* **1982**, *294*, 251–256.

Computer-Aided Correlation

Dubois, J. E.; Laurent, D. "Système DARC. Calcul des propriétés globales de molécules focalisées. Méthode de perturbation du topomodèle défocalisé". *C. R. Seances Acad. Sci., Ser. C* **1968**, *266*, 608–611.

Dubois, J. E.; Laurent, D.; Aranda, A. "Système DARC. XVI. Théorie de topologie-information. I. Méthode de perturbation d'environnements limités concentriques ordonnés (PELCO)". *J. Chim. Phys. Phys.-Chim. Biol.* **1973**, *70*, 1608–1615.

Dubois, J. E.; Sobel, Y.; Mercier, C. "Théorie des graphes chimiques: méthode DARC-PELCO. Corrélations de topologie-information et concept de regularité". *C. R. Seances Acad. Sci., Ser. 2* **1981**, *292*, 783–788.

Dubois, J. E.; Herzog, H. "Heats of Formation of Aliphatic Ketones: Structure Correlation Based on Environment Treatment". *J. Chem. Soc., Chem. Commun.* **1972**, *No. 16*, 932–933.

Dubois, J. E.; Chrétien, J. "Data Analysis Methodology: the DARC Topological System". *J. Chromatogr. Sci.* **1974**, *12*, 811–821.

Dubois, J. E. "Strain Energy Modeling of Simple and Crowded Aliphatic Ketones: Spectroscopic Properties". *Pure Appl. Chem.* **1977**, *49*, 1029–1047.

Razinger, M.; Chrétien, J. R. Dubois, J. E. "Structural Selectivity of Topological Indexes in Alkane Series". *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 23–27.

Dubois, J. E.; Carabedian, M. "Modelling of Alkyl Environment Effects on the ^{13}C Chemical Shift". *Org. Magn. Reson.* **1980**, *14*, 264–271.

Sobel, Y.; Mercier, C.; Dubois, J. E. "Méthode DARC-PELCO: QSAR de phénylalkylamines méthoxylées hallucinogènes". *Eur. J. Med. Chem.-Chim. Ther.* **1981**, *16*, 477–479.

Panaye, A.; MacPhee, J. A.; Dubois, J. E. "Steric Effects. II. Relationship between Topology and the Steric Parameter. E'_s–Topology as a Tool for the Correlation and Prediction of Steric Effects". *Tetrahedron* **1980**, *36*, 759–768.

Dubois, J. E.; Doucet, J. P.; Panaye, A. "Corrélation des effects (α-Me), en RMN^{13}C: modèle topologique par différence linéaire DARC-PULFO". *Tetrahedron Lett.* **1981**, *22*, 3521–3524.

Doucet, J. P.; Yuan, S. G.; Billon, P.; Dubois, J. E. "Functional Substituents α to ^{13}C: Topological Correlation by Perturbation of a Focus and Principal Component Factor Analysis". *Tetrahedron Lett.* **1982**, *23*, 4241–4244.

Doucet, J. P.; Panaye, A.; Dubois, J. E. "Topological Correlations of Carboon-13 Chemical Shifts by Perturbation of a Focus: DARC-PULFO Method. Attenuation and Inversion of α-Methyl Substituent Effects". *J. Org. Chem.* **1983**, *48*, 3174–3182.

Mercier, C.; Dubois, J. E. "Comparison of Molecular Connectivity and DARC/PELCO Methods: Performance in Antimicrobial, Halogenated Phenol QSARs". *Eur. J. Med. Chem.-Chim. Ther.* **1979**, *14*, 415–423.

Krawiec, Z.; Gonnord, M. F.; Guiochon, G.; Chrétien, J. R. "Gas-Solid Chromatographic Behavior of 65 Linear or Branched Alkenes and Alkanes (C_2–C_{10}) on Graphitized Thermal Carbon Black". *Anal. Chem.* **1979**, *51*, 1655–1660.

Dubois, J. E.; Chrétien, J. R.; Soják, L.; Rijks, J. A. "Topological Analysis of the Behaviour of Linear Alkenes up to Tetradecenes in Gas-Liquid Chromatography on Squalane". *J. Chromatogr.* **1980**, *194*, 121–134.

Computer-Aided Synthesis

Dubois, J. E. "Computer Assisted Modelling of Reactions and Reactivity". *Pure Appl. Chem.* **1981**, *53*, 1313–1327.

Dubois, J. E. "Conceptual Description of System Transformations and Reactions". In "Proceedings of the Eighth International CODATA Conference", Jachranka, Poland, October 4–7, 1982; Glaeser, P. S., Ed.; North-Holland: Amsterdam, 1983; pp 155–166.

Dubois, J. E.; Panaye, A. "Base-Catalysed Polyalkylation of Aliphatic Ketones. I. Graph and Topological Description of Reaction Pathways". *Tetrahedron Lett.* **1969**, *No. 19*, 1501–1504.

Dubois, J. E.; Panaye, A. "Base-Catalysed Polyalkylation of Aliphatic Ketones. II. Reaction Graph for the Polymethylation of Methylneohexyl Ketone". *Tetrahedron Lett.* **1969**, *No. 38*, 3275–3278.

Dubois, J. E.; Panaye, A.; Lion, C. "Conception assistée par ordinateur. Notion de domaine structural ordonné d'une reaction (DSOR)". *Nouv. J. Chim.* **1981**, *5*, 371–380.

Dubois, J. E.; Lion, C.; Panaye, A. "Evaluation des domaines de synthèse de cétones par alkylation et condensation magnésienne. Synthèse des cétones frontières hyper-encombrées". *Nouv. J. Chim.* **1981**, *5*, 381–391.

Dubois, J. E.; Panaye, A.; Picchiottino, R.; Sicouri, G. "Système DARC: Structure de l'invariant d'une réaction". *C. R. Seances Acad. Sci., Ser. 2* **1982**, *295*, 1081–1086.

Dubois, J. E.; Sicouri, G.; Sobel, Y; Picchiottino, R.; "Système DARC: Opérateurs localisés et co-structures de l'invariant d'une réaction". *C. R. Seances Acad. Sci., Ser. 2* **1984**, *298*, 525–530.

Sicouri, G.; Sobel, Y.; Picchiottino, R.; Dubois, J. E. "Representing, Handling and Coding Sets of Structured Entities". In "Proceedings of the Ninth International CODATA Conference", Jerusalem, Israel, June 24–28, 1984; Glaeser, P. S., Ed.; North-Holland: Amsterdam, 1985; in press.

Picchiottino, R.; Sicouri, G.; Dubois, J. E. "DARC-SYNOPSYS Expert System. Production Rules in Organic Chemistry and Application to Synthesis in Design". In "Computer Sciences and Data Banks"; Hippe, Z. S.; Dubois, J. E., Eds.; Polish Academy of Sciences: Warsaw, 1984; pp 182–193.

Computer-Aided Elucidation

Dubois, J. E.; Carabedian, M.; Ancian, B. "Elucidation structurale, automatique par RMN du carbone 13: Méthode DARC-EPIOS. Recherche d'une relation discriminante structure-déplacement chimique". *C. R. Seances Acad. Sci., Ser. C* **1980**, *290*, 369–372.

Dubois, J. E.; Carabedian, M.; Dagane, I. "Computer-Aided Elucidation of Structures by Carbon-13 NMR. The DARC-EPIOS Method: Characterization of Ordered Substructures by Correlating the Chemical Shifts of Their Bonded Carbon Atoms". *Anal. Chim. Acta* **1984**, *158*, 217–233.