

- Sieber, W. In "Computer Assisted Organic Synthesis". *ACS Symp. Ser.* 1977, No. 61, 98.
- (2) ARGIA stands for Acquisition Restitution Graphique Inter Active.
- (3) (a) Howe, W. J.; Hagadone, T. R. "Interactive Graphics-Based Substructure Searching". Presented at the 182nd National Meeting of the American Chemical Society, New York, 1981. Howe, W. J.; Hagadone, T. R. *J. Chem. Inf. Comput. Sci.* 1982, 22, 9. (b) Hounshell, W. D.; Marson, S.; Peacock, S.; Wipke, W. T. "The MACSS System Search and Retrieval Using an Automated Molecular Access System". Presented at the 182nd National Meeting of the American Chemical Society, New York, 1981. (c) Yoder, D. K.; Walker, T. J. "CAS ONLINE". Presented at the 182nd National Meeting of the American Chemical Society. (d) Attias, R. "The DARC Substructure Search System: A New Approach to Chemical Information". Presented at the 182nd National Meeting of the American Chemical Society, New York, 1981. (e) Feldman, A.; Hodes, L. *J. Chem. Inf. Comput. Sci.* 1979, 19, 125. (f) Moreau, G. *Nouv. J. Chim.* 1980, 4 (1), 17. (g) Bremser, W. *Anal. Chim. Acta* 1978, 103, 355.
- (4) Wipke, W. T.; Dyott, T. M. *J. Am. Chem. Soc.* 1974, 96, 4834.
- (5) Wipke, W. T.; Krishnan, S.; Ouchi, G. I. *J. Chem. Inf. Comput. Sci.* 1978, 18, 32.
- (6) Roos-Kozel, B. L.; Jorgensen, W. L. *J. Chem. Inf. Comput. Sci.* 1981, 21, 101.
- (7) Sussenguth, E. H. *J. Chem. Doc.* 1965, 5, 36.
- (8) Gund, P.; Andose, J. D.; Rhodes, J. B. In "Computer Assisted Organic Synthesis". *ACS Symp. Ser.* 1977, No. 61, 179.

MOLDYN: A Generalized Program for the Evaluation of Molecular Dynamics Models Using Nuclear Magnetic Resonance Spin-Relaxation Data

DAVID J. CRAIK,^{1a} ANIL KUMAR, and GEORGE C. LEVY^{*1b}

National Institutes of Health Resource for Multi-Nuclei NMR and Data Processing, Syracuse University, Syracuse, New York 13210

Received May 17, 1982

A FORTRAN 77 program is described which calculates NMR spectral relaxation parameters on the basis of user-selected models for molecular motion. The program, MOLDYN, allows the comparison of many molecular dynamics models and assists in the testing and evaluation of new models. It is designed primarily for analysis of dynamics in chemical and biological molecules, but the program structure is easily adaptable to other objectives. MOLDYN is used interactively, and extensive prompting and help options are available. Upon selection of a molecular dynamics model, the user is asked to supply or modify parameter values for that model. In one mode the program then calculates T_1 , T_2 (or line width), and NOE values by assuming dipolar relaxation of a given nuclear spin pair. However, the inverse problem, i.e., prediction of correlation times or other motional parameters consistent with the NMR observables, is usually of more importance. One of the major aims of MOLDYN is to allow such computations. The program is designed to allow the user to simultaneously vary one or more selected parameters over desired ranges or to automatically optimize modeling parameters on the basis of user-supplied experimental data. Applications range in scope from the analysis of anisotropic motion in rigid molecules to the study of complex conformational and overall molecular dynamics in biological systems.

INTRODUCTION

Nuclear magnetic resonance (NMR) relaxation measurements provide a powerful means of obtaining information related to molecular dynamics.^{2,3} With the advent of high-sensitivity NMR spectrometers, an abundance of NMR relaxation data for molecules ranging from simple systems such as substituted benzenes⁴ to macromolecules of biological significance such as DNA have been generated.⁵ The extraction of dynamics information from these data requires the formulation of mathematical models which express the relationship between experimental observables and theoretical parameters describing molecular motion. The mathematical expressions which form the basis of these models range from fairly simple functions to relatively complicated equations requiring a computer for their convenient solution. Many models are currently available to describe molecular dynamics in a variety of systems, but clear discrimination between different models is often difficult. Thus, a major task involved in the fitting or interpretation of an experimental data set in terms of molecular dynamics is the selection of a model which explains the data adequately and yet is parsimonious (i.e., not overparameterized). This paper describes a computer program, MOLDYN, which allows the comparison of many molecular dynamics models and assists in the testing and implementation of new models. Before description of this program, a short discussion of some features of molecular dynamics models is in order.

THEORY

Several mechanisms can contribute to nuclear spin relaxa-

tion.⁶ However, here we will be concerned only with the dipolar mechanism in which relaxation of a given nucleus (e.g., ¹³C) is brought about by dipolar interactions with nearby magnetic (e.g., ¹H) nuclei. Molecular motion brings about a modulation of this dipolar interaction and hence provides a fluctuating magnetic field at the observed nucleus which can induce nuclear spin relaxation. The mathematical expression which describes the time dependence of the local fluctuating magnetic field is termed the autocorrelation function, $G(t)$. The Fourier transform of this function, termed the spectral density, $J(\omega)$, describes the frequency components of molecular motion. This latter function enters directly into the calculation of observable NMR relaxation parameters such as spin-lattice relaxation times (T_1 's), spin-spin relaxation times (T_2 's), and nuclear Overhauser effects (NOE's). Expressions³ for these parameters are shown in eq 1-3.

$$\frac{1}{T_1} = \frac{2}{15} \frac{\gamma_I^2 \gamma_S^2 S(S+1) \hbar^2}{r_{IS}^6} [J(\omega_I - \omega_S) + 3J(\omega_I) + 6J(\omega_I + \omega_S)] \quad (1)$$

$$\frac{1}{T_2} = \frac{1}{15} \frac{\gamma_I^2 \gamma_S^2 S(S+1) \hbar^2}{r_{IS}^6} [J(\omega_I - \omega_S) + 3J(\omega_I) + 6J(\omega_I + \omega_S) + 4J(0) + 6J(\omega_S)] \quad (2)$$

$$\text{NOE} = 1 + \frac{\gamma_S}{\gamma_I} \left[\frac{6J(\omega_I + \omega_S) - J(\omega_I - \omega_S)}{J(\omega_I - \omega_S) + 3J(\omega_I) + 6J(\omega_I + \omega_S)} \right] \quad (3)$$

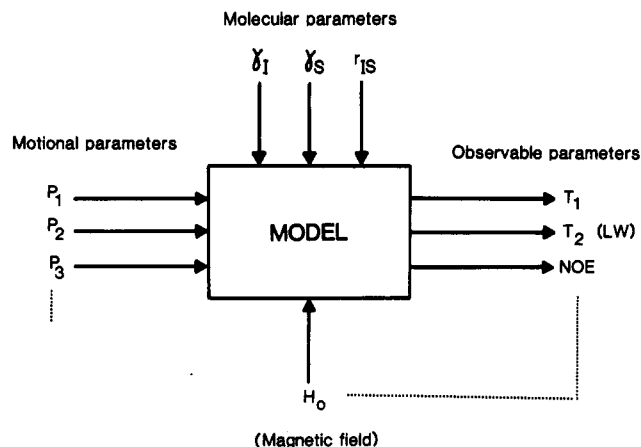


Figure 1. General illustration of the modeling approach applied to NMR studies of molecular dynamics. The observable parameters are determined by (a) the molecular system chosen (this defines γ_I , γ_S , and r_{IS}), (b) the magnetic field strength, and (c) the nature of the molecular motions present. The theoretical approach to the calculation of T_1 , T_2 (LW), and the NOE involves selecting a model and determining a set of values for the motional parameters of that model (P_1 , P_2 , P_3 , ...) that give the best fit of the calculated T_1 , T_2 (LW), and NOE values to the observed data. The dotted line joining the magnetic field and output parameters indicates that repeating the experiments at several different magnetic fields allows several sets of experimental data to be obtained.

Here \hbar is Planck's constant divided by 2π , and γ_I and γ_S are the magnetogyric ratios of the observed nucleus (spin I) and the nucleus causing relaxation (spin S), respectively. r_{IS} is the internuclear distance between the two nuclei, and ω_I and ω_S are the appropriate Larmor frequencies, which depend on the magnetic field strength at which the experiment is carried out. In the absence of significant magnetic field inhomogeneity or chemical shielding variation, observed spectral line widths (LW) at half peak height are related to T_2 via $LW = 1/\pi T_2$.

The autocorrelation and spectral density functions are dependent on the rates, amplitudes, and orientations of overall and internal motions, and the problem of deriving a model for molecular motion reduces to the formulation of an autocorrelation function and related spectral density. Depending on the complexity of the model chosen, the spectral density function may contain one or more parameters describing the rates, amplitudes, and orientations of the various motions present. Substitution of appropriate numerical values for the parameters of the model then leads to a set of calculated T_1 , T_2 , and NOE values which can be compared with experimental values to see if the model and/or motional parameters are appropriate. The general modeling approach as applied to NMR dipolar relaxation studies is illustrated in Figure 1.

The simplest model for molecular motion is that of a rigid, spherical molecule tumbling isotropically in solution. For this motional model, the spectral density reduces to the form shown in eq 4, where τ is the correlation time describing the tumbling motion and ω is the frequency in radians/second at which the function is evaluated.⁶

$$J(\omega) = \tau / (1 + \omega^2 \tau^2) \quad (4)$$

For nonsymmetrical molecules in which the overall motion is anisotropic, additional correlation times (or inversely related diffusion coefficients) are required to describe differential rates of tumbling about molecular diffusion axes. In addition, geometrical parameters describing the orientations of all significant relaxation vectors with respect to the principal axis are required,^{7,8} and the spectral density function becomes considerably more complex than eq 4. The introduction of internal motions into a model further increases its complexity and introduces the need for additional parameters.

A large number of motional models are currently available to describe overall and internal molecular dynamics in a variety of systems.⁷⁻²⁷ Since many of these models contain similar features (e.g., several parameters describing overall motions and several parameters describing internal motions), it is often useful to compare models to see if one gives a significantly better description of the experimental data for a particular molecular system. The inefficiency of writing separate programs for each of the many different models motivated us to write an *integrated general program* in which it is easy to compare results from different models. This approach has obvious advantages in terms of time saving and the reduction of programming errors when a new model is incorporated. More importantly, the implementation of many different models into an integrated package allows one to readily compare different mathematical approaches to see if they lead to similar physical descriptions.

PROGRAM DESIGN CRITERIA

(a) Overall Aims. Four major design criteria were considered when this project was initiated. First, it was considered important that the program should be designed with user interests in mind. The program always gives the user the opportunity to modify data entries, to correct typing errors, or to obtain information from help files which describe current program options. Second, we wanted to incorporate a powerful, general-purpose fitting procedure that would allow one or more variables to be automatically optimized to provide a best fit to the experimental data. This is also supplemented with a powerful module that allows the user to increment or step chosen parameters over designated ranges. Third, we anticipated developing new modeling approaches in our laboratory, and thus the program was written in such a way as to make it very easy to add new motional models. Finally, since we are primarily interested in comparing different motional models, the program flow was arranged so that it is easy to define a parameter list for a given model, perform a calculation, and then select another model and do a similar calculation.

(b) Command Structure and Program Control Flow. Two general philosophies for the command structure in this type of program can be envisaged. In the first structure, the user would have available (at any stage) a set of global commands. Upon entry of the command mnemonic, the program would check to see if the command is appropriate in terms of the current data set and/or parameters, and if so, the command would be executed. Otherwise an error message would be given. In the second philosophy, the user could be forced to follow a prearranged pathway through the program and be given only a limited number of options at every stage. Both structures have advantages and disadvantages, and so we have chosen a combination approach to optimize the command structure to the problem at hand. In this approach, the initial stage of the program forces the user to select a general class of models, followed by a specific model. The user then enters a central node of the command structure which provides more flexibility in terms of options.

PROGRAM DESCRIPTION

MOLDYN calculates NMR relaxation parameters on the basis of a user-selected model for molecular motion. Alternately, and more importantly, the program can be used to *optimize* parameter values for a given model to provide the best fit to a set of experimental data. The program has two functional aspects: (1) parameter input/output (I/O) and (2) computations of various spectral densities by using different models of molecular motion. The parameter I/O core of the program is written in such a way that it would be very easy to adapt

Table I

	Major Modules of MOLDYN
MOLDYN	the driver program which calls the appropriate sub-routines in response to the user-given commands
SETUPA	initializes program variables
HEADING	prints banner information and current revision number of the program
ASSIST	determines level of user assistance
CLASSEL	determines the class of motional model
MODELSEL	selects a specific model from the chosen motional class
PARIO	parameter input/output display module. Central program node
DCODE	service routine to decode mnemonic commands given by user
OUTRES	determines option for the presentation of output. Results may be printed on the VDT screen, line printer, plotted on a digital plotter, or stored on a disk file
PLOTIO	determines user options for plotting
EXPINP	experimental data entry package
GETFIL	determines file name of data file containing geometrical data for certain dynamics models and determines which nuclei in the data file are to be included in the calculation
SPECDEN	calls the appropriate model dependent spectral density function subroutine
RELCAL	calculates T_1 , T_2 , line width, and NOE values for single or multiple dipolar interactions by using the spectral density values provided by SPECDEN
OPTIM	driver module for optimization. Determines initial guesses and convergence criteria and calls SIMPLEX routine
SIMPLEX	optimizes one or more parameters to yield a least-squares best fit with the experimental data for a given model
FUNC	interfaces SIMPLEX and main program. Normalizes parameters and calculates deviations between theoretical and experimental data
HELP	provides the user with general or specific information relevant to current program status
Data Files	
INFO1	contains general program descriptions used in HELP option
INFO2	contains specific option descriptions used in HELP option
NUCDAT	contains nuclear symbols, atomic weights, magnetogyric ratios, and nuclear spin values
CLASSFILE	contains lists of model numbers contained in each general motional dynamics class
MODELFILE	contains cross references to the parameters contained in a given model and contains the alphabetic model name
OPTPRINT	contains command menus
NUMPRINT	contains cross-references to line numbers from OPTPRINT file

the program to modeling problems in areas other than molecular dynamics. Parameter names and other prompting information are stored in data files, and therefore users with other applications need only substitute the names and commands appropriate to their application. New routines for performing numerical calculations would of course have to be written; however, this is relatively straightforward.

(a) **Overall Structure.** MOLDYN is highly modular, making it easy to add or change specific features without changing the general program structure. Table I lists the major program subsections with their functions enumerated while Table II lists the currently available dynamics models. The general structure in Figure 2 shows how the main modules are interconnected and illustrates the most frequently used control pathways. Typically, upon reaching the central parameter display module, the user would progress through loop A (Figure 2) to modify parameter values or perhaps increment one or more parameters over a selected range. After all pa-

Table II. List of Molecular Dynamics Models

no.	description	ref
1	isotropic motion with a single correlation time	6
2	isotropic motion with multiple correlation times (spectral density is a sum of Lorentzian terms with coefficients summing to 1.0)	3
3	isotropic overall motion and internal two-state jump	18
4	isotropic overall motion and internal three-state jump	8
5	axially symmetric anisotropic overall motion	8
6	fully anisotropic overall motion	8
7	axially symmetric overall motion and internal stochastic diffusion	10
8	axially symmetric overall motion and internal three-state jump	10
9	axially symmetric overall motion and internal two-state jump	17
10	axially symmetric overall motion and internal conic diffusion	16
11	isotropic overall motion and internal conic diffusion	14
12	isotropic overall motion and internal conic diffusion	15
13	isotropic overall motion and internal conic diffusion	19
14	anisotropic overall motion with optimization of principal axis system	33 ^a
15	overall isotropic motion and internal libration	11
16	log χ^2 distribution of correlation times	27
17	Poisson distribution of correlation times	^b

^a Several new models incorporating fully anisotropic motion and a variety of internal motions will soon be incorporated.³⁶

^b Private communication with A. Ejchart.

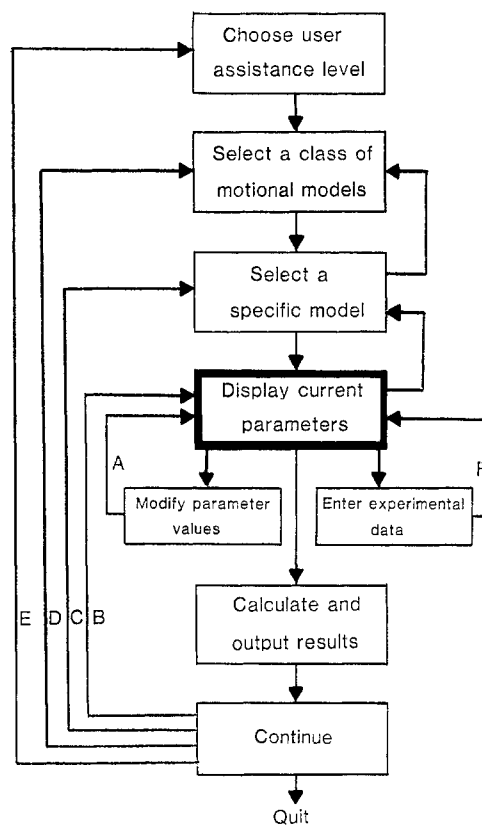


Figure 2. Simplified block diagram of MOLDYN (revision 2.02) program structure.

parameter values have been assigned correctly, the program is directed to perform a calculation and output the results in a form specified by the user. Control then returns (via loop B) to the central parameter display module, where the user may

decide to repeat the calculation using new parameter values, select a new model in the same motional class (loop C), or select a new model in a different motional class (loop D). Loop E is used to enter an experimental data set into memory to be used later for the optimization (loop F) of parameter values of the chosen model. The various modules shown in Figure 2 are accessed by the driver program in response to user directions. Exit from the program can be made at any time using the Q (Quit) command.

(b) Commands. The program is directed by user-given commands selected from a number of menus. The cursor prompt "==" is used to indicate that a command should be entered by the user. Commands are either one- or two-letter mnemonics terminated with a carriage return (CR). Certain "global commands" are always available, e.g., H for help, UL to change the user assistance level, or Q to quit the program. Other commands are only available at the point in the program relevant to their purpose. Note that the same single character mnemonic may mean different things at different points in the program. In a number of places "carriage return" by itself is regarded as a default command initiator; i.e., pressing (CR) after the prompt activates a default command. This lessens user typing.

Commands are decoded by a general purpose subroutine, DCODE (with arguments CHAR, I, R, KEY), which accepts alphanumeric input of up to 20 characters and returns a decoded character string, integer, or real variable in CHAR, I, and R, respectively. KEY is an integer variable which flags the type of input that was decoded; e.g., a value of KEY = 1 is returned if the user entered a character, KEY = 2 for an integer constant, etc. In a number of places in MOLDYN mixed character/integer input is accepted, e.g., M3 for *modify* parameter 3. This format is also handled by DCODE, which returns the integer part of the expression in I and the character part in CHAR and flags this particular type of input with a unique value of KEY. If the user enters (CR) in response to a command prompt, DCODE accepts this character as valid input and sets KEY accordingly. The calling program is thus flagged that a default option has been specified. Thus, the routine DCODE is used extensively to ensure that the user cannot crash the program by typing a response inappropriate to a FORTRAN I/O format statement.

(c) Parameters. From the theory section, it is apparent that the parameters required to calculate T_1 , T_2 , and NOE values fall into three categories. First there are the parameters which enter into the motional part (i.e., the spectral density function) of the calculation. Typically, these can be correlation times, diffusion coefficients, angular parameters describing restricted amplitude motion, or other geometrical constraints. Such entries are referred to as *primary model parameters*. Second, there are those parameters which make up the interaction part of the relaxation equations (eq 1-3), e.g., γ_I and γ_S are the magnetogyric ratios of the two spins involved in the calculation (I is the observed nucleus and S is the nucleus inducing dipolar relaxation); r_{IS} is the internuclear distance between the I and S nuclei. The program is generalized so that the user is able to examine interactions between any pairs of dipolar nuclei (and not just ^{13}C and ^1H , which is one common case). Third, the magnetic field falls into a category by itself since it is not strictly a motional parameter, but it does affect the spectral density term since the resonance frequencies ω_I and ω_S are linearly related to the magnetic field. The spectroscopist explicitly chooses the magnetic field at which the experiment is run and, by deciding to observe a particular nuclear spin system (e.g., ^{13}C - ^1H), implicitly selects γ_I , γ_S , and r_{IS} as experimental parameters. Thus, the magnetic field and interaction terms, which are collectively termed *secondary parameters*, are common to all models and are determined by

the nature of the experimental conditions used.

The program automatically generates secondary parameters for every model, while the primary parameters are encoded by a programmer when a new model is incorporated. Since each of the molecular dynamics models involves a different set of motional parameters, e.g., correlation times or geometrical constants, it was decided to have a pointer-controlled, indirect, but *generalized* method of communicating with the model-dependent spectral density subroutines. Several global arrays contain the names, default values, and other information pertaining to 100 possible parameters that might be utilized in molecular dynamics studies. Any given motional model requires only a small subset of these parameters, and indexing of this subset is readily achieved by using a cross-referencing array, ICREF, which contains pointers to the appropriate elements from the storage arrays.

(d) Addition of New Models. Because of the cross-referencing structure noted above, the parameter I/O coding associated with the inclusion of a new model is extremely simplified. The programmer needs only to select the parameters names, etc., that are appropriate to the new model from the generalized list of 100 existing parameter names and enter appropriate integers into a cross-referencing data file (MODELFILE). Thus, with the addition of two lines to this data file (one containing the cross-referencing integers and the other the alphanumeric name of the new model) the program is set up to accept values for the parameters of the new model. The general motional class into which the new model falls is flagged by adding an integer cross-reference into another data file (CLASSFILE). No recompilation of the program is necessary for these steps, which are carried out in a matter of minutes.

For completion of the addition of a new model to the program, a function subroutine must be written to evaluate the new spectral density function. To do this, one can express the spectral density in the form $J(\omega) = f(P_1, P_2, \dots, P_n, \omega)$, where P_1, P_2, \dots, P_n are the n primary motional parameters of the model and ω is the frequency at which the function is evaluated. The expression is encoded into FORTRAN, compiled, and linked to the remainder of the program. This function is evaluated by the calling routine (RELCAL) at the five frequencies (ω_I , ω_S , $\omega_I + \omega_S$, $\omega_I - \omega_S$, and 0) required for the calculation of the T_1 , T_2 , and NOE values (eq 1-3). The motional parameters are passed (via COMMON) from the I/O section of the program to the spectral density function by using generalized parameter value array PARAM(i), where i indexes the parameters in the model.

(e) Optimization. The calculation of NMR relaxation parameters such as T_1 , T_2 , and NOE is straightforward, provided that a mode for the description of the molecular motion has been chosen and values for all parameters are known. Usually the inverse problem of prediction of correlation times or other motional parameters, though more difficult, is of greater utility. MOLDYN is provided with an optimization module which finds a least-squares best fit of calculated data to available experimental data. Any number of parameters may be optimized for a given model, provided enough experimental data is available. (For example, the program allows the user to input data obtained from multiple experiments performed at varied magnetic fields.)

Since the different available dynamics models have different mathematical forms and often contain parameters of vastly different magnitudes, the optimization module was designed around the SIMPLEX²⁸⁻³⁰ algorithm, which is readily adaptable to the optimization of a variety of different functions. In our case, the function to be optimized is the normalized sum of squares of deviations between calculated and observed T_1 's, T_2 's (line widths), and NOE's. SIMPLEX is an effective and computationally compact algorithm for finding extreme points

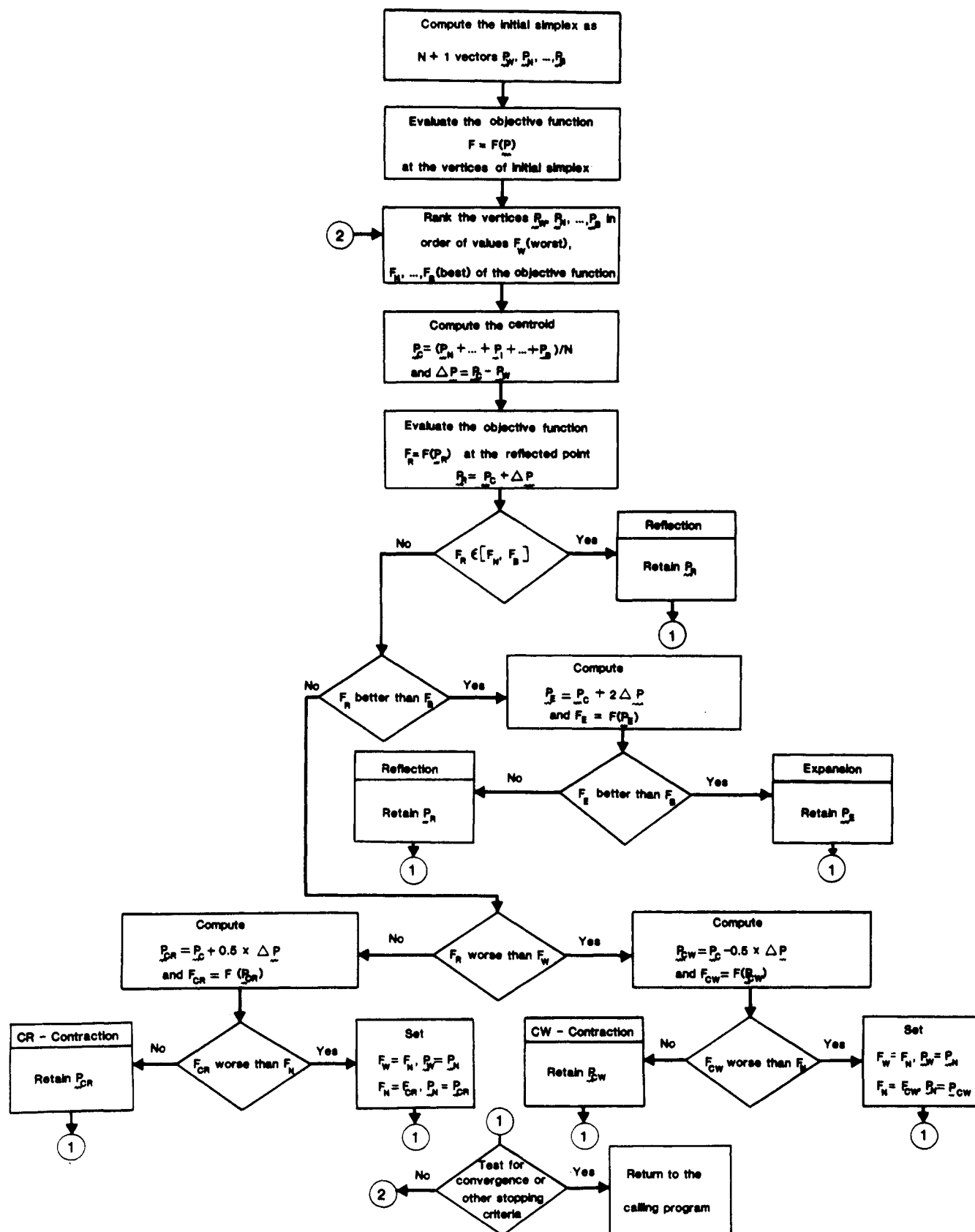


Figure 3. Flow chart of generalized SIMPLEX algorithm as implemented in MOLDYN.

of a function of n variables. It depends on the comparison of function values at $n + 1$ vertices of a general SIMPLEX algorithm, followed by the replacement of the vertex with the worst value by a new one with a better value. MOLDYN utilizes the modified SIMPLEX method^{29,30} rather than the original fixed-size SIMPLEX algorithm of Spendley et al.²⁸ In the

modified procedure, the size of the SIMPLEX algorithm is variable and depends on the local topography of the function. Figure 3 shows the SIMPLEX scheme used in MOLDYN.

In some cases the user may have reliable estimates for some parameter values in a given model but wishes to optimize the other parameter values to provide the best description of the

experimental data. Since, in general, different models have different numbers of parameters, the SIMPLEX subroutine was designed to be used with any number (up to ten and easily extendable) optimized parameters. All decisions regarding which parameters are to remain fixed, and which are to be optimized, are made by the user in the parameter display module (Figure 2). The optimization subroutine is connected to the main package by an interface routine which normalizes (via a logarithmic transformation transparent to the user) all parameter values into a convenient range for the optimization procedure.

Immediately before the optimization is commenced, the user is given the choice of supplying initial guesses for the parameters to be optimized as well as convergence criteria (or instead, default values for these parameters may be used). One very useful feature of the procedure is the option of specifying constraints on some or all of the optimized parameters. If a user chooses to constrain an optimized parameter within a set of limits, the optimization procedure will only seek solutions consistent with these limits. This provides the user with a great deal of control over the optimization process and is one means of avoiding solutions that are not physically meaningful. The constraint procedure operates by arbitrarily assigning a large error to any new vertex of the SIMPLEX subroutine which is reflected outside the selected range.

Before the optimization is initiated, the user must provide experimental data (loop E in Figure 2). After convergence is obtained, or a preset number of iterations is completed, the optimized parameters and fitting statistics are displayed. In a future revision, MOLDYN will also support graphic display of fit quality.

(f) Help. Help operates in a bifunctional manner that allows general or specific information to be obtained. Upon entering the help routine, the user is given a general description of the current stage of the program and a list of the options available. The user may then exit (E) the help routine or ask for more help (H). Since this initial general description is often sufficient to orient the user and allow a decision as to the next course to be made, the user will often exit the help routine at this stage. However, if more detailed information is required about specific options, the "more help" option may be requested. This prints detailed information about any of the currently available commands or options.

MOLDYN IN OPERATION

Upon starting the program, the user is asked to choose one of three levels of assistance. A new or less experienced user may choose LEVEL 1: the verbose level with most prompting. In this level a description of the current point in the program is printed before the user is given an opportunity to execute a command. Each command is listed in a menu, and a prompt signal is given. LEVEL 2 also lists command lists but offers no generalized description. The more experienced user, who has memorized commands and does not wish to have menu lists printed may choose LEVEL 3. In either LEVEL 2 or LEVEL 3, a user maintains the ability to obtain assistance via the HELP routines by typing the global command H. After choosing the level of assistance, a list of classes of models for molecular motion is displayed, and a user choice is requested. Upon selection of a class of motion, a number of specific models are displayed. (There is some overlap between different models, and the same model may appear under different class headings.) Although most models are formulated in terms of the motion of a single relaxation vector, it is sometimes convenient to examine several nuclei in a molecule simultaneously. In the case of multiple dipole-dipole interactions, the geometry of the molecule may be read from a disk file. This is particularly useful for examining the motion of

rigid molecules by using the Woessner⁷⁻¹⁰ formalism for anisotropic motion.

After selection of a model, the default parameters of that model are displayed, and the user enters the central node of the I/O section of the program. The main options available are M (to modify parameter values/units), D to describe the model, and X to enter experimental data. The help (H) and quit (Q) options are, of course, available, as are options to exit this node by going forward using C (proceed with calculation) or back to the previous stage using S (select new model).

If D is typed in response to the prompt a description as follows would be displayed on the terminal.

This model describes the motion of an I-S nuclear spin vector (I = observed nucleus, S = dipolar nucleus) attached to a molecule undergoing axially symmetric overall anisotropic motion, with correlation time τ_{Z} , about the Z molecular axis, and correlation time τ_{X} about the X and Y molecular axes. Internal motion of the I-S vector occurs with correlation time τ_{W} . This internal motion is restricted in amplitude to lie in a cone of half-angle THETA_CONE, having its axis aligned at angle BETA with respect to the molecular Z axis.

Reference: G. Lipari and A. Szabo, *J. Chem. Phys.* **75**, 2971 (1981).

If this model is appropriate, the user might then type M to modify the parameter values before proceeding with the calculation. The M command successively displays each parameter name, current value, and units, as shown below for the first parameter:

PARAMETER	CURRENT VALUE	NEW VALUE	UNITS
1) BETA	0.00		degrees

The cursor rests at a point under the "NEW VALUE" column adjacent to the units, and the user has the following options available: enter numeric value, enter (CR) to leave current value unchanged, I to increment parameter (IL increments logarithmically), S to step parameter, O to optimize parameter, U to modify units.

Upon execution of the I subcommand, the user is asked to supply an initial numerical value, a final numerical value, and an integer specifying the number of increments. This subcommand is useful for varying one or more parameters over selected ranges to determine their functional effect in modifying T_1 , T_2 , and NOE. The S subcommand has a similar function but operates by requesting the user to enter the number of discrete steps required (usually two to five values), followed by the individual values themselves. This option is particularly useful for simultaneously calculating T_1 , T_2 , and NOE at a number of different magnetic fields. The O subcommand takes no immediate action but flags the optimization routine by which the given parameter is to be automatically adjusted to provide a best fit to entered experimental data. The U command allows the user to change the default values of the units. For the "OBSERVED NUCLEUS" and "DIPOLAR NUCLEUS" parameters, the only allowable entries are (CR) to maintain current settings or a nuclear mnemonic symbol, e.g., C-13, H-1, N-15, etc. A number of these symbols are stored in a data file (NUCDAT) which also contains numerical information such as nuclear spin, magnetogyric ratio, etc., which are used in subsequent calculations. The user thus has complete flexibility in calculating relaxation parameters based on dipolar relaxation for any pair of interacting nuclei.

After all parameter values have been set by using the M command (or individual parameter values have been changed by using Mn, where n is the parameter number), the corrected parameters are displayed on the screen. The X option to enter

OPTIMIZED parameter values for:
 Motional Class: 4
 Model 5 : Axially symmetric + internal conic diffusion(Lipari,Szabo)
 for : Native DNA C1'

PARAMETER	OPTIMIZED VALUE	UNITS
1) BETA	115.0	degrees
2) THETA-CONE	21.60	degrees
3) TAU-X	894.3	1.E-9 seconds
4) TAU-Z	21.42	1.E-9 seconds
5) TAU-W	4.650	1.E-9 seconds
6) MAGNETIC FIELD	150.0	MHz (protons)
	270.0	
	400.0	
	500.0	
7) OBSERVED NUCL	C-13	
8) DIPOLAR NUCL	H-1	
9) BONDLENGTH	1.090	Angstrom

* Indicates Optimized Parameters

NUCLEUS NUMBER: 1

MAGNETIC FIELD	T1	T2	NOE	LW	T1	T2	NOE	LW
150.00	.218	.002	1.702	127.477	.210	-	1.620	127.000
270.00	.430	.003	1.513	126.980	.540	-	1.550	127.000
400.00	.664	.003	1.375	126.797	.680	-	1.370	127.000
500.00	.860	.003	1.315	126.724	.760	-	1.200	127.000

SIMPLEX OPTIMIZATION STATISTICS

NUMBER OF TIMES:

REFLECTION SUCCEEDED:	27
EXPANSION SUCCEEDED:	3
CW-CONTRACTION SUCCEEDED:	51
CR-CONTRACTION SUCCEEDED:	7
CW-CONTRACTION FAILED:	7
CR-CONTRACTION FAILED:	5

SUM OF SQUARES OF DEVIATIONS:	7.3281E-02
ITERATION COUNT:	100
INITIAL SIMPLEX SIZE:	4.579
FINAL SIMPLEX SIZE:	1.1497E-04
REDUCTION FACTOR:	2.5107E-05

CONVERGENCE WAS NOT ATTAINED

Figure 6. Typical results for an optimization calculation (actual text as displayed by the program).

methyl groups, it may be necessary to measure T_1 's and NOE's for many *different* carbons to uniquely define the dynamics. For systems undergoing relatively fast uncorrelated motions multifield experiments are less useful.

Note that the model shown in Figure 5 assumes that the principal axis system for molecular rotation is known. MOLDYN also supports the analyses of more complex systems where the axes are not known. In this case a reasonable starting set of axes is chosen from the inertial properties of the molecule, but the optimization program can be requested to optimize to the final axis system of best fit. Given the atomic coordinates and masses for a molecule, the program can compute the principal axes and moments of inertia and rotate the axes through the appropriate euler angles so that the inertia tensor is diagonalized. Then on the assumption that both inertia and diffusion tensors are diagonal simultaneously, the diffusion constants for anisotropic tumbling can be computed. The ease with which this procedure can be carried out follows directly from the generalized nature of our SIMPLEX optimization package and interface routine. The ability to easily optimize the molecular diffusion axes may provide an important new understanding of the way in which molecules interact in solution. In studies of anisotropic motion in nonsymmetrical molecules³²⁻³⁴ it is usually assumed that the inertial axis system can be used to characterize motion, but recent work³⁵⁻³⁸ has shown the importance of interacting substituent groups in shifting the diffusion axis system away from the inertial axis system. However, in such a case, the calculation of the relative orientation of the two axes system is not trivial and is liable to nonunique solutions and possible misinterpretation.³⁸

PROGRAM IMPLEMENTATION

MOLDYN is implemented in FORTRAN 77 on a Data General MV-8000 32-bit computer under the AOS/VS operating system, running in an interactive mode on standard video display terminals. The program consists of ~5000 lines

of code including many comments and comprises over 80 subroutines.

MOLDYN is currently available on NMR/SPECNET by dial-up to the NIH Biotechnology Research Resource for Multi-Nuclei NMR and Data Processing (phone numbers: (315) 422-3443, 422-3447). Trial users enter SPECNET with user name NEWUSER and password SPECNET.

The source code for MOLDYN is also available. For information write to G.C.L. We also plan to make the program available for wider dissemination through the Quantum Chemistry Program Exchange during 1983.

CONCLUSIONS

MOLDYN, with its command-driven interactive capabilities, is a powerful new tool for the analysis of nuclear magnetic resonance relaxation data in terms of molecular dynamics models. Extensive and generalized parameter input-output facilities make it easy to use for a nonprogrammer scientist. New molecular motional models and other capabilities for computation and display can be incorporated with relative ease.

ACKNOWLEDGMENT

This project was supported by Grant RR-01317 from the Division of Research Resources, National Institutes of Health. Further support from the National Science Foundation (Grant CHE 81 05109) is acknowledged. Support to D.J.C. from CSIRO (Australia) during part of this work is also acknowledged. Parts of the MOLDYN code were written by Daniel Aquilino, Alison Godfrey, Michael Rome, Brian Strop, and Srinivasan Sundaram. We thank Dr. R. E. London for helpful comments.

REFERENCES AND NOTES

- (1) (a) Department of Chemistry, Monash University, Clayton, Victoria, Australia 3168. (b) Camille and Henry Dreyfus Teacher Scholar, 1976-1981.
- (2) Lyster, J. R.; Levy, G. C. In "Topics in Carbon-13 NMR Spectroscopy"; Levy, G. C., Ed.; Wiley-Interscience: New York, 1974; Vol. 1, pp 86, 87.
- (3) London, R. E. In "Magnetic Resonance in Biology"; Cohen, J. S., Ed.; Wiley-Interscience: New York, 1980.
- (4) Levy, G. C.; Lichter, R. L.; Nelson, G. L. "Carbon-13 Nuclear Magnetic Resonance"; Wiley-Interscience: New York, 1980; Chapter 8.
- (5) (a) Jardetzky, O. *Acc. Chem. Res.* **1981**, *14*, 291. (b) Levy, G. C.; Hilliard, P. R., Jr.; Levy, L. F.; Inners, R.; Rill, R. L. *J. Biol. Chem.* **1981**, *256*, 9986. (c) Early, T. A.; Kearns, D. R.; Hillen, W.; Wells, R. D. *Biochemistry* **1981**, *20*, 3756. (d) Bolton, P. H.; James, T. L. *J. Am. Chem. Soc.* **1980**, *102*, 25. (e) Klevan, L.; Armitage, I. M.; Crothers, D. M. *Nucleic Acids Res.* **1979**, *6*, 1607. (f) Shindo, H. *Biopolymers* **1980**, *19*, 509. (g) Opella, S. J.; Wise, W. B.; DiVerdi, J. A. *Biochemistry* **1981**, *20*, 284.
- (6) Abragam, A. A. "The Principles of Nuclear Magnetism"; Oxford University Press: London, 1961; Chapter VIII.
- (7) Woessner, D. E. *J. Chem. Phys.* **1962**, *36*, 1.
- (8) Woessner, D. E. *J. Chem. Phys.* **1962**, *37*, 647.
- (9) Woessner, D. E. *J. Chem. Phys.* **1965**, *42*, 1855.
- (10) Woessner, D. E.; Snowden, B. S., Jr.; Meyer, G. H. *J. Chem. Phys.* **1969**, *50*, 719.
- (11) Howarth, O. W. *J. Chem. Soc., Faraday Trans. 2* **1978**, *74*, 1031.
- (12) London, R. E.; Avitabile, J. *J. Am. Chem. Soc.* **1978**, *100*, 7159.
- (13) London, R. E.; Avitabile, J. *J. Am. Chem. Soc.* **1977**, *99*, 7765.
- (14) Wang, C. C.; Pecora, R. *J. Chem. Phys.* **1980**, *72*, 5333.
- (15) Brainard, J. R.; Szabo, A. *Biochemistry* **1981**, *20*, 4618.
- (16) Lipari, G.; Szabo, A. *J. Chem. Phys.* **1981**, *75*, 2971.
- (17) London, R. E.; Phillips, M. A. *J. Magn. Reson.* **1981**, *45*, 476.
- (18) London, R. E. *J. Am. Chem. Soc.* **1978**, *100*, 2678.
- (19) Richarz, R.; Nagayama, K.; Wuthrich, K. *Biochemistry* **1980**, *19*, 5189.
- (20) Doddrell, D.; Glushko, V.; Allerhand, A. *J. Chem. Phys.* **1972**, *56*, 3683.
- (21) Wittebort, R. J.; Szabo, A. *J. Chem. Phys.* **1978**, *69*, 1722.
- (22) Ban, B. *J. Magn. Reson.* **1981**, *45*, 118.
- (23) Bull, T. E. *J. Magn. Reson.* **1978**, *31*, 453.
- (24) Connor, T. M. *Trans. Faraday Soc.* **1964**, *60*, 1574.
- (25) Tsutsumi, T. *Mol. Phys.* **1979**, *37*, 111.
- (26) Jones, A. A.; Stockmayer, W. H. *J. Polym. Sci.* **1977**, *15*, 847.
- (27) Schaefer, J. *Macromolecules* **1973**, *6*, 882.
- (28) Spendley, W.; Hext, G. R.; Himsforth, F. R. *Technometrics* **1962**, *4*, 441.
- (29) Nelder, J. A.; Mead, R. *Comput. J.* **1965**, *7*, 308.
- (30) Deming, S. N.; Morgan, S. L. *Anal. Chem.* **1973**, *45*, 278A.

- (31) Levy, G. C.; Craik, D. J.; Kumar, A.; London, R. E., submitted for publication.
 (32) Berger, S.; Kreissl, F. R.; Grant, D. M.; Roberts, J. D. *J. Am. Chem. Soc.* **1975**, *97*, 1805.
 (33) Levy, G. C.; Craik, D. J.; Norden, B.; Phan Viet, M. T.; Dekmejian, A. *J. Am. Chem. Soc.* **1982**, *104*, 25.
 (34) Beierbeck, H.; Martino, R.; Saunders, J. K. *Can. J. Chem.* **1979**, *57*, 1224.
 (35) Axelson, D. E.; Holloway, C. E. *Can. J. Chem.* **1980**, *58*, 1679.
 (36) Somorjai, R. L.; Deslauriers, R. *J. Am. Chem. Soc.* **1976**, *98*, 6460.
 (37) Levy, G. C.; Kumar, A.; Craik, D. J., unpublished data.
 (38) Levy, G. C.; Wang, D., unpublished data.

Summary of the History and Status of the System Development for the Environmental Chemicals Data and Information Network (ECDIN)

JUDITH M. HUSHON,*† JOHN POWELL,‡ and WILLIAM G. TOWN‡

Enviro Control Division, Dynamac Corporation, Rockville, Maryland 20852, and Commission of the European Communities, Joint Research Centre, I-21020 Ispra (Varese), Italy

Received June 17, 1982

The Environmental Chemicals Data and Information Network (ECDIN) has been developed over the last decade by the European Communities to provide information on all aspects of chemicals to the member countries through EURONET DIANE. This paper describes the evolution of the software for ECDIN from a homemade management system to the adoption of ADA-NATURAL. The expansion and structuring of the chemical files are also discussed, including the recent addition of the European Core Inventory (ECOIN).

The Environmental Chemicals Data and Information Network (ECDIN) developed by the European Communities provides a key source of data on chemical substances to many Europeans through EURONET DIANE. This paper is designed to document ECDIN's origin, development, and expansion to its present status as the central focal point for handling the European Inventory of Existing Commercial Chemical Substances (EINECS).

HISTORICAL BACKGROUND

The European Communities have adopted an environmental policy supported by a research program which provides technical and scientific backup to that policy and to the activities of the Member Countries (states). The research program consists of four main parts one of which is the management of environmental information.

Under this research program, the European Communities in 1973 decided to identify the problems involved in collecting and accessing data on environmental chemicals and to work toward their solution. The scope of the pilot phase was as follows:¹ to determine the availability of data; to refine the data structures in light of the available data; to assess the cost of data collection; to develop an input format; to locate data sources and potential contributors to a "network"; to demonstrate the computerized system; to test the reaction of users; to design a computer system for ECDIN.

The decision was also made to locate this effort at the EC Joint Research Centre at Ispra, Italy, which already had a large computer operation. This pilot operation was named ECDIN, the Environmental Chemicals Data and Information Network.

During this developmental phase, some basic principles were established:² (1) ECDIN would store relevant data on any chemical compound produced in sizeable quantity (>500 kg/year) irrespective of its toxicity or harmfulness. (2) ECDIN would be designed for management and retrieval of "hard" data (not document retrieval). (3) The organization

Table I. ECDIN Data Categories

no.	category
1	identification of the chemical
2	codified structure information
3	physical and chemical properties
4	chemical analysis data and methods
5	supply: production and trade
6	transport, packing, handling, storage, and hazards
7	use and disposal
8	dispersion and transformation in the environment
9	effects of the chemical on the environment (including toxicity)
10	regulatory data

would be chemically oriented. (4) The basic language of the system would be English with other languages being considered for later addition. (5) ECDIN would be available on a time-shared system. (6) The system would eventually be networked from its central location at Ispra, Italy, to at least one location in each of the EC Member states.

During the next several years, the elementary data structure of ECDIN was evolved. The structure adopted was basically hierarchical, which permitted the organization of the data elements into a tree structure. Access to the data structure would be by inverted file, and numerical data could be accessed through data ranges. Data were to be stored on a chemical-by-chemical basis in the ten data categories shown in Table I. Data structures within these categories were subsequently elaborated. The major initial emphasis by the ECDIN staff was on the development of structures in the chemical identification, physical/chemical properties, and effects on man and his environment categories.

During the pilot phase, the number of chemicals in the system was limited to 5000. In addition, a subset of 100 chemicals was selected for which more complete data sets were assembled. Following the pilot phase, expansion of the ECDIN system has been very rapid due to the merger of several existing files: compounds of interest to the Administration of the Customs Union in Brussels (16 000); compounds in the 1976 edition of the *Registry of the Toxic Effects of Chemical Substances* (RTECS) published by the U.S. National Institute

*Dynamac Corp.

†Commission of the European Communities.