Reasoning," in "Computers in Chemical Education and Research," E.V. Ludena, N.H. Sabelli, and A.C. Wahl, Eds., Plenum Press, New York, N.Y., 1977, p. 401.

(6) D.H. Smith and R.E. Carhart, "Structure Elucidation Based on Computer Analysis of High and Low Resolution Mass Spectral Data," in "High Performance Mass Spectrometry: Chemical Applications," M.L. Gross, Ed., American Chemical Society, 1978, p. 325.

(7) T.H. Varkony, D.H. Smith, and C. Djerassi, "Computer-Assisted Structure Manipulation: Studies in the Biosynthesis of Natural Products," Tetrahedron, 34, 841 (1978).

(8) D.H. Smith and P.C. Jurs, "Prediction of 13C NMR Chemical Shifts," J. Am. Chem. Soc., 100, 3316 (1978).

(9) T.H. Varkony, R.E. Carhart, D.H. Smith, and C. Djerassi, "Computer-Assisted Simulation of Chemical Reaction Sequences. Applications to Problems of Structure Elucidation," J. Chem. Inf. Comp. Sci., 18, 168 (1978).

(10) D.H. Smith, T.C. Rindfleisch, and W.J. Yeager, "Exchange of Comments: Analysis of Complex Volatile Mixtures by a Combined Gas Chromatography-Mass Spectrometry System," Anal. Chem., 50, 1585 (1978).

(11) W.L. Fitch, P.J. Anderson, and D.H. Smith, "Isolation, Identification and Quantitation of Urinary Organic Acids," J. Chrom., in press.

(12) W.L. Fitch, E.T. Everhart, and D.H. Smith, "Characterization of Carbon Black Adsorbates and Artifacts Formed During Extraction," Anal. Chem., in press.

(13) W.L. Fitch and D.H. Smith, "Analysis of Adsorption Properties and Adsorbed Species on Commercial Polymeric Carbons," Environ. Sci. Tech., in press.

(14) J.G. Nourse, R.E. Carhart, D.H. Smith, and C. Djerassi, "Exhaustive Generation of Stereoisomers for Structure Elucidation," J. Am. Chem. Soc., in press.

(15) C. Djerassi, D.H. Smith, and T.H. Varkony, "A Novel Role of Computers in the Natural Products Field," Naturwiss., in press.

(16) N.A.B. Gray, D.H. Smith, T.H. Varkony, R.E. Carhart, and B.G. Buchanan, "Use of a Computer to Identify Unknown Compounds. The Automation of Scientific Inference," Chapter 7 in "Biomedical Applications of Mass Spectrometry," G.R. Waller, Ed., in press.

(17) T.C. Rindfleisch and D.H. Smith, in Chapter 3 of "Biomedical Applications of Mass Spectrometry," G.R. Waller, Ed., in press.

(18) T.H. Varkony, Y. Shiloach, and D.H. Smith, "Computer-Assisted Examination of Chemical Compounds for Structural Similarities," J. Chem. Inf. Comp. Sci., in press.

(19) R. Carlson, et al., Bioorg. Chem., 7, in press.

(20) J.G. Nourse, "The Configuration Symmetry Group and its Application to Stereoisomer Generation, Specification and Enumeration," J. Am. Chem. Soc., in press.

3   SIGNIFICANCE

There are several results obtained during the past year which are especially significant for both the advancement of our techniques for biomolecular structure elucidation and the dissemination of results of our research. Perhaps the most significant result is in the category of dissemination because for the first time we can offer directly to the biomedical community the results of our research in the form of exportable computer software for computer-assisted structure elucidation. We have discussed in the past the differences between our work and many other forms of health-related research. The latter can generally be described in the literature in sufficient detail for others to duplicate the procedures and results and, more importantly, build on those results in extensions of the work.   When one of the primary "results" of research is a set of complex computer programs, transfer of these results becomes a formidable task.   Certainly the programs and even some of the algorithms can be described in the literature. We have done so, and the recent references at the end of the previous section represent such publications. However, unless an interested person can actually gain access to a program in an executable, or runnable, form, such descriptions are inadequate. We now have the capability of offering the new, smaller, faster version of CONGEN to the community at SUMEX on a trial basis and exported to their own computers for those who wish to make more extensive use of the program. Therefore, we can share our results with our collaborators much more easily now than in the past and persons receiving the program have a foundation on which to build.

The workshops were also significant in that we received constructive suggestions and criticisms from a   cross-section of potential users of our computational techniques, while at the same time exposing a variety of persons active in structure elucidation to CONGEN working on real problems. The success of this effort has encouraged us

to make available in the exportable CONGEN a variety of other structure manipulation tools which we and the persons at the workshop perceive as useful as adjuncts to CONGEN. Thus, as we have developed capabilities for exploration of large numbers of structures in the STRUCC program, and new ways to use and display the results of the STEREO program, we have begun incorporating these capabilities into CONGEN. This effort will continue, using CONGEN as the focal point for further developments in the area (see next section).

Other significant advances have been summarized in the sections outlining various aspects of our research, above. These include new mathematical developments which made possible the reduced size and increased efficiency of CONGEN, completion of the STEREO program and integration of it into CONGEN, new approaches to rule formation in Meta-DENDRAL and better methods for predicting mass spectra and ranking structures based on those predictions. The last item deserves further comment, because it represents a general approach to structural analysis of which we can now take advantage because of the efficiency of the new CONGEN. Spectrum prediction for a variety of spectroscopic techniques, including mass, NMR, IR, etc., including now even chiroptical methods for future work, represents a valuable method for structure determination. Now that we can deal with hundreds or even thousands of structures in intermediate stages of a problem, such spectrum prediction techniques provide powerful filters to separate plausible from implausible structures.

4   RESEARCH GOALS 1979-1980

We have several goals for CONGEN development and export during the next grant year. Because several were summarized previously, we give only a brief list here. The reprogramming effort will continue, first to incorporate aromaticity into the program and second to include capabilities for constraints interpretation and translation. These capabilities have been experimental efforts in the old, LISP version of CONGEN. Some additional development must be done and the BCPL version brought up to date with these features. Such developments will allow much more intuitive use of the program (see section on Conclusions from the Workshops) and greatly improve the ease with which structural problems can be specified to CONGEN.

We will be pushing very hard for further export of CONGEN either by developments here at Stanford or by assisting others at remote sites. We can now supply version for DEC-10 and DEC-20 under TOPS-10 or 20 or TENEX operating systems. The contract for an NIH/EPA CIS version will make CONGEN available on that system soon. Our Joint Study project will hopefully allow us to develop an IBM version at

minimal cost. A version for the CDC systems at the National Resource for Computation in Chemistry is under study. The whole area of mini-computer versions is under investigation. We expect that these efforts will make the current version widely available. However, it does leave open the question of long-term maintenance and, particularly, upgrading older versions as new developments ensue. These issues will be dealt with at length in our renewal proposal for funding subsequent to the coming year.

In further response to the workshop requests, we will be developing an extended "help" facility for the program which, together with improved documentation, will improve the ease with which new persons can become familiar with CONGEN. We have already begun investigations of how to improve the current structure drawing facilities in CONGEN, with our first goal to improve the teletype drawings so that all users can benefit. We will also be considering methods for aiding those with graphics terminals to exploit the improved structure input and output possible with such devices.

Further work on the STEREO program will be to develop further the constrained stereoisomer generator and to improve interaction with the main CONGEN program and the user. Specifically, constraints involving the chirality of structures, patterns of equivalent atoms based on either stereoisomer or topological symmetry, and undesirable structural features such as trans double bonds in small rings will be implemented. The flow of information within the program will become two-way between CONGEN and the stereoisomer generator. At present, this information only flows to the stereoisomer generator. The user interaction will be improved so the user can more easily visualize the stereoisomers and the stereochemistry of substructures and can input stereochemical information more intuitively.

We will emphasize several lines of investigation with regards to examination and evaluation of structural candidates from CONGEN, through the STRUCC program. Experiment planning will be approached from the current PLAN and LOOK commands, which search the environment of program- or user-specified structural features and give the experimentalist information on how the structural candidates vary with respect to those environments. The mass spectrum prediction and ranking functions will be completed and transferred to BCPL versions running as part of the CONGEN program. The concept of prediction and ranking will be extended to proton and carbon NMR data using, first, simple additivity rules and, later, refining the predictions based on more detailed examination of the structural relationships among the protons and carbons whose signals are being predicted. We hope that this approach will be effective in choosing a small subset of highly plausible structures based on agreement between predicted and observed spectra, just as the mass spectrum analysis functions have proven useful.

We plan no further development of the GC/MS/Computer system, or the REACT and MAXSUB programs. Rather, these will be used in applications to current structural problems in conjunction with compounds or data from other sources. Incremental improvements will be made as necessary if a particular new application demands it.

Appendix I. CONGEN Workshop Attendees, Affiliation, Research
Interests, Comments on Program and Export Status.

1) Dr. Henry Stoklosa, E.I. DuPont de Nemours.
Dr. Stoklosa has been affiliated with a group at DuPont
involved with computer applications to chemical
problems, including computer-aided organic synthesis.
He will soon be involved in another group which might
also be able to make use of the REACT program in
addition to his more general interest in CONGEN.

2) Dr. G.W.A. Milne, National Institutes of
Health. Dr. Milne is currently in charge of the
National Institutes of Health contribution to the
NIH/EPA Chemical Information System. His interests
included not only evaluation of the utility of the
program but also exploration of ways in which CONGEN
might be interfaced to the Chemical Information System.
This effort is described in more detail in Section 2.3.
Dr. Milne offered both praise and criticism. He
praised the program itself but wants good user-level
documentation and a large number of sample problems for
future workshops.

We are currently exploring with him the interest of
NIH in obtaining a version of the current program for
use on the NIH PDP-10 facility.

3) Dr. William Brugger, International Flavors and
Fragrances. Dr. Brugger represents the key person at
IFF Research responsible for computer applications in
their laboratories. Structure elucidation is a major
activity of this company not only in analysis of
natural and synthetic products but also in assessing
the relationships between chemical structure and toxic
properties affecting human health. Dr. Brugger has
access to both DEC VAX and IBM computers, the former
representing the laboratory computer. The status of a
version of CONGEN for these machines has been discussed
previously. Meanwhile Dr. Brugger and his colleagues
are evaluating CONGEN at SUMEX via the GUEST access
facility.

4) Dr. Douglas Dorman is head of the NMR
laboratory at .Lilly Research Laboratories and works
closely with mass spectroscopists and other chemists in
solving structures of a variety of compounds related to
existing or new products. Dr. Dorman has been familiar
with the "old" (non-exportable) version of CONGEN and

thus was able to critique the new program not only on its merits but also on comparison with the old version. His detailed critique is included in Appendix III. It is worthwhile to point out that we are in the process of implementing two of the important new features he mentions because of their general importance. The SURVEY and EXAMINE commands will allow both user-specified features used to explore structural possibilities and Boolean statements used to select structures with combinations of features. The MSRANK facility will perform the selection of plausible structures based on agreement of predicted and observed spectra. Dr. Dorman has access to a PDP-10 computer operating under the TOPS-10 operating system. About two weeks ago he received a copy of the exportable CONGEN and is now using the program in his own laboratory. He has agreed to provide us with continuing comments and criticism in exchange for receiving updated versions of the program.

5) Dr. Jon Clardy, Cornell University. Dr. Clardy is a recognized leader in development and applications of the technique of X-ray crystallography in structure elucidation. His attendance of the workshop was based on an interest in learning about alternative, computer-based approaches to the problem. As his letter points out, the ability to use CONGEN to help solve structures before expenditure of time and effort in X-ray analysis would be an important benefit. A more important outcome of the workshop for future research were discussions on the possibility of coupling CONGEN-suggested structures to Patterson search techniques. In principle, each of the candidates suggested by CONGEN could be used in turn to guide the search of the electron density maps for a fitting of the structure to the maps. The correct structure should yield the least ambiguous fit. Dr. Clardy will access CONGEN at SUMEX via the GUEST facility. His ability to use CONGEN at Cornell depends on the success of the efforts of Mr. In Ki Mun (next section).

6) Mr. In Ki Mun, Cornell University. Mr. Mun attended the workshop representing Prof. Fred McLafferty at Cornell. Prof. McLafferty's group has had for many years an interest in use of computer techniques to help solve structures, based primarily on mass spectral data. His research in this area has led to programs which suggest the presence of functionalities in an unknown molecule. CONGEN can, in

principle, complete such a schema for analysis by piecing together the inferred functionalities. Mr. Mun attended to explore the feasibility of this approach and to learn how best to use CONGEN on the IBM system at Cornell. He is currently evaluating the effort of modifying CONGEN for an IBM version of BCPL. If successful then this version should be available for other persons who have good interactive services available at their IBM installations.

7) Dr. Reimar Breuning, Munich. Dr. Breuning learned of the existence of the workshops from discussions with Prof. Djerassi at the IUPAC meeting on natural products. Dr. Breuning had the opportunity to attend as he was in the process of arranging a post-doctoral appointment with Prof. Nakanishi (see next section). He is actively involved in natural products structure elucidation at Munich and expects these interests to continue. Dr. Breuning can access CONGEN via Guest for the near future from Munich. At Columbia he will have the opportunity to take advantage of access to the facility there.

8) Dr. David Lynn, Columbia University. Dr. Lynn attended the workshop representing Prof. Koji Nakanishi, the latter a recognized expert in the area of structure elucidation of a number of classes of natural products of relevance to human health. Dr. Lynn is to act as the focal point for introduction of the computer methods to that research group. Prof. Nakanishi's group has access to a DEC-20 system operating under the TOPS-20 operating system. We are currently arranging for a version of CONGEN to be installed there. We anticipate no problems because of our successful experiment during the workshop of running CONGEN on a DEC-20 system at Rutgers.

9) Dr. Y. Gopichand, University of Oklahoma. Dr. Gopichand attended the workshop representing the marine natural products group of Prof. Francis Schmitz. This group specializes in structure elucidation of halogenated terpenoid molecules possessing a variety of biological activities and marine sterols representing intermediates or end products in steroid biosynthesis. As evidenced by the letter of critique, this group represents an excellent example of classical approaches to structure elucidation; sufficient data are collected such that the number of structural possibilities is reduced to a very small number. As the letter also

indicates, CONGEN should be useful at least to check the rigor of their structural assignment. What should prove more interesting is whether or not such a group, with little computer expertise, can use CONGEN earlier on in the process of structure elucidation to guide subsequent collection of data.

10) Ms. Wendy Harrison, University of Hawaii. Ms. Harrison attended the workshop representing the marine natural products group of Prof. Paul Scheuer at Hawaii. This group is engaged in structure elucidation problems similar to those encountered in Prof. Schmitz's laboratory, although focus is on different classes of organisms. Their letter of critique mentioned that due to absence of critical high resolution mass spectrometric data to establish molecular formulas, they have been unable to use CONGEN this past month to help on their problems. This situation is expected to be resolved soon. Prof. Scheuer plans to use CONGEN as an aid in a course in structure elucidation this coming semester. One difficulty is that this group only has access to a DEC PDP-11 system. This introduces all of the problems mentioned earlier on a version of the program for mini-computers, including proliferation of manufacturers and operating systems. For example, there is a version of BCPL for PDP-11 series machines, but only under the RT-11 operating system, whereas the Hawaii group runs under The RSX-11 operating system. For the near future, access from Hawaii will be to CONGEN at SUMEX running under the GUEST directory.

11) Dr. Laszlo Tokes and Dr. Michael Maddox, Syntex Research. Drs. Tokes and Maddox are, respectively, in charge of the mass spectrometry and NMR laboratories at Syntex Research. They are responsible for the majority of structure elucidation problems which rely on physical methods. Their interest in CONGEN is that it might help them solve certain problems in less time than required by manual methods. Syntex research has access only to laboratory mini-computers such as the DEC PDP-11 series machines. They, too, must await a smaller version of CONGEN suitable for mini-computers before being able to use the program in their own laboratory.

12) Dr. John Figueras, Kodak Research Laboratory. Dr. Figueras attended representing the Analytical Sciences Division of Kodak's Research Laboratory. This

division is responsible for data collection and analysis in support of the structure elucidation activities of the Laboratory including not only new developments in the photographic process but also the new technology of thin-film bound enzymes systems for clinical analyses. His role was to evaluate what part CONGEN could play in the on-going automation of of the Division. The Division will be obtaining a DEC-20 system in the near future on which CONGEN will run directly. Meanwhile we have offered GUEST access in return for continuing critique on utility of CONGEN for large, poly-heteroatomic, aromatic molecules of the types encountered in their research.

13) Dr. Charles Snelling, University of Illinois. Dr. Snelling attended the workshop representing Prof. Kenneth Rinehart in the Chemistry Dept. at Illinois. Prof. Rinehart is also an acknowledged expert in structure elucidation with emphasis on macrolide antibiotics, halogenated terpenoids and other classes of natural and synthetic products of relevance to human health problems. Dr. Snelling had several comments and criticisms about the difficulties facing a novice user of such a complicated system. Although he found the program useful and wishes to get it running at Illinois, he would like to see some changes made to simplify use of the program for the chemist. There are many computer systems available at Illinois and choice of which system on which to mount CONGEN will depend on the ease of the task and access to a particular system. This is currently under study by them.

14) Dr. Gilles Moreau, Roussel UCLAF. Dr. Moreau attended the workshop representing the French pharmaceutical concern Roussel UCLAF. This company maintains an active group in computer applications in chemistry and wished to evaluate CONGEN for its use in their structural problems. The company is quite interested and will explore use of the program via GUEST access. They are concerned about issues of secrecy for new problems and we have made it quite clear that GUEST access represents public knowledge of their problems. Therefore, assuming their interest continues, we will be arranging some alternative to SUMEX for use of CONGEN. They have access to IBM equipment and we are awaiting further description of the form of access possible.

15) Prof. Andre Dreiding, Zurich. Dr. Dreiding

has been interested in both the problem-solving and the pedagogical aspects of CONGEN for some time. He had previously used the old version and was gratified to see the improvements in the new version. He would like to see much more attention paid to the actual structures (i.e., in three dimensions) of molecules rather than simply their constitutions as CONGEN, with the exception of the STEREO command, currently represents structures. We are, of course, working very hard to introduce concepts of stereochemistry into our computational procedures. Prof. Dreiding will probably be able to access CONGEN at Zurich on an existing PDP-10 installation. If so, export to his group will be trivial.

16) Dr. James Shoolery and Dr. Michael Gross, Varian Associates. Dr. Shoolery is in charge of Varian's NMR application laboratory and Dr. Gross is in charge of computer software for Varian's NMR/computer systems. Their respective interests match their responsibilities. Dr. Shoolery feels that CONGEN could be a valuable assistant in helping solve structures based primarily on proton and carbon NMR data. In fact, he has been able to demonstrate such an approach on some recent problems of persons outside Varian. Dr. Gross is interested in a mini-computer version of CONGEN for incorporation into their existing data system. Although we cannot ourselves support such an effort, we have agreed to provide them with program listings and documentation to explore the feasibility of a small machine version.

17) Dr. Daniel F. Chodosh, Smith, Kline and French. Dr. Chodosh was not actually invited to the workshop, but happened to visit our group during one of the sessions. He was sufficiently impressed that he procured a tape to carry away a copy of the program with him. He has now been supplied with a version of CONGEN for the PDP-10 and as of last week has it running at the SKF research laboratories in Philadelphia. This is an interesting experiment because as an informal attendee of a small portion of the workshop, he is learning the program almost from scratch based on existing help facilities and documentation. He will begin introducing others to the program when he has developed sufficient familiarity to be comfortable with the program.

In addition, the following persons during the past year have asked

for information about and access to CONGEN. For the most part we have granted access through the GUEST directory, setting up an account only for those users with more than occasional log-ins.

Dr. David Cowburn
Physical Biochemistry
The Rockefeller University
New York City

We have sent Dr. Cowburn information on access to CONGEN and are currently discussing how to use some of our computational methods or extensions to them for assistance in his problems of elucidating peptide conformations.

Douglas Henry
School of Pharmacy
Oregon State University
Corvallis, Oregon

He has been sent our programs for structure drawing for use on his own computer.

The following have asked for and received information on access to CONGEN at SUMEX via the GUEST facility.

Dr. H. Kating
Institut fur Pharmazeutische Biologie
Der Universitat
Bonn, Germany

Dr. Kerber
Lehrstuhl D fur Mathematik
Aachen, Germany

Dr. Brenda J. Kimble
Radiobiology Laboratory
University of California
Davis, California

Dr. J. Neubuser
Lehrstuhl D fur Mathematic
Aachen, Germany

Dr. George Padilla
Dept. of Physiology
Duke University Medical Center
Durham, N.C.

Dr. W. Sieber
Sandoz Ltd.

Basel, Switzerland

Dr. Babu Venkataraghavan
Lederle Laboratories
Pearl River, New York

We also helped him bring up the Fortran draw program on the DEC-10
system at Lederle

Dr. Stephen Wilson
Dept. of Chemistry
Indiana University
Bloomington, Indiana

Appendix II. Sample Letter of Invitation Sent
To Prospective Workshop Attendees.

STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

DEPARTMENT OF CHEMISTRY

August 24, 1978

Professor Kenneth Rinehart
Department of Chemical Science
University of Illinois
Urbana, Illinois 61801

Dear Professor Rinehart:

I am writing to determine your interest in a mini-workshop we plan to hold at Stanford on use of computer programs for computer-assisted structure elucidation. Over the past three years we have been involved in a research effort directed in part to exploring the feasibility and utility of interactive computer programs as tools to help chemists solve unknown structures. The most highly developed of these research tools is the CONGEN program, and I know that you have in one way or another been exposed to this program.

We have made CONGEN available on the SUMEX computer here at Stanford via communications networks to a selected group of chemists because we know that the only way to improve the program and to make it useful is to apply it to a variety of real structural problems in chemistry. We have learned a great deal from this experience and have designed a production version of CONGEN with the primary goal of eliminating the deficiencies in the research version. In particular the new program is much simpler to use, much smaller and faster and can be exported to certain other computers.

There is a significant amount of work which remains to be done to make the exportable version of CONGEN a truly useful chemist's "assistant". Some of this work is underway now. But before investing a great deal of time and effort in polishing a new version to make it more useful and acceptable, we would like to expose a group of several chemists experienced in structure elucidation to this version of the program and base future research and development efforts on their perceptions of deficiencies remaining in the program and its ultimate utility. We feel its usefulness can be significant in terms of exploring alternative structural possibilities and in guaranteeing no plausible alternative has been overlooked. However, in order to demonstrate that utility we need assistance in developing a version which large numbers of chemists can use easily and productively in their own laboratories and on their own computers. We feel that by participating in this workshop you can contribute in an important way.

We plan an approximately four to five day informal workshop here at Stanford, to be attended by you or one of your experienced co-workers. During this time you would be able to use the exportable version of CONGEN here at SUMEX to work on recent structural problems encountered in your laboratory, which we hope will include some which would be

140

unknowns. In turn, we would be learning how to finish development of the first version of the program for export to your laboratory with some guarantee that major problems or deficiencies would be eliminated based on the experience of the workshop. We have not yet established a time for the workshop, but we are tentatively considering a date between October and December, 1978. We are also trying to arrange for travel support for non-industrial persons. We are seeking now only an expression of interest on your part.

There are only two requirements on your part, other than your expressed interest in participating. The first requirement is that you must be capable of accessing CONGEN at SUMEX or preferably be capable of running CONGEN on your own computer system. The second requirement is that you actually be interested in using the program in the future (assuming the workshop was worthwhile) to provide us with some feedback on whether improvements meet your criteria for acceptance.

Currently the new version of CONGEN can be used on Digital Equipment Corp. PDP-10's and 20's with little problem. We are now exploring use of IBM 360 and 370 series machines. We have already demonstrated that small segments of the program run on suitably configured IBM computers. However, for CONGEN to be used interactively requires a time-sharing operating system. If you have access to an IBM computer but are uncertain about the operating system, find out from your computer systems people the name of the operating system and call or write to us with that information.

While at Stanford you would also be able to take a close look at some of our other research efforts which are not quite so far along as CONGEN, including a recently finished stereochemical structure generator, the REACT program and a variety of tools for ranking structures based on comparison of predicted and observed spectra, facile examination of large sets of structural candidates for common and unique features and methods for assistance in experiment planning. We would appreciate your evaluation of these efforts also. The main emphasis, however, will be on CONGEN.

Obviously we do not expect a firm commitment for attendance at this time, but if at all possible I would appreciate a reply. Indications of your interest are important in this regard.

Yours sincerely,

Carl Djerassi
Professor of Chemistry