Research Proposal Submitted to the National Science Foundation

Proposed Amount: $60,132 (Direct Cost)

Proposed Effective Date: June 1, 1974

Proposed Duration: 24 Months

88,394 (Including Indirect Cost)

Title: Carbon-13 Nuclear Magnetic Resonance of Steroids

Principal Investigator: Carl Djerassi
Professor of Chemistry

Submitting Institution:
Stanford University
Department of Chemistry
Stanford, California 94305

Make grant to Stanford University

Endorsements:

<table>
<thead>
<tr>
<th>Name</th>
<th>Principal Investigator</th>
<th>Department Vice-Chairman</th>
<th>Institutional Admin. Official</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signature</td>
<td>Carl Djerassi</td>
<td>D. A. Skoog</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Professor of Chem.</td>
<td>Vice-Chairman</td>
<td></td>
</tr>
<tr>
<td>Telephone</td>
<td>415-321-2300 X2783</td>
<td>415-321-2300 X2502</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>Dec. 17, 1973</td>
<td>Dec. 17, 1973</td>
<td></td>
</tr>
</tbody>
</table>
ABSTRACT

Carbon-13 nuclear magnetic resonance (CMR) spectra show a marked sensitivity to such important features of chemical structure as carbon hybridization, electronegativity of heteroatoms, branching and steric crowding. Potentially, CMR is an extremely rich source of structural data in organic chemistry, capable of rivalling or even surpassing proton magnetic resonance. In the research proposed here, we intend to develop this potential in the field of steroid chemistry.

The factors which determine the CMR spectra of steroids are only modestly well understood. We have begun, and propose here to continue, a systematic study of families of closely-related steroids (keto- and hydroxy-androstanes and cholestanes) with the conviction that only through such a systematic study can the basic factors governing the CMR spectra of steroids be brought to light. We intend to quantify those factors as predictive rules which relate spectra to structures, and to develop computerized methods for using those rules to extract structural information from the CMR spectra of unknown steroids. We also propose to develop several chemical methods (derivatization procedures) for augmenting the information-content of such spectra.
I. BACKGROUND

Within the last twenty years, the phenomenon of nuclear magnetic resonance\(^1,2\) (NMR) has evolved from little more than a laboratory curiosity to one of the most powerful analytical tools in chemistry. The experiment itself consists of observing, in an applied magnetic field, the resonance frequencies (in the radio-frequency range) of magnetic nuclei in a liquid chemical sample. The analysis of NMR spectra yields chemical shifts and coupling constants which reflect, respectively, the chemical environments of and the bonding or spatial relationships between atoms whose nuclei are magnetic.

Because protons are magnetic, interacting particularly strongly with electromagnetic fields, and because they are present in virtually all organic compounds, proton NMR (PMR) has found broad usefulness in organic chemistry. The literature on PMR spectroscopy is huge, and frequently it is found that PMR spectra yield chemical information which would be difficult, if not impossible, to obtain by any other method. The determination of structure and conformation\(^3,4\), the analysis of mixtures\(^4,5\), the study of rate processes\(^6\) and the elucidation of reaction mechanisms\(^7\) have all been aided substantially by PMR techniques.

Other nuclei frequently observed via NMR are F-19, P-31 and C-13. The first two are not common in organic compounds and are thus used for more specialized studies. Carbon, by definition, occurs in all organic molecules, but only about 1% of all carbon nuclei are the magnetic isotope C-13. This, together with the fact that C-13 nuclei are almost two orders of magnitude less sensitive than protons to the NMR
experiment, has hampered the widespread use of C-13 NMR (CMR) as an analytical tool for organic chemists. However, recent instrumental advances such as pulsed Fourier-transform techniques and noise-modulated proton decoupling have made it possible to obtain natural-abundance CMR spectra of even large molecules (e.g., steroids) or low-concentration (ca. 0.05 M) samples within a reasonably short time (0.5 - 10 hr.).

The research to date indicates that C-13 chemical shifts (which constitute the primary data usually collected in the CMR experiment) cover a broad range (ca. 200 p.p.m.) relative to H-1 shifts (ca. 10 p.p.m.), and are highly sensitive to hybridization, electronegativity of substituents, branching and steric crowding. Thus CMR spectroscopy is potentially a rich and highly useful source of structural data. As further advances in instrumental design take place, CMR spectroscopy will become an increasingly available and informative tool in organic chemistry.
II. OBJECTIVES AND SIGNIFICANCE

The research proposed herein is directed toward understanding the factors which determine C-13 chemical shifts in steroids and toward developing computer-based methods whereby a chemist can obtain structural information from spectra of unknown steroids. This class of compounds was chosen for two reasons. First, the steroid skeleton is more or less rigid, providing a relatively controlled framework within which to study the effects of steric hindrance and other geometrical factors upon C-13 shifts. Second, a great fraction of steroid and natural products chemistry involves the identification or verification of steroid structures, and thus we expect our results to be of significant practical utility.

At the current level of understanding of C-13 chemical shifts, it is not now possible to predict the CMR spectrum of a given steroid with much certainty, although by referring to simple model systems, one can often rationalize the signs and general magnitudes of the spectral changes which take place when the nature and position of substituents are altered. It is now a challenging problem simply to assign the spectrum of a known steroid, that is to identify which observed peaks belong to which carbons.

The first definitive study of the CMR of steroids was presented only four years ago by Roberts et al.\textsuperscript{13} In that report, the assigned spectra of nearly thirty assorted steroids are presented, with the assignment task accomplished "...using specific single-frequency and off-resonance proton decoupling, hydroxyl acetylation effects on chemical shifts, deuteration, and substituent influences in analogous
compounds..." Since then, several other authors\textsuperscript{14} have reported research on the CMR of steroids, but only recently has the systematic study\textsuperscript{15,16} of families of closely related steroids begun. We have reported\textsuperscript{16} the assigned spectra of fourteen keto-substituted androstanes and cholestanes, where the keto group occupies every possible position around the skeleton. We are currently collecting data on a similar series of hydroxy-substituted steroids, and work is also in progress\textsuperscript{17} on the series of steroids containing one endocyclic double-bond. The initial stage of our proposed research is to complete the hydroxyl series and to obtain the CMR spectra of several bifunctional (e.g., dihydroxy and keto-hydroxy) steroids. With such a collection of systematic data available, we will be able to study the influence upon C-13 shifts of these three types of functionality, alone and in combination, which are by far the most common types occurring in natural steroids. From these data, we expect to be able to extract rules which will allow the accurate prediction of CMR spectra of steroids containing these groups. Zürcher\textsuperscript{18–20} has derived an extremely useful set of rules relating skeletal substitutions in steroids to changes in the PMR shifts of protons in angular methyl groups. The C-13 rules we seek will relate not only to angular methyl groups but to all carbons in the skeleton, and will thus provide a great deal more information than the Zürcher rules. These CMR rules will also form the base for our proposed work in the computerized interpretation of CMR data (\textit{vide infra}).

As a second portion of our work, we propose a study of reversible derivatization procedures which will aid in the assignment of the spectra of known steroids, and in the analysis of the spectra of unknown
steroids. Roberts and co-workers have found that acetylation of a hydroxyl group in a steroid produces characteristic changes, due primarily to steric effects, in the shifts of carbons close to that group. They have used this effect in assigning such shifts. We propose to study the effects of other hydroxyl-group derivatives, specifically, the 2,4,6-trimethylbenzoate (benzoates themselves, in our hands, have not shown any advantages over acetates), trifluoroacetate and the trimethylsilyl ether. These derivatives have different steric and electronic properties than do acetates, and should thus produce different patterns of spectral change, providing a convenient means of augmenting the information-content of ordinary CMR spectra. We also propose to observe the C-13 shift changes which take place upon cyclic ketalization of carbonyl groups, and we expect that such changes will be useful in determining the local environment of keto groups in steroids. It has been found that the presence of a paramagnetic complexing agent (e.g., a "lanthanide shift" reagent) causes large changes in C-13 chemical shifts of alcohols and ketones. These changes can be related to the geometry of the complex, which reflects the geometry of the alcohol or ketone itself. We intend to explore the use of such shift-reagents in assisting the interpretation of steroidal CMR spectra. Of particular interest will be the difference between α- and β-hydroxy steroids: It is expected that the grossly different steric environments of axial vs. equatorial hydroxyl groups will have a pronounced effect upon the geometry of the complex, and thus, very different lanthanide-shift patterns should result. If so, the effect should provide a convenient means for distinguishing the stereochemistry of sterols.
A third facet of the proposed research involves the development of computerized techniques for automatically extracting structural information from CMR spectra. This represents a logical growth of our Heuristic DENDRAL project, an eight-year joint effort between our laboratories and the Departments of Genetics and Computer Science. The purpose of the project is to develop applications of heuristic programming ("artificial intelligence") to problems in chemical inference, with the bulk of the effort directed toward the computerized interpretation of mass-spectroscopic (MS) data. In the early DENDRAL research, only saturated, acyclic, monofunctional compounds were treated, but we have recently reported the successful identification of the structures of estrogenic steroids (and mixtures thereof) via the computerized interpretation of MS data. As the complexity of compound classes has increased, we have felt a growing need for sources of structural data other than MS. CMR data show a sensitivity to stereochemistry and substituent placement which complements, rather than duplicates, MS-derived information, and thus CMR is the ideal candidate.

We have demonstrated the feasibility of using CMR data in automated structure analysis. Using a detailed and accurate set of predictive rules for saturated, acyclic amines, we have constructed a computer program which can "reason out" the structure of such an amine, starting from its empirical formula and CMR spectrum. A similar effort is proposed for the steroids (at least, those containing endocyclic double bonds, carbonyl groups and hydroxyl substituents) in which structural information would be inferred from CMR data using accurate predictive rules. This information could then be integrated with the results obtained from derivatization or special CMR techniques,
and (if necessary) from MS analysis to yield possible structures. Not only would such a system have substantial utility, but it would represent an important advance in the "state of the art" in both CMR spectroscopy and chemical information-processing.
III. PLAN OF RESEARCH

A. CMR Spectra of Steroids.

We plan to complete the series of sterols by synthesizing \(1\beta-, 2\alpha-, 4\beta-, 7\alpha-, 9\alpha-, 11\alpha-, 14\alpha-, 16\alpha-\) and \(17\alpha\)-androstanols or cholestanols, whose CMR spectra (pulsed Fourier-transform spectra, obtained at 25 Mhz. using noise-modulated proton decoupling) will be recorded and assigned. We have worked out likely synthetic pathways for the preparation of these using commonly accepted procedures and starting from compounds available in our laboratories. In order to test the extent of additivity relationships and of various interactions of substituents, we shall similarly synthesize and record the spectra of two or three dozen dihydroxy and keto-hydroxy androstanes and cholestanes. The candidates chosen will depend upon the results of the analysis of the monofunctional steroids.

Using statistical procedures similar to those of Dalling and Grant, and of Lindeman and Adams, we shall attempt to correlate structural variables with chemical shifts, the goal being the derivation of an accurate set of substituent parameters for steroids. In assessing the effects of steric crowding and skeletal distortion, we plan to utilize a computerized, classical-mechanical model of molecular structure, such as the Westheimer-type models recently reviewed by Schleyer.

B. Derivatization
We propose to analyze the changes in C-13 shifts which take place when the hydroxyl group in several of the above androstanols and cholestanols is derivatized to the 2,4,6-trimethylbenzoate, trifluoroacetate and trimethylsilyl ether. We propose similarly to investigate the effects of ketalizing (with ethylene glycol) several androstanones and cholestanones.

We propose to analyze the effects of lanthanide shift reagents (in varying concentrations) upon the CMR spectra of several of the hydroxy- and keto-steroids, with particular emphasis upon pairs of sterols which differ only in the orientation of the hydroxyl group.

These investigations will be directed toward the development of a repertoire of non-destructive, chemical methods for increasing the information available from CMR spectra.

C. Computerized Interpretation of CMR Data.

There are three phases to our proposed research in this area, all of which will make use of the heuristic programming techniques developed in our DENDRAL project.

First, we intend to develop a program to assist in the assignment of spectra obtained in parts A and B, using currently available techniques (i.e., using rules for acyclic systems together with analogies from appropriate model systems). The purpose here is twofold: On one hand, such a program will hasten a time-consuming procedure (in our work, the assignment of spectra requires about as much time as the preparation of samples and the recording of spectra, combined), while on the other hand, it will provide a context within which to develop
techniques applicable to the more difficult problem of structure identification. Specifically, we will need methods for expressing CMR rules as efficient computer code, and for deciding whether a good, unambiguous fit occurs between predicted and observed data.

Secondly, we intend to utilize the rules derived in part A, together with derivatization information from B, to write what is called a "planning" program in the DENDRAL terminology. Such a program is designed to examine the spectrum of an unknown and, referring to a set of heuristics, to attempt to verify the presence or absence specific structural features in the unknown. Whereas the predictive rules allow one to predict a spectrum from a given structure, the heuristics represent transformations of the rules which allow one to infer structural information from a given spectrum. The primary challenge in constructing the planning program will be the design of heuristics which are as informative as possible, yet which run efficiently. This program will be a useful analytic tool in itself, and will be used in the third phase of our proposed research.

This third phase will involve merging the planning program with the existing DENDRAL system, which analyzes MS data for steroids. Modifications will be made to the structure-generation program, which can construct all possible sets of acyclic substituents from a given set of atoms and attach those substituents in all possible ways to a given cyclic skeleton. The structure generator now makes use of MS-planner information, constructing only those steroids which are consistent with it. We shall modify the algorithm to make use of the output from both the MS and the CMR planners, and shall extend the algorithm to consider questions of stereochemistry, which are currently ignored. We believe
that the augmented DENDRAL system will have the capacity to identify, unambiguously, the structures of a wide variety of steroids using information from just these two spectroscopic sources.

The programs will be written in the LISP language, and will thus be compatible with the rest of the DENDRAL system. Computer time on the PDP-10 will be provided through the NIH-funded SUMEX facility at Stanford, and we request no support in this proposal for computer facilities. Programs developed in our proposed research will be available to the scientific community over the ARPA computer network.
IV. REFERENCES


17. Work in progress at the Orsted Institute in Copenhagen, Denmark by H. Eggert.


V. FACILITIES AVAILABLE

1. Instrumentation.

(a) Varian XL-100 NMR spectrometer. Departmental instrument, equipped for Fourier-transform CMR spectroscopy with or without proton decoupling (single-frequency or noise-modulated). This is among the most sensitive CMR instruments currently available.

(b) Computer Facilities. Installation of a PDP-10 computer, funded by the National Institutes of Health as part of the SUMEX resource at Stanford (a national resource devoted to the application of artificial intelligence in medicine and related fields), is scheduled for March, 1974. Computer facilities will be available to us for the computer-related aspects of the project proposed here (primarily, part III, section C).

2. Personnel associated with the project.

(a) Principal Investigator.

Academic Experience: Professor of Chemistry, Stanford University, 1959-present; Associate Professor (1952-1954) and Professor (1954-1959), Wayne State University.


Publications: Author or co-author of six books and approximately 800 publications dealing with natural products (notably steroids, terpenoids, alkaloids and antibiotics), medicinal chemistry (primarily antihistamines, oral contraceptives and anti-inflammatory agents) and applications of physical methods (mass spectrometry, optical rotatory dispersion, magnetic circular dichroism) to organic and biochemical problems.

Postdoctoral Research Associate.
Raymond Edgar Carhart, born 1946; Northwestern University, A.B. (with honors), 1968; California Institute of Technology, Ph.D. (Chem.), 1972.
Honors: Member of Phi Beta Kappa, Sigma Xi, Phi Lambda Upsilon; NIH postdoctoral Fellow, 1972-present; NSF predoctoral Fellow, 1968-1972.
Experience: Dr. Carhart has a background in Physical Organic Chemistry, with particular emphasis upon computer applications. As an undergraduate and graduate student, he has had direct research experience in organic synthesis, X-ray crystallography, quantum chemistry and PMR and CMR spectroscopy. As an NIH Fellow in our Computer Science Department, he has worked for the past year developing applications of artificial intelligence in CMR spectroscopy. He is currently analyzing the CMR data from the series of keto-substituted androstanes and cholestanes, and is also involved in some aspects of the DENDRAL program for the generation of cyclic chemical structures.

Publications:


VI. CURRENT SUPPORT AND PENDING APPLICATIONS

**Principal Investigator:**

<table>
<thead>
<tr>
<th>Field of Research</th>
<th>Source of Support</th>
<th>Annual Amount</th>
<th>Termination Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organic Mass Spectrometry</td>
<td>NIH Grant No. 5 R01</td>
<td>$52,306</td>
<td>9/30/75</td>
</tr>
<tr>
<td></td>
<td>AM 04257-13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marine Natural Products</td>
<td>NIH Grant No. 2 R01</td>
<td>75,650</td>
<td>12/31/77</td>
</tr>
<tr>
<td></td>
<td>GM AM 06840-15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pending:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnetic Circular Dichroism</td>
<td>NSF Appl. # P3P3689</td>
<td>27,640</td>
<td></td>
</tr>
<tr>
<td>in Organic Molecules</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Postdoctoral Research Associate:**

Current:

Dr. Carhart now holds an NIH postdoctoral Fellowship (# 1 F02 GM54009-01 BPS) in the amount of $7500 per year, to terminate on 9/17/74.

This Application is not being submitted to any other Agency.
VII. RESEARCH GRANT PROPOSAL BUDGET

First year beginning 6/1/74
Second year beginning 6/1/75

<table>
<thead>
<tr>
<th></th>
<th>First year</th>
<th>Second year</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. SALARIES AND WAGES:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Principal Investigator; Carl Djerassi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Postdoctoral Research Associate, Computer Science Department; Raymond E. Carhart (12 calendar man-months)</td>
<td>$14,000</td>
<td>$14,700</td>
</tr>
<tr>
<td>3. Graduate Student, Chemistry Department; Craig van Antwerp (6 calendar man-months)</td>
<td>3,180</td>
<td>3,239</td>
</tr>
<tr>
<td>TOTAL SALARIES AND WAGES:</td>
<td>17,180</td>
<td>17,939</td>
</tr>
<tr>
<td>B. STAFF BENEFITS:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17% of sal. &amp; wages 6/1/74-8/31/74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18% of sal. &amp; wages 9/1/74-5/31/75</td>
<td>3,049</td>
<td></td>
</tr>
<tr>
<td>calculated at 17.75%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18% of sal. &amp; wages 6/1/75-8/31/75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19% of sal. &amp; wages 9/1/75-5/31/76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>calculated at 10.75%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. TOTAL SALARIES, WAGES AND STAFF BENEFITS (A + B):</td>
<td>20,229</td>
<td>21,303</td>
</tr>
<tr>
<td>D. PERMANENT EQUIPMENT:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. EXPENDABLE SUPPLIES AND EQUIPMENT:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemical supplies and equipment</td>
<td>2,000</td>
<td>2,000</td>
</tr>
<tr>
<td>F. TRAVEL:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domestic, for attendance at scientific meetings</td>
<td>800</td>
<td>800</td>
</tr>
<tr>
<td>G. PUBLICATION COSTS:</td>
<td>1,500</td>
<td>1,500</td>
</tr>
<tr>
<td>H. COMPUTER COSTS:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Provided through the SUMEX facility at Stanford. No support is requested here)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I. OTHER COSTS:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hourly rental of departmental CMR instrument (Varian XL-100) 400 hr. @ $12.50/hr. (sufficient for approximately 100 CMR spectra each year)</td>
<td>5,000</td>
<td>5,000</td>
</tr>
<tr>
<td>J. TOTAL DIRECT COSTS:</td>
<td>29,529</td>
<td>30,603</td>
</tr>
<tr>
<td>K. INDIRECT COSTS:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>On Campus 47% of total direct costs</td>
<td>13,879</td>
<td>14,383</td>
</tr>
<tr>
<td>L. TOTAL COSTS (J + K):</td>
<td>43,408</td>
<td>44,986</td>
</tr>
<tr>
<td>M. TOTAL CONTRIBUTIONS FROM OTHER SOURCES:</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>N. TOTAL ESTIMATED PROJECT COST:</td>
<td>43,408</td>
<td>44,986</td>
</tr>
</tbody>
</table>

TOTAL COSTS FOR TWO YEARS: $88,394
VIII. COGNIZANT PERSONNEL

Principal Investigator (For Technical and Scientific Matters):

Carl Djerassi
Professor of Chemistry
Department of Chemistry
Stanford University
Stanford, California 94305

Telephone: (415) 321-2300, Extension 2783

For Administrative Matters (Budgets, Property, etc.):

P. M. West, Administrative Officer
Department of Chemistry
Stanford University
Stanford California 94305

Telephone: (415) 321-2300, Extension 2513

For Contractual Matters (Including Overhead and Patent Questions):

Sponsored Projects Office
Old Pavilion
Stanford University
Stanford, California 94305

Telephone: (415) 321-2300, Extension 2883