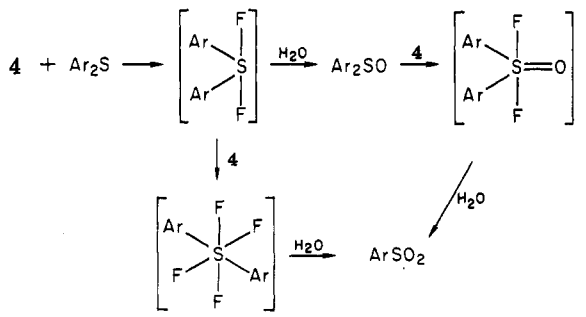
Scheme IV [a]



15

[a] Ar = *p*-chlorophenyl.

Reactions. Attempts to replace the fluorine ligands of **13** with chlorine, by treatment with 3 equiv of trimethylchlorosilane, failed to produce an isolable trichloroperiodinane.[8a] Instead, the reduced product, a chloroiodinane, was formed. Since the tridentate ligand stabilizes **4** so effectively, it was thought that the isolation of dichloroperiodinane **14** (Scheme III) might be possible. When **4** is treated with 2 equiv of trimethylchlorosilane at room temperature, however, iodinane **12**, trimethylfluorosilane, and chlorine are observed. This reaction may involve the intermediacy of **14** but did not provide direct evidence for the latter's existence in the reaction mixture.

A solution of **4** and *p*-chlorophenyl sulfide in acetonitrile does not react at room temperature. But when a mixture of solid **4** and the sulfide is heated at 135 °C, a mixture of **12**, sulfoxide, and sulfone was observed by ^1H and ^{19}F NMR spectroscopy. The presence of the sulfoxide suggests the intermediacy of a difluorosulfurane, which is hydrolyzed to give the sulfoxide. The sulfone could result either from further oxidation of the sulfoxide by **4** or by the formation of a tetrafluoropersulfurane, **15**, whcih could be hydrolyzed to sulfone (Scheme IV).

Summary

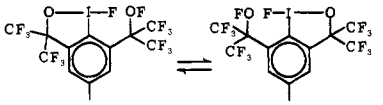
The synthesis of 4-methyl-2,6-bis[1-hydroxy-1-(trifluoromethyl)-2,2,2-trifluoroethyl]aniline (**2**) in several steps from 4-methyl-2-[1-hydroxy-1-(trifluoromethyl)-2,2,2-trifluoroethyl)aniline (**5**) provides a necessary precursor for the incorporation of a stabilizing tridentate ligand in compounds of hypervalent nonmetallic elements. One example of the usefulness of this ligand is the isolation of the exceptionally stable aryldialkoxydifluoroperiodinane **4**. It is prepared in quantitative yield by treatment of 4-methyl-2,6-bis[-1-hydroxy-1-(trifluoromethyl)-2,2,2-trifluoroethyl]iodobenzene (**10**) with 2 equiv of bromine trifluoride in Freon 113 at −20 °C. Periodinane **4** is stable for an indefinite period at room temperature and is unreactive toward atmospheric moisture. It is reduced to iodinane **12** by trimethylchlorosilane and by aqueous hydrogen chloride. It oxidizes *p*-chlorophenyl sulfide in the presence of atmospheric moisture to give a mixture of the corresponding sulfoxide and sulfone.

Further studies of the reactions and synthetic utility of periodinane **4** and applications of the tridentate ligand to other nonmetallic elements are currently underway in our laboratory.

Acknowledgment. This research was supported in part by a grant from the National Science Foundation (CHE-7905692). Mass spectra were obtained on instruments provided in part through a grant from the National Institutes of Health.

Registry No. **2**, 76220-90-1; **4**, 80360-39-0; **5**, 1992-07-0; **6**, 80360-40-3; **6**·HBr, 80360-41-4; **7**, 76220-89-8; **8**, 80375-25-3; **9**, 80360-42-5; **10**, 80360-43-6; **11**, 80360-44-7; **12**, 80360-45-8; acetone, 67-64-1; hexafluoroacetone, 684-16-2; bromine trifluoride, 7787-71-5; trimethylchlorosilane, 75-77-4; trimethylfluorosilane, 420-56-4; hydrogen chloride, 7647-01-0; *p*-chlorophenyl sulfide, 5181-10-2; *p*-chlorophenyl sulfoxide, 3085-42-5; *p*-chlorophenyl sulfone, 80-07-9.

(15) A referee has suggested that the NMR spectrum of **4** is consistent with its being a rapidly equilibrating mixture of equivalent 10-I-2 species as pictured below.



It is true that the expected averaged chemical shift for such an interconverting mixture, the average for the hypofluorite (ca. 140 ppm)[16] and the fluorine of a 10-I-3 fluoroiodinane (ca. −160 ppm),[8] −10 ppm, is very near the value of −17.7 ppm observed for **4**. The 338.7-MHz ^{19}F spectrum of **4**, taken at −35 °C, shows a relativley sharp peak for the I-F fluorines. Since the expected peak separation for the hydrofluorite and fluoroiodinane ^{19}F peaks would be about 10^5 Hz in the slow-exchange domain, this observation shows that if such a rapid equilibration is occurring it must occur with an improbably low activation barrier, i.e., considerably less than 1 kcal/mol. Since the synthesis of periodinane **13** was effected by the fluorination of the corresponding fluoroiodinane with a hypofluorite, CFOF,[8] one would expect the very similar functional groups in the above-pictured interconverting mixture to react intramolecularly to give **4**.

(16) Dungan, C. H.; Van Wazer, J. R. "Compilation of Reported F^{19} NMR Chemical Shifts"; Wiley-Interscience: New York, 1970.

Applications of Artificial Intelligence for Chemical Inference. 40.[1] Computerized Approach to the Verification of Carbon-13 Nuclear Magnetic Resonance Spectral Assignments

Matthew R. Lindley, Neil A. B. Gray, Dennis H. Smith,* and Carl Djerassi

Department of Chemistry, Stanford University, Stanford, California 94305

Received August 28, 1981

A computerized system for aiding chemists in the assignment of ^{13}C NMR spectra is described. This system checks shift assignments for consistency with known data and for internal consistencies within a set of related compounds. The analysis relies on a data base of substructural environments and correlated chemical shifts. The utility of the system is illustrated through the examination of the assignments of chemical shifts published for germacranolide sesquiterpenes.

Computer programs for the interpretation and prediction of ^{13}C NMR spectra have recently been presented.[2,3] Both programs rely upon the consistency of correlations among chemical shifts and substructural environments.

0022-3263/82/1947-1027$01.25/0

The initial step in the spectrum interpretation procedure consists of identifying standard reference substructures associated with chemical shifts similar to those of resonances in the ^{13}C spectrum of an unknown compound. The prediction process, as applied to a particular structure, derives shifts for its constituent carbon atoms on the basis of analogies among their substructural environments and reference substructures.

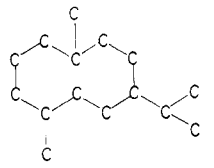
These ^{13}C analysis programs utilize a data base which contains substructures defining the molecular environments for resonating nuclei together with their associated chemical shifts. These substructural representations incorporate both the molecular constitution and configurational stereochemistry out to a four-bond radius (i.e., inclusion of δ-substituent effects). The data base has been constructed from information on known reference compounds and, at present, contains about 22 000 distinct substructure/shift combinations.

As far as possible, the initial entries in the data base used only published shift assignments based on detailed experimental data. These included structures where ^{13}C labeling and ^{13}C–^{13}C couplings had been utilized, examples where specific heteronuclear decoupling had been employed, or where a study, using shift reagents, had been made of the shifts in a set of related compounds. Many ^{13}C spectra, however, are not assigned on the basis of direct experimental evidence; instead, chemical shift trends observed in analogous compounds are utilized to derive a plausible set of assignments. As is well-known, it is quite possible to obtain a set of mutually consistent assignments for the resonances in a small set of compounds and yet have some of the assignments incorrect. As increasing numbers of spectra, whose resonance assignments were based primarily on shift-trend analysis, were entered into the data base, errors due to misassignments became apparent. Unfortunately, erroneous information in the data base can seriously degrade the performance of both the ^{13}C interpretation and prediction processes.

From the outset, the data base system incorporated some aids for checking new spectral assignments for consistency with existing information in the data base.[3] These primitive checking processes have now been expanded into a fairly comprehensive, automated approach for checking consistencies and identifying relevant reference structures in cases where discrepancies are apparent. Further, new programs have been developed to provide a mutually consistent set of resonance assignments for a new series of compounds for which there are no analogous model reference structures in the data base.

The lack of consistency among spectral assignments, as reported by different research groups, made us realize that the data base and the related checking programs have, potentially, a much wider use other than a mere support role for the experimental ^{13}C spectral prediction and interpretation programs. Specifically, these new computerized aids can help a chemist in assigning resonances by using a much more exhaustive analysis of analogous, previously assigned structures.

To illustrate the utility of these programs, a set of germacranolides and their ^{13}C spectra were collected from the literature. These spectral data were analyzed for consistency both with the information already in the data base and for mutual consistency within themselves. We show that the programs are useful in resolving ambiguities and errors that can arise in the assignment of resonances. There were several reasons for choosing germacranolides to illustrate the capabilities of the checking programs. First, during the last few years, a significant number of these compounds have been isolated and their ^{13}C spectra recorded,[4] so there was a substantial set of data available for analysis. The published shift assignments for the most part are based primarily on shift-trend analysis. Because of the flexible nature of the ten-membered germacrane ring (1) and the quite varied types of functionality and substitution, the analysis of their shifts tends to be more difficult than in more rigid sesquiterpene lactones such as the eudesmanes. Consequently, ambiguities are more likely to arise concerning the spectral assignments. Our preliminary review of data in the literature had in fact indicated that assignment inconsistencies probably existed (on the basis of the analysis presented below, the proposed reassignments are summarized in Table V). Another reason for processing the germacranolide data was that we wished to have this information in the data base so that the utility of the ^{13}C interpretation and prediction programs could be tested in compounds which are of current research interest.



1

The structural representations in the data base do not incorporate conformational factors. From this point of view, the germacranolides are a limiting case in the range of applicability of the programs. For most of the compounds currently in the data base, e.g., steroids and diterpenoids, the molecular conformations are largely dictated by the configurational stereochemistry in such rigid structures. Consequently, atoms in the same configurational environment are, generally, also in the same conformational environment. Thus, conformational effects on chemical shifts are subsumed by the configurational effects that are properly represented in our substructure coding scheme. However, a flexible ten-membered ring system allows for many conformations; the assumption

(1) For part 39 see C. W. Crandell, N. A. B. Gray, and D. H. Smith, *J. Chem. Inf. Comp. Sci.*, in press (1982).

(2) N. A. B. Gray, C. W. Crandell, J. G. Nourse, D. H. Smith, M. L. Dageforde, and C. Djerassi, *J. Org. Chem.*, **46**, 703 (1981).

(3) N. A. B. Gray, J. G. Nourse, C. W. Crandell, D. H. Smith, and C. Djerassi, *Org. Magn. Reson.*, **15**, 375 (1981).

(4) (a) R. W. Doskotch, F. S. El-Feraly, E. H. Fairchild, and C. Huang, *J. Org. Chem.*, **42**, 3614 (1977) (see Ref. 15 in ref 4j); (b) A. G. Gonzalez, J. Bermejo, H. Mansilla, A. Galindo, J. M. Amaro, and G. M. Massanet, *J. Chem. Soc., Perkin Trans. 1*, 1243 (1978); (c) P. S. Manchand and J. F. Blount, *J. Org. Chem.*, **43**, 4352 (1978) (see ref 4p); (d) W. Herz, R. de Groote, R. Murari, and J. F. Blount, *ibid.*, **43**, 3559 (1978); (e) K. Ito, Y. Sakakibara, M. Haruna, and K.-H. Lee, *Chem. Lett.*, 1469 (1979); (f) K. Ito, Y. Sakakibara, and M. Haruna, *ibid.*, 1473 (1979); (g) K. Ito, Y, Sakakibara and M. Haruna, *ibid.*, 1503 (1979); (h) N. C. Baruah, R. P. Sharma, K. P. Madhusudanan, G. Thyagarajan, W. Herz, and R. Murari, *J. Org. Chem.*, **44**, 1831 (1979); (i) W. Vichnewski, W. Herz, and N. Kumar, *ibid.*, **44**, 2575 (1979); (j) W. Herz, R. de Groote, R. Murari, N. Kumar, and J. F. Blount, *ibid.*, **44**, 2784 (1979); (k) F. S. El-Feraly, Y. M. Chan, G. A. Capiton, R. W. Doskotch, and E. H. Fairchild, *ibid.*, **44**, 3952 (1979); (l) R. N. Baruah, R. P. Sharma, K. P. Madhusudanan, G. Thyagarajan, W. Herz and R. Murari, *Phytochemistry*, **18**, 991 (1979); (m) N. Ohno and T. J. Mabry, *ibid.*, **18**, 1003 (1979); (n) W. Herz, S. V. Govindan, and J. F. Blount, *J. Org. Chem.*, **45**, 1113 (1980); (o) R. N. Baruah, R. P. Sharma, G. Thyagarajan, W. Herz, S. V. Govindan, and J. F. Blount, *ibid.*, **45**, 4838 (1980); (p) W. Herz and N. Kumar, *Phytochemistry*, **19**, 593 (1980); (q) N. Ohno and T. J. Mabry, *ibid.*, **19**, 609 (1980); (r) A. Rustaiyan, L. Nazarians and F. Bohlmann, *ibid.*, **19**, 1230 (1980); (s) W. Herz and S. V. Govindan, *ibid.*, **19**, 1234 (1980); (t) Z. S. Ferreira, N. F. Roque, O. R. Gottlieb, F. Oliveira, and H. E. Gottlieb, *ibid.*, **19**, 1481 (1980); (u) A. Romo de Vivar, E. Bratoeff, E. Ontiveros, D. C. Lankin, and N. S. Bhacca, *ibid.*, **19**, 1795 (1980).

Table I. Portion of PRED-CHECK's Analysis of the ^{13}C Data of 2[a]

Compound Name: Athanasia-25; Phyto-1979-18-995

NODE	MUL	SHELL	RESMIN	RESMAX	RESAVG	OBSRES	ZVALUE	SDEV	RES
**7	d	1	42.3	53.1	48.1	34.1	4.3	3.2	31
8	d	1	68.1	89.2	80.5	75.5	1.1	4.4	89
9	d	1	65.3	82.3	74.0	71.5	0.4	5.9	32
10	s	0	100.1	179.1	137.9	142.2	0.3	13.1	158
11	s	2	136.2	138.8	136.9	136.4	0.6	0.9	7
**12	s	2	168.8	169.8	169.2	175.8	16.9	0.3	7
*13	t	2	119.9	121.5	120.7	122.0	1.8	0.6	7
**14	q	2	11.7	16.1	13.9	21.7	5.3	1.4	12
15	q	2	23.2	25.9	24.3	25.3	0.7	1.2	4
*21	s	2	176.4	176.4	176.4	168.8	***	0.0	1
*22	d	3	34.3	34.3	34.3	45.3	***	0.0	1

[a] The data given specify the minimum (RESMIN), maximum (RESMAX), mean (RESAVG), and the standard deviation (SDEV) of the values for shifts associated with reference substructures corresponding to the configurational stereochemical environments of the carbon atoms of 2. The "SHELL" value defines the degree of correspondence between the best matching reference substructures and the substructural environments of this example. "OBSRES" is the observed resonance signal that the investigator has assigned to the specified atom. "ZVALUE" is equal to the difference between the assigned observed resonance shift and the mean predicted shift divided by the standard deviation. (A standard deviation is sometimes unavailable, as for example when a prediction is based upon a single reference example. Then a ZVALUE cannot be computed, instead a series of asterisks is printed.) The asterisks in the left column indicate that the observed shift is outside the predicted range of shifts and/or that the observed shift has a high ZVALUE, meaning that it is at least 2 times the standard deviation away from the mean shift value given.

that identical configurational environments imply similar conformations is not valid in such systems. Nevertheless, in the germacranolides we have examined, the configuration/conformation approximation can still be considered applicable because most of these structures are in well-defined conformations as a consequence of multiple bonds in the ten-membered ring or bridging ether links.

Methods

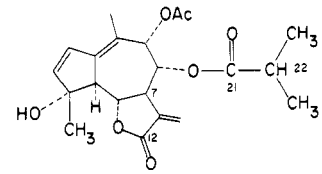
Our approach to shift assignment verification relies on the use of a set of three data checking programs. Two of these programs allow the chemist to check for consistency with analogous reference structures already in the data base and to identify relevant reference structures; the third program is designed to help the chemist maintain consistency when assigning resonance shifts throughout a series of related compounds.

Consistency with known spectral assignments, as represented by information in the data base, is achieved with the "PRED-CHECK" (*pred*iction–*check*ing) program.[3] The chemist's proposed shift assignments for the constituent atoms in a given structure are compared with the predicted shift values obtained from atoms in similar substructural environments already included in the data base. The PRED-CHECK program flags any discrepancies between assigned resonances and predicted shift ranges. The "REF-CHECK" (*ref*erence–*check*ing) program allows the chemist to identify the reference compounds used as the basis for PRED-CHECK's predictions. The chemist can determine whether an apparent discrepancy is or is not due to an erroneous assignment in the new data by considering the reference compounds both in terms of the reliability of their assignments and in terms of their appropriateness as prediction models for the new structure.

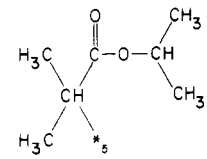
When a new series of compounds is to be processed, the new structures will typically incorporate carbon atoms in novel environments for which there are no detailed models in the data base. Consequently, spectrum predictions are imprecise, and several different permutations of assignments may prove about equally consistent with existing data. However, the various different compounds in the new series can themselves serve as models for checking the internal consistency of shift assignments. The "INTERN-CHECK" (*intern*al consistency–*check*ing) program identifies similar substructural environments in a series of compounds for which the shift or distribution of assigned shifts appears anomalous. The chemist may then, by interaction with the program, explore possibilities for reassigning resonances among the atoms of individual structures, thereby deriving a greater overall consistency.

These data-checking programs are presented in more detail through the following illustrative examples.

(1) PRED-CHECK **Program.** Typical results from the PRED-CHECK program are presented in Table I. These data, obtained from structure 2, illustrate a case in which



2



3

erroneous assignments among the proposed resonances are probable. The original assignments[5] were based upon analogies between 2 and related structures incorporating the same skeletal system and isobutyrate side chain. However, the resonances for atoms C(7) and C(22) and atoms C(12) and C(21) in 2 have probably been interchanged; the proposed resonances assigned to these atoms lie well outside the predicted ranges, and their "ZVALUES" are large. Clearly, interchanging the proposed assignments for these pairs of resonances would result in much closer agreement with the predicted data. Prior to any reassignment of resonances though, the REF-CHECK program is used.

(2) REF-CHECK **Program.** The REF-CHECK program is used to identify the reference compounds used as the basis for the predicted resonances. This may establish that the alternative assignments implied by the predictions are justified and more likely to be correct. The REF-CHECK program was applied to the spectrum of 2, and the reference compound used to predict the resonance for atom C(22) of 2 was identified. The reference compound proved to be isopropyl isobutyrate (3), with the resonance assignments taken from Couperous' collection of standard

(5) F. Bohlmann and K.-H. Knoll, *Phytochemistry*, 19, 995 (1979).

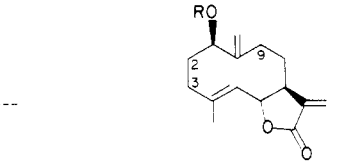
compounds.[6] Since the predicted resonance (34.3 ppm) could only belong to atom C(5) in **3**, the methine atom, C(22), of the isobutyrate side chain of **2** must have a resonance of 34.1 and not 45.3 ppm. The predicted resonances for atom C(7) were based on substructures similar at a SHELL level of one; definitive information is rarely obtained for such general substructures. Nonetheless, REF-CHECK confirmed that atoms C(7) and C(22) of **2** had been interchanged. A similar analysis performed for atoms C(12) and C(21) provided conclusive evidence that the proposed assignments should be interchanged. Once these resonances had been reassigned, the dissimilarity score (see ref 3 for definition) was reduced from 142.9 to 48.8.

There are, however, occasions when it is realized that it is the reference compound that contains the misassigned resonances. This can occur because, despite the care taken in checking for errors, it is always feasible that misassignments go unnoticed when substructures are new to the data base. It is only later, when more data of related compounds are being added, that these misassignments are recognized (utility programs can then be used to remove the erroneous data from the data base).

(3) INTERN-CHECK Program. The INTERN-CHECK program permits a complete analysis of all similar substructural environments and their assigned shifts within a set of compounds prior to entry into the data base. This analysis can be performed at each bond level; i.e., the assignment of a resonance to a carbon atom can be compared with those assigned to similar substructures described out to the two-bond (β), three-bond (γ), and four-bond (δ) levels. This enables errors to be detected in general substructures as well as in the completely specified substructures identical out to the four-bond level. The INTERN-CHECK program is controlled, interactively, by the chemist, who selects the type of resonance to be analyzed, e.g., all triplets or all doublets, and the "shell level" at which the analysis is to be performed. The chemist must also specify constraints on the characteristics permitted for shift distributions associated with specific substructures, e.g., the maximum range of shifts. INTERN-CHECK identifies substructures for which the assigned shifts violate any of the specified constraints on shift distributions.

If a particular substructure is found to be associated with an anomalous shift distribution, INTERN-CHECK identifies and produces a report on the shift data. Such reports, illustrated below, give details of the minimum, mean, and maximum shifts and the number of instances of that substructure among the set of compounds under analysis. A "SKEW" value is also given; a value of 1 for the SKEW indicates that the shifts corresponding to the set of substructures is evenly distributed over the range; larger values are indicative of tailing distributions. The chemist can obtain histograms showing the distribution of shifts and a list identifying each instance of the substructure, giving the number of the atom, the shift value, and the name of the compound. If a particular shift appears anomalous, then a list of shifts associated with other atoms of the same multiplicity in the same compound can be obtained. Examination of these alternatives may reveal other resonances that could be interchanged with the anomalous shift. The program permits further checks to see if such an interchange would also improve the consistency of shifts assigned to substructures associated with the other atom.

The following example illustrates the use of the INTERN-CHECK program in identifying an error in the assignment of two sp^3 methylene carbons. The check was performed for substructures similar at the β (two bond) level. The program returned the information that for the substructure **4**, which was present in 11 compounds (**5**–**15**, Chart I), the range of shifts was very wide (13 ppm), and the SKEW value was large. A list of the substructures



4

(6) P. A. Couperous, A. D. H. Clague, and J. P. C. M. Van Dangen, *Org. Magn. Reson.*, **11**, 590 (1978).

Chart I

5

6 R = OH
7 R = H
8 R = OCH_3
9 R = Ac

10

11 R = OH, R′ = H
12 R = H, R′ = H
13 R = Ac, R′ = H
14 R = OMe, R′ = H
16 R = OH, R′ = OAc
17 R = Ac, R′ = OAc

15

Table II. List of Substructures for Atom 9 with Their Assigned Resonances for Compounds 5–15[a]

shift	compd	shift	compd
32.8	5	34.4	11
37.2	6	34.6	12
36.1	7	33.7	13
37.0	8	24.2	14
36.3	9	34.6	15
37.4	10		

[a] The resonance for compound 14 is anomalous.

Table III. List of Resonances Associated with Atoms of the Same Multiplicity for Compound 14[a]

atom	assigned shift, δ	atom	assigned shift, δ
2	34.0	8	25.9
3	26.0	9	24.2

[a] Atom 9 could be interchanged with atom 2.

Table IV. List of Assigned Resonances for Atom 2 for Compound 14 and Related Substructures[a]

shift, δ	compd	shift, δ	compd
24.4	11	25.7	15
23.8	13	26.6	16
34.0	14	27.8	17

[a] The resonance for compound 14 is anomalous.

Table V. Summary of the Proposed Reassignments of Resonances[a]

no.	compd	atom(1)	atom(2)	program	ref
5	gallicin	1	6	INTERN-CHECK	4b
6	peroxycostunolide	2	3	INTERN-CHECK	4k
7	hydroxycostunolide	2	3	INTERN-CHECK	4k
8	methylperoxycostunolide	2	3	INTERN-CHECK	4k
9	acetylcostunolide	2	3	INTERN-CHECK	4k
10	dihydroperoxycostunolide	2	3	INTERN-CHECK	4k
14	methylperoxyparthenolide	2	9	INTERN-CHECK	4k
17	(acetyloxy)ferolide	13	14	PRED-CHECK and REF-CHECK	4a
18	ketocostunolide	2	3	INTERN-CHECK	4k
19	neurolenin B	12	1″	INTERN-CHECK	4c
20	calea (2c)	12	1″	INTERN-CHECK	4p
21	calea (2d)	12	1″	INTERN-CHECK	4p
22	calea (2e)	12	1″	INTERN-CHECK	4p
23	calea (4e)	12	1″	INTERN-CHECK	4p
24	deacetyleupaserrin	1	5	PRED-CHECK and REF-CHECK	4m
		4	10		
		10	11		
		11	4		
		14	15		
		15	4′		
		4′	14		
25	eupaserrin	1	5	PRED-CHECK and REF-CHECK	4m
		4	10		
		10	11		
		11	4		
		14	15		
		15	4′		
		4′	14		
26	detiglic-Mollisorin	1	5	PRED-CHECK	4m
27	mollisorin	1	5	PRED-CHECK	4m
		4	10		
		10	11		
		11	4		
		15	5′		
36	eupatoriopicrin	1	5	PRED-CHECK and INTERN-CHECK	4g
		4	10		
39	woodhousin	2	6	PRED-CHECK and INTERN-CHECK	4h
		4	11		
41	eupachifolin A	12	1′	PRED-CHECK	4f

[a] Atom (1) and atom(2) are those resonances which have been interchanged. In cases 24, 25, and 27, sets with braces have been interchanged (i.e., the resonance for atom 4 was reassigned to atom 10, the resonance for atom 10 was reassigned to atom 11, and so on). The programs which highlighted the erroneous assignments are given.

(Table II) indicated that the resonance assigned to atom 9 in compound 14 was far from the mean shift (34.4 ppm) and warranted further investigation. A list of shifts associated with atoms of the same multiplicity for compound 14 was given (Table III). It showed that the resonances assigned[4k] to atoms 2 and 9 could have been misassigned, and a list of substructures (Table IV), similar at the β level, for atom 2 was requested. Table IV indicates that the resonance assigned to atom 2 of compound 14 is inconsistent within this set of substructures. The reassignment was considered justifiable because interchanging the original resonances assigned to atoms 2 and 9 made them consistent within their respective sets of substructures.

Results

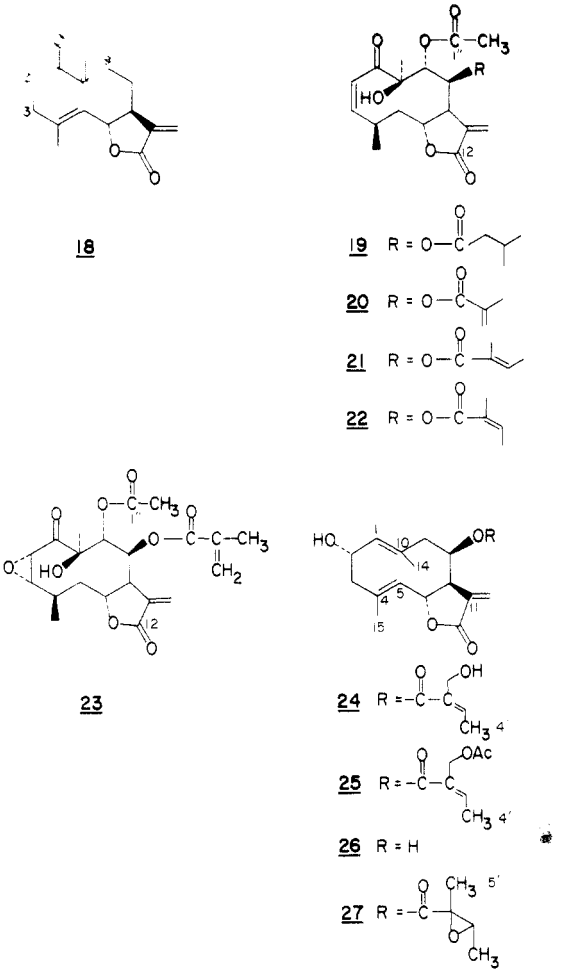
(1) Analysis of Assignments in the Set of Germacranolides. At the onset of this investigation, the data base already contained a large set of guaianolides, eudesmanes, and other sesquiterpene lactones. In addition, seven germacranolides[7,8] were present. Many of the substructural environments of the germacranolides are not well represented by detailed substructural models due to the small number of germacranolides present in the data base; therefore, the PRED-CHECK program was expected to identify few, if any, erroneous assignments of resonances in the new data. It was anticipated, though, that errors in the assignments of the resonances could be found by checking for internal consistencies with the INTERN-CHECK program due to the large set of germacranolides available.

A large group of germacradienolides and germacranolides, totaling 88 distinct compounds, were abstracted from the literature.[4] The structures were defined and the spectral data entered and assigned in accord with the published results. These initial data were analyzed by means of the PRED-CHECK program. The objectives of the analysis were to find (1) errors introduced during the reporting/publication process, (2) erroneous assignments of observed shifts to atoms of the new structures, and (3) apparent discrepancies due to invalid assignments already incorporated in the data base.

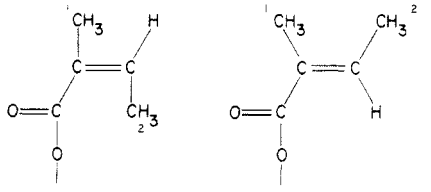
Table V summarizes the interchangeable sets of resonances as identified with the aid of the three programs. It is important to note that definitive assignments for the resonances can only be established by additional NMR experiments. We discuss several examples to illustrate how the programs were used to suggest the reassignments summarized in Table V.

(7) W. Herz, N. Kumar, and J. F. Blount, *J. Org. Chem.*, 45, 489 (1980).
(8) W. Herz, S. V. Govindan, and J. F. Blount, *J. Org. Chem.*, 45, 3163 (1980).

Chart II



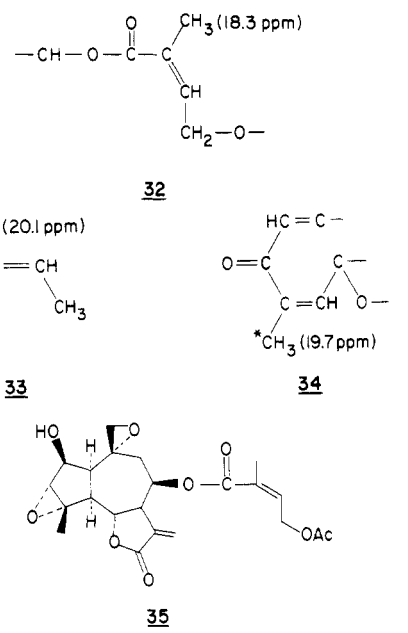
A general example of inconsistency, highlighted by the PRED-CHECK program, is the assignment of the resonances to the angelic ester side chain (**28**). This side chain is



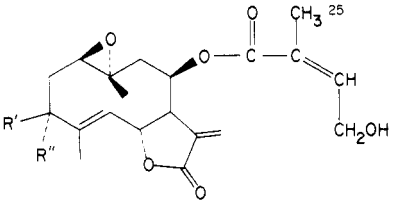
common in germacranolides (and other sesquiterpenes), and yet inconsistent assignments appear in the literature for the chemical shifts to the methyl groups of the side chain. One group[4o,p] has consistently assigned methyl C(1) of **28** the shift value of ca. 15 ppm and methyl C(2) the value of ca. 20 ppm. Other authors[4q,r] use the reversed assignment. Similar inconsistencies in the assignment of resonances to the related tiglic ester side chain (**29**) were also noted.[4j,m,p]

These inconsistencies indicated that further NMR studies such as deuterium labeling or nuclear Overhauser effect (NOE) experiments are necessary to establish the correct assignments (one example of each set of assignments has been incorporated into the data base and will remain until the inconsistency is resolved). This example of inconsistency illustrates the utility of the programs; disagreement over the assignment of resonances, as in this case, can be indicated and subsequently resolved by the chemist.

Chart III



An interesting case, involving use of all three checking programs, concerned the assignment of the resonance to the methyl group of the side chain in eupalinin B (**30**) and eupalinin D (**31**).[4e] The PRED-CHECK program had indi-



30 R' = -OAc, R'' = H

31 R' = H, R'' = -OAc

cated that the observed resonances of 12.5 and 12.7 ppm for the methyl C(25) in **30** and **31**, respectively, were not in agreement with the predicted resonance (18.3 ppm). The prediction was presumed to be accurate as it was based upon a substructure (**32**, Chart III) matching out to a three-bond environment. The INTERN-CHECK program confirmed that the observed resonance of 12.5 ppm did not agree with the resonances assigned to substructures **33** and **34** similar at the two-bond environment.

However, the observed resonances of 12.5 and 12.7 ppm for the C(25) methyl group in **30** and **31** could not have been misassigned, there being no possible alternative resonances. Since nuclear Overhauser effect (NOE) experiments had been done, the trans stereochemistry was unequivocal. The reference compound,[9] graminiliatrin (**35**), retrieved by using REF-CHECK, showed that the predicted resonance was from the same side chain as in **30** and **31**. Again, in this reference structure, the shift assignment was unambiguous. The trans stereochemistry, however, had not been unequivocally proven. In addition, we noted that the methyl group in the same side chain, but assigned with the cis stereochemistry, had been assigned the resonance of 12.4 ppm.[4j] The stereochemistry of the reference structures already in the data base had been derived by using analogies from proton NMR spectra. The capability

(9) W. Herz, R. Murari and S. V. Govidan, *Phytochemistry*, **18**, 1337 (1979).

Table VI. Portion of PRED-CHECK's Analysis of the ^{13}C Data of Eupatoriopicrin[a,c]

Compound Name: Eupatoriopicrin; Chem. Lett-1979-1503

NODE	RESMIN	RESMAX	RESAVG	OBSRES	ZVALUE[b]	SDEV	RES
**1	118.9	125.8	123.0	130.4	3.6	2.0	13
*5	128.3	135.0	131.5	127.7	1.3	2.7	14

[a] This indicates the possible pair of misassigned resonances. [b] ZVALUES are truncated by the program, not rounded off. [c] MUL = d and SHELL = 1 in both cases.

Table VII. Portion of PRED-CHECK's Analysis of the ^{13}C Data of Woodhousin[a]

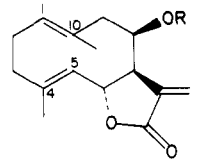
Compound Name: Woodhousin; J.O.C-1979-44-1831

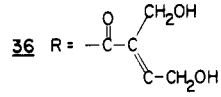
NODE	RESMIN	RESMAX	RESAVG	OBSRES	ZVALUE
*2	80.3	80.3	80.3	76.6	***
*6	76.8	76.8	76.8	80.4	***

[a] The resonances assigned to atoms 2 and 6 are inconsistent with that in the data base. [b] MUL = d, SHELL = 4, SDEV = 0.0, and RES = 1 in both cases.

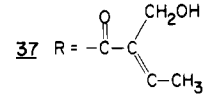
of the programs to indicate inconsistencies in the assignment of resonances is well illustrated by this example; these results also indicate the need for further NMR studies to resolve the inconsistencies and definitively characterize the structures. The new data have been incorporated into the data base to indicate that there is an inconsistency for this methyl group.

Often, the results of the prediction process (PRED-CHECK) imply that two resonances may have been interchanged, but, because the analysis has been limited through the use of generalized substructures, the implication is not in itself sufficient to justify changes. The additional results from the INTERN-CHECK program can provide confirming data, making it possible to justify an interchange of two such resonances. Thus, the PRED-CHECK program (Table VI) suggested that the resonances assigned to the two sp^2 tertiary carbon atoms C(1) and C(5) in eupatoriopicrin (**36**)[4g] might have been misassigned. However, the predicted resonances were from substructures ($CH_2C^*H{=}C(CH_3)CH_2$) similar only at the one-bond environment (see ref 3 for the definition of shell levels and bond environments), and so it was inappropriate to change the literature assignments. The results from the INTERN-CHECK program showed that the resonances assigned to carbons C(1) and C(5) of **36** were inconsistent with those assigned to similar substructures (**37** and **38**) described out to the four-bond environment.[4j] It is possible that the observed resonances assigned to carbons C(1) and C(5) in **36** have been correctly assigned and that the other two substructures (**37** and **38**) are incorrectly assigned. However, in order to maintain consistency with the predicted resonances, for the substructures similar at the one-bond level, it is reasonable to interchange the two resonances of **36** (no. 36 in Table V). The dissimilarity score was reduced from 15.7 to 9.9 once the resonances of **36** had been reassigned.

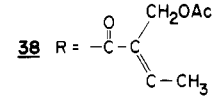




1 130.4 ppm
5 127.7 ppm
4 134.1 ppm
10 141.6 ppm



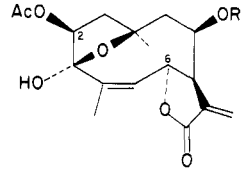
1 127.3 ppm
5 130.6 ppm
4 142.1 ppm
10 134.1 ppm

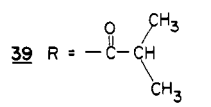


1 127.3 ppm
5 130.7 ppm
4 142.5 ppm
10 134.2 ppm

Uncertainties with respect to the assignment of resonances to atoms in new compounds are frequently acknowledged by indications that the suggested values are interchangeable. Therefore when inconsistencies are apparent among a group of similar substructures, it is easier to justify which is the correct assignment of resonances. In this example, the INTERN-CHECK program indicated that the assignments to atoms C(4) and C(10) of compounds **36–38** are inconsistent, and it is possible that either **36** or **37** and **38** had been misassigned. However, the original research group[4g] had indicated that the resonances of atoms C(4) and C(10) in **36** are interchangeable, and so it was justifiable to accept that **36** contained the misassignments (no. 36 in Table V).

Another example, illustrating the combined use of the three programs, concerns the data presented in Table VII. The PRED-CHECK program indicated that the assignment of the resonances to carbons C(2) (76.6 ppm) and C(6) (80.4 ppm) in woodhousin (**39**)[4h] were inconsistent with the predicted resonance ranges. Through REF-CHECK, it was





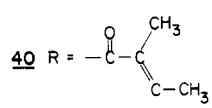


Table VIII. Portion of PRED-CHECK's Analysis of the ^{13}C Data of Eupachifolin A[a,b]

Compound Name : Eupachifolin A; Chem. Lett-1979-1473

NODE	SHELL	RESMIN	RESMAX	RESAVG	OBSRES	ZVALUE	RES
**12	4	168.8	170.3	169.3	175.1	11.0	7
**21	3	175.4	176.5	175.9	169.4	11.9	3

[a] The resonances assigned to atoms 12 and 21 have probably been misassigned. Atom 21 is the numbering given by the program and corresponds to atom 1'. [b] MUL = s and SDEV = 0.5 in both cases.

Table IX. PRED-CHECK's Analysis of the ^{13}C Data[a]

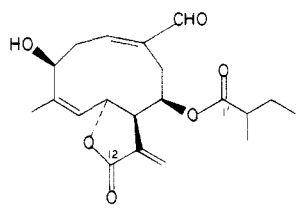
Compound Name : Budlein Analog; Phyto-1981-20-99

NODE	MUL	SHELL	RESMIN	RESMAX	RESAVG	OBSRES	ZVALUE	SDEV	RES
*1	s	4	205.1	205.4	205.2	205.7	2.5	0.2	2
*2	d	4	104.9	104.9	104.9	104.7	***	0.0	2
*3	s	4	182.8	182.8	182.8	182.9	***	0.0	2
*4	s	4	138.9	138.9	138.9	138.4	***	0.0	2
*5	d	4	133.9	133.9	133.9	133.3	***	0.0	2
*6	d	4	75.3	75.3	75.3	75.4	***	0.0	2
7	d	3	47.1	50.2	48.8	48.0	0.8	0.8	17
8	d	2	68.3	76.7	73.0	74.1	0.4	2.6	25
**9	t	3	42.6	43.1	42.8	41.9	3.0	0.2	4
*10	s	4	87.9	87.9	87.9	87.6	***	0.0	2
11	s	4	136.0	143.6	137.8	136.1	0.7	2.2	17
12	s	4	168.4	170.7	169.4	168.8	1.1	0.5	29
13	t	4	122.9	125.5	124.1	124.0	0.1	0.8	17
**14	q	4	21.3	21.4	21.3	21.0	4.6	7.0×10^{-2}	2
*15	t	4	62.3	62.3	62.3	61.9	***	0.0	2
*21	s	4	171.3	171.3	171.3	171.3	***	0.0	1
*22	t	4	43.0	43.0	43.0	42.6	***	0.0	1
23	d	4	24.6	25.4	25.0	25.2	0.4	0.3	4
24	q	4	22.2	22.4	22.3	22.2	0.7	8.9×10^{-2}	6
*25	q	4	22.2	22.4	22.3	22.0	2.3	8.9×10^{-2}	6

[a] Atoms 1-15 represent the germacranolide ring, and atoms 21-25 represent the side chain. Atoms 16-20 are oxygens.

determined that the reference compound (**40**)[8] used as a model for prediction was one of the few germacranolides already included in the data base. Consequently, it was not clear whether it was the resonances of the reference compound or the observed resonances of the new data that had been incorrectly assigned. However, with the aid of INTERN-CHECK, data from similar substructures confirmed that the observed resonances for atoms C(2) and C(6) of **39** had been misassigned (no. 39 in Table V). With these resonances exchanged, the dissimilarity score was reduced from 20.5 to 7.2.

The assignments of resonances to the carbonyls in eupachifolin A (**41**[4f], Table VIII) provide an example of where



41

misassignments can be suggested and confirmed by using PRED-CHECK and REF-CHECK. It was noted that atoms C(12) and C(1') had probably been misassigned; the predicted resonances for C(12) and C(1') represented substructural environments matched out to the four-bond and three-bond level, respectively. However, the observed resonances assigned to these atoms were well outside the predicted ranges, and the "ZVALUEs" were large. REF-CHECK indicated that the reassignments were justified, and once the observed resonances had been interchanged, they agreed with the predicted shift ranges (no. 41 in Table V). The dissimilarity score was thus reduced from 86.6 to 54.1.

(2) Routine Analysis of Assignments. To illustrate that, with a well-represented data base, the PRED-CHECK program is a useful aid for chemists in checking assignments of ^{13}C NMR spectra, we took a recent publication[10] and checked the results using our programs, with the previously analyzed germacranolides as the reference compounds for the prediction process. As is seen (Table IX), the predicted resonances are based mostly on substructures described out to four- and three-bond environments, and the observed resonances are in close agreement wiih them. The dissimilarity score was 3.1, indicating that, within the limitations of the method, all resonances were assigned correctly.

Discussion

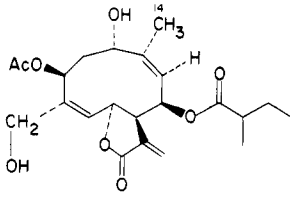
We stress that the possibly erroneous assignments identified by the programs, and their proposed reassignments, are not definite. The entire analysis is based solely on chemical shift analogies. Where ambiguities have been identified, definitive assignments are dependent on additional NMR data.

One of the disadvantages of these programs is that they are not able to analyze and compensate for the effect of the different solvents used. Spectra entered into the data base are assumed to have been run under "standard conditions", i.e., at room temperature in deuteriochloroform. New spectra run in deuteriochloroform and, for the most part, those run in pyridine-d_5 are compatible. However, spectra run in more polar solvents such as acetone-d_6 and methanol-d_4 are incompatible, exhibiting quite substantial changes in shifts. Spectra recorded in polar or mixed solvents are not utilized; the data base is only

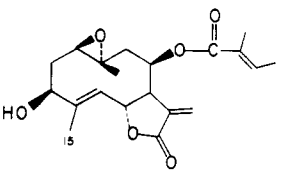
(10) W. Herz and N. Kumar, *Phytochemistry*, **20**, 99 (1981).

of marginal value as an aid to checking assignments proposed for such spectra.

Occasionally, the four-bond substructural environments represented in the data base are not sufficient to characterize properly stereochemical environments. This is illustrated by the following example: atom C(14) in deoxyeurecurvin (**42**)[4d] (which has the assigned resonance of



42



43

17.6 ppm) and atom (C15) of heliagine (**43**,[4e] 22.9 ppm) have the same stereochemical environment out to the four-bond level. However, when the structures are considered as a whole, the carbons are in quite different conformations. In such cases, a description out to a five-bond environment would be needed to distinguish these substructures. However, such an added dimension to the substructural coding process would not be sufficiently advantageous to merit its inclusion since these cases are not common. Also a substructure described out to a five-bond level would, quite often, represent the complete molecule. As such, a five-bond substructure would be unique and could not constitute an appropriate prediction model for an atom in another molecule.

Despite these limitations, the programs for assignment verification should be a great aid to chemists. Routinely, many assignments are based just on chemical shift trends and analogies to related compounds. The availability of the checking programs and their data base allows this type of analysis to be undertaken in a much more systematic and comprehensive manner.

There are currently two substantial data bases of carbon-13 spectra commercially available.[11,12] Unfortunately, neither of these incorporate stereochemistry. If configurational stereochemistry is not represented, then shift ranges are broadened because the data from substructures in different configurations are combined. Such additional broadening would inhibit our checking procedures.

Our current data base incorporates primarily data on diterpenes, steroids, and sesquiterpenes. We have previously cooperated with other research groups to expand the data base to cover particular areas such as C_{19} diterpenoid alkaloids.[13] Such cooperation with other groups is welcomed.

Experimental Section

These programs are implemented in the ALGOL-like BCPL program language[14] on a Digital Equipment Corp. KI-10 computer at the SUMEX-AIM computer facility at Stanford. The programs are available to a collaborative community of investigators via an international computer network to the limits of available resources. Please contact the authors for information concerning access to the programs.

Acknowledgment. We thank the National Institutes of Health (Grant No. RR-00612) for generous financial support. Computer resources were provided by the SUMEX facility at Stanford University under National Institutes of Health Grant RR-0785.

(11) Reference 6 in D. L. Dalrymple, C. L. Wilkins, G. W. A. Milne, and S. R. Heller, *Org. Mag. Reson.*, 11, 535 (1978).
(12) W. Bremser, L. Ernst, and B. Franke, "Carbon-13 NMR Spectral Data", Verlag Chemie, New York, 1978.
(13) S. W. Pelletier, J. Finer-Moore, N. V. Mody, N. A. B. Gray, C. W. Crandell, and D. H. Smith, *J. Org. Chem.*, 46, 3399 (1981).
(14) M. Richards and C. Whitby-Strevens, "BCPL—the language and Its Compiler", Cambridge University Press, Cambridge, London, 1979.