A Correlative Notation System for NMR Data*†

HERMAN SKOLNIK
Hercules Incorporated, Research Center, Wilmington, Delaware 19899

Received January 5, 1970

A new linear notation system which denotes carbon in terms of bonds and attached hydrogen(s) is used to correlate proton groups in organic molecules with chemical shifts. The notation system is illustrated with acyclic and cyclic examples, and the production of tables of NMR data via computer by proton group vis-a-vis neighboring groups is demonstrated. Tables of chemical shifts in ascending order is a valuable by-product of the computerized system.

In a foreword to a 1966 book, R. E. Richards[1] pointed out that "Nuclear resonance work is now being published at an ever-increasing rate, which has now reached about 2000 papers a year." Consequently, there has been considerable interest in methods for correlating the great amount of NMR data in the literature. The two most commonly used methods for relating basic proton and environmental groups with chemical shifts are the Sadtler Chemical Index[2] and the Varian Functional Group Index.[3] Both methods require manual indexing from as many viewpoints as there are proton groups and neither fully represents the molecular structure.

Recently, a new linear notation system[4] was introduced which is based on combinations of carbon and hydrogen. Although the new notation system was designed to yield unique and unambiguous notations for chemical structures, it appears to be particularly suitable for correlating proton groups with their chemical shifts. The present paper describes the new linear notation system as modified to be most useful for correlating NMR data with chemical structures and for producing tables of data in various orders with a computer.

The symbols used in the new notation system for the various combinations of carbon and hydrogen are listed in Table I, for carbons without hydrogen in Table II, and for atoms other than carbon in Table III. Tables I, II, and III show that the important proton groups are designated by the nine notation symbols shown in Table IV. As about 90% of all NMR studies reported have been concerned with ^1H resonance, the nine symbols listed in Table IV constitute the major entries or subjects in an index of proton groups relative to their neighboring groups and their chemical shifts. The use of the new notation in designating these nine proton groups, which may be located in hundreds of thousands of organic molecules in a variety of arrangements, will now be described. Most of the examples that follow were taken from the

Table I. Notation Symbols for Combinations of Carbon and Hydrogen

Single-Bonded Carbons		Double-Bonded Carbons		Triple-Bonded Carbon	
$-CH_3$	A	$=CH_2$	E	$\equiv CH$	U
$-CH_2-$	C	$=CH-$	B		
$-CH-$	Y				
$>CH-$	J (fused or bridgehead carbon)				

Table II. Notation Symbols for Carbons without Hydrogen

Single-Bonded Carbons		Double-Bonded Carbons	
$-C-$	X	$>C=$	D
$>C<$	T (fused or bridgehead carbon)	$>C=$	R (fused carbon)

Triple-Bonded Carbon		Carbonyl Group	
$\equiv C-$	V	$>C=O$	K

Table III. Notation Symbols for Atoms Other Than Carbon

Nitrogen		Oxygen		Halogen	
$>NH$	M	$-O-$	Q	$-F$	F
$-NH_2$	MH	$-OH$	QH	$-Br$	G
$>N-$	N	O_2	W	$-I$	I
$\equiv N$	Z		(nonlinear as	$-Cl$	L
$=N$	Z		in NO_2 or SO_2)		

Others		
$-H$	H	Metals are denoted by & followed by atomic symbol, e.g., &SI for silicon.
$-S$	S	Atoms between bridgeheads are preceded by :
$=S$	S	
P	P	

Table IV. The Nine Most Important Proton Groups

Notation Symbol	Proton Group
A	$-CH_3$
B	$-CH=$
C	$-CH_2-$
E	$=CH_2$
H	H
J	$>CH-$ (bridgehead)
M	$-NH-$
U	$\equiv CH$
Y	$>CH-$

* Presented at the Eastern Analytical Symposium, November 19, 1969, New York, N. Y.

† Research Center Contribution No. 1523

first 203 compounds listed in reference 3, and the compound number given is the Varian compound number.

ACYCLIC COMPOUNDS

Although the earlier paper[4] on this notation system specified the writing of notations in the same order as structural formulas are generally drawn, this order is not preferred for computer processing of an NMR data file. The original objective of designating moieties and functional groups for computerized storage and retrieval is a disadvantage for NMR, which needs an atom-to-atom emphasis. Consequently, the notation system has been modified to be compatible with the needs of users of NMR data and with the constraints of computer processing. To show NMR relationships, the preferred way to write a notation is from the lowest number atom—i.e., from the end to which the the most important functional group is attached—and to write the notation strictly linearly—i.e., from atom-to-atom without separation between a moiety and an attached functional group.

The examples listed in Table V illustrate the notation system for acyclic compounds.

In modifying the notation for NMR relationships it was important to be aware of keypunch input, computer processing (especially the constraints of alphanumeric sorting by computer), and the limitations of characters on most computer print chains. For example, the absence of subscripts and lower case letters on most computer print chains dictated the use of standard numbers and capital letters as illustrated in this paper. Alphabetization by the IBM 360 computer, which is a sort-merge operation, places a space first, special characters (such as period, comma, etc.) next in a fixed order, letters next in an alphabetical order, and numerals last in numerical order. Thus, the computer sort yields the order shown in Table VI.

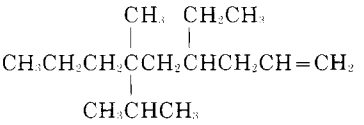
Table V. Notations for Acyclic Compounds

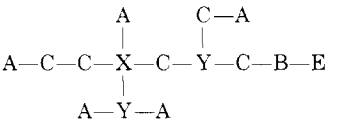
Varian No.	Compound: Structure	Notation: Visual	Notation: Written
1	CH_3OH	AQH	HQA
2	$ClCH_2CHCl_2$	LCYL2	L2YCL
4	F_3CCH_2OH	F3XCQH	HQCXF3
5	$ClCH_2CH_2Cl$	LCCL	LCCL
6	CH_3CHO	AKH	HKA
7	$CH_3\overset{O}{\overset{\|}{C}}—SH$	AKSH	HSKA
8	$CH_3\overset{O}{\overset{\|}{C}}—OH$	AKQH	HQKA
9	$HC\overset{O}{\overset{\|}{—}}OCH_3$	HKQA	AQKH
12	$ClCH_2CH_2OH$	LCCQH	HQCCL
14	CH_3CH_2OH	ACQH	HQCA
16	$HC\equiv C—CH_3$	UVA	UVA
17	$BrCH_2—\underset{\|}{\overset{Br}{C}}=CH_2$	C(G)(G)—D—E	EDGCG
24	$CH_2=CHCH_2Br$	EBCG	EBCG
25	$CH_3—\underset{\|}{\overset{Cl}{CH}}—\overset{O}{\overset{\|}{C}}—OH$	AYKQH (L)	HQKYLA
38	$CH_2=CHCH_2NH_2$	EBCMH	HMCBE
42	$CH_3CH_2CH_2NO_2$	ACCNW	WNCCA
91	$(CH_3)_2NCH_2CH_2OH$	A(A)>NCCQH	HQCCNA2
181	$CH_2(C(=O)—OCH_2CH_3)_2$	C<KQCA KQCA	C(KQCA)2

If the notation is written in the reverse order, as originally specified—i.e., as drawn in the "chemical" column—the computer sort would be as shown in Table VII.

The expanded notation is preferred as NMR chemical shifts are significant for repeating units, such as $—CH_2—$, up to the point that a proton group, such as —OH, affects it. Beyond this point, however, the repeating group can be designated by the notation symbol and the number of repeating units, as shown in Table VII for AC2CQH, AC3CQH, and AC4CQH.

Notations of more complex acyclic compounds, particularly those with branched chains, are written with parentheses to separate the attachment from the main branch, as in the following:





ACCX(A)(YA2)CY(CA)CBE (Notation as drawn from position 9)
EBCY(CA)CX(A)(YA2)CCA (Notation from position 1)

This method for denoting substituents is particularly important in the notations for cyclic compounds.

CYCLIC COMPOUNDS

Notations for cyclic compounds follow the same principles as used with notations for acyclics except that all attachments to ring atoms are placed within parentheses. This exception permits the ready recognition of the ring structure. In addition, notations for cyclics begin with a period, thus permitting the computer to separate readily cyclics from acyclics and to give the seeker for information a method for recognizing cyclics easily. The notation is written with the lowest numbered atom (based on the

Table VI. Computer Sort Order of Notations

Chemical	Compressed Notation	Expanded Notation	Sort Order
CH_3OH	HQA	HQA	1
CH_3CH_2OH	HQCA	HQCA	2
$CH_3(CH_2)_2OH$	HQC2A	HQCCA	3
$CH_3(CH_2)_3OH$	HQC3A	HQCC2A	4
$CH_3(CH_2)_4OH$	HQC4A	HQCC3A	5
$CH_3(CH_2)_5OH$	HQC5A	HQCC4A	6
$(CH_3)_2CHOH$	HQYA2	HQYA2	7

Table VII. Computer Sort Order of Notations

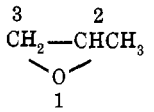
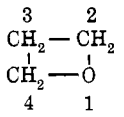
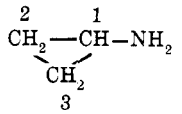
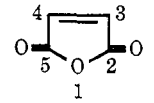
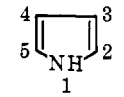
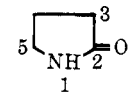
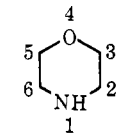
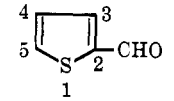
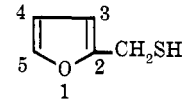
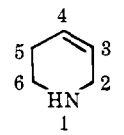
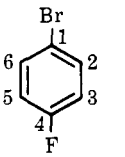
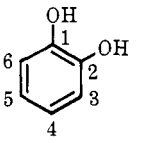
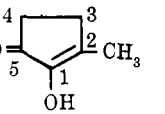
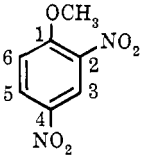
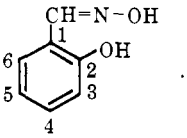
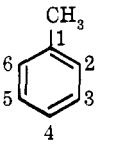
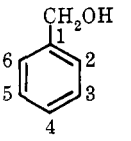
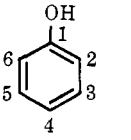
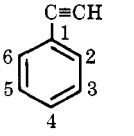
Chemical	Compressed Notation	Sort Order	Expanded Notation	Sort Order
CH_3OH	AQH	6	AQH	6
CH_3CH_2OH	ACQH	1	ACQH	2
$CH_3(CH_2)_2OH$	AC2QH	2	ACCQH	1
$CH_3(CH_2)_3OH$	AC3QH	3	AC2CQH	3
$CH_3(CH_2)_4OH$	AC4QH	4	AC3CQH	4
$CH_3(CH_2)_5OH$	AC5QH	5	AC4CQH	5
$(CH_3)_2CHOH$	A2YQH	7	A2YQH	7

Ring Index[5] numbering) cited first or as near to the left end as possible. A few examples are illustrated in Table VIII.

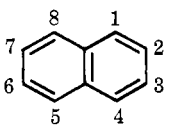
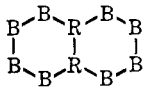
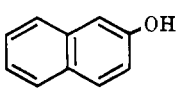
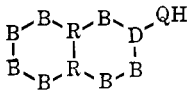
Bicyclic compounds, which contain bridgehead atoms, are assigned notations beginning with a period to designate the cyclic structure and ending with a colon to designate the atoms between the bridgeheads; the first notation symbol represents position 1 in accordance with the Ring Index. Examples are shown in Table IX.

The lozenge, ⌑, is used to designate condensed ring

Table VIII. Notations for Cyclic Compounds

Compound: Varian No.	Compound: Structure	Notation: Visual	Notation: Written
32		.QYC. | A	.QY(A)C.
33		.QCCC.	.QCCC.
37		.YCC. | MH	.Y(MH)CC.
48		.QKBBK.	.QKBBK.
55		.MBBBB.	.MBBBB.
68		.MKCCC.	.MKCCC.
83		.MCCQCC.	.MCCQCC.
94		.SDBBB. | KH	.SD(KH)BBB.
101		.QDBBB. | CSH	.QD(CSH)BBB.
115		.MCBBCC.	.MCBBCC.
121		.DBBDBB. | | G F	.D(G)BBD(F)BB.
124		.D—DBBBB. | | QH QH	.D(QH)D(QH)B3B.
129		.D—DCCK. | | QH A	.D(QH)D(A)CCK.
149		.D—D—B—DBB. | | | QA NW NW	.D(QA)D(NW)BD (NW)BB
156		.D——DB4. | | BZQH QH	.D(BZQH)D(QH)B3B.
157		.DB5. | A	.D(A)B4B.
161		.DB5 | CQH	.D(CQH)B4B.
162		.DB5. | QH	.D(QH)B4B.
186		.DB5. | VU	.D(VU)B4B.

structures and R is the symbol used for the carbons common to condensed aromatic rings, as illustrated below:

Structure	Notation: Visual	Notation: Written	
 naphthalene		B4R2	B4
 β-naphthol		BD(QH)B2R2	B4

Spiro compounds are designated with □ T □, as illustrated below:

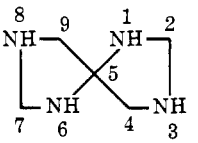
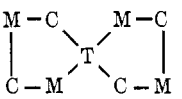
Structure	Notation: Visual	Notation: Written		
 1,3,6,8-Tetraaza-spiro[4.4]nonane		MCMC	T	MCMC

Table IX. Notations for Bicyclic Compounds

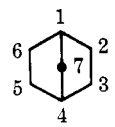
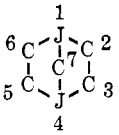
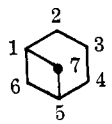
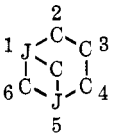
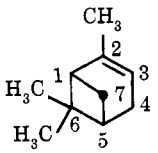
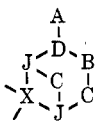
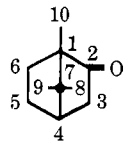
Structure	Visual	Written
 norbornane		.JCCJCC:C.
 norpinane		.JCCCJC:C.
 α-pinene		.JD(A)BCJX(A2):C.
 camphor		.T(A)KCJCC:XA2.

Table X. Computer Permutation of HQCCMCA

Fixed Position ↓

	H	QCCMCA
HQ	C	CMCA
HQC	C	MCA
HQCC	M	CA
HQCCM	C	A
HQCCMC	A	

Table XI. Compurter Permutation of Notation and Chemical Shifts for HQCCMCA

Notation

Fixed Position ↓

	H	QCCMCA
HQ	C	CMCA
HQC	C	MCA
HQCC	M	CA
HQCCM	C	A
HQCCMC	A	

Chemical Shift

Fixed Position ↓

	3.55	-3.65-2.73-3.53-2.68-1.12
3.53-	3.65	-2.73-3.53-2.68-1.12
3.53-3.65-	2.73	-3.53-2.68-1.12
3.53-3.65-2.73-	3.53	-2.68-1.12
3.53-3.65-2.73-3.53-	2.68	-1.12
3.53-3.65-2.73-3.53-2.68-	1.12	

CORRELATION OF PROTON GROUPS WITH CHEMICAL SHIFTS

The need to interpret NMR spectra in terms of molecular structure of a single chemical or a mixture of chemicals requires a body of data by which proton group assignments in a variety of compounds are relatable to proton chemical environments. The Varian "NMR Spectra Catalog" and the Sadtler data system have been valuable tools in this respect. The notation system described in this paper is particularly suitable for a computer processing by which each proton group in a compound can be related to its chemical shift.

As already pointed out, the following proton groups constitute the most important NMR elements in the great majority of chemicals: A (CH_3—), B (—CH=), C ($—CH_2—$), E ($=CH_2$), H, J (bridgehead >CH—), M (—NH—), U (≡CH), and Y (>CH—). Permutation of notations on these symbols for each chemical produces automatically a nine-part index. Our permutation program is essentially a wrap-around mechanism on a fixed position, such as Column 25 on a tab card, on which a notation is shifted by the computer as many times as there are any of the nine NMR elements. For example, the computer processes the notation for 2-ethylaminoethanol, HQCCMCA, as shown in Table X. The same permutation mechanism is also applied to the chemical shifts associated with each proton group in a molecule, as illustrated in Table XI. After each notation with its chemical shifts is permuted by the nine important proton groups as inputed to the computer, all inputed records are then alphabetized.

The results of this process are illustrated in Table XII for six alcohols which contain the moiety $HOCH_2CH_2—$. The permutation-alphabetization process yields a printout

Table XII. Alphabetized Computer Printout of Six Alcohols Containing the Moiety $HOCH_2CH_2$— (See Under H QCC Below)

Alphabetized Permuted Notation (Fixed Field)			Permuted Chemical Shift (Fixed Field)			Varian No.
HQCC	A		2.28-3.58-1.57-	0.92		43
HQCCMC	A		3.53-3.65-2.73-3.53-2.68-	1.12		92
HQCCN	A	2	3.60-3.60-2.45-2.25-	2.25		91
HQCCY(QA)	A		3.10-3.73-1.72-3.33-3.55-	1.18		120
HQCCY(QA)	A		3.58-3.80-1.68-3.65-4.03-	1.23		86
HQC	C	A	2.28-3.58-	1.57	-0.92	43
HQC	C	L	2.80-3.68-	3.68		12
HQC	C	MCA	3.53-3.65-	2.73	-3.53-2.68-1.12	92
HQC	C	NA2	3.60-3.60	2.45	-2.25-2.25	91
HQC	C	Y(QA)A	3.10-3.73-	1.72	-3.33-3.55-1.18	120
HQC	C	Y(QH)A	3.58-3.80-	1.68	-3.65-4.03-1.23	86
HQCCM	C	A	3.53-3.65-2.73-3.53-	2.68	-1.12	92
HQ	C	CA	2.28-	3.58	-1.57-0.92	43
HQ	C	CL	2.80-	3.68	-3.68	12
HQ	C	CMCA	3.53-	3.65	-2.73-3.53-2.68-1.12	92
HQ	C	CNA2	3.60-	3.60	-2.45-2.25-2.25	91
HQ	C	CY(QA)A	3.10-	3.73	-1.72-3.33-3.55-1.18	120
HQ	C	CY(QH)A	3.58-	3.80	-1.68-3.65-4.03-1.23	86
input data for this table	H	QCCA	$HOCH_2CH_2CH_3$	2.28	-3.58-1.57-0.92	43
	H	QCCL	$HOCH_2CH_2Cl$	2.80	-3.68-3.68	12
	H	QCCMCA	$HOCH_2CH_2NHCH_2CH_3$	3.53	-3.65-2.73-3.53-2.68-1.12	92
	H	QCCNA2	$HOCH_2CH_2N(CH_3)_2$	3.60	-3.60-2.45-2.25-2.25	91
	H	QCCY(QA)A	$HOCH_2CH_2CH(OCH_3)CH_3$	3.10	-3.73-1.72-3.33-3.55-1.18	120
	H	QCCY(QH)A	$HOCH_2CH_2CH(OH)CH_3$	3.58	-3.80-1.68-3.65-4.03-1.23	86
HQCC	M	CA	3.53-3.65-2.73-	3.53	-2.68-1.12	92
HQCC	Y	(QA)A	3.10-3.73-1.72-	3.33	-3.55-1.18	120
HQCC	Y	(QH)A	3.58-3.80-1.68-	3.65	-4.03-1.23	86

Table XIII. Computer Printout of —CH_2— Chemical Shifts by Increasing Value for the Six $HOCH_2CH_2$— Compounds of Table XII

Notation (Fixed Field)			Chemical Shift (Fixed Field)			Varian No.
HQC	C	A	2.28-3.58-	1.57	-0.92	43
HQC	C	Y(QH)A	3.58-3.80-	1.68	-3.65-4.03-1.23	86
HQC	C	Y(QA)A	3.10-3.73-	1.72	-3.33-3.55-1.18	120
HQC	C	NA2	3.60-3.60-	2.45	-2.25-2.25	91
HQCCM	C	A	3.53-3.65-2.73-3.53-	2.68	-1.12	92
HQC	C	MCA	3.53-3.65-	2.73	-3.53-2.68-1.12	92
HQ	C	CA	2.28-	3.58	-1.57-0.92	43
HQ	C	CNA2	3.60-	3.60	-2.45-2.25-2.25	91
HQ	C	CMCA	3.53-	3.65	-2.73-3.53-2.68-1.12	92
HQ	C	CL	2.80-	3.68	-3.68	12
HQC	C	L	2.80-3.68-	3.68		12
HQ	C	CY(QA)A	3.10-	3.73	-1.17-3.33-3.55-1.18	120
HQ	C	CY(QH)A	3.58-	3.80	-1.68-3.65-4.03-1.23	86

by each proton group and its proton environment within the full notation; similarly, the chemical shift of the proton group is displayed within the complete shift data. Thus, with the single input of data (notation and chemical shift of each proton group in each of the six molecules), the computer yields complete information under each of the five proton groups in the six molecules for a total of 27 lines of data from the six lines of input. In addition, the computer yields printouts by increasing value of the chemical shift of any proton group, as illustrated in Table XIII for —CH_2—. The over-all output of information by this method is indeed a rich one, and the process is relatively easy with a minimum input.

LITERATURE CITED

(1) Emsley, J. W., J. Feeney, and L. H. Sutcliffe, "High Resolution Nuclear Magnetic Resonance Spectroscopy," Vol. **2**, Foreword by R. E. Richards, Pergamon Press, New York, 1966.

(2) "Chemical Shift Index Chart," Sadtler Research Laboratories, Inc., Philadelphia, Pa., 1967.

(3) "NMR Spectra Catalog," pp. 2-5, Varian Associates, Palo Alto, Calif., 1962.

(4) Skolnik, H., "A New Linear Notation System Based on Combinations of Carbon and Hydrogen," *J. Heterocyclic Chemistry*, **6**, 689-95 (1969).

(5) Patterson, A. M., L. T. Capell, and D. F. Walker, "The Ring Index," 2nd ed., American Chemical Society, Washington, D. C., 1960.