Some Experience with the Hayward Linear Notation System*

H. WINSTON HAYWARD, HELEN M. S. SNEED, JAMES H. TURNIPSEED,
Office of Research and Development, U. S. Patent Office, Washington, D. C. 20231

and STEPHEN J. TAUBER
Information Technology Division, National Bureau of Standards, Washington, D. C. 20234

Received February 3, 1965

In written documents information about chemical compounds is generally stated by reference to structure diagrams or names. Each structure diagram is an undirected connected graph which represents, usually unambiguously, the topology of the molecule to which it refers. In digital computers (or punched card or tape machines) molecular structures must be represented linearly. If precisely the information in the structure diagram is to be present in the computer, then the linear representation must be isosemantic to the original structure diagram; *i.e.*, it must be unambiguous. If all information about a given substance must be readily interrelated, then it is advantageous to have a unique machine representation of the structure.

Part of the cooperative research effort of the U. S. Patent Office and the National Bureau of Standards during the past several years has been the development of a unique and unambiguous linear notation system for chemical structures. The notation system used, essentially that described by Hayward (*1*), was restricted to individual, completely determinate, organic structures in which no atom is connected to more than four bonds occurring in cycles. Rules have also been devised for certain types of Markush structures (2) and for coordination compounds (*3*). Various computer operations upon the notations have been described elsewhere (*4*).

"Ciphering Project." With the services of 19 students hired for the summer of 1963, we conducted a project with two major objectives:

I. to produce a sizeable file (50,000–100,000) of linear notations of wide variety for subsequent experimenting and

II. to determine the sufficiency of the notation rules to represent a large number of organic structures of varying complexity

and several subordinate objectives:

III. to determine whether training in organic chemistry is necessary to learn a useful part of the notation rules,

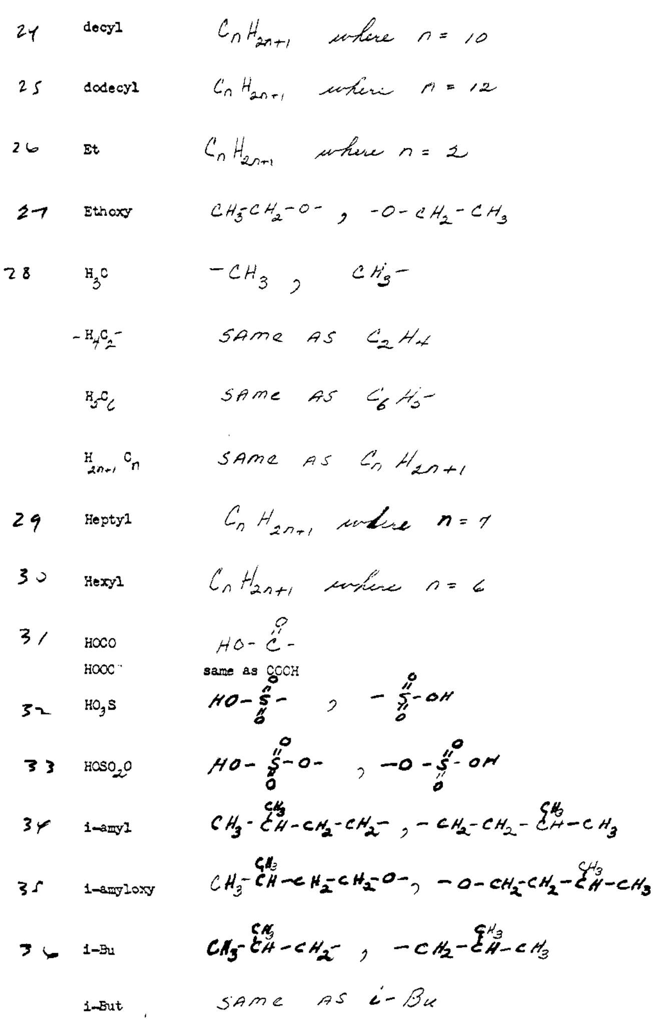
IV. to determine whether such a part of the notation rules can readily be taught in formal classes, and

V. to determine the approximate production rates and error rates involved in the several operations necessary for building a file of linear notations.

* Presented in part before the Division of Chemical Literature, 147th National Meeting of the American Chemical Society, Philadelphia, Pa., April 5-10, 1964.

PROCEDURE

The students learned the notation rules (with the restrictions noted below) in a series of four to six lectures lasting two to four hours and spread over one to two weeks, interspersed with time for "homework" assignments and followed by one to three weeks for practice. During the latter time the students also learned to use a dictionary of names for about 60 frequently occurring organic radicals (*cf.* Figure 1). Because the students



24	decyl	C_nH_{2n+1} where $n = 10$
25	dodecyl	C_nH_{2n+1} where $n = 12$
26	Et	C_nH_{2n+1} where $n = 2$
27	Ethoxy	CH_3CH_2-O- , $-O-CH_2-CH_3$
28	H_3C	$-CH_3$, CH_3-
	$-H_4C_2-$	same as C_2H_4
	H_5C_6	same as C_6H_5-
	$H_{2n+1}C_n$	same as C_nH_{2n+1}
29	Heptyl	C_nH_{2n+1} where $n = 7$
30	Hexyl	C_nH_{2n+1} where $n = 6$
31	HOCO	$HO-\overset{O}{\overset{\Vert}{C}}-$
	HOOC	same as COOH
32	HO_3S	$HO-SO_2-$, $-SO_2-OH$
33	$HOSO_2O$	$HO-SO_2-O-$, $-O-SO_2-OH$
34	i-amyl	$CH_3-CH(CH_3)-CH_2-CH_2-$, $-CH_2-CH_2-CH(CH_3)-CH_3$
35	i-amyloxy	$CH_3-CH(CH_3)-CH_2-CH_2-O-$, $-O-CH_2-CH_2-CH(CH_3)-CH_3$
36	i-Bu	$CH_3-CH(CH_3)-CH_2-$, $-CH_2-CH(CH_3)-CH_3$
	i-But	same as i-Bu

Figure 1. A page from the "Dictionary of Common Organic Radicals."

reported for work at various times, it was necessary to conduct three training cycles, partly overlapping, and the amount of training varied among the students. Table I indicates the amount of schooling, college major, amount of formal chemistry training, and amount of training in the Hayward notation system, as well as the age and sex of the students.

Specifically omitted from consideration during this project were stereochemical and isotopic features, and sandwich compounds, catenanes, bridged ions, and any other structures with bonds other than between two specific atoms, because at the time, under objectives I and II above, the features of particular interest were those inherent in structure diagrams when considered as graphs. Much of the training time was spent on the rules dealing with the priorities of structural features in determining the order of citing various parts of the structures.

During the training period, structure diagrams were used from the cumulative *Index Chemicus* (*5*), the Merck Index (*6*), Heilbron (*7*), and a Fieser and Fieser text (*8*); during the production period, from the first named publication and a few from "The Ring Index" (*9*). For the production period the students were divided into two physically remote groups (*cf.* Table I).

Group A, which consisted of students who had reported to work sooner and who had generally stronger chemistry background, was assigned to copying structures from *Index Chemicus*, sequentially from the first entry in each of the six volumes, and to enciphering them (writing linear notations for them). The structures were copied onto 5 × 8 in. blank cards (Figure 2a), identified by the *Index Chemicus* serial number, initialed, and dated. Copying chemical structures turned out not to be as simple as we had hoped. Figure 3, illustrates a typical page from *Index Chemicus*. In order to obtain complete structure diagrams, it was generally necessary to interpret terms such as Me, Et, Ph, *p*-anisyl, and mercapto. For this reason learning the dictionary of common organic radicals was included in the training. Structures including terms not known to the students were either interpreted by one of the present authors or were not copied. Also not copied were compounds whose structures were not given (unless easily interpreted nomenclature was given), partially indeterminate structures (including polymers), sandwich and other π-bonded structures specifically omitted from consideration, structures having atoms attached to more than four bonds in cycles, and inorganic compounds. Dotted lines conveying stereochemical information were rendered as solid lines; other stereochemical and any isotopic designations were omitted.

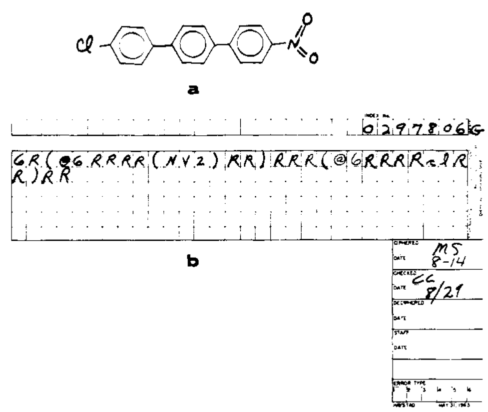


Figure 2. (a) Structure card. (b) Ciphering card. The area for the linear notation provides for 180 characters; the identifying number goes into the upper right corner; the initials of the person writing the notation and of those doing various types of checking go into the right margin together with the date. The blank space at the bottom is provided for writing the structure upon deciphering, and the boxes at the top are for the eventual "empirical structure."

Enciphering involved writing the linear notation corresponding to a structure onto a 5 × 8 in. card form (Figure 2b) and adding the appropriate identifying information. The six volumes of the cumulative *Index Chemicus*, supplemented by "The Ring Index," were circulated so that on the average each student of group A spent about

Table I. Background and Training of Students

Student	Group	Sex	Age	Years of college	Major	Years chemistry: High school	Years chemistry: College	Years chemistry: Organic	Ciphering training: Lecture, hr.	Ciphering training: Practice, days
1	A	F	20	3	Chem.	1	3	1	10	10
2	A	F	21	3	Chem.	1	1	1	11	6.5
3	A	M	19	1	Pre-med.	1	1	0	12.5	11.5
4	A	M	19	2	Chem.	2	2	1	11	6.5
5	A	M	24	2	Math.	1	0	0	11	6.5
6	A	F	20	2	Chem.	1	1	1	11	6.5
7	A	M	20	2	Pre-med.	1	1	0	10	11.5
8	A	F	20	3	Chem.	1	2	1	17.5	16.5
9	A	M	21	2	Chem.	1	1	1	17.5	16.5
10	A	M	20	3	Pre-med.	1	2	1	12.5	11.5
11	A	F	20	3	Math.	1	2	0	12.5	11.5
12	B	M	18	0	Math.	1	0	0	11	6.5
13	B	M	20	3	Physics	1	0	0	11	6.5
14	B	M	17	0	Math.	1	0	0	11	11.5
15	B	M	19	1	Physics	1	0	0	11	6.5
16	B	M	21	2	Math.	1	1	0	17.5[a]	5.5
17	B	M	20	4	Engnrng.	1	1	0	12.5	4.5
18	B	M	19	2	Econmcs.	1	0	0	12.5	4.5
19	B	M	17	0	Chem.	1	0	0	22.5[a]	6.5

[a] Attended part of one lecture series and all of another.

half the day copying structures and about half the day enciphering.

Group B was assigned to enciphering, checking, and deciphering. During the first several weeks of production, the students spent part of their time working from the structure cards generated by group A, and writing linear notations onto cards such as those shown in Figure 2b, but distinguishable by color. (A fraction of these structures were enciphered onto the two colors of the card forms by two different students in group B instead of by one in each group.) During this period they also compared corresponding notations character by character and sorted the notation cards into "match" and "mismatch" piles. (No student compared his own notations.) Later these students reviewed the structure and notation cards produced by group A, decided whether the notations were canonical, and sorted the notation cards into "accept" and "reject" piles. The students of group B also deciphered notations on cards in the "match" and "accept" piles.

RESULTS

Production Data. The production data are summarized in Table II. The numbers of production hours were calculated on the basis of the number of hours paid. Sick leave and time spent to attend seminars were subtracted, but other interruptions were included in the hours of production. The outputs of 5 × 8 in. cards were measured by the inch, at 109 cards per inch. [Calibrations by three persons of 6-in. stacks of cards were 109, 110, and 109 cards per inch. The two instances of identically the same batches of cards being measured (in Table IV, the subtotals for the notations reviewed and notations rejected of those generated by group A and for the notations reviewed by and notations rejected by group B) suggest that the manner of measuring cards was reasonably precise. All measurements of numbers of cards were by the inch unless otherwise stated.]

The students worked essentially at their own rates, without express incentives for high production. The threefold variation between minimum and maximum production rates therefore reflects differences both in aptitude and in diligence among the several students. Most of the students worked two to a room. The method of collecting data does not permit direct calculation of separate work rates for the several tasks.

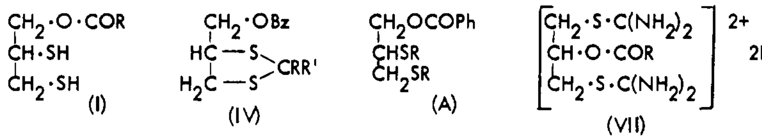
When asked individually whether the drawing of structures or the enciphering went faster, students of group A unanimously asserted that the enciphering was faster. This assertion is supported by the times required by each of three students to copy and to encipher 100 structures from "The Ring Index." Since the structures as shown in "The Ring Index" required essentially no interpretation, in contrast to those shown in *Index Chemicus*, the proportion of time required for copying structures to that for enciphering them may be considered as a lower limit for this ratio for structures in *Index Chemicus*. (The absolute times for working with the two sources cannot be compared directly, because short term rates are cited for enciphering from "The Ring Index" and because the collection in "The Ring Index" is not random.) The total production rate of group A (essentially only structure copying and enciphering) working from *Index Chemicus* (Table II) and the ratio of times for the two tasks working from "The Ring Index" (Table III) indicate a sustained enciphering rate of at least 30 notations per hour.

Table II. Production Figures

			Structures drawn[b]			Structures ciphered[b]							
Student	Group	Production. hr.[a]	*Index Chemicus*	"Ring Index"	Total	*Index Chemicus*	"Ring Index"	Total	Notations compared[b]	Notations reviewed[b]	Notations deciphered[b]	Total work units[b]	Work units per hr.
1	A	361	7,710	. . .	7,710	8,290	. . .	8,290			39	16,000	44.4
2	A	315	7,050	. . .	7,050	6,830	. . .	6,830			4	13.900	44.1
3	A	218	5,070	. . .	5,070	5,190	. . .	5,190			52	10,300	37.1
4	A	419	5,640	664	6,300	5,360	672	6,030			35	12,400	29.6
5	A	387	5,330	157	5,490	5,340	150	5,490			32	11,000	28.4
6	A	315	5,290	. . .	5,290	5,220	. . .	5,220			49	10,600	33.7
7	A	288	3,140	. . .	3,140	3,560	. . .	3,560			. . .	6,700	23.3
8	A	323	3,050	. . .	3,050	3,760	. . .	3,760			. . .	6,810	21.1
9	A	323	2,980	. . .	2,980	2,940	. . .	2,940			. . .	5,920	18.3
10	A	355	6,430	. . .	6,430	710	. . .	710			. . .	7,140	20.1
11	A	345	522	. . .	522	5,210	. . .	5,210			53	5,790	16.8
Subtotal	A	3,709	52,200	821	53,000	52,400	822	53,200			264	107,000	28.8[d]
12	B	341		. . .	. . .	3,110	. . .	3,110		8,120	6,130	17,400	51.0
13	B	390		300	300	2,220	300	2,520		5,530	5,650	14,000	35.9
14	B	357		. . .	. . .	1,480	. . .	1,480		3,830	3,920	9,230	25.9
15	B	399		302	302	1,040	302	1,340		4,650	3,560	9,850	24.7
16	B	407		837	837	1,070	837	1,910		3,380	3,190	9,310	22.9
17	B	236		199	199	524	199	723		3,630	1,790	6,340	26.9
18	B	281		203	203	529	203	732		2,340	2,000	5,280	18.8
19	B	311		. . .	. . .	1,070	. . .	1,070		1,640	3,010	5,720	18.4
Subtotal	B	2,722		1,840	1,840	11,000	1,840	12,900		33,100	29,300	77,100	28.3[d]
Total		6,431	52,200	2,660	54,800	63,400	2,660	66,100	20,400[c]	33,100	29,600	204,000[c]	31.7[c,d]

[a] Hours paid, less official leave, during production period. [b] Data are given to three significant figures, and totals are rounded off to three significant figures. [c] No record was kept which would permit crediting individual workers for comparing notations, but notation comparison was included in the total number of work units. [d] Average.

2977 ANTITUBERCULOUS SULPHUR COMPOUNDS. PART IV. SOME DIMERCAPTOPROPYL ESTERS AND RELATED DITHIO-URONIUM BROMIDES. E.P. Adams, F.P. Doyle, et al. (Beecham Res. Lab. Ltd., Brockham Park, Betchworth Surrey) Recd. Nov. 24, 1959. J. Chem. Soc. 1960(6),2674-80.

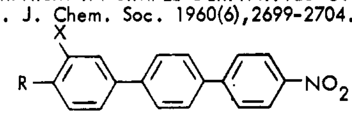


(I) (IV) (A) (VII)

			R=
1)	C19 H38 O2 S2	(I)	$(CH_2)_{14}Me$ p. 2677
2)	C10 H18 O2 S2	"	$(CH_2)_5 >$ CH-
3)	C5 H9 CLO2 S2	"	CH_2Cl
4)	C9 H16 O4 S2	"	$(CH_2)_2CO_2Et$
5)	C7 H14 O2 S3	"	$(CH_2)_2SMe$
6)	"	"	$CH_2 \cdot S \cdot Et$
7)	C12 H16 O2 S3	"	CH_2SCH_2Ph
8)	C11 H14 O2 S2	"	CH_2Ph
9)	C10 H12 O2 S2	"	Ph
10)	C11 H14 O3 S2	"	p-MeO-Ph
11)	C10 H11 CLO2 S2	"	o-Cl-Ph
12)	"	"	p-Cl-Ph
13)	C10 H10 CL2 O2 S2	"	2,4-diCl-Ph
14)	C14 H14 O2 S2	"	α-naphthyl
15)	C17 H19 NO4 S3	"	p-tos-NH-Ph
16)	C14 H16 O4 S2	(A)	Ac p. 2676
17)	C42 H72 O4 S2	"	palmitoyl
18)	C22 H18 N2 O4 S2	"	nicotinoyl

			R=	
19)	C10 H12 O2 S2	(A)	H	
20)	C12 H18 BR2 N4 O2 S2	(VII)	Ph	
21)	C13 H20 BR2 N4 O2 S2	"	p-tolyl	
22)	C13 H20 BR2 N4 O3 S2	"	p-anisyl	
23)	C12 H17 BR2 CLN4 O2 S2	"	p-Cl-Ph	
24)	C12 H16 BR2 CL2 N4 O2 S2	"	2,4-diCl-Ph	
25)	C12 H17 BR2 N5 O4 S2	"	$p-NO_2-Ph$	
26)	C12 H16 BR2 N6 O6 S2	"	$2,4-diNO_2-Ph$	
27)	C7 H16 BR2 N4 O2 S2	"	Me	
28)	C13 H20 BR2 N4 O2 S2	"	$PhCH_2$	
29)	C13 H16 O2 S2	(IV)	Me	R'=Me
30)	C17 H16 O2 S2	"	Ph	R'=H
31)	C10 H12 O2 S2	2-mercapto-1-(mercapto-Me)-Et benzoate		
32)	C13 H16 O2 S2	5-BzO-2,2-diMe-1,3-dithian		
33)	C17 H16 O2 S2	5-BzO-2-Ph-1,3-dithian		
34)	C11 H12 BR2 O2	(2-Br-1-Br-Me-Et) Ph-acetate		
35)	C11 H20 N4 S2	2,2-di(N'-isopropylidenethioureido) propane		

2978 MESOMORPHISM AND POLYMORPHISM IN SIMPLE DERIVATIVES OF p-TERPHENYL. P. Culling, G.W. Gray, et al. (Univ. Hull) Recd. Jan. 19,1960. J. Chem. Soc. 1960(6),2699-2704.

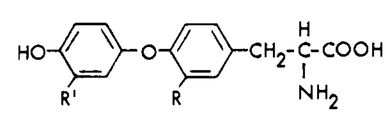


(A)

			R=	X=	
1)	C18 H14 N2 O2	(A)	NH_2	H	p. 2702
2)	C20 H16 N2 O3	"	AcNH-	"	p. 2703
3)	C18 H13 NO3	"	OH	"	
4)	C19 H15 NO3	"	MeO	"	
5)	C19 H12 N2 O2	(A)	CN	H	
6)	C18 H12 CLNO2	"	Cl	"	
7)	C20 H15 N3 O5	"	AcNH-	NO_2	p. 2704
8)	C18 H13 N3 O4	"	NH_2	"	

2979 THE SYNTHESIS OF THYROXINE AND RELATED SUBSTANCES. PART XIV. THE PREPARATION AND CHROMATOGRAPHY OF SOME IODINATED THYRONINES. J.S. Varcoe and W.K. Warburton. (Glaxo Lab. Ltd., Greenford, Middlesex) Recd. Jan. 7, 1960. J. Chem. Soc. 1960(6),2711-15.

		R=	R'=	
1)	C15 H14 INO4	3-I	H	p. 2714
2)	C15 H14 INO4	H	3-I	p. 2715
3)	C15 H13 I2 NO4	3-I	"	p. 2714
4)	C15 H13 I2 NO4	3',5'-diI-L-thyronine		p. 2715
5)	C15 H12 I3 NO4	3,3',5'-triI-L-thyronine		p. 2714



2980 2-PYRROLIDINO- AND 2-PIPERIDINO-PURINE. G. Levin and M. Tamari. (Hebrew Univ. Hadassah Med. School, Jerusalem) Recd. Jan. 20,1960. J. Chem. Soc. 1960(6),2782-2783.

1)	C8 H11 N5 O2	$4-NH_2-5-NO_2$-2-pyrrolidino-pyrimidine	p. 2782
2)	C9 H11 N5	2-pyrrolidino-purine	
3)	C9 H13 N5 O2	$4-NH_2-5-NO_2$-2-piperidino-pyrimidine	p. 2783
4)	C10 H13 N5	2-piperidino-purine	

Figure 3. A typical page from "Encyclopedia Chimica Internationalis."

Table III. Drawing and Ciphering Rates

Student	Drawing: No. of structures	Drawing: Time[a]	Drawing: Structures per hr.	Ciphering: No. of notations	Ciphering: Time[a]	Ciphering: Notations per hr.	Ratio drawing to ciphering
8	100	1 hr. 40 min.	60	100	1 hr. 30 min.	67	1.11
9	100	2 hr. 5 min.	48	100	1 hr. 50 min.	55	1.14
17	100	2 hr.	50	100	2 hr.	50	1.00
Total	300	5 hr. 45 min.	52[b]	300	5 hr. 20 min.	56[b]	1.08[b]

[a] To nearest 5 min. [b] Average.

According to the estimate of the editor (5, introduction) the full 17,099 articles in *Index Chemicus* contain about 180,000 compounds. The 52,000 structures enciphered were taken from 8682 articles and thus represent 57% of the approximately 91,000 compounds in these articles. A random sample of 200 articles contained 1752 compounds, of which 498 (28%) contained nomenclature terms not in the special dictionary. There were an additional 225 (13%) (structures not given or incomplete, 138; inorganic, 56; polymers, 2; sandwich and other π-bonded compounds, 25; mesomeric structures without Kekulé formulas, 4) which would have been excluded from enciphering for the other reasons stated above.

Reliability Data. The gross data from error checking are shown in Table IV. Among the duplicate notations generated, the percentage of notations which mismatched indicates an average error rate of at least 12.7%, since no more than half the mismatched notations could be correct. Those notations which mismatched were deciphered, when possible, and the structures obtained compared, with the results shown in Table V. These figures indicate that about half the mismatched notations resulted from misapplication of precedence or contraction rules and thus are unambiguous, though noncanonical, notations.

Table V. Comparisons from Mismatched Notations.

Structures agree	2610	50.3%
Structures disagree	1100	21.3
One notation not directly decipherable	1470	28.4
Total	5180	100.0%

Detailed comparison by one of the present authors, with the structures intended to be represented, of a 5.3% sample of the notations generated by group A which did not match the corresponding notations generated by group B gave the results shown in Table VI. These figures indicate that group B made at least 2.7 times as many errors in this sample as did group A, who in turn had an error rate of 6.6%. No specific cause for this difference in error rate has been established. The data do set limits on the error rate among the notations that were compared character by character of 12.7 and 16%. [The lower limit derives from the assumption that one cipher of each pair mismatched was correct (*i.e.*, one-half of 25.3%), the upper limit from the assumption that in 73% of the cases one cipher and in 27% of the cases neither cipher of the mismatched pair was correct.]

Table IV. Notation Checking

Student	Group	Notations written[a]	Notations[a]: Compared	Mismatched		Notations[a]: Reviewed	Rejected	
1	A	8,290	1,380	306	22%	4,860	128	2.6%
2	A	6,830	1,060	200	19	4,620	90	1.9
3	A	5,190	919	218	24	3,980	189	4.7
4	A	6,030	915	316	35	3,140	264	8.4
5	A	5,490	976	257	26	3,100	295	9.5
6	A	5,220	924	277	30	3,370	182	5.4
7	A	3,560	674	194	29	2,460	433	18
8	A	3,760	646	139	22	2,480	192	7.7
9	A	2,940	599	178	30	1.800	167	9.3
10	A	706	607	124	20	...	...	...
11	A	5,210	684	97	14	3,410	102	3.0
Subtotal	A	53,200	9,300	2,300	24.5[b]	33,200	2,040	6.14[b]
						Reviewed by	Rejected by	
12	B	3,110	3,110	695	22	8,120	492	6.1%
13	B	2,520	2,220	532	24	5,530	416	7.5
14	B	1,480	1,480	383	26	3,830	248	6.5
15	B	1,340	1,040	345	33	4,650	217	4.7
16	B	1,910	1,070	290	27	3,380	217	8.2
17	B	723	524	70	13	3,630	213	5.9
18	B	732	529	115	22	2,340	66	2.8
19	B	1,070	1,070	443	41	1,640	103	6.3
Subtotal	B	12,900	11,000	2,870	26.0[b]	33,100	2,030	6.13[b]
Total		66,100	20,400	5,180	25.3[b]			

[a] *Cf.* Table II, footnote *b*. Percentages for individuals are rounded to two significant figures; for the groups, to three.
[b] Average.

Table VI. Checking of Mismatched Notations.

Canonical notations	90[a]	73%
Unambiguous notations, but noncanonical	11[a]	9
Wrong notations	22[a]	18
Total	123	100%

[a] Determined by direct count.

Table VII. Checking of Accepted Notations

Sample size	779[a]	
Reviewing errors (enciphering errors)	9[a]	1.2%
Structure copying errors	32[a]	4.1
Deciphering errors	6[a]	0.8
Not directly decipherable	0[a]	0.0

[a] Determined by direct count.

Among the notations reviewed and accepted as correct, a sample of *ca.* 2.5% was taken from each day's output. These notations were deciphered and each resulting structure was compared with the corresponding structure given in the original source. If the structures agreed, the notation was examined for canonical form. If the structures disagreed, then the structure copied onto the corresponding blank card was compared with the source structure and the notation checked for canonical form against the copied structure. Deciphered structures were also checked against the corresponding notations. It was thus possible to assess the number of errors in reviewing (*i.e.*, accepting a noncanonical notation as correct), copying structures, and deciphering. The procedure is summarized in Figure 4, and the results are given in Table VII. These data indicate that *ca.* 98.8% of the notations accepted as correct were indeed canonical for the structures enciphered.

About one-third of the notations rejected as incorrect upon review were checked similarly and in addition a separate tally was made of those noncanonical notations which deciphered to the correct structure, giving the results shown in Table VIII. During review a systematic attempt had been made to sort incorrect notations according to the type of error made. Although this sorting turned out not to be sufficiently reliable for statistical analysis, it did produce distinctly nonrandom subpopulations. The data from each of these were extrapolated separately to give the extrapolated totals for all rejected notations. Of the notations stated to be incorrect, nearly 6% were actually canonical. A further 42%, although not canonical, nevertheless unambiguously represented the intended structures.

If the data of Tables VII and VIII are extrapolated over the entire file of notations reviewed, then the breakdown shown in Table IX emerges.

The types of errors noticed were almost entirely formal (leading to unambiguous but noncanonical notations) and clerical (omitted, extraneous, transposed, and wrong symbols). A very few other errors were also noted: salts enciphered as though covalent, a free acid enciphered instead of its anion, inorganic ions represented by their empirical formulas, and a bond intersection interpreted as an atom.

The enciphering error rate in this sample (7.0%) corresponds closely to the error rate (6.6%) in enciphering earlier by the same individuals (see above). The rate of falsely rejecting canonical ciphers when reviewing was used [0.35%, *i.e.*, 5.7% of 6.14% (Tables VIII and IV)] was

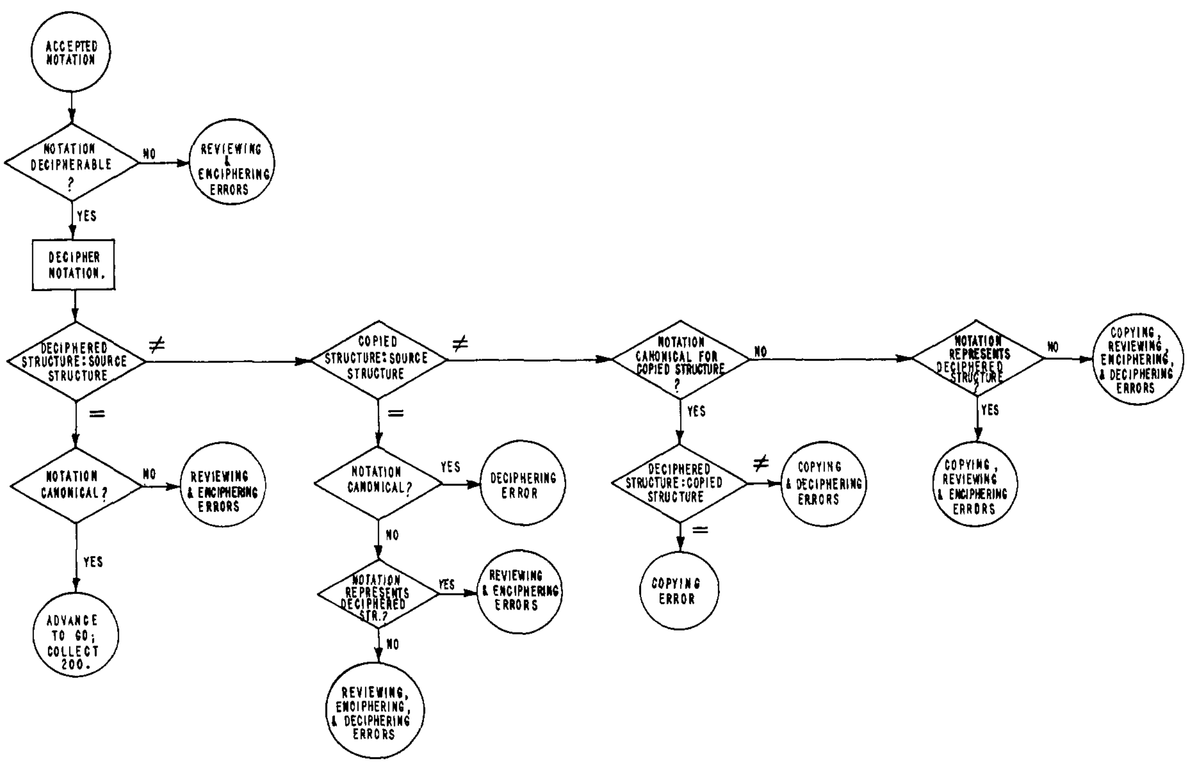


Figure 4. A schema of the analysis for errors among notations accepted as correct upon review.

Table VIII. Analysis of Rejected Notations

Alleged error type[a]	I	II	III	IV	V	VI	Total	Extrapolated total[c]	
Population	460	174[b]	536	755	117[b]	19[b]	2060		
Sample size[b,c]	179	167	99	105	112	16	678	(2060)	
Structure copying errors[b]	32	14	13	7	4	1	71	222	10.8%
Enciphering errors[b]	172	145	87	105	103	14	626	1940	94.3
Deciphering errors[b]	11	1	1	0	0	1	14	35	1.7
Reviewing errors[b] (canonical notations)	7	22	12	0	9	2	52	117	5.7
Structures agree[b] (except canonical notations)	159	135	26	12	81	10	423	879	42.5
Structures disagree[b]	4	3	41	31	13	3	95	475	23.0
Not directly decipherable[b]	9	7	20	62	9	1	108	594	28.8

[a] *Cf.* text. [b] Data determined by direct count. [c] Taken serially from the beginning of unordered stacks of cards.

Table IX. Reliability of Notations

Structure copying errors	4.5%
Canonical notations	93.0
Structures agree (except canonical notations)	3.2
Structures disagree	1.7
Not directly decipherable	2.1

much lower than the rate at which canonical ciphers were not matched by ciphers generated by the checking group [(17.5%, *i.e.*, 73% of 24.5% (Tables IX and IV)].

Characteristics of the Notation. It was found that very few structures were such that the more complex rules had to be invoked. For example, no structure was found which contained a ring system so complex that there was any atom not in a peripheral ring, although such structures do occur in the literature (*10*).

A sample of 1338 notations chosen from various segments of the output was analyzed for the lengths of the notations. The extremes found were 4 and 116 characters, the average *ca.* 25 characters, and the median *ca.* 23 characters.

CONCLUSIONS

Relying on experience with 19 college and precollege students who enciphered about 55,000 recently reported organic compounds of widely varying complexity, we conclude that a sustained enciphering rate of 30 notations per hour, under production conditions without an incentive system, and a rate of 93% error-free notations and 96% unambiguous notations indicate that the Hayward notation system can be readily and consistently applied to a very large fraction of the structure diagrams of known organic compounds. No unexpected difficulty was encountered in enciphering.

Prior training in organic chemistry is not necessary for learning the basic notation system. However, some special training was required to enable the students to recognize enough organic nomenclature terms to work with the quasi-structural formulas usual for communication among specialists. Considerable time and a nontrivial amount of intellectual effort was required to transform these into true structural formulas.

Lack of any appreciable change in the enciphering error rate after the training period indicates that a learning plateau was reached within 8–18 days. Checking notations by inspection proved more reliable than checking by duplicate enciphering, at least to the extent that only one-fiftieth as many canonical ciphers were incorrectly rejected by the former procedure. This advantage need not be intrinsic to the method, since the students checking had had more experience by the time they began inspecting notations.

Average and median lengths of 25 and 23 characters, respectively, we consider to be sufficiently concise for a notation that is essentially an atom by atom and bond by bond representation. It is so seldom necessary to use the more esoteric rules, which are included in the Hayward system for completeness, that they could be looked up in a manual as needed.

Since those rules dealing with precedence of citation of atoms and bonds, required for uniqueness, demand a major fraction of the effort in teaching and learning the notation rules, and since almost one-half of the enciphering errors were made in applying the precedence rules, it would be advantageous automatically to convert human-generated, non-unique but unambiguous notations into the unique, canonical notations required for file maintenance and complete-match searching (*4*).

LITERATURE CITED

(1) Hayward, H. W., "A New Sequential Enumeration and Line Formula Notation System for Organic Compounds," U. S. Patent Office Research and Development Reports, No. 21, Washington, D. C., 1961 (available from Superintendent of Documents, Washington, D. C. 20402, $1.25).

(2) Turnipseed, J. H., Sneed, H. M. S., U. S. Patent Office Research and Development Report (unpublished).

(3) McDonnell, P. M., Pasternack, R. F., *J. Chem. Doc.*, **5**, 56 (1965).

(4) Tauber, S. J., Fraction, G. F., Hayward, H. W., "Chemical Structures as Information—Representations, Transformations, and Calculations," in B. F. Cheydleur, Ed., "Technical Preconditions for Retrieval Center Operation," Spartan Books, Washington, D. C., in press.

(5) Garfield, E., Ed. "Encyclopedia Chimicus Internationalis (Cumulative Index Chemicus)," Institute for Scientific Information, Philadelphia, Pa., 1962.

(6) Stecher, P. G., Ed., "The Merck Index of Chemicals and Drugs," 7th Ed., Merck & Co., Inc., Rahway, N. J., 1960.

(7) Heilbron, I. M., "Dictionary of Organic Compounds," Oxford University Press, New York, N. Y., 1934.

(8) Fieser, L. F., Fieser, M., "Advanced Organic Chemistry," Reinhold Publishing Corp., New York, N. Y., 1961.

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

(10) *Cf.*, *e.g.*, U. S. Patent 3,066,099 (Nov. 27, 1962).

SLA OFFICERS

Alleen Thompson, Head Librarian, Atomic Power Equipment Department, General Electric Company, San Jose, Calif., assumed her duties as President of Special Libraries Association on June 9, 1965, at the close of the Annual Meeting at the 56th Annual Convention in Philadelphia.

Other newly elected officers for 1965–1966 are: President-Elect Dr. **Frank E. McKenna**, Supervisor, Information Center, Central Research Laboratories, Air Reduction Company, Inc., Murray Hill, N. J.; Chairman, Advisory Council, **Herbert S. White**, Director, NASA Facility, Documentation, Inc., Bethesda, Md.; Chairman-elect, Advisory Council, Mrs. **Helen F. Redman**, Head Librarian, Technical Library, Los Alamos Scientific Laboratory, Los Alamos, N. M.; and two new Directors: **Phoebe F. Hayes**, Director, Bibliographical Center for Research, Denver Public Library, Denver, Colo.; and **Ruth M. Nielander**, Librarian, Lumbermens Mutual Casualty Company, Chicago.

Continuing on SLA's Board of Directors are: Immediate Past-President, **William S. Budington**, Associate Librarian, The John Crerar Library, Chicago; Treasurer (serving the second year of a three-year term), **Jean E. Flegal**, Librarian, Business Library, Union Carbide Corporation; and Directors: **William K. Beatty**, Librarian and Professor of Medical Bibliography, Northwestern University Medical School, Evanston, Ill.; **Helene Dechief**, Head Librarian, Canadian National Railways, Montreal; **Kenneth N. Metcalf**, Librarian and Archivist, Henry Ford Museum and Greenfield Village, Dearborn, Mich.; and Mrs. **Dorothy B. Skau**, Librarian, Southern Regional Research Library, U. S. Department of Agriculture, New Orleans.

TRANSLATORS AND TRANSLATIONS: SERVICES AND SOURCES

Special Libraries Association has published the second edition of "Translators and Translations: Services and Sources in Science and Technology." Like the 1959 volume, this volume was edited by Frances E. Kaiser. The 224-page reference contains information on translating personnel and activities. It sells for $14.50.

Containing more than three times as many entries as the 1959 edition, the present book provides the names, addresses, telephone numbers, educational background, professional experience, volume of translating performed annually, subject and language proficiencies, and other vita on 470 free-lance translators and 87 commercial translating firms located principally in the United States but also in Canada and Europe. It describes 342 pools and other sources of translation information, and it cites 194 bibliographies and lists of translated literature.

CBS LABORATORIES INTRODUCES NEW DIAZO MICRODUPLICATORS

Two new diazo microfilm duplicating machines for information handling and distribution were shown by the Columbia Broadcasting System Laboratories at the National Microfilm Association Convention in Cleveland, May 11–13.

The two machines introduced by CBS Laboratories were the Model 303 Roll Microduplicator which produces diazo roll film duplicates at speeds up to 50 feet per minute and holds resolution to 500 lines per millimeter; the Model 601 Card Microduplicator which is designed for the production of the newer microfiche cards and other nonrigid cut films wherever high resolution duplication of micro-images is required.

NEW DURA TYPEWRITER FOR CHEMICAL FORMULAS

Dura Business Machines demonstrated a high-speed typewriter capable of automatically reproducing chemical formulas, or printed material containing such formulas, either from punched paper tape or edge-punched cards. It can also produce the same material when operated manually.

The formula-writing typewriter is a special adaptation of Dura's MACH 10 175-word-per-minute automatic typewriter, introduced two years ago for use in a variety of applications varying from personalized letter writing to paper tape data input to computers.

The new machine will be able to take paper tape prepared by a computer or otherwise and produce a hard-copy print-out, all automatically and at high speed.

The Dura MACH 10 features a stationary carriage, standard keyboard, and interchangeable typing spheres; operates on-line or off-line; and provides automatic punch control for one or two outputs. It also reads both tape and edge-punched cards in a wide range of programmed operations.

U. S. PATENTS

According to figures released by the Patent Office, U. S. Department of Commerce, a total of 41,022 patents were issued during the calendar year 1964 to residents of the United States, as compared with 40,039 patents issued in 1963.

Residents of foreign countries received 9,168 patents during 1964. Germany leads the list with 2,444 followed by the United Kingdom with 1,885; France with 1,041; Canada with 690; and Switzerland with 679. Currently, over four-fifths of the patents issued by the U. S. Patent Office go to citizens of this country.

Patents are now being issued at the rate of about 1,000 each week and patent applications are being received by the Patent Office at the rate of approximately 350 each working day.