ESSESA: An Expert System for Structure Elucidation from Spectra. 5. Substructure Constraints from Analysis of First-Order ^{1}H-NMR Spectra

Hong Huixiao* and Xin Xinquan

Chemistry Department, Nanjing University, Nanjing 210008, People's Republic of China

Received May 20, 1994®

This paper describes the knowledge base of first-order ^{1}H-NMR spectral analysis and the interpretation program for analysis of ^{1}H-NMR spectral data in ESSESA. Logical representation and the production system rules concerning analysis of ^{1}H-NMR spectra are discussed as well as inferential models useful in first-order ^{1}H-NMR spectral analysis. The unsaturation and the atomic composition of an unknown compound as well as the substructure constraints from the analysis of infrared spectrum, are passed to the interpretation program that develops the substructure constraints from its analysis of first-order ^{1}H-NMR spectral data by inference from the knowledge base of spectral analysis. The knowledge base contains 531 substructures.

INTRODUCTION

The elucidation of the structure of organic compounds based on spectroscopic methods, like ^{1}H-NMR, ^{13}C-NMR, IR, mass spectrometry, and UV/vis spectroscopy, is still essentially an empirical procedure. This is due to the fact that the derivation of constitution from the spectra[1,2] contains steps whose complexity prohibits formalization. Therefore the practical application of automatic interpretation programs[3] is far more limited than sometimes postulated in the literature.[4] Artificial intelligence, especially the expert systems method deriving rules out of a knowledge base, is generally applicable. ESSESA is an expert system for structure elucidation from spectral analysis.[5−8]

One of the most powerful tools for the determination of an unknown organic structure is the ^{1}H-NMR spectroscopy. From the first-order ^{1}H-NMR spectrum the types of hydrogen and structural environment in a structure can be derived. In some cases, the information from ^{1}H-NMR spectrum can lead to a unique structure, but in many cases there remains a large number of structural possibilities. In ESSESA the complete structure of a molecule is obtained by computer when data from several spectroscopic techniques are used simultaneously. The analysis of first-order ^{1}H-NMR spectral data gives the secondary substructure constrains which will be used in the generation of the complete structure.

KNOWLEDGE BASE OF FIRST-ORDER ^{1}H-NMR SPECTRA ANALYSIS

The conventional approach taken by chemists to first-order ^{1}H-NMR spectral interpretation is based upon models, often simplified, of the physical processes underlying resonance and the resulting spectral absorption. The physical models can be used to relate specific spectral signals to particular structural components of the molecule. Usually, there are several factors that together determine the detailed characteristics (such as chemical shift, multiplicity, and area of peak) of a first-order ^{1}H-NMR spectral signal and which are often in some sort of a hierarchical relationship that defines their relative importance. The initial analysis of a spectral signal may identify the presence of a specific type of hydrogen in the unknown molecule, and more detailed analysis of the form of the signal may determine aspects of the larger environment of that type of hydrogen.

The chemist's knowledge of first-order ^{1}H-NMR spectral analysis that has been incorporated into ESSESA is encoded in the form of spectral feature-substructure relationship rules written in PROLOG, which comprise the knowledge base for ^{1}H-NMR spectral analysis.

If a set of specific ^{1}H-NMR absorption peaks given by an unknown compound is $\mathbf{W}_h$, such that

$$\mathbf{W}_h = [W_{hi}(c, m, r)] \quad i = 1, 2, 3, ..., n \tag{1}$$

here c is the chemical shift, m is the multiplicity, and r is the area of the peak. The set of substructures in the knowledge base of first-order ^{1}H-NMR spectra analysis is $\mathbf{S}_h$

$$\mathbf{S}_h = [S_{hj}] \quad j = 1, 2, 3, ..., k \tag{2}$$

The set of specific absorption peaks that correspond to a substructure S_{hj} is $\mathbf{W}_{shj}$ that is expected to be

$$\mathbf{W}_{shj} = [W_h(c1, m, r), ..., W_h(c2, m, r)] \tag{3}$$

And if the set of substructures from IR spectral analysis is $\mathbf{S}_{ir}$

$$\mathbf{S}_{ir} = [S_1, S_2, ...Sp] \tag{4}$$

In order to analyze a first-order ^{1}H-NMR spectrum of an unknown compound, it is necessary to pick out the subset S_{hh} from the set S_h, such that

$$\exists W_{hi} \exists W_{shj} \exists \mathbf{S}_{ir}\{W_{hi} \subseteq W_{shj} \wedge S_j \subseteq \mathbf{S}_{ir}\} \rightarrow \mathbf{S}_{hh} \supseteq S_j \tag{5}$$

According to this procedure, the construction of the knowledge base for first-order ^{1}H-NMR spectral analysis requires that the logical representation formula (eq 3) of the set of substructures S_h be found and encoded in PROLOG rules. In ESSESA, the information in eq 3 is expressed by the production system rule—i.e., the spectral feature−substructure relationships. Given a sufficiently large number of spectra, or alternatively the appropriate information derived from published observations, it is possible to develop the ability to associate certain absorption peaks with the corresponding hydrogen types and their structural environment. Thus a set of rules can be developed to derive the

® Abstract published in *Advance ACS Abstracts,* October 15, 1994.

0095-2338/94/1634-1259$04.50/0

Table 1

no.	substructure	chem. shift	multiplicity	area	IR constraint	no.	substructure	chem. shift	multiplicity	area	IR constraint
1	**$CH_3C\leqslant$**	0.1−2.2	1	3		52	**$>CHCH_2C(=O)O-$**	2.2−3.0	2	2	−C(=O)O−
2	**CH_3CH_2-**	0.6−1.6	3	3		53	**$CH_3CH_2C(=O)O-$**	2.3−3.1	4	2	−C(=O)O−
3	**$CH_3CH<$**	0.7−2.1	2	3		54	**$\geqslant CCH_2C(=O)O-$**	2.5−3.0	1	2	−C(=O)O−
4	**$CH_3C\equiv C-$**	1.6−2.6	1	3		55	**$CH_3CH_2N<$**	2.4−3.4	4	2	−N<
5	**$CH_3C(=O)-$**	1.7−2.7	1	3	−C(=O)−	56	**$-CH_2CH_2N<$**	2.4−3.8	3	2	−N<
6	**CH_3S-**	1.8−2.8	1	3	−S−	57	**$>CHCH_2N<$**	2.4−2.9	2	2	−N<
7	**$CH_3C(=O)O-$**	1.9−2.6	1	3	−C(=O)O−	58	**$\geqslant CCH_2N<$**	2.4−3.5	1	2	−N<
8	**CH_3Ar**	1.9−2.8	1	3	Ar	59	**CH_3CH_2NH-**	2.4−3.4	5	2	−NH−
9	**$CH_3N<$**	2.2−3.2	1	3	−N−	60	**$-CH_2CH_2NH-$**	2.4−3.8	4	2	−NH−
10	**$CH_3OC(-)=O$**	3.5−4.0	1	3	−OC(=O)−	61	**$>CHCH_2NH-$**	2.4−2.9	3	2	−NH−
11	**CH_3O-**	3.2−4.1	1	3	−O−	62	**$\geqslant CCH_2NH-$**	2.4−3.5	2	2	−NH−
12	**$CH_3C=C$**	1.6−2.0	1	3	C=C	63	**$-CH_2CH_2NH_2$**	2.4−3.8	5	2	$-NH_2$
13	**$CH_3CH=C$**	1.6−2.0	2	3	C=C	64	**$>CHCH_2NH_2$**	2.4−2.9	4	2	$-NH_2$
14	**CH_3NH-**	2.5−2.7	2	3	−NH−	65	**$\geqslant CCH_2NH_2$**	2.4−3.5	3	2	$-NH_2$
15	**$CH_3C=N$**	1.8−2.0	1	3	C=N	66	**CH_3CH_2Ar**	2.4−3.0	4	2	Ar
16	**$CH_3CH=N$**	1.8−1.9	2	3	C=N	67	**$-CH_2CH_2Ar$**	2.4−3.0	3	2	Ar
17	**$CH_3CH_2CH_2-$**	0.9−2.1	6	2		68	**$>CHCH_2Ar$**	2.9−3.7	2	2	Ar
18	**$>CHCH_2CH<$**	1.0−1.5	3	2		69	**$\geqslant CCH_2Ar$**	2.9−3.7	1	2	Ar
19	**$CH_3CH_2C\leqslant$**	1.1−1.7	4	2		70	**$>NCH_2N<$**	2.9−3.3	1	2	−N<
20	**$>CHCH_2C\leqslant$**	1.1−2.1	2	2		71	**$>NCH_2NH-$**	2.9−3.3	2	2	−N<, −NH−
21	**$CH_3CH_2CH<$**	1.2−2.4	5	2		72	**$>NCH_2NH_2$**	2.9−3.3	3	2	−N<, $-NH_2$
22	**$-CH_2CH_2CH_2-$**	1.2−2.4	5	2		73	**$-NHCH_2NH-$**	2.9−3.3	3	2	−NH−
23	**$-CH_2CH_2C\leqslant$**	1.2−2.4	3	2		74	**$-NHCH_2NH_2$**	2.9−3.3	4	2	−NH−
24	**$\geqslant CCH_2C\leqslant$**	1.2−1.9	1	2		75	**CH_3CH_2Cl**	3.1−3.7	4	2	
25	**$>CHCH_2CH_2-$**	1.2−2.4	4	2		76	**$-CH_2CH_2Cl$**	3.1−4.0	3	2	
26	**$C\equiv CCH_2N<$**	3.1−4.1	1	2	C≡C	77	**$>CHCH_2Cl$**	3.2−4.1	2	2	
27	**$C\equiv CCH_2NH-$**	3.1−4.1	2	2	C≡C	78	**$\geqslant CCH_2Cl$**	3.2−4.2	1	2	
28	**$C\equiv CCH_2NH_2$**	3.1−4.1	3	2	C≡C, NH_2	79	**$-O(O=)CCH_2S-$**	3.0−3.9	1	2	−C(=O)O−
29	**$CH_3CH_2C\equiv C$**	1.9−2.5	4	2	C≡C	80	**$-O(O=)CCH_2SH$**	3.0−3.9	2	2	−C(=O)O−
30	**$-CH_2CH_2C\equiv C$**	1.9−2.8	3	2	C≡C	81	**$-O(O=)CCH_2N<$**	3.1−4.0	1	2	−C(=O)O−
31	**$C\equiv CCH_2CH<$**	1.9−3.0	2	2	C≡C	83	**$-O(O=)CCH_2NH-$**	3.1−4.0	2	2	−C(=O)O−
32	**$\geqslant CCH_2C\equiv C$**	1.9−3.0	1	2	C≡C	84	**$-O(O=)CCH_2NH_2$**	3.1−4.0	3	2	−C(=O)O−
33	**CH_3CH_2CN**	2.0−3.0	4	2	CN	85	**$-O(O=)CCH_2C(=O)O-$**	3.2−3.5	1	2	−C(=O)O−
34	**CH_2CH_2CN**	2.0−3.0	3	2	CN	86	**$ArCH_2S-$**	3.2−4.0	1	2	Ar
35	**$>CHCH_2CN$**	2.0−3.0	2	2	CN	87	**$ArCH_2SH$**	3.2−4.0	2	2	Ar
36	**$\geqslant CCH_2CN$**	2.0−3.0	1	2	CN	88	**$-CH_2CH_2O-$**	3.3−4.3	3	2	−O−
37	**CH_3CH_2S-**	2.1−3.0	4	2	−S−	89	**$>CHCH_2O-$**	3.4−4.4	2	2	−O−
38	**$-CH_2CH_2S-$**	2.4−2.9	3	2	−S−	90	**$\geqslant CCH_2O-$**	3.4−4.2	1	2	−O−
39	**$>CHCH_2S-$**	2.4−3.1	2	2	−S−	91	**$-CH_2CH_2OH$**	3.4−4.4	4	2	−OH
40	**$\geqslant CCH_2S-$**	2.4−3.1	1	2	−S−	92	**$>CHCH_2OH$**	3.4−4.2	3	2	−OH
41	**$-CH_2CH_2SH$**	2.4−2.9	4	2	SH	93	**$\geqslant CCH_2OH$**	3.4−4.2	2	2	−OH
42	**$>CHCH_2SH$**	2.4−3.1	3	2	SH	94	**$-CH_2CH_2N<$**	2.4−3.8	3	2	−N−
43	**$\geqslant CCH_2SH$**	2.4−3.1	2	2	SH	95	**$-CH_2CH_2NH-$**	2.4−3.8	4	2	−NH−
44	**$-CH_2CH_2C(=O)-$**	2.4−3.5	3	2	−C(=O)−	96	**$-CH_2CH_2NH_2$**	2.4−3.8	5	2	$-NH_2$
45	**$\geqslant CCH_2C(=O)-$**	2.2−3.2	1	2	−C(=O)−	97	**$O=CCH_2C=O$**	3.3−3.9	1	2	−C=O
46	**$CH_3CH_2C(=O)-$**	2.3−3.0	4	2	−C(=O)−	98	**$-(O=)CCH_2CH=O$**	3.3−3.9	2	2	−C=O
47	**$>CHCH_2C(=O)-$**	2.2−2.8	2	2	−C(=O)−	99	**$-O(O=)C-CH_2CN$**	3.4−3.6	1	2	CN, −C(=O)O−
48	**$-CH_2CH_2CH=O$**	2.2−3.5	4	2	−CH=O	100	**$ArCH_2N<$**	3.4−4.0	1	2	Ar
49	**$>CHCH_2CH=O$**	2.2−2.8	3	2	−CH=O	101	**$ArCH_2NH-$**	3.4−4.0	2	2	Ar, −NH−
50	**$\geqslant CCH_2CH=O$**	2.2−3.2	2	2	−CH=O	102	**$ArCH_2NH_2$**	3.4−4.0	3	2	Ar, $-NH_2$
51	**$-CH_2CH_2C(=O)O-$**	2.1−2.9	3	2	−C(=O)O−	103	**$ArCH_2C(=O)O-$**	3.4−3.8	1	2	Ar, −C(=O)O−

104	>NCH_2CN	3.4−3.8	1	2	−CN
105	−$NHCH_2CN$	3.4−3.8	2	2	−NH−, −CN
106	−CH_2CH_2O−	3.3−4.3	3	2	−O−
107	CH_3CH_2O−	3.4−4.1	4	2	−O−
108	>$NCH_2C(=O)$−	3.4−4.0	1	2	−C=O
109	>$NCH_2CH=O$	3.5−4.0	2	2	−CH=O
110	−NH−$CH_2C(=O)$−	3.5−4.0	2	2	−C=O, −N−
111	−$NHCH_2CH=O$	3.5−4.0	3	2	−CH=O, −N−
112	$NH_2CH_2C(=O)$−	3.5−4.0	3	2	−C(=O)−, −NH
113	$NH_2CH_2CH=O$	3.5−4.0	4	2	−CH=O, −NH
114	ArCH_2CN	3.5−4.0	1	2	Ar, −CN
115	$ClCH_2C(=O)$−	3.6−4.7	1	2	−C=O
116	$ClCH_2C(=O)O$−	3.6−4.2	1	2	−C(=O)O−
117	Ar$CH_2C(=O)$−	3.6−4.2	1	2	Ar, −C(=O)−
118	Ar$CH_2CH=O$	3.6−4.2	1	2	Ar, −CH=O
119	$ClCH_2C{\equiv}C$−	3.7−4.0	1	2	−C≡C−
120	−$OCH_2CH=O$	3.7−4.2	2	2	−O−, −CH=O
121	−$OCH_2C(=O)$−	3.7−4.2	1	2	−O−, −C(=O)−
122	$HOCH_2C(=O)$−	3.7−4.2	2	2	−OH, −C(=O)−
123	ArCH_2Ar	3.7−4.2	1	2	Ar
124	−$CH_2CH_2OC(=O)$−	4.0−4.4	3	2	−O−C(=O)−
125	$CH_3CH_2OC(=O)$−	4.0−4.6	4	2	−O−C(=O)−
126	>$CHCH_2OC(=O)$−	3.8−4.5	2	2	−O−C(=O)−
127	⩾$CCH_2OC(=O)$−	3.8−4.6	1	2	−O−C(=O)−
128	−$OCH_2C(=O)O$−	3.8−4.7	1	2	−O−, −C(=O)O−
129	$HOCH_2C(=O)O$−	3.8−4.7	2	2	HO−, −C(=O)O−
130	−OCH_2CN	4.0−4.5	1	2	−O−, −CN
131	−$OCH_2C{\equiv}C$−	4.0−4.5	1	2	−O−, −C≡C−
132	$HOCH_2C{\equiv}C$−	4.0−4.5	2	2	HO−, C≡C−
133	$-CH_2CH_2NO_2$	4.3−5.3	3	2	$-NO_2$
134	>$CHCH_2NO_2$	4.2−5.2	2	2	$-NO_2$
135	⩾CCH_2NO_2	4.2−5.2	1	2	$-NO_2$
136	ArCH_2O−	4.3−4.8	1	2	−O−, Ar
137	ArCH_2OH	4.3−4.8	2	2	HO−, Ar
138	ArCH_2Cl	4.4−4.9	1	2	Ar
139	−$C{\equiv}CCH_2OC(=O)$−	4.4−5.0	1	2	−O−C=O, C≡C−
140	−OCH_2O−	4.5−6.1	1	2	−O−
141	$HOCH_2O$−	4.5−6.1	1	2	−OH
142	Ar$CH_2OC(=O)$−	4.9−5.4	1	2	−O−C=O, Ar
143	(−CH_2)$CHCH_2$−	0.9−1.4	7	1	
144	$(CH_3)CHCH_2$−	1.0−1.9	9	1	
145	(−CH_2)(>CH)CHCH<	1.0−1.9	5	1	
146	>OH(⩾C)CHCH<	1.1−2.5	3	1	
147	(CH_3)CHCH<	1.2−2.0	8	1	
148	(−CH_2)$CHCH_3$	1.2−1.9	8	1	
149	(CH_3)CHC⩽	1.2−2.1	7	1	
150	>$CH(CH_3)CHCH_2$−	1.4−1.9	7	1	
151	−$CH_2(CH_3)$CHC⩽	1.4−2.0	6	1	
152	>$CH(CH_3)$CHCH<	1.5−2.0	6	1	
153	$(CH_3)CHCH_3$	1.6−1.8	10	1	
154	CH_3(>CH)CHC⩽	1.5−2.0	5	1	
155	(−CH_2)CHCH<	1.0−1.6	6	1	
156	(−CH_2)CHC⩽	1.0−1.6	5	1	
157	−CH_2(>CH)CHC⩽	1.0−1.9	4	1	
158	(>CH)CHCH<	1.1−2.5	4	1	
159	(CH_3)CHC(=C−)−	1.8−2.6	7	1	>C=C<
160	(CH_3)CHCH=C−	1.8−2.6	8	1	>C=C<
161	(−CH_2)CH_3CHC(=C−)−	1.8−2.6	6	1	>C=C<
162	−CH_2CH_3CHCH=C−	1.8−2.6	7	1	>C=C<
163	CH_3(>CH)CHC=C−	1.8−2.6	5	1	>C=C<
164	CH_3(>CH)CHCH=C−	1.8−2.6	6	1	>C=C<
165	⩾CCH_3CHC(=C−)−	1.8−2.6	5	1	>C=C<
166	⩾CCH_3CHCH(=C−)−	1.8−2.6	6	1	>C=C<
167	(−CH_2)CHC=C−	1.8−2.6	5	1	>C=C<
168	(−CH_2)CHCH(=C−)−	1.8−2.6	6	1	>C=C<
169	>CH<(−CH_2)CHC(=C−)−	1.8−2.6	4	1	>C=C<
170	>CH(−CH_2)CHC(H=C−)−	1.8−2.6	5	1	>C=C<
171	−CH_2(⩾C)CHC(=C−)−	1.8−2.6	3	1	>C=C<
172	−CH_2(⩾C)CHC(H=C−)−	1.8−2.6	4	1	>C=C<
173	(>CH)CHC(=C−)−	1.8−2.6	3	1	>C=C<
174	(>CH)CHC(H=C−)−	1.8−2.6	4	1	>C=C<
175	⩾C⩽(>CH)CHC(=C−)−	1.8−2.6	2	1	>C=C<
176	⩾C(>CH)CHC(H=C−)−	1.8−2.6	3	1	>C=C<
177	(CH_3)CHC≡C−	2.4−3.2	7	1	−C≡C−
178	−$CH_2(CH_3)$CHC≡C−	2.4−3.2	6	1	−C≡C−
179	>$CHCH_3$CHC≡C−	2.4−3.2	5	1	−C≡C−
180	⩾CCH_3CHC≡C−	2.4−3.2	4	1	−C≡C−
181	(−CH_2)$_2$CHC≡C−	2.4−3.2	5	1	−C≡C−
182	−CH_2(>CH)CHC≡C−	2.4−3.2	4	1	−C≡C−
183	⩾C(−CH_2)CHC≡C−	2.4−3.2	3	1	−C≡C−
184	(>CH)CHC≡C−	2.4−3.2	3	1	−C≡C−
185	⩾(>CH)CHC≡C−	2.4−3.2	2	1	−C≡C−
186	(−CH_2)CHAr	2.0−3.0	5	1	Ar
187	−$CH_2(CH_3)$CHC(=O)−	2.2−3.0	6	1	−C(=O)−
188					
189	−$CH_2(CH_3)$CHCH=O	2.2−3.0	7	1	−CH=O
190	>CH(−CH_2)CHAr	2.3−2.7	4	1	Ar
191	−CH_3(−CH_2)CHAr	2.2−30	6	1	Ar
192	>$CHCH_3$CHAr	2.2−3.0	5	1	Ar
193	⩾CCH_3CHAr	2.2−3.0	4	1	Ar
194	−CH_2(⩾C)CHAr	2.2−3.0	3	1	Ar
195	(>CH)CHAr	2.2−3.0	3	1	Ar
196	>CH(⩾C)CHAr	2.2−3.0	2	1	Ar
197	−$CH_2(CH_3)$CHAr	2.2−3.0	6	1	Ar
198	>$CH(CH_3)$CHC(=O)−	2.2−3.0	5	1	−C(=O)−
199	>$CH(CH_3)$CHO=O)−	2.2−3.0	6	1	−CH=O
200	⩾$C(CH_3)$CHC(=O)−	2.2−3.0	4	1	−C(=O)−
201	⩾$C(CH_3)$CHC(=O)−	2.2−3.0	5	1	−CH=O
202	(−CH_2)CHC(=O)−	2.2−3.0	5	1	−C(=O)−
203	(−CH_2)CHCH=O	2.2−3.0	6	1	−CH=O
204	(>CH)CHC(=O)−	2.2−3.0	3	1	−C(=O)−
205	(>CH)CHCH=O	2.2−3.0	4	1	−CH=O
206	⩾C(>CH)CHC(=O)−	2.2−3.0	2	1	−C(=O)−
207	⩾C(>CHCHCH=O	2.2−3.0	3	1	−CH=O
208	−$CH_2(CH_3)$CHCN	2.2−3.0	6	1	−CN
209	(−CH_2)CHCN	2.2−3.0	5	1	−CN
210	>$CH(CH_3)$CHCN	2.2−3.0	5	1	−CN
211	⩾CCH_3CHCN	2.2−3.0	4	1	−CN

Table 1 (Continued)

no.	substructure	chem. shift	multiplicity	area	IR constraint	no.	substructure	chem. shift	multiplicity	area	IR constraint
212	>CH(−CH_2)**C**HCN	2.2−3.0	4	1	−CN	264	(−CH_2)**C**HSH	2.8−3.4	6	1	−SH
213	>C(−CH_2)**C**HCN	2.2−3.0	3	1	−CN	265	>CH(−CH_2)**C**HS−	2.8−3.4	4	1	−S−
214	(−CH)**C**HCN	2.2−3.0	3	1	−CN	266	>CH(−CH_2)**C**HSH	2.8−3.4	5	1	−SH
215	>C(>CH)**C**HCN	2.2−3.0	2	1	−CN	267	>C(−CH_2)**C**HS−	2.8−3.4	3	1	−S−
216	(CH_3)**C**HCN	2.6−2.8	7	1	−CN	268	>C(−CH_2)**C**HSH	2.8−3.4	4	1	−SH
217	>CH(−CH_2)**C**HC(=O)−	2.3−3.0	4	1	−C(=O)−	269	(>CH_2)**C**HS−	2.8−3.4	3	1	−S−
218	>CH(−CH_2)**C**HCH=O	2.3−3.0	5	1	−CH=O	270	(>CH_2)**C**HSH	2.8−3.4	4	1	−SH
219	>CH(−CH_2)**C**HC(=O)O−	2.3−3.0	4	1	−C(=O)O−	271	>CH(>C)**C**HS−	2.8−3.4	2	1	−S−
220	−CH_2(CH_3)**C**HC(=O)O−	2.3−3.0	6	1	−C(=O)O−	272	>CH(>C)**C**HSH	2.8−3.4	3	1	−SH
221	>CH(CH_3)**C**HC(=O)O−	2.3−3.0	5	1	−C(=O)O−	273	>CH(>C)**C**HO−	3.2−4.2	2	1	−O−
222	>C(CH_3)**C**HC(=O)O−	2.3−3.0	4	1	−C(=O)O−	274	>CH(>C)**C**HOH	3.2−4.2	3	1	−OH
223	(−CH_2)**C**HC(=O)O−	2.3−3.0	5	1	−C(=O)O−	275	>C(CH_3)**C**HO−	3.3−4.5	4	1	−O−
224	(>CH)**C**HC(=O)O−	2.3−3.0	3	1	−C(=O)O−	276	>C(CH_3)**C**HOH	3.3−4.5	5	1	−OH
225	>CH(>C)**C**HC(=O)O−	2.3−3.0	2	1	−C(=O)O−	277	(−CH_2)$_2$**C**HO−	3.2−4.2	5	1	−O−
226	>C(−CH_2)**C**HC(=O)O−	2.3−3.0	3	1	−C(=O)O−	278	(−CH_2)$_2$**C**HOH	3.2−4.2	6	1	−OH
227	(CH_3)**C**HC(=O)O−	2.4−2.8	7	1	−C(=O)O−	279	(>CH)$_2$**C**HO−	3.3−4.5	3	1	−O−
228	>C(−CH_2)**C**HC(=O)−	2.6−3.4	3	1	−C(=O)−	280	(>CH)$_2$**C**HOH	3.3−4.5	4	1	−OH
229	>C(−CH_2)**C**HCH=O	2.6−3.4	4	1	−CH=O	281	−O(O=)C**C**(−CH_3)HC(=O)O−	3.3−4.1	4	1	−C(=O)O−
230	(CH_3)**C**HAr	2.7−3.3	7	1	Ar	282	−O(O=)C**C**(−OH_2)HC(=O)O−	3.3−4.1	3	1	−C=O
231	−CH_2(CH_3)**C**HN<	2.7−3.5	6	1	−N−	283	(−OH_2)(O=)C**C**(>CH)HC(=O)O−	3.3−4.4	2	1	−C=O
232	−CH_2(CH_3)**C**HNH−	2.7−3.5	7	1	−NH−	284	−O(O=)C**C**(>C)HC(=O)O−	3.3−4.1	1	1	−C=OO−
233	−CH_2(CH_3)**C**HNH_2	2.7−3.5	8	1	−NH_2	285	(CH_3)**C**HC(=O)−	2.4−3.6	7	1	−C=O−
234	>CH(CH_3)**C**HN<	2.7−3.8	5	1	−N−	286	(−CH_2)(CH_3)**C**HO−	3.4−3.9	6	1	−O−
235	>CHCH_3**C**HNH−	2.7−3.8	6	1	−NH−	287	(−CH_2)(CH_3)**C**HOH	3.4−3.9	7	1	−OH
236	>CH(CH_3)**C**HNH_2	2.7−3.8	7	1	−NH_2	288	−O(O=C)**C**(−CH_2)HCl	4.1−5.1	3	1	−C(=O)O−
237	>C(CH_3)**C**HN<	2.7−3.8	4	1	−N−	289	−O(O=C)**C**(>C)HCl	4.2−5.2	1	1	−C=OO−
238	>C(CH_3)**C**HNH−	2.7−3.8	5	1	−NH−	290	−O(O=C)**C**(−CH_3)HS−	3.4−3.8	4	1	−C(=O)O−, −S−
239	>C(CH_3)**C**HNH_2	2.7−3.8	6	1	−NH_2	291	−O(O=C)**C**(−CH_3)HSH	3.4−3.8	5	1	−C(=O)O−, −SH
240	(−CH_2)**C**HN<	2.7−3.8	5	1	−N−	292	−O(O=C)**C**(−CH_2)HS−	3.4−3.8	3	1	−C(=O), −S−
241	(−CH_2)**C**HNH−	2.7−3.8	6	1	−NH−	293	−O(O=C)**C**(−CH_2)HSH	3.4−3.8	4	1	−C(=O)O−, −SH
242	(−CH_2)**C**HNH_2	2.7−3.8	7	1	−NH_2	294	−O(O=C)**C**(>CH)HS−	3.4−3.8	2	1	−C(=O), −S−
243	>CH(−CH_2)**C**HN<	2.7−3.8	4	1	−N−	295	−O(O=C)**C**(>CH)HSH	3.4−3.8	3	1	−C(=)O−, −SH
244	>CH(−CH_2)**C**HNH−	2.7−3.8	5	1	−NH−	296	−O(O=C)**C**(>C)HS−	3.4−3.8	1	1	−C(=O)O−, −S−
245	>CH(−CH_2)**C**HNH_2	2.7−3.8	6	1	−NH_2	297	−O(O=C)**C**(>C)HSH	3.4−3.8	2	1	−C(=O)O−, −SH
246	>C(−CH_2)**C**HN<	2.7−3.8	3	1	−N−	298	(−CH_2)(O=C)**C**HN<	3.4−50	3	1	−OC=O, −N−
247	>C(−CH_2)**C**HNH−	2.7−3.8	4	1	−NH−	299	−O(O=C)**C**(−CH_2)HNH−	3.4−50	4	1	C=O, −NH−
248	>C(−CH_2)**C**HNH_2	2.7−3.8	5	1	−NH_2	300	−O(O=CC(−CH_2)HNH_2	3.4−50	5	1	C=O, −NH_2
249	(>CH)**C**HN<	2.7−3.8	3	1	−N−	301	−O(O=C)**C**(>CH)HN<	3.4−5.0	2	1	C=O, −N−
250	(>CH)**C**HNH−	2.7−3.8	4	1	−NH−	302	−O(O=C)**C**(>CH)HNH−	3.4−5.0	3	1	C=O, NH−
251	(>CH)**C**HNH_2	2.7−3.8	5	1	−NH_2	303	−O(O=C)**C**(>CH)HNH_2	3.4−5.0	4	1	C=O, −NH_2
252	>CH(>C)**C**HN<	2.7−3.8	2	1	−N−	304	−O(O=C)**C**(>C)HN<	3.4−5.0	1	1	C=O, −N−
253	>CH(>C)**C**HNH−	2.7−3.8	3	1	−NH−	305	−O(O=CC(>C)HNH−	3.4−5.0	2	1	C=O, −NH−
254	>CH(>C)**C**HNH_2	2.7−3.8	4	1	−NH_2	306	−O(O=C)**C**(>C)HNH_2	3.4−5.0	3	1	C=O, −NH_2
255	(CH_3)**C**HS−	3.3−3.4	7	1	−S−	307	(CH_3)**C**HN<	2.8−4.5	7	1	−N−
256	>CH(−CH_2)**C**HAr	2.7−3.8	4	1	Ar	308	(CH_3)**C**HNH−	2.8−4.3	8	1	−NH−
257	−CH_2(CH_3)**C**HS−	2.8−3.4	6	1	−S−	309	(CH_3)**C**HO−	3.5−4.9	7	1	−O−
258	−CH_2(CH_3)**C**HSH	2.8−3.4	7	1	−SH	310	>CH(CH_3)**C**HO−	3.5−4.1	5	1	−O−
259	>CH(CH_3**C**HS−	2.8−3.4	5	1	−S−	311	>CHCH_3**C**HOH	3.5−4.1	6	1	−OH
260	>CH(CH_3)**C**HSH	2.8−3.4	6	1	−SH	312	−O(O=)CC(CH_3)HN<	3.5−4.5	4	1	−OC=O, −N−
261	>C(CH_3)**C**HS−	2.8−3.4	4	1	−S−	313	−O(O=C)C(CH_3)HNH−	3.5−4.5	5	1	−OC=O, −NH−
262	>C(CH_3)**C**HSH	2.8−3.4	5	1	−SH	314	−O(O=C)C(CH_3)HNH_2	3.5−4.0	6	1	−OC=O, −NH_2
263	(−CH_2)**C**HS−	2.8−3.4	5	1	−S−	315	>CH(−CH_2)**C**HO−	3.5−4.7	4	1	−O−

316	>CH(−CH_2)**C**HOH	3.5−4.6	5	1	−OH
317	Ar(CH_3)**C**HOC(=O)−	5.7−.61	4	1	−OC(=O)−, Ar
318	Ar(−CH_2)**C**HOC=O−	5.7−.61	3	1	−OC(=O)−, Ar
319	Ar>CH−**C**HOC=O−	5.7−6.1	2	1	−OC(=O)−, Ar
320	Ar(>C)**C**HOC=O−	5.7−6.1	1	1	0OC(=O−, Ar
321	−O(CH_3)**C**HC(=O)−	4.1−4.9	4	1	−O−, −C(=O)−
322	HO(CH_3)**C**HC(=O)	4.1−4.9	5	1	−OH, −C(=C)−
323	−O(CH_3)**C**HCH=O	4.1−4.9	4	1	−O−, −CH=O
324	−O(>CH)**C**HC=O	4.2−5.0	2	1	−O−, −C(=O)
325	HO(>CH)**C**HC=O	4.2−5.0	3	1	−OH, −C(=O)−
326	HO(>CH)**C**HCH=O	4.2−5.0	3	1	−O−, −CH=O
327	HO(>CH)**C**HCH=O	4.2−5.0	4	1	−OH, −CH=O
328	HO(>C)**C**HC(=O)−	4.2−5.0	1	1	−O−, −C(=O)−
329	HO(>C)**C**HC(=O)−	4.2−5.0	2	1	−OH, −C(=O)−
330	−O9>C)**C**HCH=O	4.2−5.0	2	1	−O−, −CH=O
331	HO(>C)**C**HCH=O	4.2−5.0	3	1	−OH, −CH=O
332	>CHCH_3**C**HOC(=O)−	3.9−5.0	5	1	−C(=O)O−
333	>CH(−CH_2)**C**HOC(=O)−	4.0−5.0	4	1	−C=(O)O−
334	>C(CH_3)**C**HOC(=O)−	4.0−5.0	4	1	−C=(O)O−
335	>C(−CH_2)**C**HOC(=O)−	4.0−5.0	3	1	−C(=O)O−
336	>C(>CH)**C**HOC(=O)−	4.0−5.0	2	1	−C(=O)O−
337	−O(O=)CC(−CH_2)HC(=O)−	3.6−4.3	3	1	−OC=O, −C(=O)−
338	−O(O=)CC(−CH_3)HC(=O)−	3.6−4.3	4	1	−OC=O, −C(=O)−
339	−O(O=)CC(−CH_3)HCH=O	3.6−4.3	5	1	−OC=O, −CH=O
340	−O(O=)CC(−CH_2)HCH=O	3.6−4.3	4	1	−OC=O, −CH=O
341	−O(O=)CC(>CH)HC(=O)−	3.6−4.3	2	1	−OC=O, −C(=O)−
342	−O(O=)CC(>CH)HCH=O	3.6−4.3	3	1	−OC=O, −CH=O
343	−O(O=)CC(<C)HC=O)−	3.6−4.3	1	1	−OC=O, −C(=O)−
344	−O(O=)CC(<)HCH=O	3.6−4.3	2	1	−OC=O, −CH=O
345	−O(O=)CC(>CH_2)HO−	3.6−4.2	2	1	−OC=O, −O−
346	−O(O−dbd)CC(>CH_2)HOH	3.6−4.2	3	1	−OC=O, −OH
347	−O(O=)CC(−CH_3)HO−	3.6−4.2	4	1	−OC=O, −O−
348	−O(O=)CC(−CH_3)HOH	3.6−4.2	5	1	−OC=O, −OH
349	−O(O=)CC(−CH_2)HO−	3.6−4.2	3	1	−OC=O, −O−
350	−O(O=)CC(−CH_2)HOH	3.6−4.2	4	1	−OC(=O), −OH
351	O=CCH−O−	3.6−4.2	1	1	−OC(=O, −O−
352	−OO=CCHOH	3.6−4.2	2	1	−O−C(=O), −OH
353	(>CH)C(<CH)HCl	3.8−4.8	3	1	
354	(CH_3)(>CH)CHCl	3.8−4.8	5	1	
355	>C(CH_3)CHCl	3.8−4.8	4	1	
356	>C(CH_2)CHCl	3.8−4.8	3	1	
357	>C(>CH)CHCl	3.8−4.8	2	1	
358	>CH(−CH_2)CHCl	3.8−4.8	4	1	
359	−O(O=)CC(−OC=O)HC(=O)O−	4.0−4.5	1	1	−OC(=O)
360	Ar(CH_3)CHN<	3.9−4.4	4	1	Ar, −N−
361	Ar(CH_3)CHNH−	3.9−4.4	5	1	Ar, −NH−
362	Ar(CH_3)CHNH_2	3.9−4.4	6	1	Ar, −NH_2
363	Ar(−CH_2)CHN<	3.8−4.5	3	1	Ar, −N−
364	Ar(−CH_2)CHNH−	3.8−4.5	4	1	Ar, −NH−
365	Ar(−CH_2)CHNH_2	3.8−4.5	5	1	Ar, −NH_2
366	Ar(>CH)CHN<	3.8−4.5	2	1	Ar, −N−
367	Ar(>CH)CHNH−	3.8−4.5	3	1	Ar, −NH−
368	Ar(>CH)CHNH_2	3.8−4.5	4	1	Ar, −NH_2
369	(−CH_2)CHCl	3.9−4.5	5	1	
370	(>CH)CHOC(=O)−	3.9−4.7	3	1	−O−C(=O)−
371	−CH_2CH_3CHCl	4.0−4.3	6	1	
372	>C(−CH_2)CH−O−	4.0−5.0	3	1	−O−
373	>C(−CH_2)CHOH	4.0−5.0	4	1	−OH
374	HO(>CH)CHC≡C−	4.0−4.7	3	1	−O−, −C≡C−
375	−O(>CH)CHC≡C−	4.0−4.7	2	1	−OH, −C≡C−
376	−O(CH_3)CHC≡C−	4.0−4.7	4	1	−O−, −C≡C−
377	HO(CH_3)CHC≡C−	4.0−4.7	5	1	−OH, −C≡C−
378	HO(CH_2)CHC≡C−	4.0−4.7	3	1	−O−, −C≡C−
379	HO(−CH_2)CHC≡C−	4.0−4.7	4	1	−OH, −C≡C−
380	HO(>C)CHC≡C−	4.0−4.7	1	1	−O−, −C≡C−
381	HO(>C)CHC≡C−	4.0−4.7	2	1	−OH, −C≡C−
382	Ar(CH_3)CHAr	4.0−5.5	4	1	Ar
383	Ar(−CH_2CHAr	4.0−5.5	3	1	Ar
384	Ar(>CH)CHAr	4.0−5.5	2	1	Ar
385	Ar(>C)CHAr	4.0−5.5	1	1	Ar
386	−S(−S)CHS−	4.0−4.5	1	1	−S−
387	−S(−S)CHSH	4.0−4.5	2	1	−S−, −SH
388	HS(−S)CHSH	4.0−4.5	3	1	−S−, −SH
389	Cl(−CH_2)CHC(=O)−	4.1−5.0	3	1	−C(=O)−
390	Cl(−CH_2)CHCH=O	4.2−4.8	4	1	−CH=O
391	Cl(CH_3)CHC(=O)−	4.5−5.5	4	1	−C(=O)−
392	Cl(CH_3)CHCH=O	4.5−5.5	5	1	−CH=O
393	Cl(>CH)CHC(=O)−	4.5−5.5	2	1	−C=O
394	Cl(>CH)CHCH=O	4.5−5.5	3	1	−CH=O
395	(>C)CHC(=O)−	4.5−5.5	1	1	−C(=O)
396	Cl(>C)CHCH=O	4.5−5.5	2	1	−CH=O
397	(CH_3)CHCl	4.2−4.7	7	1	−C(=O)O−
398	Cl(CH_3)CHC(=O)O−	4.2−5.0	4	1	−C(=O)O−
399	Cl(>CH)CHC(=O)O−	4.2−5.2	2	1	−C(=O)O−
400	O=CO(>CH)CHC(=O)O−	4.2−5.6	2	1	−C(=O)
401	O=CO(CH_3)CHC(=O)O−	4.2−5.6	4	1	−C(=O)O−
402	(CH_2)CHC=O	4.2−5.6	3	1	−C(=O)O−
403	O=CO(>CCHC=OO−	4.2−5.6	1	1	−C(=O)O−
404	−CH_2CHNO_2	4.3−47	6	1	−NO_2
405	−O(−CH_2)CHO−	4.4−5.0	3	1	−O−
406	−O(−CH_2)CHOH	4.4−5.0	4	1	−O−, −OH
407	HO(−CH_2)CHOH	4.4−5.0	5	1	−OH
408	−O(>C)CHO−	4.4−5.0	1	1	−O−
409	−O(>C)CHOH	4.4−5.0	2	1	−O−, −OH
410	HO(>C)CHOH	4.4−5.0	3	1	−OH
411	−O(CH_3)CHO−	4.4−4.9	4	1	−O−
412	−O(CH_3)CHOH	4.4−4.9	5	1	−O−, −OH
413	Ar(CH_3)CHO−	4.3−5.0	4	1	Ar, −O−
414	Ar(CH_3)CHOH	4.3−5.0	5	1	Ar, −OH
415	Ar(−CH_2)CHO−	4.3−5.2	3	1	Ar, −O−
416	Ar(−CH_2)CHOH	4.3−5.2	4	1	Ar, −OH
417	−O(O=)CCHO−	4.5−5.0	1	1	−O(O=)C, −O−
418	−O(O=)C−C(O−)HOH	4.5−5.0	2	1	−O(O=)C, −O−, −OH
419	−O(O=)C−C(OH)HOH	4.5−5.0	3	1	−OO=C, −OH
420	(CH_3)CHNO_2	4.6−4.8	7	1	−NO_2
421	(−CH_2)CHOC(=O)−	4.6−5.3	5	1	−O−C(=O)−
422	(CH_3)CHOC(=O)−	4.7−5.2	7	1	−O−C(=O)−
423	Ar(>C)CHO−	4.3−5.5	1	1	Ar, −O−
424	Ar(>C)CHOH	4.3−5.5	2	1	Ar, −OH
425	−O(>CH)CHO−	4.8−5.7	2	1	−O−
426	−O(>CH)CHOH	4.8−5.7	3	1	−O−, −OH
427	HO(>CH)CHOH	4.8−5.7	4	1	−OH

Table 1 (Continued)

no.	substructure	chem. shift	multiplicity	area	IR constraint	no.	substructure	chem. shift	multiplicity	area	IR constraint
428	Ar(>CH)**CH**O−	4.8−5.6	2	1	Ar, −O−	480	>**CH**−**SH**	0.9−2.5	2	1	−SH
429	Ar(>CH)**CH**OH	4.8−5.6	3	1	Ar, −OH	481	>**C**−**SH**	0.9−2.5	1	1	−SH
430	(Ar)**CH**C(=O)−	4.8−5.3	1	1	−OH	482	Ar−**SH**	3.0−4.0	1	1	Ar, −SH
431	(Ar)**CH**CH=O	4.8−5.3	2	1	−CH=O, Ar	483	**HC**≡C−CH_∞−	1.7−2.6	1	1	C≡C
432	(Ar)**CH**C(=O)O−	4.8−5.3	1	1	−C(=O)O−, Ar	484	**HC**≡C−CH<	1.7−2.6	1	1	C≡C
433	(Ar)**CH**N<	4.8−5.4	1	1	−N<, Ar	485	**HC**≡C−C>	1.7−2.6	1	1	C≡C
434	(Ar)**CH**NH−	4.8−5.4	2	1	−NH−, Ar	486	**HC**≡C−Ar	2.7−3.4	1	1	Ar, C≡C
435	(Ar)**CH**NH_2	4.8−5.3	3	1	Ar, $-NH_2$	487	**HC**≡C−C=C<	2.6−3.1	1	1	C=C, C≡C
436	−O(Ar)**CH**C(=O)O−	4.9−5.5	1	1	−O−, −C(=O)O−, Ar	488	**HC**≡C−C=O	2.1−3.3	1	1	−C(=O)−, C≡C
437	HO(Ar)**CH**C(=O)O−	4.9−5.5	2	1	−OH, C(=O)O− Ar	489	**HC**≡C−C≡C−	1.7−2.4	1	1	C≡C
438	CH_3(−CH_2)**CH**OC(=O)−	5.0−5.6	6	1	−O−C(=O)−	490	**HC**≡C−O−	1.1−1.5	1	1	−O−, C≡C
439	Cl(CH_3)**CH**Ar	5.0−5.6	4	1	Ar	491	**HO**−CH_2−	3.5−5.5	3	1	−OH
440	Cl(−CH_2)**CH**Ar	5.0−5.6	3	1	Ar	492	**HO**−CH<	3.5−5.5	2	1	−OH
441	(>CH)**CH**Ar	5.0−5.6	2	1	Ar	493	**HO**−C≤	3.5−5.5	1	1	−OH
442	Cl(>C)**CH**Ar	5.0−5.6	1	1	Ar	494	**HO**−Ar	4.5−7.5	1	1	Ar, −OH
443	−O(−CH_2)**CH**C(=O)−	5.0−5.9	3	1	−O−, −C(=O)−	495	Ar−**NH**−Ar	2.9−4.8	1	1	Ar, −NH−
444	−O(−CH_2)**CH**CH=O	5.0−5.9	4	1	−O−, −CH=O	496	Ar−**NH**−CH_3	2.9−4.8	4	1	Ar, −NH−
445	HO(−CH_2)**CH**C(=O)−	5.0−5.9	4	1	−OH, −C(=O)−	497	Ar−**NH**−CH_2−	2.9−4.8	3	1	Ar, −NH−
446	HO(−CH_2)**CH**CH=O	5.0−5.9	5	1	−OH, −CH=O	498	Ar−**NH**−CH<	2.9−4.8	2	1	Ar, NH−
447	Cl(>CH)**CH**Cl	5.0−6.3	2	1		499	Ar−**NH**−C≤	2.9−4.8	1	1	Ar, −NH−
448	(Ar)**CH**Ar	5.0−5.7	1	1	Ar	500	>C−NH−CH_2−	0.4−3.5	3	1	−NH−
449	Cl(CH_3)**CH**Cl	5.0−6.1	4	1		501	>C−**NH**−CH<	0.4−3.5	2	1	−NH−
450	Cl(>C)**CH**Cl	5.0−6.1	1	1		502	>C−**NH**−C≤	0.4−3.5	1	1	−NH−
451	Cl(−CH_2)**CH**OH	5.2−6.0	4	1	−OH	503	>CH−**NH**−CH_2−	0.4−3.5	4	1	−NH−
452	Cl(−CH_2)**CH**O−	5.2−6.0	3	1	−O−	504	>CH−**NH**−CH<	0.4−3.5	3	1	−NH−
453	Cl(CH_3)**CH**O−	5.2−6.0	4	1	−O−	505	−CH_2−**NH**−CH_2−	0.4−3.5	5	1	−NH−
454	Cl(CH_3)**CH**OH	5.2−6.0	5	1	−OH	506	−CH_2−**NH**−CH_3	0.4−3.5	6	1	−NH−
455	Cl(>CH)**CH**O−	5.2−6.0	2	1	−O−	507	>CH−**NH**−CH_3	0.4−3.5	5	1	−NH−
456	Cl(>CH)**CH**OH	5.2−6.0	3	1	−OH	508	>C−**NH**−CH_3	0.4−3.5	4	1	−NH−
457	Cl(>C)**CH**O−	5.2−6.0	1	1	−O−	509	Ar−**NH**−C(=O)−	7.8−9.4	1	1	Ar, −NH−
458	Cl(>C)**CH**OH	5.2−6.0	2	1	−OH	510	CH_3−NH−C(=O)−	6.0−8.2	4	1	−NH−, −C(=O)−
459	Cl(−CH_2)**CH**Cl	5.0−6.1	3	1		511	−CH_2−**NH**−C(=O)−	6.0−8.2	3	1	−NH−, −C(=O)−
460	−O(O=)**CCH**−C(=O)O−	4.9−5.7	1	1	−C(=O)O−	512	>CH−**NH**−C(=O)−	6.0−8.2	2	1	−NH−, −C(=O)−
461	NC(>N)**CH**C(=O)O−	5.3−5.7	1	1	−N−, −C=OO−	513	>C−**NH**−C(=)−	6.0−8.2	1	1	−NH−, −C(=O)−
462	NC(−NH)**CH**C(=O)O−	5.3−5.7	2	1	−NH−, −C=OO−	514	−CH_2−**NH**−CH=O	6.0−8.2	4	1	−NH−, −CH=O
463	NC(NH_2)**CH**−C(=O)O−	5.3−5.7	3	1	NH_2, −C(=O)O−	515	>CH−**NH**−CH=O	6.0−8.2	3	1	−NH−, −CH=O
464	−O(−O)**CH**O−	5.4−5.8	1	1	−O−	516	>C−**NH**−CH=O	6.0−8.2	2	1	−NH−, −CH=O
465	−O(−O)**CH**OH	5.4−5.8	2	1	−OH, −O−	517	−CH_2−**NH_2**	0.4−3.5	3	2	−NH_2
466	Ar(−O)**CH**C=O	5.4−6.1	1	1	Ar, −O−, −C(=O)−	518	>CH−**NH_2**	0.4−3.5	2	2	−NH_2
467	Ar(HO)**CH**C=O	5.4−6.1	2	1	Ar, −OH, −C(=O)−	519	>C−**NH_2**	0.4−3.5	1	2	−NH_2
468	Ar(−O)**CH**CH=O	5.6−6.0	2	1	Ar, −O−, −CH=O	520	Ar−**NH_2**	2.9−4.8	1	2	Ar, −NH_2
469	Ar(HO)**CH**CH=O	5.6−6.0	3	1	Ar, −OH, −CH=O	521	−(O=)C−**NH_2**	5.0−6.5	1	2	−C=O, −NH_2
470	Ar(−O)**CH**Ar	5.6−6.0	1	1	Ar, −O−	522	O=**CH**−O−	7.7−8.3	1	1	O=**CH**−, −O−
471	Ar(HO)**CH**Ar	5.6−6.0	2	1	Ar, −OH	523	O=**CH**−CH_2−	9.3−10.6	3	1	O=**CH**−
472	Cl(CH_3)**CH**NO_2	5.6−6.0	4	1	−NO_2	524	O=**CH**−CH<	9.3−10.6	2	1	O=**CH**−
473	Cl(−CH_2)**CH**NO_2	5.6−6.0	3	1	−NO_2	525	O=**CH**−C≤	9.3−10.6	1	1	O=**CH**−
474	Cl(>CH)**CH**NO_2	5.7−6.1	2	1	−NO_2	526	−(O=)C−**OH**	10.0−13.5	1	1	−(O=)C−Oh
475	Cl(>C)**CH**NO_2	5.7−6.1	1	1	−NO_2	527	Ar−**H**	6.0−8.5	1	1	Ar
476	(Cl)**CH**C(=O)O−	5.7−6.1	1	1	−C(=O)O−	528	**CH_2**=C<	4.8−8.0	1	2	CH_2=C
477	−(O=)COC(Ar)**H**C(=O)O−	5.7−6.1	1	1	Ar, −C(=O)O−	529	−**CH**=C<	4.8−9.8	1	1	−CH=C
478	(Ar)**CH**Cl	5.9−6.1	1	1	Ar	530	**CH_2**=CH−	4.8−8.0	2	2	CH_2=C
479	−CH_2**SH**	0.9−2.5	3	1	−SH	531	−**CH**=CH−	4.8−9.8	2	1	−CH=C

subset S_{hh} of substructures that are indicated by a specific first-order ^{1}H-NMR spectrum.

In ESSESA, the first-order ^{1}H-NMR spectrum is not the only one to derive the substructure constraints, IR and ^{13}C-NMR spectra are used also. The IR spectrum is the first one to be analyzed, and the result (eq 4)[5] from the IR spectral analysis is used as the constraint in the first-order ^{1}H-NMR spectral analysis.

On the basis of eq 5, to derive the subset S_{hh} of substructures that are consistent with the IR and the first-order ^{1}H-NMR spectral data and other chemical information is to confirm that the expecting absorption peaks of S_{hh} exist in the spectral data. Equation 5 can be written in production system rules; for example, the production system rule that is derived the identification of the substructure CH_3-Ar is as follows:

IF in the first-order ^{1}H-NMR spectrum of an unknown compound there are peak with chemical shift 1.9–2.8 ppm, multiplicity 1, and area equivalent to three hydrogen atoms, and there exists the aromatic substructure from the IR spectral analysis

Then the substructure CH_3-Ar may be present in the structure of the unknown compound

In ESSESA this rule can be written in PROLOG such as follows:

```
Peakh("CH3-Ar", C, M, R):-C>=1.9,
                          C<=2.8, !,
                          M=1(Single),
                          R=3(H atoms).
Subh("CH3-Ar"):-Peakh("CH3-Ar", C, M, R), !,
                Subir("Ar").
```

There are 531 substructures in the knowledge base used by ESSESA for the first-order ^{1}H-NMR spectrum analysis. It exceeds the databases used by Sasaki[9] and Miller.[10] These 531 substructures are shown in Table 1. There exists overlap within this table, for example substructures 17, 21, and 19 overlap substructure 2 and so on. This problem will be solved before the complete structure candidates are generated on the basis of information derived from the IR, ^{1}H-NMR, and ^{13}C-NMR spectral analysis. The mutually consistent set of substructures that will be used by the structure generator can be achieved by means of a procedure that finds those combinations of permitted substructures that are compatible with the overall composition of the molecule and with the constraints derived from all of the spectral data. Each such combination of substructures then may be used to define a distinct problem that can be referred to a subsequent structure generation program.

INTERPRETER PROGRAM

After the ring and double-bond characteristics of the entered molecular formula and the digitized first-order ^{1}H-NMR spectrum as well as the analysis result of IR spectrum are passed to the interpreter program of first-order ^{1}H-NMR spectral analysis, the interpretation of the first-order ^{1}H-NMR spectrum is started. Using eq 5 and the knowledge base the interpreter program begins to identify the various substructural fragments that may be present in the structure. The result of interpretation is a substructure list that is used in the structure generation. The structure of this interpreter program is shown in Figure 1.

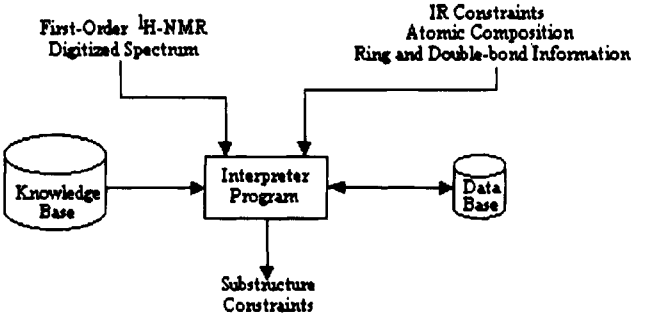


Figure 1. Overview of the structure of the interpreter program.

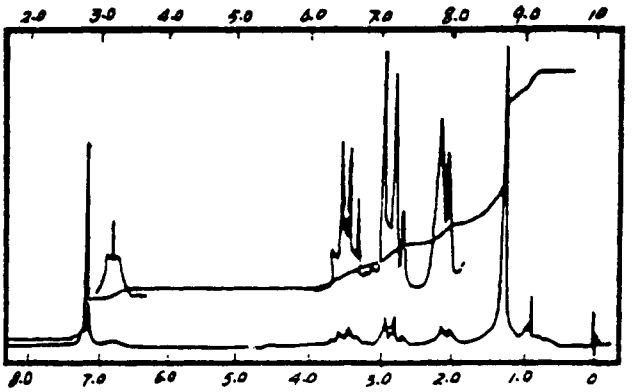


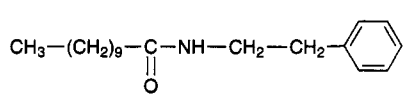
Figure 2. The ^{1}H-NMR spectrum of $C_{19}H_{31}NO$.

The interpreter program works with the set of 531 substructures. Each of these substructures is correlated with a defined pattern of first-order ^{1}H-NMR spectral absorption. The spectral features that characterize substructural fragments normally consist of spectral ranges within which specific types of peaks are expected. The initial set of substructures is, in effect, screened against the entered spectral data, and any substructure whose requisite spectral pattern is absent is discarded. The result of this analysis is a subset of the 531 substructures. Each of the members of the subset is related to absorption data, which is consistent with the first-order ^{1}H-NMR spectrum of the unknown compound.

In the interpretation, goal-driving inference tactics were used. In such an inference model, substructures from the knowledge base are used as the goals, and the program will seek spectral patterns that fulfill the premises of the goals, as defined by the production rules. If any single premise of a goal is not satisfied, that goal is determined to be false, that is to say, that a particular substructure cannot be contained in the unknown structure. If all the premises of a goal are satisfied, then the goal is considered to be true, and the substructure corresponding to the goal may be embedded within the complete structure of the unknown compound.

EXAMPLE

N-Phenethylundecanamide $C_{19}H_{31}NO$, structure 1, is taken as an example to illustrate the procedure in which the constraints from first-order ^{1}H-NMR spectral analysis are obtained in ESSESA. Figure 2 is the first-order ^{1}H-NMR spectrum of *N*-phenethyundecanamide.



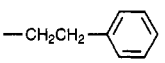
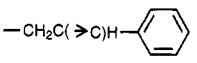
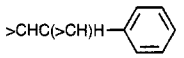
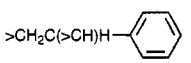
At first, the first-order ^{1}H-NMR is entered. The spectral data must be presented in digital form, but the means of

Table 2. The Digitized ^{1}H-NMR Spectrum of $C_{19}H_{31}NO$

chemical shift	multiplicity	area	chemical shift	multiplicity	area
0.89	3	3	1.23	100[a]	16
2.13	3	2	2.80	4	2
3.47	4	2	6.79	100[a]	1
7.19	100[a]	5			

[a] The integer 100 indicates the multiplet structure.

Table 3. The Analysis Result of ^{1}H-NMR Spectrum of $C_{19}H_{31}NO$

no.	LNSCS code	substructure
1	CC.	CH_3CH_2-
2	CCC.	$CH_3CH_2CH_2-$
3	C.2CC.2	$>CHCH_2CH<$
4	CCC.3	$CH_3CH_2C\leqslant$
5	C.CC.3	$-CH_2CH_2-C\leqslant$
6	CCC.2	$CH_3CH_2CH<$
7	C.CC.	$-CH_2CH_2CH_2-$
8	C.2CC.3	$>CHCH_2C\leqslant$
9	C.3CC.3	$\geqslant CH_2C\leqslant$
10	C.CC.2	$-CH_2CH_2CH<$
11	C.CC.=O	$-CH_2CH_2C(=O)-$
12	C.CN.	$-CH_2CH_2NH-$
13	C.CC1=CC=CC=C1	
14	C.2CN.	$>CHCH_2NH-$
15	C.2C(C.2)2	$>CHC(>CH)HCH<$
16	C.2C(C.2)C.3	$>CHC\geqslant CHCH<$
17	C.C(C.)C.2	$-CH_2C(>CH)HCH_2-$
18	C.C(C.)C.3	$-CH_2C(>CH)HCH_2-$
19	C.C(C.2)C.3	$-CH_2C(>CH)HCH<$
20	C.2C(C.2)2	$>CHO(>CH)HCH<$
21	C.C(C.3)C1=CC=CC=C1	
22	C.2C(C.2)Cl=CC=CC=C1	
23	C.2C(C.3)C.=O	$>CHC\geqslant CHC(=O)$
24	O=C.C(C.2)2	$>CHC(>CH)HC(=O)$
25	C.2C(C.3)N.	$>CHC(\geqslant CHNH-$
26	N.C(C.2)2	$>CHC(CH)HNH-$
27	C.C(C.3)N.	$-CH_2C\geqslant CHNH-$
28	C.C(C.2)Cl=CC=CC=Cl	
29	C.NC.=O	$-CH_2NHC(=O)-$
30	C.2NC.=O	$>CHNHC(=O)-$
31	C.3NC.=O	$\geqslant CNHC(=O)-$
32	Cl=CC=CC=Cl.	

digitization is unimportant, as long as it is accurate. ESSESA accepts the following ranges for the digitized data: chemical shifts of 0.0−11.0 ppm; multiplicity of 0−10, the integer 100 is input if the peak is a broad multiplet group of absorption peaks; relative area of peaks is an integer of 1−20. The digital data from the spectrum in Figure 2 are given in Table 2. The constraints from IR spectral analysis and the ring and double-bond characteristics and the atomic composition, which are obtained in the IR spectral analysis, of *N*-phenethylundecanamide are passed to the interpreter program to help the analysis of first-order ^{1}H-NMR.

As the interpreter program acquires these data, it makes use of the rules in the knowledge base to compare the stored spectral patterns with the digital input data to identify the substructures that might be contributing to the *N*-phenethylundecanamide structure. The resulting constraints from the first-order ^{1}H-NMR spectral analysis of *N*-phenethylundecanamide are listed in Table 3. The structure of *N*-phenethylundecanamide can be compared with the substructures shown in Table 3, which ESSESA decided should be present in the structure of *N*-phenethylundecanamide, based upon the first-order ^{1}H-NMR spectrum. The substructures 1, 2, 7, 12, 13, 29, and 32 exist really in the structure of *N*-phenethylundecanamide. Most of the substructures from the analysis of the first-order ^{1}H-NMR spectra do not exist in the structure of *N*-phenethylundecanamide. These substructures are generated from multiplet explanation of the peaks, especially the multiplet peaks. They can be deleted from the analysis of ^{13}C-NMR spectrum and substructural consistent analysis.

DISCUSSION

Chemical shift in NMR spectra is the physical property related with structural environment. The chemical shifts of two peaks in a ^{1}H-NMR spectrum will be very near if the structural environments of the two hydrogen atoms in the structure are very similar. The peaks of their chemical shifts are very near and will generate a complicated shape peak. The assignment of such peaks is very hard for first-order ^{1}H-NMR spectra. In ESSESA all possible explanations of such peaks are used as the substructure constraints in order to generate all possible candidate structures for an unknown compound. The correct substructures that really exist in the structure of an unknown compound will be obtained by analysis of all substructure constraints from IR, ^{1}H-NMR, and ^{13}C-NMR spectral data.

Some substructures will generate peaks that have similar chemical shifts and the same multiplet as well as the areas of peaks. These peaks can be assigned to some different substructures, but not all of these different substructures are really included in the unknown structure. The incorrect substructure will be deleted by subsequent ^{13}C-NMR spectral analysis and substructures consistent analysis. The details about this process will be presented in a future paper.

REFERENCES AND NOTES

(1) Silverstein, R. M.; Bassler, G. C.; Morrill, T. C. *Spectrometric identification of organic compounds*; 5th ed.; J. Wiley: New York, 1991.

(2) Clerc, J. T.; Pretsch, E.; Seibl, J. *Structural analysis of organic compounds by combined application of spectroscopic methods*; Elsevier: Amsterdam, 1981.

(3) Gray, N. A. B. *Computer-assisted structure elucidation*; J. Wiley: New York, 1986.

(4) Gray, N. A. B. Chem. Intell. Lab. Sys. 5:11−32, 1986.

(5) Huixiao, H.; Xinquan, X. ESSESA: An Expert System for Structure Elucidation from Spectra. 1. Knowledge Base of Infrared Spectra and Analysis and Interpretation Program. *J. Chem. Inf. Comput. Sci.* **1990**, *30(3)*, 203−210.

(6) Huixiao, H.; Xinquan, X. ESSESA: An Expert System for Structure Elucidation from Spectra. 2. Novel algorithm of perception of the linear independent smallest set of smallest rings. *Anal. Chim. Acta* **1992**, *262*, 179−191.

(7) Huixiao, H.; Xinquan, X. ESSESA: An Expert System for Structure Elucidation from Spectra. 3. LNSCS for Chemical Knowledge Representation. *J. Chem. Inf. Comput. Sci.* **1992**, *32(1)*, 116−120.

(8) Huixiao, H.; Xinquan, X. ESSESA: An Expert System for Structure Elucidation from Spectra. 4. The Canonical Representation of Structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*(4), 730−734.

(9) Sasaki, S.; Kudo, Y.; Ochiai, S.; Abe, H. Automated Chemical Structure Analysis of Organic Compounds: An Attempt to Structure Determination by the Use of NMR. *Mikrochimica Acta* **1971**, *5*, 726−742.

(10) Beech, G.; Jones, R. T.; Miller, K. Structural Interpretation of Proton Magnetic Resonance Spectra by Computer: First-Order Spectra. *Anal. Chem.* **1974**, *46(6)*, 714−718.