Construction of Functional Group Reactivity Database under Various Reaction Conditions Automatically Extracted from Reaction Database in a Synthesis Design System

Akio Tanaka*

Organic Synthesis Research Laboratory, Sumitomo Chemical Co., Ltd., 1-98, Kasugade-naka 3-chome, Konohana-ku, Osaka 554-8558, Japan

Hideho Okamoto

Center for Research and Advancement in Higher Education (Hakozaki Campus), Kyushu University, 6-10-1, Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan

Malcolm Bersohn

Department of Chemistry, University of Toronto, Ontario M5S 3H6, Canada

Received November 7, 2009

To be able to estimate the reactivity of functional groups under certain reaction conditions, we have stored three types of data: (1) data of change or destruction of the functional groups by the conditions of the reaction conditions; (2) data showing no influence of the reaction conditions on the functional groups; and (3) data showing the relative reactivity of two functional groups in the presence of certain reaction conditions. These three types of data, considered together, form entities that are referenced as "interaction data". These interaction data are used in a synthesis design system called SYNSUP. A new module in our system has been constructed that automatically generates interaction data from the reaction databases. From 15 265 reactions in the database, our program selected 2763 useful reactions with yields of ≥90% and one functional group change. From these useful reactions, data regarding 465 interferences, 815 cases of inert functional groups (under the reaction conditions), and 62 relative rate data could be extracted. In addition, with the use of multiple relative rate datasets, the reactivity of more than two functional groups could be deduced.

1. INTRODUCTION

The achievement of efficient syntheses of organic compounds with more than one functional group requires careful selection of reaction conditions, to avoid side reactions. Having side reactions usually necessitates column chromatography of the product mixture. Chemical companies hope to omit the purifications of products or intermediates, because of the cost of separating away undesired coproducts. To find high-yield and efficient reactions, the first choice is reaction retrieval by a reaction database system. Also, another potential tool for solving the problem of finding efficient routes is a computerized synthesis design program. Our system generates plausible synthetic routes via a retrosynthetic search.[1] Such a system must be able to estimate the reactivity of functional groups under various reaction conditions.

So far, many synthesis design systems have been reported, and some are still being continuously developed.[2,3] The first system for computer-aided organic synthesis (CAOS) was published by Corey and Wipke in 1969; this system was identified as organic chemical simulation of synthesis (OCSS),[4] which was the predecessor of the better-known logic and heuristics applied to synthetic analysis (LHASA) system.[5] In 1972, Bersohn published his synthesis design program, which did not interact with the user.[6] The program is known as SYNSUP.[7] The Sumitomo Chemical Company has taken part in the development of SYNSUP from 1984 to 2007, and now the Sumitomo Chemical Company and its related subsidiary companies use the system. The targets are pharmaceutical and agricultural compounds, including their intermediates, electric materials, additive agents, and compounds involved in the study of industrial processes.[8] The system has now more than 5800 organic reactions, the number of which is steadily increasing.

Refined synthesis design must consider the possibility of side reactions. To cope with the task, a synthesis design system needs a module to get the reactivity of the functional groups present under the reaction conditions. When a functional group away from the reaction center is noted to be affected by a reaction condition, the functional group must be converted to a different group; in such a case, either the functional group must be inert under the reaction conditions or the functional group must be protected with a protecting group. In LHASA, the reactivity table of functional groups and reagents has been applied to direct the use of protecting groups.[9,10] The database consists of a functional groups/reagents (FG/RGNT) cross-reference table, which describes

* Author to whom correspondence should be addressed. E-mail: tanakaa1@sc.sumitomo-chem.co.jp.

10.1021/ci9004332
Published on Web 03/01/2010

three types of reactivity—i.e., high, medium, and low—for 66 functional groups and 18 subclassified groups to 138 relatively independent reagents. With the use of FG/RGNT, the interfering groups are identified and possible protecting groups may be then considered. The database was powerful but reactivity differences of more than one functional group were ambiguous. Furthermore, the table was manually prepared and was not updated by means of automatic extraction from reaction databases.

Systems to propose possible products under certain reaction conditions have been reported. The system, which is called CAMEO, given the users' reactants and reagents, will then predict the reaction products.[11] When more than one reaction product is proposed, the system judges major/minor products by the heat of formation. The system stores other physical parameters such as pK_a, bond dissociation energy, and HOMO/LUMO levels. Another system (EROS6.0) defines each reaction using physicochemical parameters on the reaction center atoms, and the system can construct model equations for the reaction rate constant.[12] However, the system corresponds to specific reactions and does not correspond to general chemical reactions. The system, named SOPHIA, predicts possible products, given arbitrary reactants and reaction conditions.[13] The predictions are based on some knowledge bases. One knowledge base is referenced to recognize reaction sites, and another is used to check the converted products, relative to the conditions.

We do not embed a similar reaction prediction system in SYNSUP. However, we do need to evaluate all steps in multistep routes. SYNSUP accomplishes this task with a database that estimates functional group reactivity under reaction conditions. In this paper, we wish to report the format of the reactivity data and a module to extract the data from reaction databases.

2. INTERACTION DATA AND MODULES

In general, the purpose of using reaction database is to recognize the reaction centers in reported reactions in order to synthesize a goal molecule. Based on the database retrieval, the appropriate reaction condition or reagent is decided. In a synthesis design system, proposal of reactions to generate precursors is one of the key modules. If any functional group in the molecule, other than the reactant functional group, would be destroyed then SYNSUP declines to generate the reactant(s). SYNSUP is a database-oriented and empirical system, so a module to store the influence of reaction conditions plays an important role in our system.

We lose some yield when the function group outside of the main functional group has some reactivity in the expected conditions. We lose most or all the yield when the side structure reacts faster than the main structure. Three types of data sets have been prepared, interfering data, inert data, and relative rate data, collectively called interaction data. In the interaction data a term, functional group, is used instead of structure. The interfering data describes a reaction condition affecting a functional group, the inert data describes a functional group not affected by a condition, and the relative rate data indicates reactivity differential between two functional groups that are reactive to a condition. The interaction data sets are extracted from reaction databases.

Table 1. Contents of Interaction Data

name	contents
Interfering Data	ID number (**D1**) condition number (**D2**) functional group number (**D3**) converted functional group number (**D4**) first relative rate ID number (**D5**) reference number (**D6**)
Inert Data	same members as Interfering Data, but with the following definitions: converted function group number = 0 (**D4**) first relative rate ID number = 0 (**D5**)
Relative Rate Data	ID number (**D7**) condition number (**D8**) less-reactive functional group number (**D9**) more-reactive functional group number (**D10**) converted more-reactive functional group number (**D11**) next relative rate ID number (**D12**) reference number (**D13**)

The details of members in the data sets are described in the following paragraphs.

In SYNSUP functional groups are considered to have "central" atoms. For example the carbon atom in the center atom of the functional group $-CF_2-$. Two atoms of the same element which are doubly or triply bonded to each other are both considered to be central atoms of a functional group. Thus the nitrogen atoms of $N=N$ are both considered to be central atoms of functional group.[14] In groups of the type C−X, where X is a heteroatom, the C atom is considered to be the central atom. So in RCH_2Br the CH_2 carbon is the central atom of that functional group. Hydrogen atoms are never viewed as center atoms. A saturated carbon with a heteroatom neighbor is not a central atom. Each type of central atom, based on its environment, is given a characteristic number. The central atom numbers of primary, secondary and tertiary alcohols, for example, are all different.

Reaction conditions in the interaction data consist of reagents and catalysts. The following reactive functional groups in reactants are additionally reflected to reaction conditions; acid chlorides, chloroformates, chlorosulfonyl isocyanates, enamines, guanidines, hydrazines, hydroxyamines, imino chlorides, isocyanate, primanry amines, phosphines, phosphites, secondary amines, sulfenyl chloride, tertiary amines, thiols, and trialkylsilyl chlorides. On the other hand,

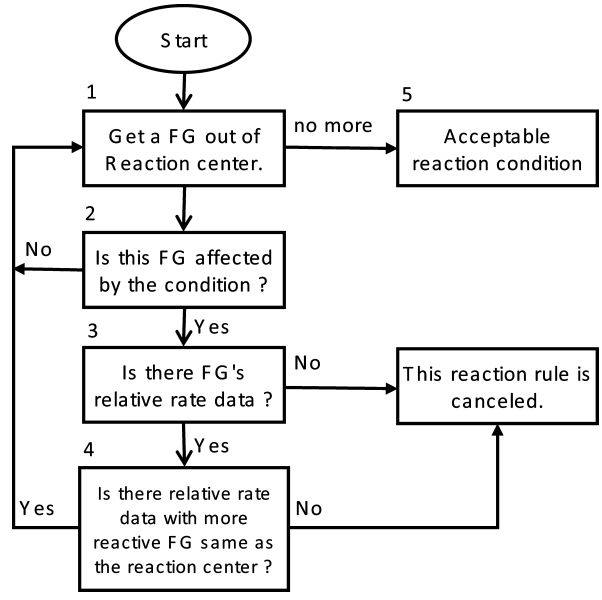


Figure 1. Flow of reaction check with interaction data.

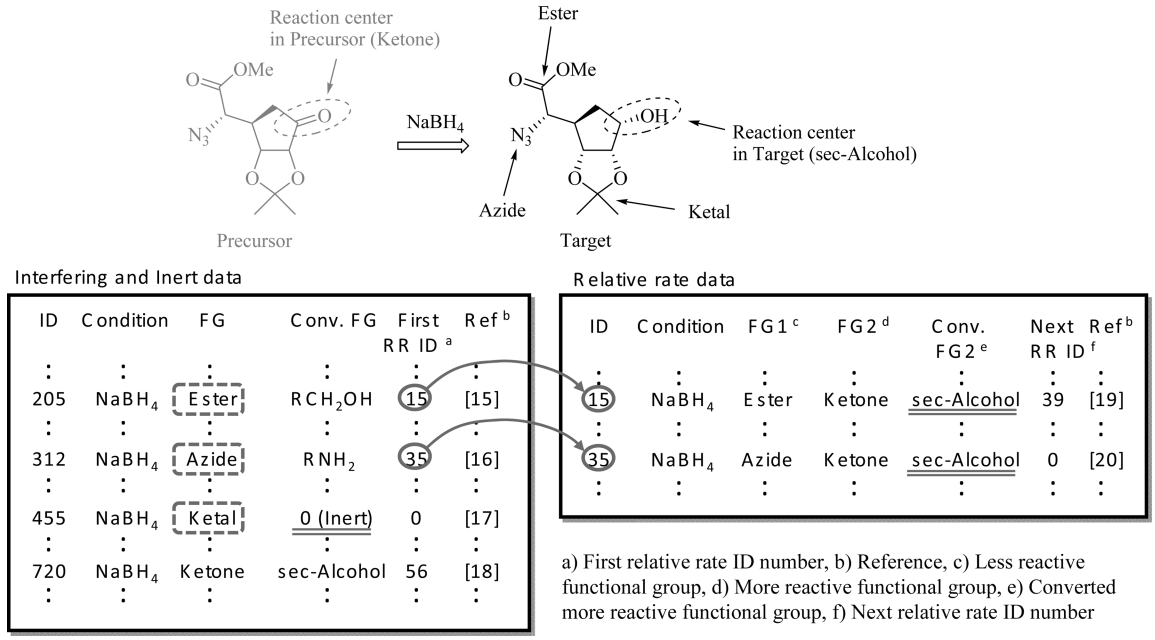


Figure 2. Illustration of the reaction condition, checked using interfering, inert, and relative rate data.

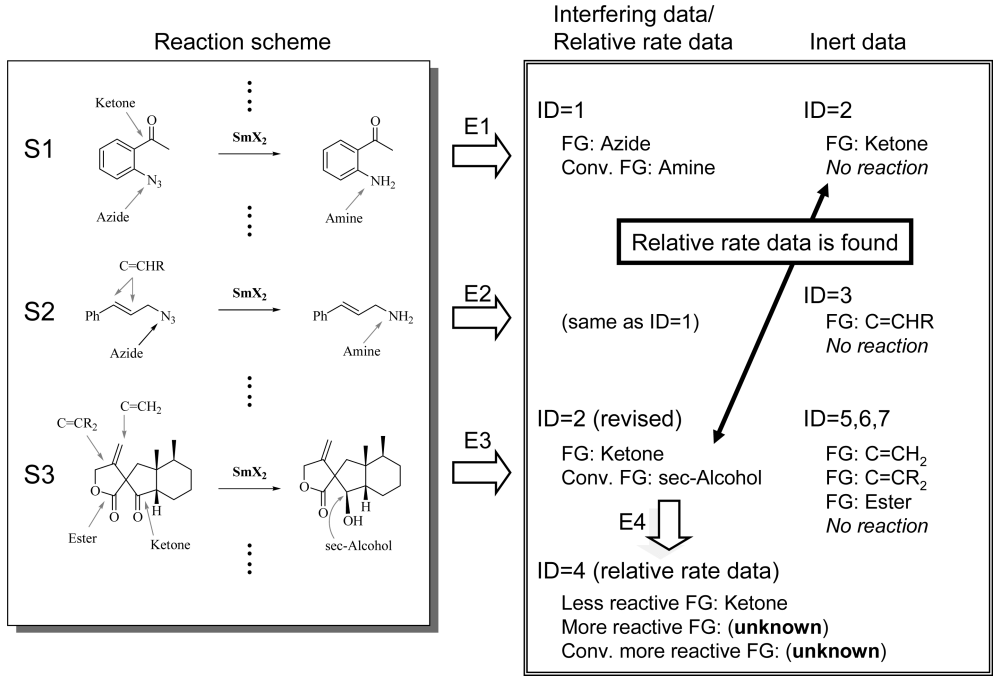


Figure 3. Flow of the extracting interaction data by one scan. FG denotes a functional group.

reaction temperatures, solvents, pressures, and the concentrations of the conditions are not defined clearly. Although, the purpose of interaction data is to warn about the possibilities of side reactions, the purpose is not proposal of the most appropriate reaction conditions for syntheses of a certain compound. At this time, we think that the current definition is enough.

The contents of the interaction data—i.e., interfering data, inert data, and relative rate data—are shown in Table 1.

"Interfering data" contain a functional group (**D3**), which is affected by a condition (**D2**) to be converted into a functional group (**D4**), and an ID number of the first relative rate (**D5**) linked to a relative rate data with the same functional group (**D2**) and condition (**D3**). "Inert data" has the same datasets as interfering data, but specifically the functional group (**D3**) is not affected by the condition (**D2**), so both the converted functional group (**D4**) and the first relative rate ID number (**D5**) are always zero. "Relative rate data" contain two functional groups—one less-reactive (**D9**) and one more-reactive (**D10**)—to represent two different reactive functional groups under a condition (**D8**). A converted more-reactive functional group (**D11**) corresponds to the functional group (**D10**) of the product side. The relative rate data are linked from the interfering data's first relative rate ID number (**D5**) or the next relative rate ID number (**D12**) linked from other relative rate data. The data are linked by the same (less-reactive) functional group ID number (**D9**) and the same condition ID number (**D8**).

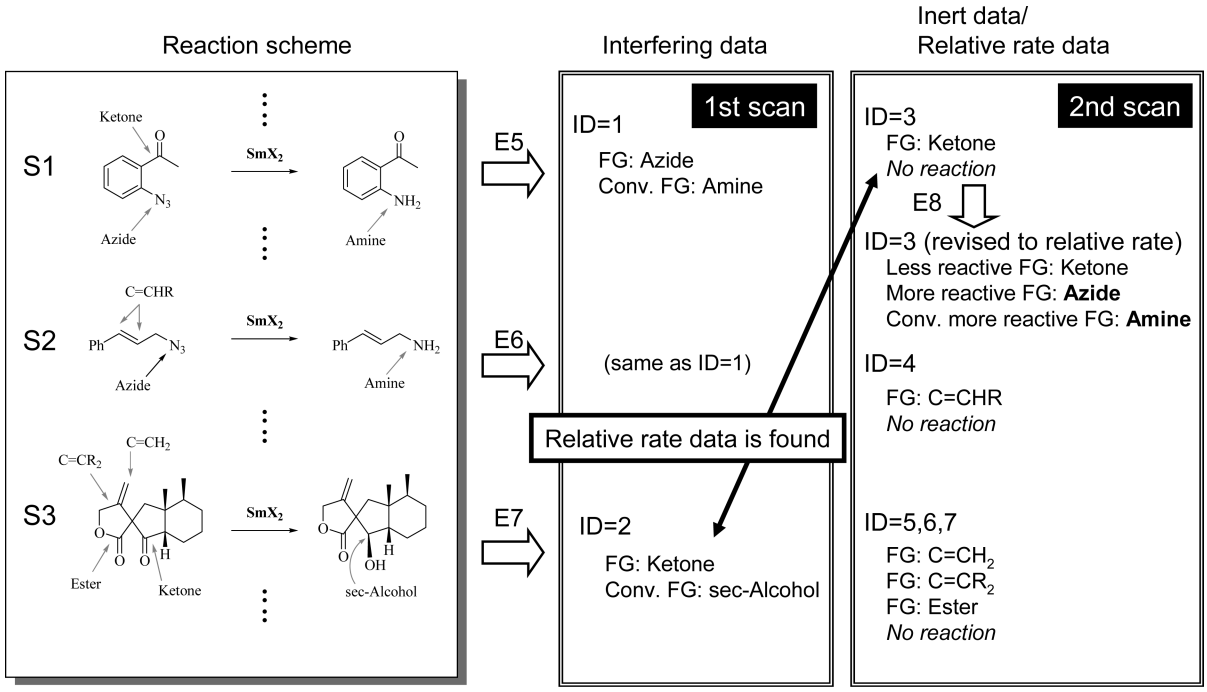


Figure 4. Flow of the extracting interaction data by double scan. FG denotes a functional group.

Table 2. Number of Reactions to Extract Interaction Data

name of database set	CCR1995	CCR1996	CCR1997	CCR1995−7
number of reactions in database	4203	5688	5374	15265
number of exploitable reactions	831	987	945	2763
ratio	19.77%	17.35%	17.58%	18.10%

Datasets for the interaction data are referenced before reactant structures are generated. The flow to check the datasets for the interaction data is described as follows (see Figure 1):

(1) If an unchecked functional group exists, pick it up and proceed to step 2. Otherwise, go to step 5.

(2) The functional group is examined with the condition using the interfering and inert data. If the functional group is survives, go back to step 1. Otherwise, go to step 3.

(3) If the functional group is reactive to the condition, relative rate data are examined using the first relative rate to go to step 4. Otherwise, cancel the reaction rule.

(4) If one of the converted more-reactive functional groups in the linked relative rate data is the same as the reaction center of the product side, the relative rate data that involve the reactive functional group out of the reaction center are less reactive than that of the reaction center obtained under the reaction condition. The reactive functional group then is judged to have survived, and one proceeds to step 1. Otherwise, cancel the reaction rule.

(5) All functional groups out of the reaction center are examined under the reaction condition. If all functional groups from the reaction center are determined to have not survived, the module is passed on to generating precursors.

For the purpose of comparison using relative rate data, **D11** is prepared for comparison with the interfering function group from the reaction center. Some reaction rules also have a functional group ID number of precursor reaction centers, which are also usable to be examined.

Figure 2 illustrates an example of checking interaction datasets. A reaction rule is now applied to prepare a secondary alcohol from a ketone with $NaBH_4$, while the target contains three other functional groups:, an ester, an azide, and a ketal. The precursor is indicated in gray, because it has not been generated yet. The reaction condition (the use of $NaBH_4$) affects the ester and the azide, according to the interfering data ID = 205 and 312 (Figure 2). On the other hand, from the inert data, the ketal is not affected by $NaBH_4$ (ID = 455, Figure 2). The interfering data of the ester with $NaBH_4$ (ID = 205) has a first relative rate ID number of 15, which describes a secondary alcohol as a converted more-reactive functional group being the same as the reaction center of product. The relative rate data response indicated that the ester survives. The azide group also survives, because of the existence of relative rate data (ID = 35) similar to that of the ester. All three functional groups from the reaction center are finally judged to survive under the reaction conditions.

The module of automatic extraction of interaction data from reaction database consists of the following four steps:

(1) Conversion of condition ID number in reaction database into SYNSUP,

(2) Recognition of functional group conversion in reaction center of reaction database,

(3) Extraction of interaction data from the functional group conversion, and

(4) Extraction of inert and relative rate data.

A word dictionary has been prepared to acknowledge SYNSUP condition ID number (**D2**) used in the reaction rules from the reaction database (step 1). Next, all functional groups of a reactant and a product in a reaction scheme are recognized; the reaction center then is recognized using

Table 3. Number of Appearances of Functional Groups and Reactants of the Exploitable Reactions in CCR1995−7[a]

		Number				Number				Number				Number	
No.	name	FG	react	No.	name	FG	react	No.	name	FG	react	No.	name	FG	react
1	C═CHR	1072	621	31	ArF	78	38	61	ArI	24	22	91	hemiacetal	8	8
2	ether	992	609	32	ArBr	69	65	62	imide	23	22	92	hydrazine	8	4
3	ester	817	643	33	RNH_2	69	68	63	C≡CH	20	20	93	RN═NR′	8	4
4	ketone	672	620	34	$ROCH_2OR'$	69	57	64	RN−NRR′	19	17	94	acid anhydride	7	7
5	$C{=}CR_2$	545	383	35	tetraalkyl silane	59	55	65	hydrazone	18	17	95	RR′C═N−N	7	7
6	*sec*-alcohol	354	290	36	*tert*-alcohol	56	49	66	C−C(−N)═N	17	17	96	iminoether	7	6
7	sulfide	217	184	37	RCF_3	56	46	67	aza ketone	16	16	97	RCSSR′	7	7
8	RCH_2OH	195	191	38	azide	51	51	68	C−O−N (O)	16	14	98	RHC═N−N (C)	7	7
9	carbamate	164	155	39	enol ether, no H	50	40	69	N−CH(R)−N	16	12	99	C═C(S)(N)	7	7
10	C≡CR	159	91	40	N−CHOR−C	47	44	70	RR′CHCl	15	15	100	guanidine	7	7
11	$C{=}CH_2$	149	141	41	RCH_2Br	46	46	71	aldimine	14	14	101	$N{-}CH_2OR$	7	7
12	ketal	145	128	42	sulfone	46	44	72	RR′CHBr	13	13	102	ROOR′	6	3
13	ArCl	144	112	43	SC(C)═C	45	33	73	alkynyl silane	13	12	103	aziridine	6	6
14	aldehyde	134	134	44	urea	43	42	74	C═CRBr	13	13	104	C═C(−S)(−S)	6	6
15	alkylsilyl ether	128	107	45	$C{-}CF_2{-}C$	39	11	75	ArSH	13	13	105	pyridine *N*-oxide	6	6
16	nitron	123	117	46	SCH═C	39	32	76	formamidine	13	13	106	R_3SiN	6	5
17	RCONHR′	120	107	47	ketimine	38	35	77	RR′CHI	12	12	107	C═C(−N)(−N)	6	6
18	*sec*-amine	119	113	48	sulfoxide	37	37	78	tetraalkyl stannane	12	12	108	S−CH(−C)(−N)	6	6
19	carboxylic acid	119	116	49	ketenamine (C)	36	35	79	phosphane oxide	12	11	109	phosphane	5	5
20	SiC(R,R′)O	119	103	50	RR′NSR″	35	35	80	selenide	12	12	110	$RR'CCl_2$	5	5
21	RCONR′R″	118	117	51	RCH_2Cl	35	33	81	*N*-sulfonamide	11	11	111	$S{-}C{=}N(NR_2)$	5	5
22	acetal	117	110	52	phosphonate	32	28	82	C−C(Cl)−N	11	9	112	thiourea	5	5
23	*tert*-amine	117	110	53	*N*-alkyl imide	30	29	83	*N*-alkyll sulfonamide	10	10	113	C−S(═O)−N	5	5
24	C═N−C N	106	97	54	$RCONH_2$	30	30	84	RCH_2SH	10	10	114	$C{-}CBr_2{-}C$	5	5
25	ketenamine (N)	101	78	55	Sn−C (C)	30	10	85	formamide	10	10	115	hemimercaptal	5	5
26	aldenamine	98	87	56	sulfonate	29	27	86	RCH_2I	10	10	116	thioester	5	4
27	P−O−C	96	42	57	*tert*-sulfonamide	29	29	87	*tert*-bromide	9	9	117	C−N═N−N	5	5
28	epoxide	94	93	58	enol ether, with H	28	25	88	C−C(═N)−S	9	9	118	O−C(═S)−O	5	5
29	PhOH	93	78	59	acid chloride	26	26	89	oxime	8	8	119	$N{-}CF_3$	5	3
30	nitrile	83	78	60	furan	24	23	90	hemiketal	8	8	120	vinyl chloride	5	5

		Number				Number				Number				Number	
No.	name	FG	react	No.	name	FG	react	No.	name	FG	react	No.	name	FG	react
121	trichloromethyl	5	5	151	S−C(Br)═C	2	2	181	episulfide	2	2	211	O−C(═C)−S	1	1
122	RR′CH−B	4	2	152	SiC(−R)−O	2	2	182	NC(═O)S	2	2	212	sulfonyl chloride	1	1
123	NC(═S)−S	4	4	153	thioketone	2	2	183	RR″CCl	2	2	213	trialkyl silane	1	1
124	RC(═S)−O	4	4	154	*tert*-thiol	2	2	184	$C{-}CH_2F$	1	1	214	RCF_2I	1	1
125	$R_3Si{-}NR_2$	4	4	155	$C{=}C(NO_2)R$	2	2	185	phosphate	1	1	215	$C{-}CF_2{-}S$	1	1
126	trialkylsilyl chloride	4	4	156	RR′NOH	2	2	186	RHgR′	1	1	216	$B(OR)_3$	1	1
127	sulfonyl imine	4	4	157	RCH(−S)(−S)	2	1	187	NC(═S)O	1	1	217	S−C(═S)(−S)	1	1
128	RR′CHSH	4	4	158	RR′C═NH	2	2	188	RCONHOH	1	1	218	RR′R″C−I	1	1
129	pyrimidine	4	4	159	R−CH(OMe)OX	2	2	189	sulfinate ester	1	1	219	$C{=}CI_2$	1	1
130	vinyl fluoride	4	4	160	alkyl boronic acid	2	2	190	trialkylthiophosphate	1	1	220	$N{-}SO_2{-}N$	1	1
131	oxime ether	4	4	161	acylsilane	2	2	191	$RR'C(OH)_2$	1	1	221	C═CR(−Si)	1	1
132	RR′CHF	4	4	162	xanthate ester	2	2	192	Br(RO)CRR′	1	1	222	N−CHBr−C	1	1
133	C═C═C	4	4	163	nitrone	2	2	193	$RSiCl_3$	1	1	223	alkynyl bromide	1	1
134	orthoester	4	4	164	hydroxyperoxide	2	2	194	dithiocarboxylic acid	1	1	224	sulfenyl chloride	1	1
135	$NC(NH_2)N$, aromatic C	3	3	165	$C{-}CF_2{-}Cl$	2	2	195	$RS{-}SO_3H$	1	1	225	C−CBrCl−C	1	1
136	$ROSO_2OR'$	3	3	166	C═CCl(−N)	2	2	196	N−C(═O)S−C	1	1	226	C−C(═N)Cl	1	1
137	Si−NH−C	3	1	167	$O{-}CF_2{-}C$	2	1	197	SCH_2Br	1	1	227	CR(OH)═NR′	1	1
138	aminoxidie	3	3	168	C═CH(−Si)	2	2	198	$(RO)_2P(=O){-}N$	1	1				
139	carbonate	3	3	169	RB(−O)(−O)	2	2	199	C−N(−O)═N	1	1				
140	thioamide	3	3	170	O−CRR′−S	2	2	200	Cl−CRR′−N	1	1				
141	RC(═O)−NR−C═X, X not oxygen	3	3	171	C═CHCl	2	2	201	dichloromethyl	1	1				
142	ROCHBr−C	3	3	172	diamidophosphate	2	2	202	isocyanide	1	1				
143	phosphite	3	3	173	C═CHBr	2	2	203	RCH_2Li	1	1				
144	nitroso	3	3	174	acid fluoride	2	2	204	$C{=}CCl_2$	1	1				
145	N−CH(−R)−N	3	3	175	C═CR(−I)	2	2	205	trialkylborane	1	1				
146	isothiocyanate	3	3	176	$SiCH_2Cl$	2	1	206	C−CHClBr	1	1				
147	C═CBr(−N)	3	3	177	$RCHF_2$	2	2	207	$RBCl_2$	1	1				
148	RCH═N−S	3	3	178	C═CH(−I)	2	2	208	RR′P(═O)OR″	1	1				
149	RSC(═N)SR′	3	3	179	C═N═N	2	2	209	N═N(−O)	1	1				
150	SCF_3	3	3	180	ArSeH	2	2	210	acyl azide	1	1				

[a] Table legend: "FG" = functional group; "react" = reacting conditions.

mapping data (step 2). This is followed by extraction and storage of an interfering data from changes in the reaction center functional group (step 3). In the reaction center, the reactant functional group is stored as **D3**, and the product side is stored as a converted functional group **D4**. Finally, all functional groups located in more than two nodes far from the reaction center are recognized as inert data under the reaction condition. Here, in the case where the functional group and condition pair has already been stored as part of the interfering data, the pair will be stored as new relative rate data, instead of inert data. The reaction condition is stored as **D8**. The interfering functional group is stored as a less-reactive functional group **D9**. The functional group in the reactant reaction center is stored as a more-reactive functional group **D10**. The functional group in the product reaction center is stored as a converted more-reactive functional group **D11** (step 4). The operations of steps 1− 4 were programmed in SYNSUP. Interaction data are auto-

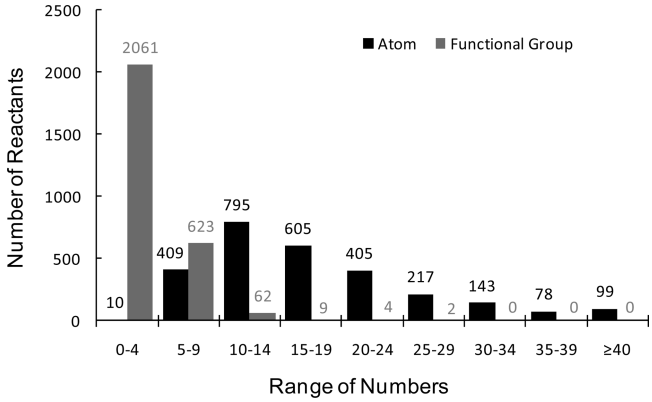


Figure 5. Distribution of reactants in the CCR1995−7 database, based on the numbers of atoms and functional groups.

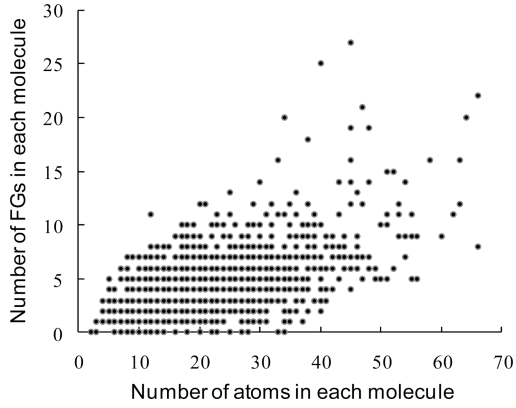


Figure 6. Plot of the number of atoms versus the number of FGs in each molecule.

matically collected if only RDFile,[21] including reaction schemes, is input.

When extracting interaction data from the reaction database, the database is scanned twice. First, all of the interfering data is extracted, then all of the inert and relative rate data are done. The reason for this is that undefined relative rate data are sometimes found by extracting three types of data at one scan.

Figure 3 describes the flow to extract interaction data by a scan from three reactions schemes using a samarium dihalide (denoted as S1, S2, S3).[22−24] From S1, one interaction dataset (ID = 1) and one inert dataset (ID = 2) are obtained (E1). Next, from S2, the interfering data are the same as that of S1 (ID = 1) and should be skipped. One inert dataset (ID = 3) is newly obtained (E2). Then, from S3, one interfering dataset is found but the data conflict with the already extracted inert data (ID = 2) (E3). After revising ID = 2 as the interfering data, the preparation of new relative rate data is attempted; however, the attempt failed (E4). The reason for the failure is that the more-reactive functional group cannot be decided in S3, and, additionally, the revised interfering data (ID = 2) are independent of the interfering data ID = 1. To solve the problem, the number of scans of the reaction database must be two. Figure 4illustrates the flow of extracting interaction data by double scan.

On the first scan, only two interfering datasets (IDs = 1, 2) are extracted from S1, S2 and S3 (E5, E6, E7). The second scan tries to collect inert data and relative rate data. From S1, one inert datum (ID = 3) is found, but the data conflict with already-found interfering data (ID = 2). The inert datum then is deleted to make a new relative rate dataset (E8), which consists of a ketone as the less-reactive functional group, an azide as the more-reactive functional group, and an amine as the converted more-reactive functional group (rev. ID = 3). This is followed by getting an inert dataset (ID = 4) from S2, and three inert datasets from S3. As the result of the double scans, two interfering datasets (IDs = 1, 2), four inert datasets (IDs = 4, 5, 6, 7), and one relative rate datum (ID = 3) are successfully extracted.

Table 4. Number of Interaction and Relative Rate Data Preliminarily Extracted from the CCR1995−7 Database

data type	value
number of interfering data, NIFD	461
number of inert data, NIND	748
number of relative rate data	123
number of conditions, NC	133
number of functional groups, NFG	197
[(NIND + NIFD)/(NC × NFG)] × 100	4.61%

Table 5. Top Ten Conditions Including Many Interaction Data by Preliminary Extraction

		Number of Interaction Data		
order	condition	total	interfering	inert
1	Pd(0)	82	41	41
2	$NaBH_4$	47	17	30
3	CF_3CO_2H	41	10	31
4	Na_2S	38	12	26
5	strong base	36	4	32
6	HCl	29	13	16
7	MnO_2/CH_2Cl_2	26	7	19
8	Ag^+	25	5	20
9	LiOH	23	1	22
10	$LiAlH_4$	22	7	15

Table 6. Top Ten Functional Groups Including Many Interaction Data by Preliminary Extraction

		Number of Interaction Data		
order	functional group	total	interfering	inert
1	ether	74	22	52
2	C═CHR	58	9	49
3	ester	57	28	29
4	ketone	51	33	18
5	*sec*-alcohol	43	28	15
6	$C{=}CH_2$	28	1	27
6	sulfide	28	13	15
8	ArCl	27	0	27
8	$C{=}CR_2$	27	3	24
10	nitro	26	9	17

3. REACTION DATASETS

With respect to source reaction datasets, commercially available *Current Chemical Reaction* (CCR) reaction databases released in 1995, 1996, and 1997 were used on trial.[25] In this paper, the three datasets are called CCR1995, CCR1996, and CCR1997, respectively. Because of the preliminary trial and elimination of unreliable or unclear reaction data, we used reactions in databases that were satisfied by the following conditions:

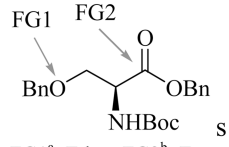
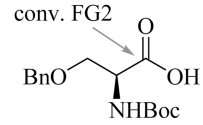
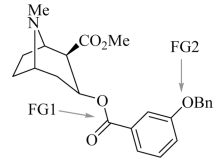
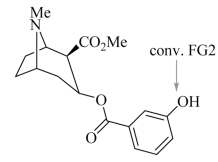
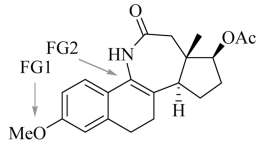
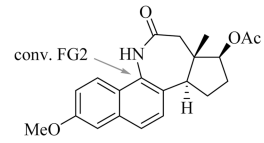
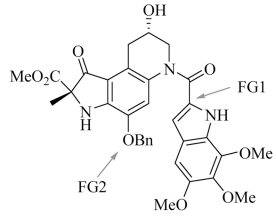
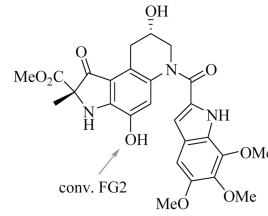
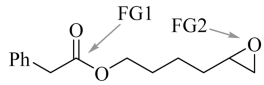
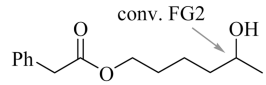
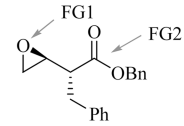
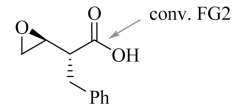
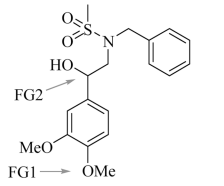
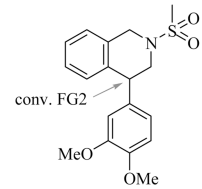
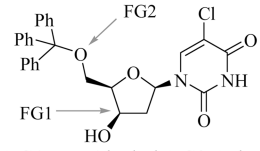
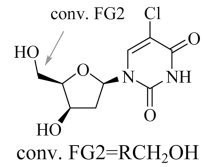
(a) Yield being ≥90%,

(b) One reactant and one product,

(c) One functional group change or one bond order change,

(d) Reaction condition being provided, and

Table 7. Contradictory Relative Rate Data of Preliminary Extraction from *Current Chemical Reactions* (CCR)

Entry	Reagent/ Catalyst	Reactant	Product	Ref.
1	Pd(0)	 FG1=Ether, FG2=Ester	 conv. FG2=Carboxylic acid	[26]
2	Pd(0)	 FG1=Ester, FG2=Ether	 conv. FG2=Phenol	[27]
3	Pd(0)	 FG1=Ether, FG2=Enamine	 conv. FG2=Aniline	[28]
4	Pd(0)	 FG1=Enamine, FG2=Ether	 conv. FG2=Phenol	[29]
5	Pd(0)	 FG1=Ester, FG2=Epoxide	 conv. FG2=sec-Alcohol	[30]
6	Pd(0)	 FG1=Epoxide, FG2=Ester	 conv. FG2=Carboxylic acid	[31]
7	CF_3CO_2H	 FG1=Ether, FG2=sec-Alcohol	 conv. FG2=Ar-R	[32]
8	CF_3CO_2H	 FG1=sec-Alcohol, FG2=Ether	 conv. FG2=RCH_2OH	[33]

[a] Less-reactive functional group. [b] More-reactive functional group. [c] Converted more-reactive functional group.

(e) Compatible structure in SYNSUP.

For accentuating and focusing on simple functional group change reactions, high-yield and no-byproduct reactions were employed. Variation of the reaction conditions is confined to the available SYNSUP conditions. The maximum number of atoms (besides H) in a molecule is 72, according to the SYNSUP specifications.

Eighteen percent (18%) of the reactions in the CCR1995−7 database, representing a total of 2763 reactions, are exploitable (see Table 2). It is observed that 525 reaction conditions are defined in SYNSUP, and 319 conditions were found to be translatable from the reaction database. Furthermore, 244 conditions in the CCR1995−7 database were ignored for data extraction.

Among the reactants of 2763 exploitable reactions, 227 types of functional groups were recognized (see Table 3). There were many reactants with functional groups, olefins,

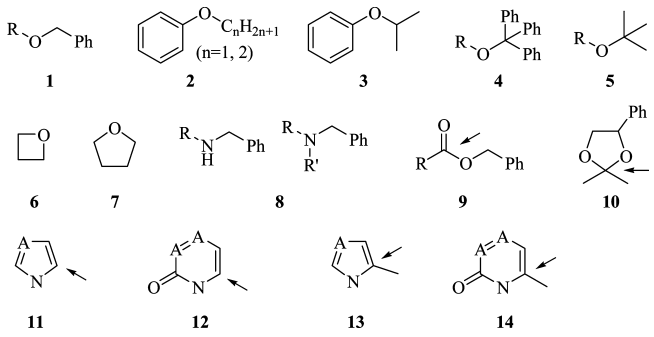


Figure 7. Redefined functional groups. Atoms highlighted by arrows are the center of functional groups; the symbol "A" represents an arbitrary atom.

ethers, esters, and ketones, which comprised more than 20% of all reactants.

The distributions of the number of atoms and functional groups in the 2763 reactants from the CCR1995−7 database are shown in Figure 5. The number of atoms in each molecule excludes the number of H atoms. The minimum number of atoms was 2, the maximum was 66, and the average was 18.0. The minimum number of functional groups was zero for two reactants, the maximum was 27, and the average was 3.5.

The correlation between the number of atoms and functional groups was low ($R^2 = 0.4041$; see Figure 6). By speculating that one reactant contains an average of 3.5 functional groups, one interfering dataset derived from a

Table 8. Results of Interaction Data Extraction from CCR1995, CCR1996, and CCR1997

	Results				
data type	CCR1995	CCR1996	CCR1997	average	CCR1995−7
number of exploitable reactions	831	987	945	921.00	2763
number of interfering data, NIFD	198	184	206	196.00	465
number of inert data, NIND	326	331	340	332.33	813
number of relative rate data	13	8	13	11.33	62
number of conditions, NC	83	78	88	83.00	133
number of functional groups, NFG	137	140	121	132.67	209
$[(NIND + NIFD)/(NC \times NFG)] \times 100$	4.61%	4.72%	5.13%	4.82%	4.60%
execution time	43 s	55 s	45 s	47.67 s	151 s

Table 9. Number of Interactions for Each Condition

	condition	Number of Interactions				condition	Number of Interactions				condition	Number of Interactions		
No.	name	total	interfering	inert	No.	name	total	interfering	inert	No.	name	total	interfering	inert
1	Pd(0)	92	40	52	46	H_2S	8	3	5	91	R_3Si-OTf	3	2	1
2	$NaBH_4$	53	17	36	47	*N*-bromoacetamide	8	4	4	92	$HgCl_2$	3	1	2
3	CF_3CO_2H	45	11	34	48	CDI	8	6	2	93	$SbCl_5$	3	2	1
4	Na_2S	42	12	30	49	$H_2/RuCl_2$	8	3	5	94	PX_3	3	2	1
5	strong base	40	5	35	50	Baker's yeast	8	4	4	95	$HClO_4$	3	1	2
6	HCl	31	13	18	51	RNHOH	8	3	5	96	BR_3	3	1	2
7	MnO_2/CH_2Cl_2	28	7	21	52	mild base	8	7	1	97	$LiBF_4$	3	2	1
8	Ag^+	26	5	21	53	HCOOH	7	2	5	98	HI	3	2	1
9	SmX_2	24	12	12	54	$NaIO_4$	7	2	5	99	O_3	3	1	2
10	LiOH	24	2	22	55	$SnCl_4$	7	2	5	100	NBS	3	1	2
11	$PdBr_2/L$	24	5	19	56	NaOCl	7	5	2	101	CBr_4	3	1	2
12	$LiAlH_4$	23	7	16	57	BF_3	7	2	5	102	hot alumina	3	3	0
13	alkoxide	22	10	12	58	Me_2SO_4	6	2	4	103	$NiCl_2$	3	1	2
14	CrO_3/pyridine	22	5	17	59	RSH	6	2	4	104	HBr	2	1	1
15	DIBAL	21	6	15	60	I^-	6	1	5	105	$K_2Cr_2O_7/H_2SO_4$	2	2	0
16	NH_3, RNH_2	20	7	13	61	KH	6	1	5	106	RuO_2	2	2	0
17	I_2	19	6	13	62	$Na_2S_2O_4$	6	1	5	107	$SnCl_2$/pyridine	2	1	1
18	$FeCl_3$	19	7	12	63	Br_2	6	2	4	108	^{t}BuOCl	2	1	1
19	RLi	19	9	10	64	NIS	6	1	5	109	$Al(O\text{-}^i Pr)_3$	2	2	0
20	B_2H_6	18	10	8	65	$LiClO_4$/Mg(II)	6	3	3	110	$TiCl_4$	2	1	1
21	Na	17	5	12	66	nitration conditions	6	3	3	111	DMAP *N*-oxide	2	1	1
22	aqueous acid	17	6	11	67	$ClSiMe_2CMe_3$	6	1	5	112	NaSH	2	2	0
23	R_3N	16	3	13	68	$Me_4NBH(OAc)_3$	5	1	4	113	NO_2	2	2	0
24	H_2/lindlar	16	6	10	69	Ce(IV)	5	2	3	114	MoO_3/HMPA	2	1	1
25	NH_2SO_3H	15	3	12	70	AlH_3	5	2	3	115	$NaBrO_3$	2	2	0
26	benzoyl peroxide	15	4	11	71	$SOCl_2$	5	2	3	116	Cu^+/RS^-	2	1	1
27	$Zn(BH_4)_2$	15	4	11	72	$RNHNH_2$	5	2	3	117	$ClSiMe_3$	2	2	0
28	ZrO_2/ThO_2	15	2	13	73	TsCl	5	2	3	118	R_3P	2	1	1
29	Fe	14	2	12	74	PPA	5	2	3	119	$(Me_3Si)_2NH$	2	1	1
30	NaH	14	6	8	75	$NaBO_3$	5	3	2	120	$Pd_2(dba)_3$	1	1	0
31	$(COCl)_2$	14	3	11	76	$Pd(OAc)_2/Cs_2CO_3$	5	5	0	121	$ZnCl_2$	1	1	0
32	N_3^-	14	6	8	77	$In(CF_3SO_3)_3$	5	3	2	122	$HSiCl_3$	1	1	0
33	alkaline H_2O_2	14	7	7	78	$KMnO_4$	4	2	2	123	CuO/Cr_2O_3	1	1	0
34	BBr_3	13	3	10	79	R_2AlCl	4	1	3	124	P_2O_5	1	1	0
35	OsO_4	12	2	10	80	$KHSO_5$	4	2	2	125	SeO_2	1	1	0
36	DDQ	12	4	8	81	F_2	4	3	1	126	$Ba(OH)_2$	1	1	0
37	Raney nickel	12	6	6	82	DBU	4	2	2	127	$RuCl_2/PPh_3$	1	1	0
38	^{t}BuOOH	12	3	9	83	CH_2N_2	4	2	2	128	TsNHCl	1	1	0
39	CN^-	11	2	9	84	DABCO	4	2	2	129	$ZrCl_4$	1	1	0
40	$AlCl_3$	11	5	6	85	SCN^-	4	1	3	130	Na(Hg)	1	1	0
41	LDA	10	5	5	86	Me_3Al	4	1	3	131	AcCl	1	1	0
42	Cu^{2+}/pyridine	10	8	2	87	$POCl_3$	4	1	3	132	Cp_2ZrCl	1	1	0
43	Zn	10	4	6	88	Na/naphthalenide	4	1	3	133	9-BBN	1	1	0
44	R_2Zn	10	1	9	89	Ipc_2BCl	4	1	3					
45	RMgX	10	7	3	90	K	3	2	1					

Table 10. Number of Interactions for Each Functional Group

No.	functional group name	Number of Interactions			No.	total	Number of Interactions		
		total	interfering	inert			functional group name	interfering	inert
1	C═CHR	58	9	49	54	imide (NH)	6	0	6
2	ester	57	27	30	55	C−O−N	6	1	5
3	ketone	51	33	18	56	heteraryl CH(11)	6	2	4
4	*sec*-alcohol	43	28	15	57	ArO^{i} Pr(3)	6	1	5
5	ArOMe, ArOEt (**2**)	40	5	35	58	heteroaryl CR (**13**)	5	0	5
6	ether	33	0	33	59	^{t}BuOR(5)	5	1	4
7	ROBn (**1**)	33	13	20	60	SCH═C	5	1	4
8	$C═CH_2$	28	1	27	61	formamide	5	1	4
9	sulfide	28	13	15	62	furan	5	0	5
10	ArCl	27	0	27	63	$RCONH_2$	5	4	1
11	$C = CR_2$	27	3	24	64	enol ether (no H)	5	0	5
12	nitron	26	9	17	65	formamidine	5	1	4
13	ketal	24	8	16	66	sulfonate	5	0	5
14	RCONR″	24	11	13	67	imide (N without H)	4	4	0
15	carbamate	21	2	19	68	phosphonate	4	1	3
16	aldehyde	21	17	4	69	SC(C)═C	4	0	4
17	acetal	21	13	8	70	tet raalkyl stannane	4	2	2
18	RCH_2OH	20	12	8	71	RR′NSR″	4	0	4
19	SiCRR′OR″	20	0	20	72	C−C(Cl)−N	4	0	4
20	alkyl silyl ether	20	0	20	73	C≡C−C	4	4	0
21	epoxide	19	16	3	74	*N*-alkyl sulfonamide (N)	3	0	3
22	$ROCH_2OR′$	19	0	19	75	ketenamine	3	0	3
23	nitrile	19	11	8	76	$N−C(NH_2)−N$	3	0	3
24	azide	16	16	0	77	$RR′CCl_2$	3	2	1
25	RCONHR′	16	3	13	78	CH═CHCH═CHC(═O)R	3	3	0
26	ArBr	15	0	15	79	C≡C−(C,H)	3	3	0
27	carboxylic acid	15	9	6	80	thiourea	3	2	1
28	*tert*-amine	14	4	10	81	RC(═O)−NR−*C* = X	3	0	3
29	C═N−C	12	5	7	82	acid anhydride	3	1	2
30	PhOH	12	0	12	83	*tert*-bromide	3	3	0
31	enol ether	12	0	12	84	*N*-sulfonamide	3	0	3
32	*sec*-amine	11	4	7	85	RR′CHI	3	3	0
33	N−CHOR−C	11	1	10	86	RR′CHBr	3	3	0
34	P−O−C	10	4	6	87	C═CRBr	3	2	1
35	urea	10	0	10	88	pyridine	3	2	1
36	RCF_3	10	0	10	89	C−C(−N)═N	3	0	3
37	THF (**7**)	9	1	8	90	RHC═N−N	3	0	3
38	sulfoxide	9	7	2	91	(C,H)−C═C−(C,H)	3	3	0
39	aldenamine	8	2	6	92	$ArCH═C−CO_2R$ (trans)	3	3	0
40	tetraalkyl silane	8	1	7	93	N−CH(R)−N	3	0	3
41	RNH_2	8	3	5	94	C═C═C	3	0	3
42	RCH_2Br	8	4	4	95	hydrazone	3	0	3
43	*tert*-alcohol	8	2	6	96	C═C−C	2	2	0
44	RCO_2Bn (**9**)	8	4	4	97	ArCH═C−EWG	2	2	0
45	C≡CR	7	0	7	98	RR′NOH	2	1	1
46	ArF	7	0	7	99	CH═C−C(═O)C	2	2	0
47	RR′CHCl	7	5	2	100	carbonate	2	0	2
48	RR′NBn,RNHBn (**8**)	7	5	2	101	R−CH(OMe)OX	2	0	2
49	aza ketal	7	1	6	102	ketimine	2	1	1
50	RCH_2Cl	7	6	1	103	C═C−Ar	2	2	0
51	sulfone	7	1	6	104	aziridine	2	0	2
52	*tert*-sulfonamide	7	0	7	105	C═C(−S)(−S)	2	0	2
53	RN−NR″	6	1	5	106	$ROCAr_3$ (**4**)	2	2	0

No.	functional group name	Number of Interactions			No.	total	Number of Interactions		
		total	interfering	inert			functional group name	interfering	inert
107	phosphite	2	2	0	160	$C−CH═CH_2$	1	1	0
108	thioester	2	1	1	161	RR′C═CHR″	1	1	0
109	nitroso	2	2	0	162	$C−CBr_2−C$	1	1	0
110	acid chloride	2	2	0	163	hydroxyperoxide	1	1	0
111	RCSSR′	2	1	1	164	cyclopentene	1	1	0
112	C═C(S)(N)	2	1	1	165	ROCHBr−C	1	1	0
113	phosphane	2	1	1	166	alkynyl ketone	1	1	0
114	cyclic ketone	2	2	0	167	hemimercaptal	1	0	1
115	C═C−C−OH	2	2	0	168	C═C−CH−EWG	1	1	0
116	acid fluoride	2	2	0	169	$CH−C≡C−CO_2R$	1	1	0
117	C−CR═C−EWG	2	2	0	170	iminoether	1	0	1
118	C≡CH	2	0	2	171	CH−C═C−C(═O)C	1	1	0
119	ArI	2	0	2	172	$O−CF_2−C$	1	0	1
120	C═CRCl	2	0	2	173	C−C≡C−C═C	1	1	0
121	trichloromethyl	2	0	2	174	$RBCl_2$	1	1	0
122	Sn−C (C)	2	0	2	175	$C═C(CN)_2$	1	1	0
123	guanidine	2	0	2	176	C═CRF	1	1	0
124	$N−CH_2−OR$	2	0	2	177	oxime ether (N)	1	0	1
125	SCF_3	2	0	2	178	RR′P(═O)OR″	1	0	1
126	RCOCH═CMeR′	1	1	0	179	enone	1	1	0
127	thioketone	1	1	0	180	C−N═N−N	1	1	0
128	$RR′C(OH)C(═CH_2)EWG$	1	1	0	181	$RCH_2CH_2C≡CCH_2OH$	1	1	0
129	*N*-alkylpyridine	1	1	0	182	hemiacetal	1	0	1
130	C≡C−C≡C	1	1	0	183	hydrazine	1	1	0
131	RCH_2F	1	1	0	184	C═CHCl	1	0	1
132	$C−C═C−CO_2R$	1	1	0	185	isothiocyanate	1	1	0
133	RHgR′	1	1	0	186	diamidophosphate	1	1	0

Table 10. Continued

No.	functional group name	Number of Interactions total	interfering	inert	No.	total	Number of Interactions functional group name	interfering	inert
134	R−C≡C−(elext rophile)	1	1	0	187	O−C(=S)−O	1	1	0
135	NC(=S)O	1	0	1	188	5-hydrofuran-2-one	1	1	0
136	oxime	1	1	0	189	trialkyl silane	1	1	0
137	amine oxide	1	1	0	190	RCF_2I	1	1	0
138	RC(=S)−O	1	1	0	191	$RCH_2-C-C\equiv C-CO_2R'$	1	1	0
139	ROOR′	1	0	1	192	C=CBr(−N)	1	0	1
140	RCONHOH	1	0	1	193	R_3SiN	1	0	1
141	sulfinate ester	1	1	0	194	C=CR−COMe	1	1	0
142	hemiketal	1	0	1	195	$CH_2=CR-CH_2-CR'-COR''$	1	1	0
143	RR′C=NH	1	1	0	196	RC(=O)−C=CR″	1	1	0
144	$S-C=N(NR_2)$	1	1	0	197	$C=CI_2$	1	1	0
145	$RR'C(OH)_2$	1	1	0	198	orthoester	1	0	1
146	C−S(=O)−N	1	1	0	199	cyclic C−CH=CH−COR	1	1	0
147	CHCH=CH	1	1	0	200	C=CHCHO (cis)	1	1	0
148	CCF_2C	1	0	1	201	oxetane (**6**)	1	1	0
149	$ArCH=CHCO_2R$	1	1	0	202	C≡C−C−CH−EWG	1	1	0
150	thioamide	1	0	1	203	selenide	1	0	1
151	C=C−OR	1	1	0	204	NC(=O)S	1	0	1
152	N−C(=O)S−C	1	0	1	205	cyclohexene	1	1	0
153	C=C−N in indole	1	1	0	206	RN=NR′	1	0	1
154	sulfonyl imine	1	1	0	207	$RR'N-C-CH-CO_2R''$	1	1	0
155	C=C	1	1	0	208	RR″CCl	1	0	1
156	ketenamine	1	1	0	209	S−CH(−C)(−N)	1	0	1
157	C=C in enol lactone	1	1	0					
158	benzyl ketal (**10**)	1	1	0					
159	C=C−CHO	1	1	0					

reaction center and 2.5 inert datasets from functional groups from the reaction center were expectedly extracted in a reaction scheme.

4. RESULTS AND DISCUSSION

4.1. Preliminary Extraction of Interaction Data. The extraction of interaction data from the CCR1995−7 database (2763 reactions) was first performed. The result is summarized in Table 4. The reaction conditions used and the top-10 functional groups are listed in Tables 5 and 6, respectively.

Using preliminary extraction, 461 interfering data, 748 inert data, and 123 relative rate data were found. The number of functional groups and reaction conditions included in the extracted data were 197 and 133, respectively. The combination of the functional groups and reaction conditions is 26201, whose 4.61% was complied into database. The result describes that the sum of inert and relative rate data per an interfering data was 1.9, which was slightly smaller than the expected number 2.5.

In the interaction data, the most referred condition was Pd(0) and the number of datasets was 82, including 41 inert datasets. This was followed by 47 datasets with $NaBH_4$ and 41 datasets with CF_3CO_2H (see Table 5). Most functional groups (74) included an ether, including 22 interfering datasets. Along with the ether, olefin (58) and ester (57) also were ranked (see Table 6). In this preliminary data extraction, four relative rate datasets (in other words, four pairs of reaction schemes) were determined to be contradictory, as shown in Table 7.

In Table 7, entries 1 and 2 describe the contradiction of relative reactivity between ester and ether under the Pd(0) condition. Entries 3 and 4 describe the contradiction between enamine and ether under the Pd(0) condition, entries 5 and 6 describe the contradiction between epoxide and ester under the Pd(0) condition, and entries 7 and 8 describe the contradiction between alcohol and ether under the CF_3CO_2H condition. All of the contradictory relative rate data contain a deprotecting reaction, and ether groups that protect the reactive alcohol are mainly included. Generally, an ether bond, especially a dialkyl ether, is relatively stable under various reaction conditions, so it is used as a protecting group.

The ethers of contradictory relative rate data—BnO, ^{t}BuO, TrO, etc.—were beyond the default definition of a functional group. The default was based on the center O and neighboring atoms, the number of H atoms, and bond types. Here, new expanded functional groups were defined like subclass in LHASA. The seven functional groups for ether were defined: benzyl ether, **1**; methoxy(or ethoxy) aryl, **2**; *iso*-propoxy aryl, **3**; trityl ether, **4**; *t*-butyl ether, **5**; oxetanyl, **6**; and tetrahydrofuranyl, **7**. In the same manner, for general secondary and tertiary amino groups, a general ester and ketal, a new secondary and tertiary benzyl amino group (**8**), benzyl ester (**9**), and benzyl ketal (**10**) were defined. In addition, imine groups in heteroaromatic rings (**11**, **12**, **13**, and **14**) were considered separately, because the chain structures of aldenamine (N−CH=C), ketenamine (N−CR=C), and heteroaromatic rings were not distinguished in SYNSUP (see Figure 7).

The current definitions of functional groups have limitations for the presentation of all reactive sites. In addition, depending on reaction databases, ambiguous definitions of reaction conditions are sometimes encountered. For example, entries 1, 2, 4, 5, and 6 in Table 7 are concerned with hydrogenation, but entry 3 involves dehydrogenation in the presence of palladium. These definitions must be revised manually. At the least, the revision must be done so that contradictory data are not extracted.

4.2. Results of Extraction of Interaction Data. After revising some functional groups, four types of interaction datasets have been extracted from the reaction datasets: CCR1995, CCR1996, CCR1997, and CCR1995−7 (see Table 8). CCR1995, CCR1996, and CCR1997 have been prepared to inspect the number of datasets that are newly added every year.

The number of interaction data from the CCR1995−7 database (see Table 8) increased, in comparison to the preliminary extraction (Table 4), because the number of functional groups increased. On the other hand, the number of relative rate data decreased. The reason for this decrease was that relative rate data did not include any contradictory datasets. Tables 9 and 10 summarize 465 interfering and 813 inert datasets from CCR1995−7, including 133 reaction conditions and 209 functional groups.

The most frequently found reaction condition was Pd(0), which was referred by 82 interfering or inert data. This means that the reactivity of Pd(0) with 82 types of functional groups was observed. This is 39.2% of all functional groups. We see the unique average of 4.6% in Table 8. The frequency of Pd(0) was followed by $NaBH_4$ (53 times), and CF_3CO_2H (45 times). The frequency of Pd(0) means that Pd(0) provides a high reaction yield in functional group changes and deprotecting reactions.

No interfering data of ether was extracted using redefined functional groups. It was also found that ether groups were reactive only when they were protected with groups such as ROBn (**1**) and ArOMe (**2**). The average number of exploitable reactions in the CCR1995 to CCR1997 databases was 921.0, and the average number of extracted interfering and inert data were, respectively, 196.0 and 332.3. Broadly speaking, every year, out of 1000 reaction schemes, 20% of reactions are converted to interfering data, and 1.6% of the combinations of two interfering data are converted to relative rate data.

The sum of relative rate data from databases CCR1995, CCR1996, and CCR1997 were 34 (= 13 + 8 + 13). The number is less than the 62 relative rate datasets from the CCR1995−7 database. This is due to the extraction method. First, interfering datasets are only extracted by focusing on the reaction centers of all reaction schemes stored in the database. Next, from the first scheme in the databases, relative rate and inert data are collected, focusing on the reaction centers (see Figure 4). Suppose an interfering datum of a functional group in the reaction center under a reaction condition is found, but the corresponding inert data already exist. The data then are classified as the interfering data, but the prospective new relative rate data are never stored, because of a lack of information (ID=4 in Figure 3).

All relative rate data extracted from the CCR1995−7 database are shown in Table 11.

Despite the fact that a relative rate datum consists of two functional groups and a reaction condition, relative rate reactivity among of more than two functional groups under a specific reaction condition may be predicted by considering more than one relative rate data (see Table 12).

The extracted relative rate data suggest that, against $NaBH_4$, an aldenamine is less reactive than an ester,[34] an ester is less reactive than a ketone,[19] and a ketone is less reactive than an aldehyde,[35] which are stored individually. The relation deductively suggests the relative rate reactivity between an aldemaine and a ketone, between an aldenamine and an aldehyde, and between an ester and an aldehyde, which are not stored in the relative rate data. SYNSUP accommodates the deduced suggestion.

This time, some functional groups were redefined only to avoid a contradiction in the extracted relative rate data. Other functional groups concerned with protecting groups may be reconsidered for the purpose of a more-precise protecting group proposal. The ratio of extracted interfering or inert data is <5%, compared to the combination of functional groups and reaction conditions. In a synthesis design procedure, it may be found that the influence of the reaction

Table 11. Relative Rate Data Extracted from the CCR1995−7 Database

conditions	reactive FG[a]	more-reactive FG
aqueous H^+	carboxylic acid	nitrile
mild base	nitro	sulfide
strong base	epoxide	ester
NaH	ester	RCH_2Br
LDA	**2**	**1**, RCONR″
alkoxide	ester	RCH_2Cl
amine	nitro	ketal
H_2S	**1**	azide
$LiA1H_4$	imine	ester
	carbamate	ester
B_2H_6	C=CR_2	CH=CHCH=CHCOR
	RCONHR′	**1**
$NaBH_4$	imine	aldehyde
	aldenamine	ester
	azide	ketone
	ketone	aldehyde
	ester	ketone, RCH_2Cl, imine
DIBAL	**2**	ester, epoxide
CrO_3/Pyridine	C=CHR	*sec*-alcohol, RCH_2OH
I_2	C=CHR	phosphite, RCH_2OH, *tert*-amine
	RCONR″	C=CHR
MnO_2	*C* = CHR	*sec*-alcohol, RCH_2OH
	sec-alcohol	RCH_2OH
ZnO_2/ThO_2	ester	ketone
Pd(0)	*C* = CR_2	*C* = CHR,9, Cyclopentene
	epoxide	**9**, $R_2C(OH)C=CH_2$
	aldenamine	C=C
	1	CH_2−CH=CH, azide, **9**, alkynyl ketone, nitro, ArCH=C−EWG[b]
Na	ketone	ester
HCl	nitrile	nitro
	ester	**1**, ketal, tetraalkyl silane,
	nitro	imide
	1	C−N=N−N
^{t}BuOOH	C=CHR	phosphite
CF_3CO_2H	*sec*-alcohol	**4**
	N−CHOR−C	**4**
	ester	OP(OR)(−N)$_2$
	RCONR″	**9**, ester
	11	ester
Na_2S	carbamate	ester
$FeCl_3$	ketal	**4**
SmX_2	ketone	azide

[a] Functional group. [b] EWG = electron withdrawing group.

Table 12. Reactivity of Interfering Functional Groups Derived from Relative Rate Data

No.	conditions	reactivity orders of interfering functional groups[a]
1	$NaBH_4$	aldenamine < ester < RCH_2Cl
2		aldenamine < ester < imine < aldehyde
3		aldenamine < ester < ketone < aldehyde
4		azide < ketone < aldehyde
5	I_2	RCONR″ < C=CH_2 < RR″N
6		RCONR″ < C=CH_2 < phosphite
7		RCONR″ < C=CH_2 < RCH_2OH
8	MnO_2	C=CHR < *sec*-alcohol < RCH_2OH
9	HCl	nitrile < nitro < *N*-alkyl imide
10		ester < ROBn (**1**) < R−N=N−N
11	CF_3CO_2H	RCONR″ < ester < OP(OR)(−N)$_2$
12		CH in heterocycles (**11**) < ester < OP(OR)(−N)$_2$

[a] Components on the right-hand sides are more reactive.

conditions on some functional groups outside of the reaction center is not clear, because of a lack of the necessary interaction data. To handle this inconvenience, it is necessary to extract as large of a reaction database as possible.

5. CONCLUSION

For the purpose of synthesis design for multifunctional group targets in SYNSUP, interaction data to determine the influence of reaction conditions on functional groups have been developed. It consists of interfering, inert, and relative rate data. The interfering and inert data describe the reactivity of a functional group under reaction conditions, and the relative rate data describes the reactivity of two functional groups against a reaction condition. The three types of data are automatically collected from the reaction database. It was found that the definition of functional groups should be expanded for the protected functional groups, especially protected functional groups. To determine the interaction, the reaction database should be scanned twice. The first scan is performed to obtain interfering data, and the second scan is performed to obtain inert and relative rate data. By preparing not only interfering and inert data but also relative rate data, we can precisely recognize the reactivity of functional groups under reaction conditions. With the use of multiple relative rate datasets, the reactivity of more than two functional groups could be deduced.

REFERENCES AND NOTES

(1) Zass, E. *Encyclopedia of Computational Chemistry*; Schleyer, R. R., Ed.; Wiley: Chichester, U.K., 1998; pp 2402−2420.

(2) Todd, M. H. Computer-aided organic synthesis. *Chem. Soc. Rev.* **2005**, *34*, 247–266.

(3) (a) Gelernter, H. L.; Sanders, A. F.; Larsen, D. L.; Agarwal, K. K.; Boivie, R. H.; Spritzer, G. A.; Searleman, J. E. Empirical Explorations of SYNCHEM. *Science* **1977**, *197*, 1041–1049. (b) Hendrickson, G. J. B. Organic Synthesis in the Age of Computers. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 1286–1295. (c) Funatsu, K.; Sasaki, S. Computer-Assisted Organic Synthesis Design and Reaction Prediction System AIPHOS. *Tetrahedron Comput. Method.* **1988**, *1*, 27–37. (d) Ihlenfeldt, W. D.; Gasteiger, J. Computer-Assisted Planning of Organic Syntheses: The Second Generation of Programs. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 2613–2633. (e) Barone, R.; Attolini, M.; Arbelot, M.; Chanon, M. What To Do With This Reaction? A New Program in the Field of Computer-Aided Synthesis Design - Application to the Diels-Alder Reaction. *Eur. J. Org. Chem.* **2006**, 5184–5190. (f) Law, J.; Zsoldos, Z.; Simon, A.; Reid, D.; Liu, Y.; Khew, S. Y.; Johnson, A. P.; Major, S.; Wade, R. A.; Ando, H. Y. Route Designer: A Retrosynthetic Analysis Tool Utilizing Automated Retrosynthetic Rule Generation. *J. Chem. Inf. Model.* **2009**, *49*, 593–602.

(4) Corey, E. J.; Wipke, W. T. Computer-Assisted Design of Complex Organic Syntheses. *Science* **1969**, *166*, 178–192.

(5) Corey, E. J.; Petersson, A. An Algorithm for Machine Perception of Synthetically Significant Rings in Complex Cyclic Organic Structures. *J. Am. Chem. Soc.* **1972**, *94*, 460–465.

(6) Bersohn, M. Automatic Problem Solving Applied to Synthetic Chemistry. *Bull. Chem. Soc. Jpn.* **1972**, *45*, 1897–1903.

(7) Dogane, I.; Takabatake, T.; Bersohn, M. Computer-executed synthesis planning, a progress report. *Recl. Trav. Chim. Pays−Bas* **1992**, *111*, 291–296.

(8) Tanaka, A.; Kawai, T.; Takabatake, T.; Oka, N.; Okamoto, H.; Bersohn, M. Synthesis of an azaspirane via Birch reduction alkylation prompted by suggestions from a computer program. *Tetrahedron Lett.* **2006**, *47*, 6733–6737.

(9) Corey, E. J.; Orf, H. W.; Pensak, D. A. Computer-Assisted Synthetic Analysis. The Identification and Protection of Interfering Functionality in Machine-Generated Synthetic Intermediates. *J. Am. Chem. Soc.* **1976**, *98*, 210–221.

(10) Corey, E. J.; Long, A. K.; Greene, T. W.; Miller, J. W. Computer-Assisted Synthetic Analysis. Selection of Protective Groups for Multistep Organic Syntheses. *J. Org. Chem.* **1985**, *50*, 1920–1927.

(11) Jorgensen, W. L.; Laird, E. R.; Gushurst, A. J.; Fleischer, J. M.; Gothe, S. A.; Helson, H. E.; Paderes, G. D.; Sinclair, S. CAMEO: A program for the logical prediction of the products of organic reactions. *Pure Appl. Chem.* **1990**, *62*, 1921–1932.

(12) Röse, P.; Gasteiger, J. Automated Derivation of Reaction Rules for the EROS 6.0 System for Reaction Prediction. *Anal. Chim. Acta* **1990**, *235*, 163–168.

(13) Satoh, H.; Funatsu, K. SOPHIA, a Knowledge Base Guided Reaction Prediction System−Utilization of a Knowledge Base Derived from a Reaction Database. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 34–44.

(14) Bersohn, M. The computer derivation of stereochemical relationships from the chirality of ring atoms. *J. Chem. Soc., Perkin Trans. 1* **1979**, 1975–1977.

(15) Lewis, N.; Taylor, R. J. K.; Watoson, R. J.; McKillop, A. A Simple and Efficient Procedure for the Preparation of Chiral 2-Oxazolidinones from α-Amino Acides. *Synth. Commun.* **1995**, *25*, 561–568.

(16) Soai, K.; Yokoyama, S.; Ookawa, A. Reduction of Azides to Amines with Sodium Borohydride in Tetrahydrofuran with Dropwise Addition of Methanol. *Synthesis* **1987**, 48–49.

(17) Takikawa, H.; Shimbo, K.; Mori, K. Pheromone Synthesis, CLXXXIII. Synthesis of (1R,2R,5S,7R)- and (1R,2S,5S,7R)-2-hydroxy-exo-Brevicomin, the Components of the Male-Produced Volatiles of the Mountain Pine Beetle, *dendroctonus ponderosae*. *Liebigs Ann.* **1997**, 821–824.

(18) Varma, R. S.; Kabalka, G. W. Allylic Alcohols Via the Chemoselective Reduction of Enone Systems with Sodium Borohydride in Methanolic Tetrahydrofuran. *Synth. Commun.* **1985**, *15*, 985–990.

(19) Martin, S. F.; Clark, C. W.; Corbett, J. W. Applications of Vinylogous Mannich Reactions. Asymmetric Synthesis of the Heteroyohimboid Alkaloids (−)-Ajmalicine, (+)-19-epi-Ajmalicine, and (−)-Tetrahydroalstonine. *J. Org. Chem.* **1995**, *60*, 3236–3242.

(20) Kapeller, H.; Griengl, H. Synthesis of methyl 5-azido-5-deoxy-2,3-*O*-isopropylidenecarba-α-D-allo-hexafuranuronate, the sugar part of carbapolyoxins and carbanikkomycins. *Tetrahedron* **1997**, *53*, 14635–14644.

(21) Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J. Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 244–255.

(22) Benati, L.; Montevecchi, P. C.; Nanni, D.; Spagnolo, P.; Volta, M. Reduction of azides to amines by samarium diiodide. *Tetrahedron Lett.* **1995**, *36*, 7313–7314.

(23) Maiti, S. N.; Singh, M. P.; Micetich, R. G. Facile conversion of azides to amines. *Tetrahedron Lett.* **1986**, *27*, 1423–1424.

(24) Hamelin, O.; Greene, A. E.; Depres, J. P.; Declercq, J. P.; Tinant, B. Highly Stereoselective First Synthesis of an A-Ring-Functionalized Bakkane: Novel Free-Radical Approach to 9-Acetoxyfukinanolide. *J. Am. Chem. Soc.* **1996**, *118*, 9992–9993.

(25) Current Chemical reactions, which is one of reaction databases in ISIS/Base. Information of the database and system is available via the Internet at http://www.symyx.com/products/databases/synthesis/index.jsp.

(26) Sajiki, H. Selective inhibition of benzyl ether hydrogenolysis with Pd/C due to the presence of ammonia, pyridine or ammonium acetate. *Tetrahedron Lett.* **1995**, *36*, 3465–3468.

(27) Tamagnan, G.; Gao, Y. G.; Bakthavachalam, V.; White, W. L.; Neumeyer, J. L. An efficient synthesis of *m*-hydroxycocaine and *m*-hydroxybenzoylecgonine: two metabolites of cocaine. *Tetrahedron Lett.* **1995**, *36*, 5861–5864.

(28) Back, T. G.; Ghau, J. H. L.; Parvez, M. A Convenient Synthesis of 11-aza-C-homoestranes. *Synthesis* **1995**, 162–164.

(29) Boger, D. L.; Mckie, J. A.; Nashi, T.; Ogiku, T. Enantioselective Total Synthesis of (+)-Duocarmycin A, epi-(+)-Duocarmycin A, and Their Unnatural Enantiomers. *J. Am. Chem. Soc.* **1996**, *118*, 2301–2302.

(30) Dragovich, P. S.; Prins, T. J.; Zhou, R. Palladium Catalyzed, Regioselective Reduction of 1,2-Epoxides by Ammonium Formate. *J. Org. Chem.* **1995**, *60*, 4922–4924.

(31) Kim, Y. M.; Kim, D. H. Convenient Preparation of All Four Possible Stereoisomers of 2-Benzyl-3,4-epoxybutanoic Acid, Pseudomechanism-based inactivator for Carboxypeptidase A via α-Chymotrypsin-Catalyzed Hydrolysis. *Bull. Korean Chem. Soc.* **1996**, *17*, 967–969.

(32) Zhao, W. M.; Wan, X. Y. A Practical Synthesis of 4-(3′,4′-Dihydroxylphenyl)-1,2,3,4-tetrahydroisoquinoline. *Org. Prep. Proced. Int.* **1995**, *27*, 513–516.

(33) Xia, X. Y.; Wang, J. Y.; Hager, M. W.; Sisti, N.; Liotta, D. C. Stereocontrolled synthesis of β-2′-deoxypyrimidine nucleosides via intramolecular glycosylations. *Tetrahedron Lett.* **1997**, *38*, 1111–1114.

(34) Baldwin, J. E.; Fryer, A. M.; Spyvee, M. R.; Whitehead, R. C.; Wood, M. E. Stereocontrol in the synthesis of kainoids. *Tetrahedron Lett.* **1996**, *37*, 6923–6924.

(35) Martin, S. F.; Clark, C. W.; Corbett, J. W. Applications of Vinylogous Mannich Reactions. Asymmetric Synthesis of the Heteroyohimboid Alkaloids (−)-Ajmalicine, (+)-19-epi-Ajmalicine, and (−)-Tetrahydroalstonine. *J. Org. Chem.* **1995**, *60*, 3236–3242.

CI9004332