Computer-Assisted Synthetic Analysis. The Identification and Protection of Interfering Functionality in Machine-Generated Synthetic Intermediates

E. J. Corey,* H. W. Orf, and David A. Pensak

Contribution from the Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138. *Received May 7, 1975*

Abstract: The Harvard program for computer-assisted synthetic analysis (LHASA) has been given the ability to identify those functional groups in a structure which will interfere with the effective operation of a synthetic reaction and also to discern whether or not those groups can be protected. The program module (FGR) which is responsible for this function contains data on the level of reactivity of each functional group toward a large number and diversity of chemical reagents. Functional groups are subclassified, when appropriate, according to the environmental features within a molecule which can affect group reactivity. Further, functional groups are characterized either as protectable (e.g., hydroxyl, carbonyl) or unprotectable (e.g., nitro). During the operation of the FGR program module, the transform evaluation process is conducted so that an examination is made of both the target (product) molecule and the precursor (reactant) molecule. Current capabilities, limitations, and further extensions of this program module are outlined.

The deactivation or masking of functional groups is a very important part of synthetic methodology and practice. It may involve the use of an externally derived "protecting" group or take advantage of a connection to another unit within the molecule. The use of these "control operations"[1] and others (e.g., activation of functional groups or the introduction of directing groups) often contributes crucially to the experimental realization of a synthetic plan. Generally, the larger the number of functional groups in the molecule to be synthesized, the more likely will be the need for and the greater the importance of functional group protection. This area of organic synthesis is characterized less by the availability of a small number of "ideal" protecting groups than by the alternative of a large arsenal of protecting groups each with a definitely restricted range of applicability. A protecting group must convert some functional group to a form which will not cause interference with reactions aimed at modifying other units in the molecule. Further, there is a reciprocal requirement that the various units in the molecule not interfere with the attachment or removal of a protecting group. The extension of computer-assisted synthetic analysis to sophisticated levels necessitates the detection of functional group interference in synthetic reactions and the use of this information in a manner which is useful to the chemist. This paper[2] deals with the handling of functional group interference and protection in the current version of the Harvard program for computer-assisted synthetic analysis (LHASA). The sections which follow outline the chemical data contained in the program, including those relevant to chemical reagents, reaction conditions and functional group reactivity, and the way in which these data are used by the program.

I. The Functional Group/Reagent Cross-Reference Table

A. Group Reactivities. As indicated previously,[3-5] "trees" of synthetic intermediates are generated in LHASA from a target molecule by antithetic (retrosynthetic) analysis. The conversion of a structure in the "synthetic tree" to another structure corresponding to a synthetic precursor is accomplished by the application of a "transform" (retroreaction). Associated with each transform in LHASA are sets of reaction conditions which permit the realization of the synthetic step corresponding to that transform.

The assessment of the consequences of applying a specific set of reaction conditions to a given precursor molecule is carried out in LHASA by a Functional Group Reactivity (FGR) module. This module evaluates the chemical reactivity of each functional group present toward the individual reagents which constitute the set of reaction conditions being considered. In order for a general evaluation of the chemical reactivity of *any* functional group toward *any* particular reagent to be made by this program module, information must be available concerning the reactivity of every recognized functional group toward a great number and diversity of chemical reagents. The required information has been gathered and assembled into a large functional group/reagent (FG/RGNT) cross-reference table. This table, a portion of which is shown in expanded form in Figure 1 and which is contained in its entirety in condensed form in Tables I and II, constitutes the chemical data base of the FGR module. It includes the reactivity levels of all recognizable groups toward 60 different reagents. These reagents, which are collectively referred to as the reagent library, taken individually or in combinations, suffice to describe all reaction conditions presently used by existing LHASA transforms (vide infra). New reagents and hence new reaction conditions can be added easily whenever the need arises.

Because the reactivity of many functional groups can depend dramatically on their molecular environment, it was necessary to divide the FG/RGNT table into two sub-tables. One sub-table, REACTB (REACtive functional group TaBle), consists of functional group types whose reactivities are to a reasonable approximation independent of their molecular environment and are determined more by the reactive nature of the groups themselves. Functional groups such as isocyanates, oximes, disulfides, etc., are represented in this sub-table (Table I).

The other sub-table, ENVRTB (ENViRonment influenced functional group TaBle), consists of functional groups whose reactivity depends markedly on their molecular environment (see Table II). Halides, alcohols, electron-withdrawing groups, olefins, and many of the more common organic functional groups fall into this category. Groups of this type are subclassified in ENVRTB according to the structural and chemical features within their molecular environments which are apt to influence reactivity toward a particular chemical reagent. The utilization of a suitable number of subclasses for a functional group permits a rea-

	pH<1	pH2-4	pH4-6	pH9-10	pH>10	*RLi/RMgX*	Orgzinc	R_2Cu	Wittig	Str Nu:
Phosphonate	High	Low	Low	Low	High	High	Low	Low	Low	Low
Ether	High	Low	Low	Low	Low	Low	Low	Low	Low	Low
Peroxide	High	Low	Low	Low	High	High	Low	*Moderate*	Low	High
Nitrite	High	Low	Low	High	High	High	High	High	High	High
Dihalide	High	Low	Low	*Moderate*	High	High	*Moderate*	High	Low	*Moderate

...continued

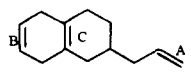
	pH<1	pH2-4	pH4-6	pH9-10	pH>10	*RLi/RMgX*	Orgzinc	R_2Cu	Wittig	Str Nu:
CH_2 epoxides	High	*Moderate*	Low	Low	*Moderate*	High	*Moderate*	High	Low	*Moderate
Non-CH_2 epoxides	High	*Moderate*	Low	Low	*Moderate*	*Moderate*	Low	*Moderate*	Low	Low
Alpha D or alpha C(π) epoxides	High	High	*Moderate*	Low	*Moderate*	High	*Moderate*	High	Low	*Moderate
Alpha W epoxides	High	Low	Low	*Moderate*	High	High	High	High	*Moderate*	*Moderate

...continued

Figure 1. Portions of the functional group/reagent (FG/RGNT) cross-reference table. The upper excerpt is from the REACTB sub-table (Table I), while the lower is from the ENVRTB sub-table (Table II). The entire FG/RGNT table is reproduced in condensed form in Tables I and II.

sonable assessment of the chemical reactivity of that group in a great diversity of molecular environments (vide infra).

B. The Designation of Reactivity Levels. Under the present scheme, the reactivity level of a functional group or group subclass toward a particular chemical reagent has been designated as low, moderate, or high. The selection of three levels of reactivity was derived as a compromise between the use of some larger number which would necessitate making finer reactivity distinctions than are feasible, and the use of just two levels which would not allow for the degree of selectivity achievable in practice with many synthetic reagents. In structure I, for example, although all the



1

double bonds are reactive toward catalytic hydrogenation over platinum, it is still possible to hydrogenate selectively bond A or bonds A and B in the presence of C, indicating the need for more than just a two level, reactive-nonreactive scheme.

The use of a three level reactivity scheme is by no means a rigorous solution. As more information concerning relative reactivities among different functional groups becomes available and as new, more discriminatory reagents continue to be developed, it will become possible and appropriate to make finer reactivity distinctions by increasing the number of reactivity levels in the scheme.

During the course of compiling the FG/RGNT table, several instances arose where conclusive information concerning the reactivity of a particular functional group toward a specific reagent was not available. For example, the reactivity at ambient temperature of an isocyanide toward Pb(IV) or the reactivity of an azide toward reagents such as peracid, CrO_3/pyr, or CrO_3/H^+ is not unequivocally known. In such instances, tentative reactivity levels (generally low) were designated to complete the table.

C. The Reagent Library. The 60 reagents listed in the FG/RGNT cross-reference table comprise the basic units of LHASA's reagent library (Figure 2) which is composed of 11 categories, A-K, according to chemical type. Although a very large number of individual chemical reagents is available to the chemist, it is unnecessary to enumerate every reagent available within each condition category since a number of reagents of similar type can be adequately characterized with respect to reactivity and selectivity by a representative or "prototype" reagent. For example, the three reagents listed under condition category F—Na/NH_3, Al(Hg), and $SnCl_2$—are respectively prototypical representatives of strong, moderate, and mild basic reducing reagents. Similarly, the B_2H_6 subclass of condition category G is taken as representative of a large and diverse family of versatile borane reductants.[6]

A. Aqueous
1. pH<1
2. pH 2-4
3. pH 4-6
4. pH 9-10
5. pH>10

B. Organometallic
6. Organometallic (RLi, RMgX)
7. Organozinc
8. Organocopper
9. Wittig

C. Nucleophilic
10. CN^-, SR^- (strongly NU:)
11. NH_3 (NU:)
12. OAc^- (mildly NU:)

D. Catalytic Redn
13. NiH (Raney Ni)
14. $H_2/Pt/H^+$
15. H_2/Pd
16. H_2/Lindlar

E. Acidic Redn
17. Zn/HCl
18. Zn/HOAc
19. Cr(II)

F. Basic Redn
20. Na/NH_3
21. Al(Hg)
22. $SnCl_2$

G. Hydride Sources
23. $LiAlH_4$
24. $NaBH_4$
25. AlH_3
26. B_2H_6

H. Lewis Acids
27. $AlCl_3$, BF_3
28. $SnCl_4$

I. Soft Acids
29. Hg(II)
30. Pb(IV)/25°

J. Oxidizing Agents
31. OsO_4
32. $KMnO_4$
33. O_3/-50°
34. RCO_3H
35. CrO_3/pyr
36. CrO_3/H^+
37. H_2O_2/OH^-
38. Quinone
39. DMSO
40. NaOCl
41. NBS
42. I_2
43. Cl_2/Br_2
44. MnO_2
45. HIO_4
46. SeO_2/H^+
47. SeO_2/pyr

K. Miscellaneous
48. Ag(I)
49. Radical
50. R_3SnH
51. $Ni(CO)_4$
52. $NaNR_2$
53. CH_2N_2
54. $SOCl_2$
55. Ac_2O/25°
56. Ac_2O/80°
57. DCCD
58. Pyrolysis
59. CH_3I
60. Pb(IV)/80°

Figure 2. LHASA's reagent library.

Using these prototype reagents, either individually or in combination, realistic sets of reaction conditions can be expressed for a great variety of chemical transformations. Consider, for example, two of the methods commonly employed for the complete reduction of ketones, Clemmensen and Wolff-Kishner reductions. Conditions for the Clemmensen reduction correspond exactly to prototype reagent 17 (Zn/HCl), while the NH_2NH_2/KOH conditions of the Wolff-Kishner reduction are expressed adequately by prototype reagents 11 (Nu:) and 5 (pH >10).

D. Derivation of Group Subclasses. As discussed above, the chemical reactivity of many functional groups can be strongly influenced by their immediate molecular environment. Allylic, benzylic, and primary, secondary, and tertiary aliphatic alcohols differ markedly in their reactivity toward reagents such as Ac_2O, MnO_2, and HCl. Carbon-carbon double bonds differ in reactivity depending on degree of alkylation, attachment of electron-withdrawing or -donating groups, and strain. In order to evaluate the chem-

Table I. The REACTB Sub-table

REAGENT / NUMBER / FG	1 1234567890	1111111112 1234567890	2222222223 1234567890	3333333334 1234567890	4444444445 1234567890	5555555556 1234567890
ACID	LLLLLMLLLL	LLLLLLLLLL	LLHLLHLLLL	LLLLLLLLML	LLLLLLLLLL	LLHHLLMLLM
AMIDE*1	HLLLHHLLLL	LLLLLLLLLL	LLHLHHLLLL	LLLLLLLLLH	HLLLLLLLLL	LLLHLHLLLL
AMIDE*2	HLLLHHLLLL	LLLLLLLLLL	LLHLHHLLLL	LLLLLLLLLH	HLLLLLLLLL	LLLHLHLLLL
AMIDE*3	HLLLHHLLLL	LLLLLLLLLH	LLHLHHLLLL	LLLLLLLLLL	LLLLLLLLLL	LLLLLLLLLL
CARBONIUM	HHHHHHHHHH	HHHHHHHHHH	HHHHHHHHHH	HHHHHHHHHH	HHHHHHHHHH	HHHHHHHHHH
ISOCYANATE	HHHHHHHHHH	HHHHHLHHHH	HLHHHHHLLL	LHHLHHHLLH	LLLHHHHLLH	LHHLLLLHLM
ACID*HALIDE	HHHHHHHHHH	HHHHHHHHHH	HHHHHHHHHH	HHHHHHHHHH	HLLHHHHHHH	HHHLHHLHLH
THIOESTER	HLLHHHHLHH	HLHHHLHLLH	LLHHHHLLLL	LHHLLLHLLH	HLHLLLLLLL	LHLLLLLHLM
AMINE*3	LLLLLLLLLL	LLLLLLLLLL	LLLLLLLLHH	HHHHHLHHLL	HLHHHLLLHL	LLLLLLLLHH
AZIRIDINE	LLLLLLLLLL	LLLLLLLLLH	LLLLLLLLHH	HHHHHHHHLH	HLHLHMLLHL	LLLLMHLLHH
AMINE*2	LLLLLLLLLL	LLLLLLLLLL	LLLLLLLLHH	HHHHHMHHHH	HLHHHMLLHL	LLLHHHLLHH
AMINE*1	LLLLLLLLLL	LLLLLLLLLL	LLLLLLLLHH	HHHHHMHHHH	HLHHHMLLHL	LLLHHHLLHH
NITROSO	LLLLHHHHHL	LLHHHHHHHH	HHHHHHLLLL	MHHHLHHLLH	HLLLLMMLHH	LHLHLLLHLM
DIAZO	HHHLLHHHHL	LLHHHHHHHH	HHHMHHHHHH	MHHHHHLHHH	HHHLHHHHHH	HLLHHHHHMH
HALOAMINE	LLLHHHHHHH	HHHHHHHHHH	HHHHHHHHLL	LMHHHHHHLL	HLLHHHHHHH	LHLLLLLLLM
HYDRAZONE	HLLLHHLLLH	LLHHHLHHLH	LLHLHHLLHH	HHHHHHMHHH	HHHHHHMHHL	LLLHHHLHMH
OXIME	HLLLLLLLLL	LLHHHLHHHH	HLHLHHHLLH	LHHLLHLLHH	HMHLMLLLHL	LLHHHHHHLH
IMINE	HHLLLHHMHH	LLHHHLHHHH	MLHLHHLLLH	HHHHHHMHHH	HHHHHHHLHL	LLLHHHHHMH
THIOCYANATE	LLLLHHHLLH	LLHHHLHLLH	LLHLLHLLLL	LHHHLMLLLL	LLLLLLLLLL	LHLLLLLHLM
ISOCYANIDE	HHHLLHHHLL	LLHHHHHHHH	HLHLHHHHLL	LHHHHHLHHH	HHHHHHLLHH	HHLHHHHHLM
AZO	LLLLHHHLLL	LLHHHHHHHH	HHHHHHLLLL	LLHHLHHLLL	LLLLLLLLHH	LHLLLLLHLM
HYDROXYLAMINE	HLLLLLLLLL	LLHHHMHMMH	MLHLHMHMMH	HHHHHHMHMH	HHHHHMMLHL	LLLHHHLMHH
AMINE*OXIDE	HLLLLHLLLL	LLHHHHHHHH	HHHMHHHLLL	LLLLLLLLLL	LLLLLLLLLH	LMLHHHLHLL
THIOL	LLLLLLLLLL	LLHHHHLLLL	LLLLLLLLLH	LHHHLHLHLH	HHHHHHHLHL	LLHHHHMHHH
EPISULFIDE	HLLLHHHLLH	LLHHHHHHMH	MLHLHHHLMH	LHHHLHHLLH	HMHLHHHHHH	HHLLLHLHHH
SULFIDE	LLLLLLLLLL	LLHHHHLLLH	LLLLLLLLLH	LHHHLHLLLH	HLHLHMLLLL	LLLLLLLLHH
C*SULFONATE[a]	LLLLLMLLLL	LLLLLLLLLL	LLHLMMLLLL	LLLLLLLLLL	LLMLLLLLLL	LLLLLLLLLL
PHOSPHINE	LLLLLLLLLL	LLLLLLLLLL	LLLLLLLLLH	HHHHHHHMLM	HHHMHMMLLL	LLLLLLLLHH
PHOSPHONATE	HLLLHHLLLL	LLLLLLLLLH	LLHLLLLLLL	LLLLLLLLLL	LLLLLLLLLL	LHLLLLLLLL
ETHER	HLLLLLLLLL	LLLLLLLLLL	LLLLLLLLLL	LLLLLLLLLL	HLLLLLLLML	LLLLLLLLLL
PEROXIDE	HLLLHHLMLH	LLHHHHHHHH	HHHHHHHLLL	LLLLLLLLLL	HHHLLLHLHH	HHLHLHLHLM
NITRITE	HLLHHHHHHH	HLHHHLHHHH	LHHHHHLLLL	LHHHHHHLLH	HHHHMHHLHH	HHLHLHLHLM
DIHALIDE	HLLMHHMHLM	LLHHHLHMMH	MLMLLLMLLL	LLLLLLLLLL	LLLLLLLLMM	LHLLLLLLLL
TRIHALIDE	HLLMHHMHLM	MLHHHLHMMH	MLMLLLMLLL	LLLLLLLLLL	LLLLLLLLHH	MHLLLLLMLL
HYDRATE	HHLHHHHHHH	HLHHLLHHHH	HHHHHHHHHH	LHLHHHHLLH	HHHLHHHHMH	HHHHHHHHHH
HEMIKETAL	HHLHHHHHHH	HLHHLLHHHH	HMHHHHHHHH	HHLHHHHLLH	HHHLHHHLLL	LHLHHHHHLH
KETAL	HHLLLLLLLL	LLLHLLHHLL	LLLLLLHHLL	LLLLLHLLLL	LLLHLLHLLL	LLLLLHLLLL
HEMIACETAL	HHLHHHHHHH	HLHHLLHHHH	HMHHHHHHHH	HHHHHHHLLM	HHHLHHHLLL	LHLHHHHHLH
ACETAL	HHLLLLLLLL	LLLHLLHHLL	LLLLLLHHLL	LLHLLHLLLL	LLMLLHLLLL	LLLLLHLLLL
AZIDE	HLLLLHHLLM	LLHHHHHHHH	HHHHHHHHMM	LMHLLLLLLL	MLHLLLLLLL	LLLLLMLHLL
DISULFIDE	HLLLLHHMHH	LLHHHHHHHH	HHHLHHHLHH	MHHHLHLHLH	HHHHHHHMHH	HHLLLLLHMH
ALLENE	HLLLLLLLLL	LLHHHHHLLH	LLLLHHHLHH	HHHHLLLLLL	HLHLLHMLHH	LLLLLLLLLH
VINYLW[b]	HLLHHHHHHH	HLHHHLHHHH	HLHHHHHHLH	HHHLLLHLLH	HLHLLHMLHH	LLLLLLLLLH
VINYLD[c]	HHLLLLLLLL	LLHHHLHHHL	LLLLHHHHHH	HHHHHHLHLL	HHHHHHHMHH	LLLHHHHLLH
ESTERX[d]	LLLLLLLLLL	LLLLLLLLLL	LLLLLLLLLL	LLLLLLLLLL	LLLLLLLLLL	LLLLLLLLLL
AMIDZ[e]	LLLLLLLLLL	LLLLLLLLLL	LLLLLLLLLL	LLLLLLLLLL	LLLLLLLLLL	LLLLLLLLLL

[a] C*SULFONATE refers to the carbon atom in a sulfonate group which has directly bonded to sulfur.

[b] VINYLW ≡ a vinyl extended withdrawing group. [c] VINYLD ≡ a vinyl extended donating group.

[d] The ESTERX functional group entry refers to that portion of the ester derived from an alcohol.

[e] The AMIDZ functional group entry refers to that portion of the amide derived from an amine.

ical reactivity of functional groups of this sort, those chemical and structural features which influence reactivity must be identified and taken into account in the subclassification scheme. In the case of halides and sulfonates, such considerations led to the assignment of six subclasses: vinyl, aryl, SN1 type (tertiary, benzylic, and those with an α donating group), allylic, those with an α withdrawing group, and "others" (primary and secondary alkyl halides and sulfonates). The other functional group subclasses, which are enumerated in Table III, were derived similarly. Although these subclasses will not serve to define rigorously the reactivity associated with every conceivable occurrence of a functional group in an organic molecule, they do provide a workable scheme which can classify a great diversity of functional groups intelligently with respect to their chemical reactivity.

It is not unusual for a particular functional group in the target molecule to be characterized by more than one of its group subclasses. Such multisubclassed functional groups are assumed to have a reactivity level toward a particular

Table II. The ENVRTB Sub-table

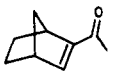
REAGENT/ FG		1	1111111112	2222222223	3333333334	4444444445	5555555556
NUMBER/SUBCLASS		1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
KETONES:	1	HMLMHHHMHH	MLMHLLHLLH	MLHHHHHMMM	LMLHLLMLLH	HLHLLHMLLM	LHHMLMLMLH
	2	HMLMHHMMHH	MLMHLLHLLH	MLHHHHHMMM	LLLMLLLLLH	HLHLLHMLLM	LHMMLMLLLH
	3	MLLLMHMLMM	LLLMLLMLLH	LLHMMMMLLL	LLLMLLLLLM	MLHLLMLLLL	LHLLLLLLLM
	4	HMLMHHMMHH	MLMHLLHLLH	MLHHHHHMMM	LLLMLLLLLH	HLHLLHMLLM	LHMMLMLLLH
	5	MLLLMHMLHM	LLLMLLMLLH	LLHMMMMLLL	LLLMLLLLLM	MLHLLMLLLL	LHLLLLLLLM
	6	MLLLMHMLMM	LLLMLLMLLH	LLHMMMMLLL	LLLMLLLLLM	MLHLLMLLLL	LHLLLLLLLM
	40	HMLMLLLLLM	LLMMLLHLLH	LLHHMMLLHH	LHMMMMLLLH	HLHLMHHLMM	LHHHMHLLLH
	20	HLLMHHMMMH	MLLMLLMLLH	LLHMMMMLLL	LLLMLLLLLM	MLHLLMLLLL	LHLLLLLLLM
	7	HLLMHHHMHM	LLHHMLHMMH	LLHLMHHMMM	LHLLMMMMLM	HLHLLHHLHM	LHLMLMLMLH
	10	HMLMHHHHHM	HLHHMLHHHH	HMHMMHHMLM	LHLHHHHLLH	MMMLHMMLMM	MHMMLLLHLM
	1	HMLMHHHMHH	MLMHLLHLLH	MLHHHHHMMM	LMLHLLMLLH	HLHLLHMLLM	LHHMLMLMLH
	9	LLLMHHMMMH	MLLMLLMLLH	LLHLLMLLLL	LLLLLLLLLH	MMMLHMMLMM	MHMMLLLHLM
	10	HMLMHHHHHM	HLHHMLHHHH	HMHMMHHMLM	LHLHHHHLLH	MMMLHMMLMM	MHMMLLLHLM
	8	MLLLMHMLMM	LLLMLLMLLH	LLHMMMMLLL	LLLMLLLLLM	MLHLLMLLLL	LHLLLLLLLM
ALDEHYDES:	4	HMLMHHHMHH	MLHHMLHLLH	LLHHHHHMML	LHLLLMMLLH	MLMLLHHLHM	LHMMLMLMLH
	5	HMLMHHHMHH	MLHHMLHLLH	LLHHHHHMMM	LHLLLMHLLH	HMHLMHHLHM	LHMMMHLHLH
	6	HMLHHHLLLH	MLHHMLHLLH	LLHHHHHMMM	LHLLHHMLLH	HHHLHHHLHM	LLMMHHLHLH
	1	LLLLLHHMHH	MLHHMLHLLH	LLHHHHMLLL	LHLLLMLLLM	LLLLLLLLHM	LHMMLLLLLH
	2	LLLLMHHMHH	MLHHMLHLLH	LLHHHHLLLL	LHLLLMMLLM	LLLLLLLLHM	LHMMLLLLMH
	3	HMLMHHHHHM	HLHHMLHHHH	HMHMMHHMLM	LHLHHHHLLH	MMMLHMMLMM	MHMMLLLHLM
W GROUPS:	2	LLLLLLLLLL	LLHHLLHMMH	LLHLMHMLLL	LLLLLLLLLL	LLLLLLLLLL	LLLLLLLLLM
	3	LLLLLLLLLL	LLHHLLHMMH	LLHLMHMLLL	LMLLLLLLLL	MLMLLMMLLL	LLLLLLLLLM
	4	HLLMHHHMHM	LLHHMLHMMH	LLHLMHHMMM	LHLLMMMMLM	HLHLLHHLHM	LHLMLMLMLH
	10	LLLLLLLLLL	LLHHLLHMMH	LLHLMHMLHH	LHMMLMLHLH	HMHLMHHLHM	LLMMMHMMLH
	1	HMLMHHHHHM	HLHHMLHHHH	HMHMMHHMLM	LHLHHHHLLH	MMMLHMMLMM	MHMMLLLHLM
LACTAMS &	1	MLLLMMLLLL	LLLLLLLLLM	LLHLLMLLLL	LLLLLLLLLL	LLLLLLLLLL	LLLMLLLLLL
LACTONES:	2	HMLHHHMMMM	HLMMLLHMLH	MLHHHHHMLL	LLLLLMHLLH	LLLLLLLLLM	LHLMLMLHLM
EPOXIDES:	10	HMLLMHMHLM	MLMHLLHMMH	MLHLMMHHLL	LLLLLLLLLL	MLLLLMLLMH	LHLHLLLHLM
	1	HMLLMMLMLL	LLLHLLHMMH	LLHLLLHHLL	LLLLLLLLLL	MLLLLMLLMH	LHLHLLLHLM
	20	HHMLMHMHLM	MLMHMLHMHH	MLHLMMHHLL	LLLLLMLLLL	MMMLLHLLMH	LHLHLMLHLM
	40	HLLMHHHHMM	MLMHMLHMHH	HMHLMMHHLL	LLLLLLLLLL	MLMLLMLLMH	LHLHLLLHLM
ALCOHOLS:	2	HMLLLLLLLL	LLMMMLHMLH	MLLLLLHMLM	LHLLHHLHLM	MLMHLMLLLL	LLLHHHLMLH
	40	HMLLLHMLML	LLMMMLHMMH	HLLLLLHMLM	LHLLHHMHLH	HMHHHHLMLM	LHLHHHLHLH
	1	HLLLLLLLLL	LLLLLLMLLL	LLLLLLLLLL	LMLLHHLLLL	LLLLLLLLLL	LLLHHHLLLH
	7	HHMLLLLLLL	LLMHLLHHMH	MLLLLLHHLM	LLLLLLLLLL	LMLLLLLLLL	LLLHHHLHLH
	6	LLLLLLLLLL	LLLLLLLLLL	LLLLLLLLLM	LHLLMHMMLH	HMHMMMMLHL	LLMLHHLLLH
	5	HMLLLLLLLL	LLLLLLHMLL	LLLLLLMMLL	LLLLLLLLLL	LMLLLLLLLL	LLLHHHLMLH
	10	HMLMHHHLMM	LLLLLLMLLH	MMLLLLLLLL	LMLLLMMLLM	LLLLLLLLLL	LHLHHHLMLH
	20	HMLLLLLLLL	LLLLLLMLLL	LLLLLLMLLH	LHLLLMLLLL	LLMLHLLLLL	LLLHHHLLLH
HALIDES &	2	HLLLHMLMLL	LLHHHHLLLH	LLLLLMLLLL	LLLLLLLLLL	LLLLLLLLML	LHLLLLLLLL
SULFONATES:	3	LLLLLMLMLL	LLLLLLLLLH	LLLLLLLLLL	LLLLLLLLLL	LLLLLLLLLL	LHLLLLLLLL
	4	HMLMMHHHHH	HHHHHHHHHH	HHHMMHHHML	LMLLHHHLHH	LLLLMMHHHL	MHMLLMLHLH
	5	HMLMMHHHHH	HHHHHHHHHH	HHHMMHHHML	LMLLHHHLHH	LLLLMMHHHL	HHMLLMLHLH
	10	LLLMHHHHHH	HMHHHHHHHH	HHHMMMMLLL	LMLLHHHLMH	LLHLLHHLML	MHLLLLLHLH
	1	MLLLMMLMMM	MLMMMMMMMH	MMMLLLMLLL	LLLLLLLLLL	LLLLLLLMLL	LHLLLLLLLL
ALKYNES:	1	MLLMHHHHHH	HMHHHHHMHH	HMHHMHHMMM	HHHMLLHLLH	HLHLLHHLHH	LHMLLLLMLH
	2	HLLLMMLLLL	LLHHHHHMMH	MLMLMHHHHH	HHHHLMLLLL	HLHLLHMMHH	LHLMLLLMLH
	3	MLLLLLLMLL	LLHHHHMLMH	LLLLMHMLHH	HHHMLLLLLL	HLHLLHLMHH	LLLLLLLLLH
	5	MLLMHHHHHH	HLHHHHHMMH	HMHHMHHMMM	HHHMLLHLLH	HLHLLHHLHH	LHMLLLLMLH
	40	HLLLMMLLLL	LLHHHHHMMH	MLMLMHHHHH	HHHHLMLLLL	HLHLLHMMHH	LHLMLLLMLH
	4	MLLLLLLLLL	LLHHHHMLMH	LLLLMHMLHH	HHHHLLLLLL	HLHLLHLMHH	LLLLLLLLLH
	10	MLLLLMLLLL	LLHHHHHLMH	MLMLMHHHHH	HHHMLLLMLL	HLHMLHLMHH	LLLLLLLLLH
	20	HHMLLLLLLL	LLHHHHHHMH	LLLLHHHHHH	HHHHLMLMLM	HMHLLHMMHH	LLLHMHLHLH
OLEFINS:	3	HLLLLLLLLL	LLHHHLLLLM	LLLLLHMLLM	HHHMLLLLLL	HLHLLMLLML	LLLLLLLLLH
	2	HLLLLLLLLL	LLHHHLLLLL	LLLLLHMLLM	HHHMLLLLLL	HLHLLMLLML	LLLLLLLLLH
	1	HLLLLLLLLL	LLMHMLLLLL	LLLLLMMLLM	MHHHLMLLLL	HLHLLMLLML	LLLLLLLLLH
	400	MLLLLLLLLL	LLLMLLLLLL	LLLLLMMLLM	MMMHLMLLLL	HLHLLLLLLL	LLLLLLLLLH
	4	LLLHHHHHHH	HLHHHHHHHH	HMHMMMLLLL	MMMLLLHLLH	MLMLLLMLHH	LHHLLLLLLM
	5	HLLMHHMHMM	LLMHMLHMMH	MLMLMHMLLM	MHHLLLMLLM	HLHLLLMLHM	LHMLLLLLLH
	6	MLLMHHMHMM	MLHHHLHMMH	MLMLMMLLLL	MHMLLLMLLM	HLMLLLMLHM	LHMLLLLLLM
	7	HHLLLLLLLL	LLLHLLHLLL	LLLLMHMLMH	HHHHMHLMLH	HMHLMMMLHL	LLHMHHLLHH
	8	HHLLLLLLLL	LLMHMLHMLL	LLLLMHHMMH	HHHHLHLMLM	HMHLLMMLHM	LLLMMHLLLH
	9	HLLLLLLLLL	LLMHMLMLLH	LLLLLMMLLM	MHHMLMLLLL	HLHLLMLLHM	LMLLLLLLLH
	10	HLLLLHMLLL	LLLHLLHLLH	MLMLMHMLLM	MMHMLLLLLM	MLMLLLLLLL	LLLLLLLLLM
	20	HLLLLLLLLL	LLHHHLLLLH	LLLLLHHMLM	MHHMLMLMLL	HLHLLMLLHM	LMLLLLLLLH
	40	HLLLMHLLLL	LLMHMLLLLL	LLLLLMMLLM	HHHMLLLLLL	HLHLLMLLML	LHLLLLLLLH
	100	HLLLMLLLLL	LLHHHLMLLH	LLLLLHHLLM	MHHMLLLLLL	MLHLLMLLMM	LLLLLLLMLH
	200	HMLLLLLLLL	LLHHHHMLLM	LLMLMHHMHH	HHHHLMLLLL	HLHLLMLLHM	LLLLLLLMLH

Table III. Functional Group Subclasses

Functional Group	Subclass	Subclass Rating
Ketones	I. Enolizable	
	A. Cyclic	
	1. Strained (Those in a 3 or 4 membered ring, or those in a 5 membered ring which is part of a [2.1.1], [2.2.1],or [3.2.1] bicyclic system.)	1
	2. Alpha CH_2	2
	3. No alpha CH_2 but alpha CH	3
	B. Acyclic	
	1. Alpha CH_3 or alpha CH_2	4
	2. No alpha CH_3or alpha CH_2 but 2 alpha CH	5
	3. No alpha CH_3or alpha CH_2 but 1 alpha CH	6
	C. Others	
	1. W-CH-C(=O)-C	40
	2. W-CR_2-C(=O)-CH	20
	3. C=C-CH-C(=O)-C or L-C-CH-C(=O)-C	7
	4. C-C(=O)-C(=O)-C	10
	II. Nonenolizable	
	A. Strained (see above)	1
	B. W-CR_2-C(=O)-CR_3	9
	C. C-C(=O)-C(=O)-C	10
	D. Others	8
Aldehydes	I. Enolizable	
	A. R(sat)-CH-CHO	4
	B. R(unsat)-CH-CHO or L-C-CH-CHO	5
	C. W-CH-CHO	6
	II. Nonenolizable	
	A. Ar-CHO	1
	C. W-CHO	3
	B. R_3-CHO	2
Withdrawing Groups (ester, cyano,sulfone, nitro,and sulfoxide)	I. Ar-W or R_3-W	2
	II. R(sat)-CH-W	3
	III. R(unsat)-CH-W or L-C-CH-W	4
	IV. W-CH-W	10
	V. W-C=O	1
Alcohols	I. Attached hydrogen	
	A. Primary or secondary allylic or benzylic	2
	B. Primary or secondary with alpha W group	40
	C. Primary or secondary alkyl	1
	II. No attached hydrogen	
	A. Tertiary benzylic or allylic	7
	B. Ar-OH or H(unsat)-OH	6
	C. Tertiary alkyl	5
	III. Others	
	A. W-CH-C-OH	10
	B. D-C-C-OH	20

Functional Group	Subclass	Subclass Rating
Alkynes	I. Terminal	
	A. H-C≡C-W	1
	B. H-C≡C-C=C	2
	C. H-C≡C-R	3
	II. Nonterminal	
	A. R-C≡C-W	5
	B. R-C≡C-C=C	40
	C. R-C≡C-R	4
	III. Others	
	A. C≡C-C-OH or C≡C-C-C-OH	10
	B. C≡C-D	20
Halides and Sulfonates	I. Vinyl	2
	II. Aromatic	3
	III. S_n1 (tertiary,benzylic, and alpha D group)	4
	IV. Alpha W group	10
	V. Allylic	5
	VI. Others	1
Epoxides	I. CH_2 epoxides	10
	II. Non-CH_2 epoxides	1
	III.	20
	IV.	40
Lactones and Lactams	I. Normal	1
	II. Strained (Those in a 3 or 4 membered ring or those in a 5 membered ring which is part of a [2.1.1], [2.2.1], or [3.2.1] bicyclic system.)	2
Olefins	1. Simple alkyl	
	A. Mono or gem-disubstituted	3
	B. Cis or trans disubstituted	2
	C. Trisubstituted	1
	D. Tetrasubstituted	400
	II. Others	
	A. W-C=C-W or C=C(W)(W)	4
	B. C=C(W)(D)	5
	C. C=C-W	6
	D. C=C-N or C=C-O	7
	E. C=C-D (D≠N or O)	8
	F. C=C-Ar	9
	G. D-C=C-W	10
	H. C=C-C=C	20
	I. C=C-CH-C_π	40
	J. C=C-C≡C	100
	K. Strained (Those in or exo to a 3 or 4 membered ring, in a 5 membered ring which is part of a [2.1.1],[2.2.1], or [3.2.1] bicyclic system, or any trans olefns in rings smaller than 10 membered.)	200

reagent equivalent to that of its most highly reactive subclass. The C=C in structure II, for example, fits into two



II

subclasses, one consisting of "strained" olefins and another consisting of olefins containing an α withdrawing group. While the reactivity of the strained olefin subclass toward strong base is low, that of olefins with an α withdrawing group is high. Therefore, the reactivity of the C=C in II toward strong base is taken as high.

Reasonable extensions to this initial version of the subclassification scheme will include both the refinement of existing group subclasses and the designation of new functional group subclasses for groups currently not subclassified (i.e., those groups presently in the REACTB sub-table, Table I).

II. Internal Representation of the FG/RGNT Table

The FG/RGNT table contains a large amount of chemical information and is consulted with great frequency during the process of transform evaluation. The internal storage of this information must thus be as compact and readily accessible as possible if the FGR module is to operate efficiently. To this end, the FG/RGNT sub-tables, REACTB (Table I) and ENVRTB (Table II), have been stored internally as one-dimensional Fortran arrays. The generation of these numerical arrays is accomplished by means of a separate computer program which takes as input the English-readable, condensed sub-tables, Tables I and II. The one-dimensional array structure was chosen because individual words within the array can be readily accessed by means of a simple and direct address calculation.

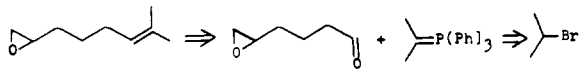
Each functional group in the REACTB sub-table (Table I) and each group subclass in the ENVRTB sub-table (Table II) are allocated four computer words in the array. These words are composed of 16 two-bit[7] fields, each of which contains the reactivity of one group or group subclass toward one chemical reagent. The reactivity levels of low, moderate, and high are stored in the two-bit fields as patterns of 00, 01, and 10, respectively. In this manner, reactivity levels toward as many as 64 (4 words × 16 two-bit fields per word) different chemical reagents can be compactly stored in the four array words allocated to each functional group or group subclass.

III. Condition Statements. The Internal Representation of Reaction Conditions

As outlined previously,[4] there is associated with each individual transform in the data base of LHASA a collection

of queries containing information which bears on the applicability of that transform to any particular target molecule. These queries or "qualifiers" are written in a language based on "chemical English" which is both computer readable and intelligible to a chemist with little or no programming experience.[8] In the data table entry for each transform is collected information regarding the scope and limitations of the equivalent synthetic reaction. Included among these qualifiers, which are assessed during the transform evaluation phase of processing, are lists of the prototype reagents which constitute sets of reaction conditions capable of bringing about the desired transformation. The qualifiers which contain these prototype reagents (termed "condition statement" qualifiers) contain the word CONDITIONS followed by the list of individual reagents, separated by the keyword AND, which constitute that set of conditions. The previously mentioned Clemmensen and Wolff-Kishner reductive conditions, for example, would be internally represented in the ketone reduction transform entry by the condition statement qualifiers "CONDITIONS ZN/HCL" and "CONDITIONS NU: AND PH >10," respectively.

When a disconnective transform (i.e., one which retrosynthetically produces two or more fragments) is being considered, some of the reagents specified are usually meant to apply to just one of the reacting fragments. For example, consider the application of the Wittig transform to structure III.



III IV V VI

The halide VI must first be treated with triphenylphosphine and butyllithium to produce the ylide V, which is subsequently allowed to react with the aldehyde IV. In this case, it is not sufficient merely to list the conditions as "CONDITIONS NU: (Ph_3P) AND ORGANOMETALLIC (BuLi) AND WITTIG (>=PPh_3)" because this implies that the epoxide moiety in IV is subjected to all three reagents. Such an implication leads erroneously to the designation of the epoxide as an interfering group since it is susceptible to attack by organometallic reagents. In reality, however, the epoxide is stable to the only reagent (Wittig) to which it is actually exposed.

To handle situations of this sort, a method for specifying which reagents apply to which fragments is required. This is done in the condition statement qualifier itself by using the keywords FRAGMENT*1 and/or FRAGMENT*2. The condition statement for the Wittig transform shown above thus becomes "CONDITIONS WITTIG AND IN FRAGMENT*1 NU: AND ORGANOMETALLIC". This is interpreted as indicating that all fragments are subjected to the Wittig reagent, while only the halide fragment (fragment*1) is subjected to the other reagents. The method by which the program decides which fragments are to be designated as fragment*1 and fragment*2 is straightforward. During the evaluation of a transform, there is one and only one carbon in the target molecule which is designated as carbon*1.[8] The fragment that contains carbon*1 after the target molecule has been transformed into the precursor fragments is defined as fragment*1. Fragment*2 refers to the other fragment(s).

IV. Computer Derivation and Internal Representation of Subclassified Functional Groups

The task of functional group subclassification in LHASA is handled completely by one Fortran subroutine, FGSUB (Functional Group SUBclassification routine), which is outlined in Figure 3. A brief summation of the flow of control is given below, with the accompanying numbers corresponding to the numbered boxes in Figure 3.

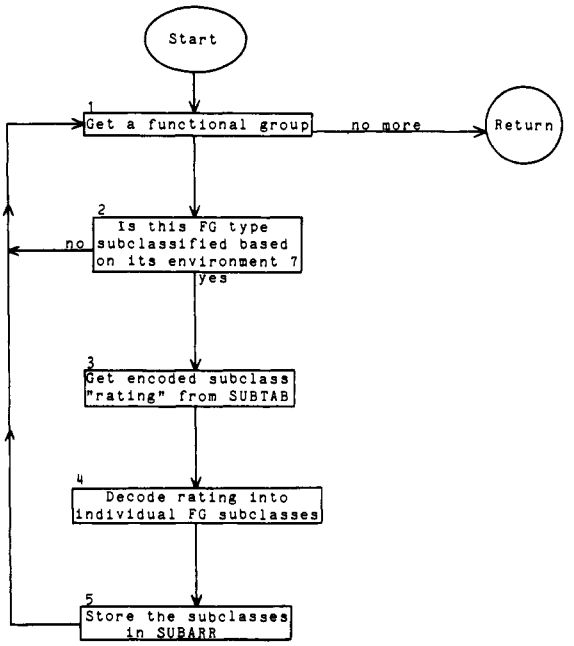


Figure 3. FGSUB, LHASA's functional group subclassification routine.

```
TRANSFORM 2
NAME HALIDE & SULFONATE SUBCLASSIFICATION TRANSFORM
RATING 1 (initial rating upon entering this transform)
GROUP*1 CAN BE FLUORIDE OR CHLORIDE OR BROMIDE &
        OR IODIDE OR SULFONATE (groups allowed to
                                enter this transform)
....
....
ADD 1 IF OLEFIN ON CARBON*1 THEN GO TO BLOCK2
ADD 2 IF CARBON*1 IS AROMATIC THEN GO TO BLOCK2
IF CARBON*1 IS A BRIDGEHEAD THEN GO TO BLOCK2
IF WITHDRAWING GROUP ON CARBON*1
        BEGIN BLOCK1
        SET*THE*FLAG
        ADD 10
        BLKEND BLOCK1
ADD 4 IF CARBON*1 IS ALLYLIC THEN GO TO BLOCK2
ADD 3 IF CARBON*1 IS TERTIARY THEN GO TO BLOCK2
ADD 3 IF CARBON*1 IS BENZYLIC THEN GO TO BLOCK2
ADD 3 IF DONATING GROUP ON CARBON*1 THEN GO TO BLOCK2
IF THE FLAG*IS*SET NOT THEN GO TO BLOCK2
SUBTRACT 1
BLOCK2 FINISHED
```

Figure 4. A portion of the SUBTAB functional group subclassification table which deals with the subclassification of halides and sulfonates. The keyword CARBON*1 refers to the carbon atom bearing the halide or sulfonate.

(1) The functional groups are picked up one at a time.

(2) Certain functional group types are currently not subclassified. These are groups whose chemical reactivities are to a reasonable approximation independent of their molecular environment and are determined more by the reactive nature of the groups themselves (i.e., those groups contained in the REACTB sub-table, Table I). Such groups are ignored by FGSUB.

(3) If the group is subclassified based on its molecular environment, then it must be determined into which subclass(es) this particular group belongs. Queries concerning the molecular environment of a functional group are made by reading through a small data table, SUBTAB (SUBclassification TABle), which is similar in many respects to the chemistry data tables used by the program to evaluate transforms.[8] A portion of this table, dealing with the classification of halides and sulfonates, is reproduced in Figure 4. The table is entered with an initial "rating" (in this case 1) and various questions are asked which alter the value of this rating. When the table reading for a particular functional group entry is complete, the final value of the rating is an encoded indication of the subclasses to which this functional

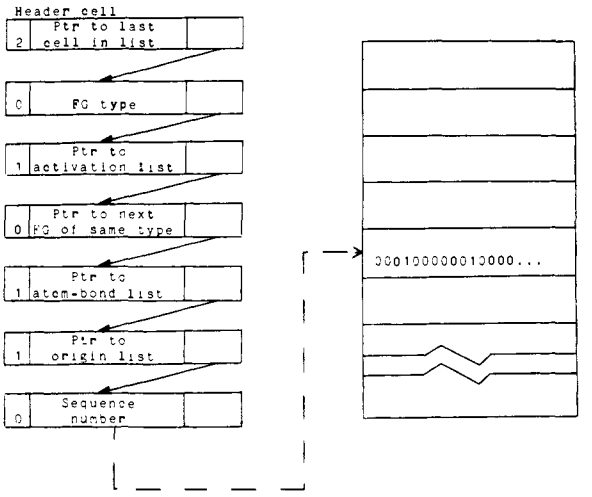


Figure 5. The internal storage of functional group (FG) subclasses.

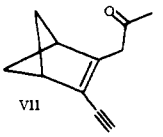
group belongs. The rating correlations for all functional groups currently subclassified in LHASA are given in Table III.

(4) The rating is decoded next and from it are extracted individual subclasses categorizing the group under consideration. The rating number itself can consist of up to three digits. Individual subclasses of a functional group can correspond to the following numbers: 1–9 in the ones digit, and 1, 2, and 4 in the tens and hundreds digits. The reason for allowing just 1, 2, and 4 to correspond to separate subclasses in the tens and hundreds digits is that any additive combination of 1, 2, and/or 4 produces a unique single digit number itself. In this manner, the tens and hundreds digits can represent up to three functional group subclasses *each* and, when combined with any ones digit (representing a unique subclass), up to seven different group subclasses can be represented by a single three-digit number. To illustrate this point, consider the rating breakdown schemes shown:

162 (100) + (20 + 40) + (2) ≡ 4 subclasses

340 (100 + 200) + (40) ≡ 3 subclasses

778 (100 + 200 + 400) + (10 + 20 + 40) + (8) ≡ 7 subclasses



VII

The 340 rating, for example, would serve to represent a C=C of the type shown in structure VII, which according to Table III, is characterized by the three olefin subclasses II-I (rating 40), II-J (rating 100), and II-K (rating 200). The 778 rating demonstrates how a group characterized by as many as seven different subclasses can be uniquely expressed using just three digits.

Since the ones digit is taken as a unique number itself, this system permits up to 15 different subclassifications for any one functional group type, 9 from the ones digit and 3 each from the tens and hundreds digits.

(5) Once the individual subclasses have been decoded from the rating, they are stored in a single "subclassification" computer word. This is done by turning on a bit[7] in the word for each subclass found. Subclasses corresponding to 1 through 9 in the ones digit of the rating are represented by bits 1 through 9 in the computer word. The 1, 2, and 4 tens digit subclasses correspond to bits 10, 11, and 12, while bits 13, 14, and 15 represent the 1, 2, and 4 hundreds digits. The C=C in structure VII thus would have bits 12, 13, and 14 turned on in its subclassification word.

The internal storage of subclassification words is shown diagrammatically in Figure 5. Each functional group has a "parent" list associated with it containing a variety of chemical and computational information concerning the group.[9] The last cell in each list contains the "sequence" number of that parent group, a number which is unique to that group and can be used to distinguish it from other functional groups. The sequence number of each group serves as a subscript pointer to the group's subclassification word. All subclassification words are contained in a one-dimensional array, SUBARR (SUBclassification ARRay), and are thus readily accessible. The sequence numbers of groups not currently subclassified by LHASA will point to a completely zeroed word in SUBARR.

When all the functional groups have been examined, group subclassification is complete and control is returned from FGSUB.

V. Protectable Functional Groups

In synthetic sequences involving multi-functional intermediates, selective modification of a particular functional group in the presence of other groups with higher or comparable chemical reactivities is often required and is accomplished generally by the use of protecting groups.[10]

For the protection of a specific functional group in a molecule, it is necessary to choose a particular protecting group, the point in the synthesis where the group is applied, and the point at which the group is removed. During antithetic analysis the need for protection is made apparent by the presence of functionality in the immediate precursor(s) of a target structure which would interfere with the operation of a transform. Whether, how, and at what point an interfering group can be protected is a complex matter which depends on the following considerations:

1. Whether any technique for the protection of the group in question is available in current synthetic practice.

2. Whether an available protected form of the group has the stability required to permit the operation of the transform in question.

3. Whether the required protection step itself would be complicated by competing reactivity at other sites (from groups of the same or different type) in the structure and, if so, whether the protection is possible further down in the synthetic tree (i.e., earlier in the synthesis itself).[11]

4. Whether the functional group to be protected is accompanied in the offspring structure by one or more groups of the same type already in protected form. In such a case, differential protection using *different* protecting techniques will be necessary.

5. Whether two or more groups can be protected by the same masking unit.

During the generation of a tree of synthetic intermediates by antithetic analysis, each step in which protection of a group must be in effect is defined in an order corresponding to descending down the tree. Clearly, for the synthetic direction of analysis, the steps requiring protection are recognized in the reverse order. The optimal points for protection and deprotection of that group, however, are not defined in either direction of analysis but become determinable only after a complete synthetic pathway has been generated. The same holds true for the choice of an optimal protecting group when several alternatives are feasible.

In view of the above-mentioned considerations concerning functional group protection, only a small number of functional group types have been designated as protectable in the current version of LHASA. These groups, which are listed in Table IV along with the conditions under which they are considered protectable, are those for which a pro-

Table IV. Protectable Functional Groups

Reagent	Aldehydes Ketones	Amine*1 Amine*2	Alcohols	Thiols	Reagent	Aldehydes Ketones	Amine*1 Amine*2	Alcohols	Thiols
pH<1	UP	S	UP	S	OsO_4	S	P	S	S
pH2-4	P	S	UP	S	$KMnO_4$	P	P	P	UP
pH4-6	S	S	UP	S	O_3/-50	P	P	S	UP
pH9-10	P	S	P	S	RCO_3H	P	P	S	UP
pH>10	P	S	P	S	CrO_3/pyr	P	P	P	S
RLi/RMgX	P	S	P	S	CrO_3/H^+	UP	P	UP	UP
Orgzinc	P	S	P	S	H_2O_2/alk	P	P	P	S
R_2Cu	P	S	S	S	Quinone	S	P	P	P
Wittig	P	S	P	S	DMSO	S	P	S	S
Str Nu:	P	S	P	S	NaOCl	P	P	P	UP
Nu:	P	S	S	S	NBS	P	P	UP	UP
Mild Nu:	S	S	S	S	I_2	P	S	UP	P
Raney Ni	P	S	UP	UP	Cl_2/Br_2	P	P	UP	UP
$H_2/Pt/H^+$	P	S	UP	UP	MnO_2	S	P	P	P
H_2/Pd	P	S	UP	UP	HIO_4	P	P	P	P
H_2/Lind	S	S	S	UP	SeO_2/H^+	UP	P	UP	P
Zn/HCl	UP	S	UP	S	SeO_2/pyr	P	S	P	P
Zn/HOAc	P	S	UP	S	Ag(I)	S	S	S	S
Cr(II)	P	S	P	S	Radical	P	P	P	UP
Na/NH_3	P	S	UP	S	R_3SnH	P	S	P	S
Al(Hg)	P	S	UP	S	$Ni(CO)_4$	P	S	S	S
$SnCl_2$	P	S	S	S	$NaNR_2$	P	S	P	S
$LiAlH_4$	P	S	S	S	CH_2N_2	P	S	P	P
$NaBH_4$	P	S	S	S	$SOCl_2$	P	P	P	P
R_2AlH	P	S	S	S	AcOAc/25	P	P	P	P
B_2H_6	P	S	S	S	AcOAc/80	P	P	P	P
$AlCl_3$	UP	S	UP	S	DCCD	S	S	S	S
$SnCl_4$	P	S	UP	S	Pyrol	P	S	UP	P
Hg(II)	P	P	S	S	CH_3I	S	P	S	P
Pb(IV)/25	P	P	P	UP	Pb(IV)/80	P	P	P	UP

UP=unprotectable, P=protectable, S = stable

tecting group exists which is stable to a *large* diversity of chemical reagents and for which the protection-deprotection steps are particularly facile.

When a transform has been successfully evaluated and a precursor structure is displayed to the chemist, those groups in the precursor deemed unstable but protectable toward the conditions required for the transformation are initially displayed enclosed in a box as an indication of their need for protection. Unstable, *unprotectable* groups are initially displayed enclosed in a dashed box. Presently, no attempt is made to assign specific protecting groups or to define a point for the introduction or removal of a protecting group.

VI. The Revised Transform Evaluation Process

The transform evaluation process in LHASA functions as a screen to filter out transforms corresponding to synthetic steps with little likelihood of success.[4,5] In the past, such evaluation has been performed by an examination of the target molecule *only*. This procedure is the most expedient since, when the transform evaluation phase is entered, the structural and chemical information available in core memory pertains to the input target structure. Considering the type of structural and chemical information required by the FGR module, however (information concerning the number of reacting fragments and the immediate molecular environments of functional groups *in the precursor*), it became apparent that a thorough examination of the precursor molecule as well as the target molecule was necessary. To this end, the entire transform evaluation process of LHASA has been redesigned so that both the target and precursor molecules are scrutinized when evaluating a potential reaction.

The cost of allowing potential precursors to be examined before being accepted as valid precursors is that a complete structural and chemical perception must be done on a molecule which later may be discarded. The benefits, however, clearly justify the cost. With the perceptual information pertaining to the precursor at hand, the environments of all functional groups in the precursor are determined easily.

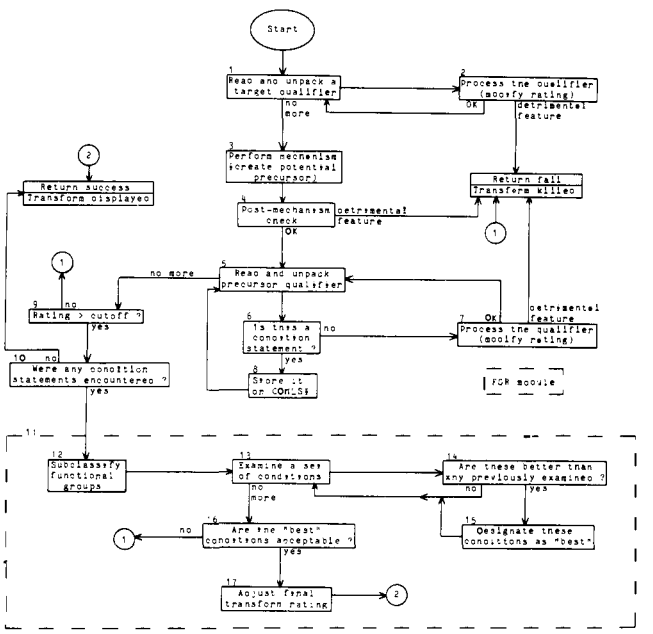


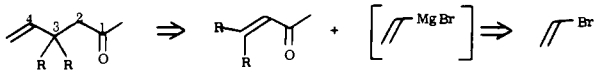
Figure 6. RXEVAL, the main transform evaluation routine of LHASA.

Also, the presence of unstable or undesirable features in the precursor (trans small ring double bonds, unstable aromatic tautomers, etc.) can be found directly instead of having to be "predicted" from looking only at the target molecule. Furthermore, many qualifiers which are difficult to evaluate having only the target molecule can now be deferred until the precursor is generated.

The controlling transform evaluation routine, RXEVAL (RXn. EVALuation), is outlined in Figure 6. A brief summary of each of the steps in the evaluation process is given below with the accompanying numbers corresponding to the numbered boxes in Figure 6.

(1-2) A preliminary examination of the target molecule

is made. Here, a series of questions (qualifiers) is asked concerning the structural features of the target molecule which can be expected to help or hinder the progress of the reaction in the synthetic direction. These qualifiers are based on current chemical knowledge, pertaining to the scope and limitations of that reaction. Most qualifiers specify a numerical value which is to be added to or subtracted from a basic transform "rating" (whose initial value reflects the *general* scope of the transform) if the structural unit specified in the qualifier is present in the target molecule. Consider, for example, the application of the organometallic Michael transform to structure VIII.[11] The qualifier



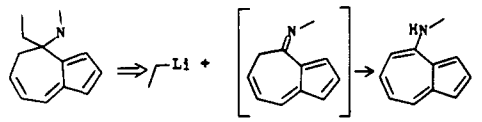
VIII

"SUBTRACT 10 FOR EACH ALKYL GROUP ON CARBON*3 OFFPATH" reflects the fact that β-branching in the enone substrate (position R) generally hinders the reaction.[12]

Other qualifiers may check for detrimental features which, if present in the target, are sufficient to cause the entire transform to be rejected or "killed". Such a qualifier, applicable to the above shown Michael transform, would be "KILL IF NO HYDROGEN ON CARBON*2" since a quaternary carbon*2 precludes the retrosynthetic production of the C=C in the precursor.

(3) If the target molecule passes the initial examination, then the transform "mechanism" is performed. This creates a precursor molecule which, if subjected to the reaction conditions of the current transform, should generate the target molecule. Referring to the atom numbering in structure VIII, the Michael transform mechanism would be "BREAK BOND*3" (the bond connecting carbons 3 and 4), "JOIN CARBON*2 AND CARBON*3" (i.e., increase the bond order of the 2-3 bond by one), "ATTACH A BROMIDE TO CARBON*4".

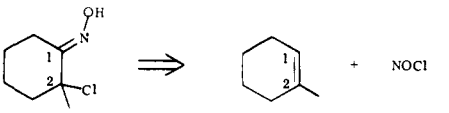
(4) Once a potential precursor has been generated, it is examined by a separate module, EVL01 (EVaLuation subroutine 01), to see if it contains any unstable or undesirable features which were not foreseen in the preliminary examination. For example, a preliminary examination of structure IX indicates that it could be generated from an organo-



IX X XI

lithium addition to the imine X. Once X has actually been generated, however, the fact that it is really an aromatic tautomer of structure XI and not a true imine is detected. Because this feature is detrimental to the successful operation of the transform, a "kill" return is taken from EVLØ1.[13]

(5-8) If no detrimental features were found by EVLØ1, further queries about the precursor structure can be made. If the qualifier encountered is a condition statement qualifier (vide supra), that set of conditions is temporarily stored on a list, CDNLST (ConDitioN LiST), which is examined later. If the qualifier is not a condition statement, then it is generally a query which would have been difficult to evaluate accurately using the target molecule. Consider, for example, the qualifier "KILL IF PLUS CHARGE BETTER ON CARBON*1 THAN ON CARBON*2", referring to the nitrosyl chloride addition transform shown:



XII XIII

Here it is necessary to determine the relative stabilities of positive charges on carbons 1 and 2 in structure XIII. While a direct and straightforward comparison can be made in structure XIII (in this case, simply by counting the number of alkyl groups at each site), an accurate comparison using structure XII cannot be made without also knowing what modifications about carbons 1 and 2 will be made by the "mechanism".

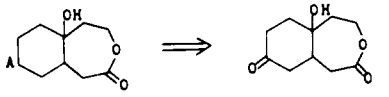
(9) When the end of the precursor qualifiers is reached, the present value of the rating is queried. If it is below a predetermined "cutoff" value, the transform under consideration is rejected. The initial rating and rating cutoff values are adjusted so that structures with any reasonable chance of undergoing the synthetic reaction successfully will have final qualifier ratings above the cutoff value.

(10) At this point, a check is made to see if any condition statement qualifiers were encountered. If none were, as is the case with "conditionless" transforms (reactions which proceed spontaneously or which require only moderate heating), then the transform evaluation program returns successfully. The potential precursor then becomes a valid precursor and is displayed to the chemist.

(11) If condition statements were encountered during transform evaluation, then control is transferred to the FGR module which examines the reactivity of all functional groups present in the precursor toward the reaction conditions of the current transform. An abbreviated discussion of the FGR module will be given at this point to facilitate an understanding of its role in transform evaluation. Details concerning its mode of operation are outlined in the following section of this paper.

(12) The molecular environments of the functional groups present in the precursor are now examined, and those containing environmental features apt to influence their chemical reactivity are subclassified accordingly.

(13-15) Next, the sets of reaction conditions which were stored on CDNLST are retrieved and examined individually. Consider the retrosynthetic introduction of a ketone on atom A in structure XIV, which synthetically corresponds



XIV XV

to the complete reduction of the ketone in structure XV. For this reaction, the program has listed the following sets of reaction conditions: "ZN/HCL" (Clemmensen reduction), "NU: AND PH >10" (Wolff-Kishner NH_2NH_2/KOH reduction), "NU: AND PH2: 4 AND $NABH_4$"[14] (modified Wolff-Kishner reduction involving tosylhydrazone formation and reduction by hydride[15]), and "NU: AND PH2: 4 AND NI/H2" (thioketal formation and reduction by Raney nickel).

When one set of the above conditions is encountered by the program, the reactivities of all functional groups in XV toward those conditions are examined. The "best" sets of conditions are taken as those which interfere least with other groups in the molecule. In the case of structure XV, the thioketal desulfurization and modified Wolff-Kishner conditions are considered "best" because, while both the lactone and tertiary alcohol are stable toward these reductive conditions, the lactone is unstable to the strongly basic

Wolff-Kishner conditions and the tertiary alcohol is unstable to the strongly acidic Clemmensen conditions.

(16) Finally, the "best" conditions found are checked to see if they are acceptable or if the degree of interference (expressed as a percentage, with 100% being unacceptable) imposed by other functional groups is too high to permit the "best" conditions to be used. In the case of the ketone reduction in structure XV, both "best" sets of conditions (thioketal desulfurization and modified Wolff-Kishner reduction) would have 0% degrees of interference and thus clearly would be acceptable. Had the choice of reductive conditions been restricted only to the Clemmensen and strongly basic Wolff-Kishner conditions, however, a high degree of interference would have been encountered and the "best" set of conditions in this case would have had less acceptability. Whether or not this degree of interference is deemed *un*acceptable (i.e., whether or not it will be 100%) is controlled by the chemist since the program permits the manual specification of the amount of emphasis that is placed on interfering functionality (vide infra).

(17) Having found an acceptable "best" set of conditions, the program adjusts the final transform rating to reflect the degree of functional group interference expected from the use of these conditions. This rating adjustment is simply to subtract from the current rating the product of the percent interference and the difference of the current rating and the rating "cutoff" value (initially set at −25). Thus, a reaction with a rating of 75 and a 50% degree of interference gets a final rating of (75 − (75 − (−25))(0.50)) = 25. A reaction rated at 25 and also with a 50% degree of interference gets a final rating of (25 − (25 − (−25))(0.50)) = 0. It is important to note that this rating decrement calculation is such that no decrement will ever be large enough to kill a transform which has less than a 100% degree of interference. However, transforms with a 100% degree of interference are always killed.

If the rating decrement is less than 100%, the precursor structure is displayed to the chemist signifying that transform evaluation has been successfully completed.

VII. Computer Evaluation of Interfering Functionality

The previous section dealt with the operation and role of the FGR module in transform evaluation. It is now appropriate to consider the central FGR routine, EVLØ3 (EVaLuation subroutine Ø3), which determines the total degree of functional group interference expected when a target molecule is subjected to a particular set of reaction conditions.

An outline of the EVLØ3 routine is given in Figure 7. A brief discussion of the important steps in the routine follow, with the accompanying numbers corresponding to the numbered boxes in Figure 7.

(1) The variable MLTPLR (MuLTiPLieR) is initialized to 100. This variable will reflect the total degree of interference to be expected from the "best" set of reaction conditions found.

(2) There are certain variables which pertain only to a single set of reaction conditions. These include: LOCAL, which reflects the degree of interference encountered for the set of conditions currently being examined; PROKNT (PROtected group KouNT), which is a count of the functional group types deemed protectable toward the current conditions; and LOOPFG (LOOP FlaG) which is used to by-pass remaining components of the current condition set when those already examined are deemed unacceptable. These variables are initialized to zero.

(3) A reagent is read from CDNLST. This list contains individual reagents constituting those sets of reaction conditions applicable to the transform currently being evaluated.

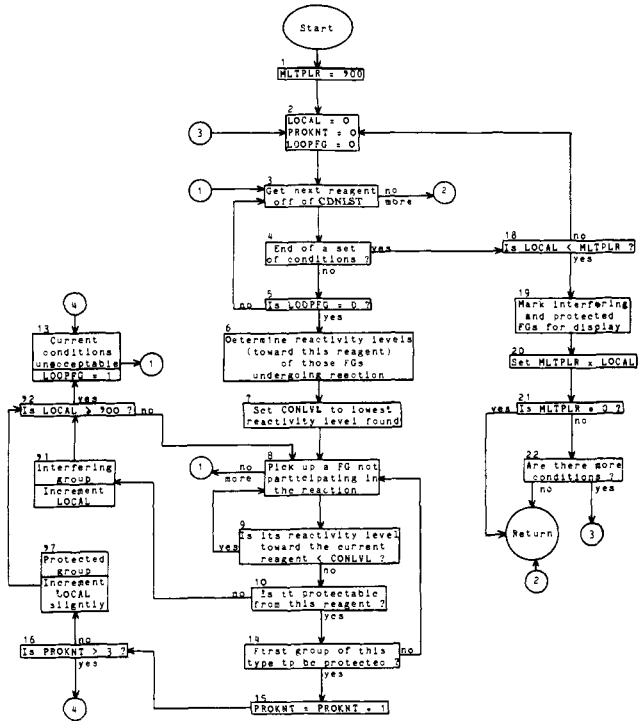


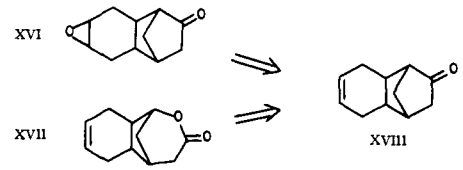
Figure 7. EVL03: the evaluation of interfering functionality.

Each CDNLST cell contains a reagent number, corresponding to one prototype reagent, and a fragment specifier denoting which fragment, if any, that the reagent is to act upon. Since the CDNLST list can contain reagents corresponding to different sets of reaction conditions, a "separator" cell is inserted between the last reagent cell of one set of conditions and the first reagent cell of another set.

(4) If a separator cell has been encountered, then the end of a set of conditions has been reached.

(5) If the end of a condition set has not been encountered, the value of LOOPFG is queried. If LOOPFG is nonzero, then the present reagent is to be by-passed.

(6-7) If this reagent is to be examined, then the reactivity levels of those groups participating in the reaction toward this reagent must be extracted from the FG/RGNT table. A variable, CDNLVL (ConDitioN reactivity LeVeL), is set equal to the lowest reactivity level found, which will be either moderate or high. This is done because in some cases it is useful to know whether the functional group undergoing the reaction is highly reactive or just moderately reactive toward the reagent being employed. Consider, for example, the retrosynthetic conversions of XVI and XVII back to XVIII. These conversions synthetically correspond to C=C



epoxidation and Baeyer-Villiger oxidation, respectively, and as such can be considered competitive in that both require peracid. However, the reactivity of the cis cyclic C=C in XVIII toward peracid is listed as moderate while the ketone's reactivity is listed as high due to its "strained" nature. This eventually permits the program to realize that the XVII to XVIII transform would proceed with no interference from the C=C, while the ketone would interfere with the XVI to XVIII epoxidation transform. (The XVI to XVIII transform is still reasonable in this case since the ketone's interference could be removed by protecting it as a

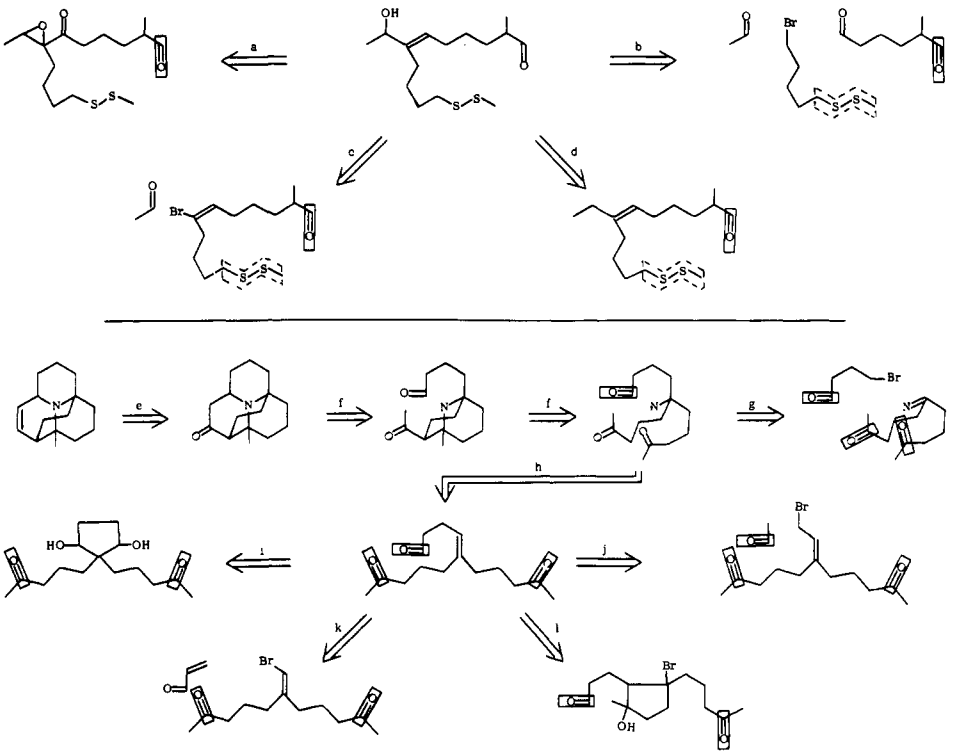


Figure 8. Synthetic sequences generated by LHASA using the new transform evaluation scheme including the FGR module. Functional groups deemed reactive but protectable toward the conditions of a reaction are enclosed in a box. Unstable, unprotectable groups are enclosed in a dashed box. (a) Wharton reaction (NH_2NH_2); (b) "double" Wittig reaction[16] (1. Ph_3P; 2. BuLi; 3. BuLi); (c) organometallic addition (Li); (d) allylic oxidation (SeO_2); (e) reduction-dehydration (1. $NaBH_4$; 2. $SOCl_2$);[17] (f) Mannich reaction (mild H^+); (g) organometallic addition to imine (Li); (h) Markovnikov amination (via BrN_3);[18] (i) Grob fragmentation (1. TsCl; 2. *t*-BuOK); (j) α-alkylation ($LiNR_2$); (k) organometallic Michael addition (Li); (l) Grob fragmentation (*t*-BuOK).

ketal. As will be seen shortly, the program also realizes this fact, brings it to the attention of the chemist, and adjusts its interference evaluation accordingly.)

(8) Those functional groups not meant to participate in the reaction are now individually considered. If a fragment specifier is in effect for the current reagent, only those groups in the proper fragment will be considered.

(9) If this functional group's reactivity level toward the current reagent is less than CDNLVL, then the group is regarded as stable and unreactive toward the reagent. (Recall that a multisubclassed functional group will be assumed to have a reactivity level equivalent to that of its most highly reactive subclass.)

(10) If the functional group is reactive toward this reagent, than a check is made to see if the group is protectable. The six functional group types considered protectable by LHASA under various conditions are listed in Table IV.

(11) If the group is not considered protectable under the current conditions, then it is designated as an interfering group. In this case LOCAL, which reflects the total degree of interference encountered toward the current set of reaction conditions, is incremented. The amount that LOCAL is incremented when an interfering group is encountered can be manually specified by the chemist by setting the value of the "interference weight" variable on the graphical input display. Although its value is initially set at 45, the chemist may vary its value from 0, which causes the FGR module to be disregarded entirely, to 100, which makes the occurrence of a single interfering group sufficient to deem a set of reaction conditions unacceptable. In this way, the chemist is given full control over the amount of emphasis that is placed on interfering functionality.

(12-13) If at any time the value of LOCAL equals or exceeds 100, then the set of conditions currently being considered is deemed unacceptable and LOOPFG is set to 1 in order that the remaining reagents comprising this set of conditions may be by-passed.

(14-17) If the group is considered protectable under the current conditions, then a check is made to see if this group is the first of its type to be designated as protected from the current conditions. If it is, then the value of LOCAL is incremented slightly (by one-quarter the value of the "interference weight" variable) and the count of protected group types is increased by one. If a single set of conditions requires four or more different functional group types to be protected, then this set of conditions is deemed unacceptable. Although it is, at times, perfectly reasonable to perform a reaction in the presence of four or more protected group types, it is also usually true that the reaction could be performed more appropriately at a different point in the synthetic scheme, one where there are fewer group types requiring protection from its reaction conditions.

(18-20) Once every reagent constituting a particular set of conditions has been examined, the value of LOCAL is compared with MLTPLR. If LOCAL is smaller, then the set of conditions just examined was better than any previously encountered. In this case, the value of MLTPLR is reset to LOCAL. Also, any groups found unstable and/or protectable to the just-examined conditions are marked for later display. (Recall that, if the "best" set of conditions found is deemed acceptable, groups which were unstable but protectable toward those conditions are displayed to the chemist enclosed in a box, and unstable, unprotectable groups are displayed enclosed in a dashed box.)

(21) If MLTPLR is zero, then a set of conditions has been found for which there is no interference. In this case, there is no need to examine any other condition sets so control is returned to RXEVAL.

(22) If MLTPLR is not zero, but the final set of conditions has been examined, then control is also returned to

RXEVAL. At this time, the return value of MLTPLR reflects the total degree of interference to be expected from applying the "best" set of reaction conditions. If this value is 100, then no conditions examined were acceptable and the transform is killed. If MLTPLR is less than 100, however, acceptable conditions were found so the final transform rating is adjusted (vide supra) and the precursor resulting from the successfully evaluated transform is displayed to the chemist.

Sample synthetic sequences generated by LHASA using the new transform evaluation scheme containing the FGR module are shown in Figure 8.

VIII. Discussion

The addition of a functional group reactivity module to LHASA has enabled the program to identify and bring to the chemist's attention occurrences of interfering functionality and situations where such interferences can be removed by the use of protecting groups. Straightforward extensions *keeping within* the present framework of the FGR module will enable a more sophisticated treatment of interfering functionality and functional group protection. These include the expansion of the set of prototype reagents, an increase in the number of functional group types subclassified, and the extension of the protectable functional group scheme to include additional group types. Perhaps more exciting are extensions to the FGR module in which the identification of interfering functionality and the need for functional group protection are themselves used for a deeper analysis of the problem of protection. For example, when highly reactive, unprotectable functionality continually interferes with the operation of important, simplifying transforms, requests to remove the interfering group can be generated automatically. This would correspond to this first instance of the program using results of its own derivation to generate subgoal requests de novo, i.e., without those requests having been formulated previously in the program. The automatic generation of subgoals in this manner can be seen as an initial step toward the development of higher level strategies in which the synthetically challenging aspects of a particular target molecule are themselves used to generate strategies for its synthesis.

It is also desirable, as an extension to the functional group protection scheme, for the program to be capable of specifying not only *when* functional group protection is needed, but also *which* protecting groups are eligible, and *where* in the synthetic sequence the protection and deprotection steps would be appropriate. Such an extension is clearly feasible since from a chemical standpoint it involves only the gathering of well-documented information concerning the types of available protecting groups and the conditions required for the incorporation and removal of those groups.

Acknowledgments. We are grateful to the National Institutes of Health for financial assistance. Helpful suggestions from other members of the LHASA project, especially W. L. Jorgensen, J. W. Vinson, and W. J. Howe, are greatly appreciated.

Appendix. The Functional Group/Reagent Cross-Reference Table

The functional group/reagent (FG/RGNT) cross-reference table lists the reactivity level of every functional group recognized by LHASA toward 60 prototype chemical reagents. Because the reactivity of many functional groups can depend dramatically on their molecular environment, the FG/RGNT table consists of two sub-tables, REACTB and ENVRTB (Tables I and II, respectively). The REACTB sub-table contains functional groups whose reactivities are to a reasonable approximation independent of their molecular environment and are determined more by the intrinsic nature of the groups themselves. In the ENVRTB sub-table, the functional groups are subclassified based on molecular environment, and reactivity levels are listed separately for each subclass.

The reagent numbers listed in each sub-table correspond to the respectively numbered prototype reagents in Figure 2. In the ENVRTB sub-table, functional group subclasses are listed by their respective subclassification ratings. Descriptions of the subclasses corresponding to these ratings are given in Table III. Details concerning the subclassification of functional groups, the selection of prototype reagents, and the designation of reactivity levels are discussed in the body of this paper.

References and Notes

(1) E. J. Corey, *Pure Appl. Chem.*, **14**, 19 (1967).
(2) For a preceding paper in this series, see E. J. Corey, W. J. Howe, H. W. Orf, D. A. Pensak, and G. Petersson, *J. Am. Chem. Soc.*, **97**, 6116 (1975).
(3) E. J. Corey, *Q. Rev., Chem. Soc.*, **25**, 455 (1971).
(4) E. J. Corey, R. D. Cramer, III, and W. J. Howe, *J. Am. Chem. Soc.*, **94**, 440 (1972).
(5) The following doctoral dissertations may be consulted for details and additional information not included herein: (a) W. J. Howe, Ph.D. Thesis, Harvard University, 1972; and (b) D. A. Pensak, Ph.D. Thesis, Harvard University, 1973.
(6) H. C. Brown, "Boranes in Organic Chemistry", Cornell University Press, Ithaca, N.Y., 1972.
(7) The individual bits comprising a computer word are considered as having values of 0 or 1 corresponding respectively to that bit being "off" or "on" in the word.
(8) For recent accounts of the transform evaluation process, the syntax of the data table language, and sample data table entries, see ref 5b.
(9) Further details concerning the list structures associated with functional groups in LHASA appear in ref 5b.
(10) J. F. W. McOmie, "Protective Groups in Organic Chemistry", Plenum Press, New York, N.Y., 1973.
(11) Clearly such a determination can be made only after a string of lower level precursor structures in the synthetic tree is defined, either by actual generation or by a type of look-ahead process.
(12) G. F Wright in "Steric Effects in Organic Chemistry", M. S. Newman, Ed., Wiley, New York, N.Y., 1956, pp 406–416.
(13) EVLØ1 is also called before processing any molecule input by the chemist. Were the chemist to draw in structure X, for example, processing would not be allowed. Instead, X would automatically be tautomerized to XI, and control would be returned to the chemist.
(14) The actual hydride reagent employed in the modified Wolff–Kishner reduction (ref 15) is $NaBH_3CN$ (100°) which is represented by the prototype reagent $NaBH_4$.
(15) R. O. Hutchins, C. A. Milewski, and B. E. Maryanoff, *J. Am. Chem. Soc.*, **95**, 3662 (1973).
(16) E. J. Corey, J. I. Shulman, and H. Yamamoto, *Tetrahedron Lett.*, 447 (1970).
(17) E. J. Corey and R. D. Balanson, *J. Am. Chem. Soc.*, **96**, 6516 (1974).
(18) D. Van Ende and A. Krief, *Angew. Chem., Int. Ed. Engl.*, **13**, 279 (1974).