(15), 50 (17); ^{19}F NMR ($CDCl_3$) ϕ^* −113.45 ppm (doublet, J = 51.4 Hz) (lit. value −71.5 ppm[29]); HRMS calcd for $C_7H_5F_2NO_2$ 173.0295, found 173.0291.

General Procedure for the Fluorination of Electron-Rich Aromatic Systems Using *N*-Iodosuccinimide To Avoid Aromatic Ring Bromination. 1-(Difluoromethyl)-2,4,6-trimethylbenzene (5g). *N*-Iodosuccinimide (2.01 g, 8.92 mmol) was dissolved in 18 mL of dry CH_2Cl_2 and allowed to stir under nitrogen. The mixture was cooled to −30 °C, and 1.10 mL (4.83 mmol, i.e., 44.6 equiv of F^-) of pyridinium poly(hydrogen fluoride) were added via polypropylene/polyethylene syringe, followed by the dropwise addition of dithiolane **4g** (0.50 g, 2.23 mmol) in 1.0 mL of CH_2Cl_2. After 15 min, the reaction mixture (which was dark purple) was diluted with CH_2Cl_2 (30 mL) and filtered through a column (a 50-mL polypropylene/polyethylene syringe with a cotton plug) of basic alumina. The organic solution was dried (Na_2SO_4) and concentrated. Preparative TLC (10% diethyl ether/hexane) afforded 55% of the difluoro compound **5g**: ^1H NMR ($CDCl_3$) δ 2.42 (s, 3 H), 2.43 (s, 6 H), 6.86 (s, 2 H), 6.94 (t, 1 H, J = 54.4 Hz); mass spectrum, m/z (rel intensity) 170 (47, M^+), 155 (21), 119 (100), 91 (23), 51 (33), 39 (29); ^{19}F NMR ($CDCl_3$) ϕ^* −111.78 ppm (doublet, J = 54.1 Hz). Anal. Calcd for $C_{10}H_{12}F_2$: C, 70.57; H, 7.11; F, 22.32. Found: C, 70.87; H, 7.04; F, 22.10.

General Procedure for the Fluorination of Ketone-Derived Dithianes. 6,6-Difluoroundecane (5a). 1,3-Dibromo-5,5-dimethylhydantoin (1.65 g, 5.76 mmol) was dissolved in 12 mL of dry CH_2Cl_2 and allowed to stir under nitrogen. The mixture was cooled to −78 °C and 0.95 mL (4.16 mmol i.e., 38.4 equiv of F^-) of pyridinium poly(hydrogen fluoride) was added via polypropylene/polyethylene syringe, followed by the dropwise addition of dithiane 8 (0.50 g, 1.92 mmol). After 45 min, the reaction mixture was diluted with hexane (30 mL) and filtered through a column of basic alumina. Flash chromatography (2.0% diethyl ether/hexane) yielded 258 mg (70%) of difluoro compound **5a**.

(29) Wielgat, J.; Wozniacki, R. *J. Fluorine Chem.* **1984**, *26*, 211–214.

Acknowledgment. We are grateful for support of this research from the National Institutes of Health (PHS 5R01 CA 25836). High field NMR spectra were obtained on instruments in the NSF Regional Instrumentation Facility supported by a grant from the National Science Foundation (NSF CHE 79-16100) and high resolution mass spectrometry was performed on instruments supported by a grant from the National Institutes of Health (PHS GM 27029).

Registry No. **3a**, 927-49-1; **3b**, 112-54-9; **3c**, 15600-08-5; **3d**, 119-61-9; **3e**, 830-13-7; **3f**, 555-16-8; **3g**, 487-68-3; **4a**, 103383-69-3; **4b**, 103383-70-6; **4c**, 29575-88-0; **4d**, 6317-10-8; **4e**, 16775-67-0; **4f**, 41159-02-8; **4g**, 41159-04-0; **5a**, 103383-71-7; **5b**, 62127-45-1; **5c**, 16319-73-6; **5d**, 360-11-2; **5e**, 27415-48-1; **5f**, 29848-57-5; **5g**, 103383-72-8; **8**, 5849-09-2; 1,2-ethanedithiol, 540-63-6; 1,3-propanedithiol, 109-80-8; 1,3-dibromo-5,5-dimethylhydantoin, 77-48-5; pyridinium poly(hydrogen fluoride), 62778-11-4.

Computer-Assisted Mechanistic Evaluation of Organic Reactions. 12. pK_a Predictions for Organic Compounds in Me_2SO

Alan J. Gushurst and William L. Jorgensen*

Department of Chemistry, Purdue University, West Lafayette, Indiana 47907

Received March 4, 1986

A new algorithm for pK_a prediction has been implemented in the computer synthesis program CAMEO. This algorithm is based on a few organizing principles extracted from a large number of experimentally determined acidities in Me_2SO. It considers such molecular factors as field and electron-delocalization effects, hydrogen bonding, hybridization, and steric effects. Useful generalizations have also been developed for hetero–hetero activated systems, benzenoid compounds, and others. These organizing principles coupled with a small data base make it possible to predict the acidities of a vast number of organic compounds to within 2 pK_a units.

I. Introduction

An interactive computer program, CAMEO, which predicts the products of organic reactions given starting materials and conditions, is under continued development. A key aspect of the program is that it does not use large, internally stored data bases, but rather general organizing principles and algorithms. The recognition of such principles is fundamentally important in addition to permitting greater efficiency in the program. Mechanistic rules have been developed to handle base-catalyzed and nucleophilic chemistry[1] including some organometallic reactions,[2] organosilicon chemistry,[3] acid-catalyzed and electrophilic reactions,[4] nucleophilic[5] and electrophilic[6] aromatic substitution, and thermal pericyclic processing including cycloadditions and sigmatropic and electrocyclic rearrangements.[7,8]

A knowledge of pK_a values for organic compounds is important to the program as well as to the chemist. It is essential to have this information in order to determine the feasibility of proton transfer. Since proton transfer is often the first step in a base-catalyzed reaction, prediction of reaction products would not be possible without knowledge of acidities. The continued development of the program has identified a need for a more generalized and comprehensive algorithm for pK_a determination. The old version of the program used a data driven algorithm in which functional group numbers were stored in an array of acidity levels.[1,2] This array was then used following functional group perception to determine the most acidic hydrogens. Although the old algorithm covered a large number of compounds, various limitations such as the inability to handle triactivation and substituted aromatic compounds prompted the need for a change.

(1) Salatin, T. D.; Jorgensen, W. L. *J. Org. Chem.* **1980**, *45*, 2043.
(2) Salatin, T. D.; McLaughlin, D.; Jorgensen, W. L. *J. Org. Chem.* **1981**, *46*, 5284.
(3) Peishoff, C. E.; Jorgensen, W. L. *J. Org. Chem.* **1983**, *48*, 1970.
(4) McLaughlin, D. Ph.D. Thesis, 1983, Purdue University.
(5) Peishoff, C. E.; Jorgensen, W. L. *J. Org. Chem.* **1985**, *50*, 1056.
(6) Bures, M. B.; Roos-Kozel, B. L.; Jorgensen, W. L. *J. Org. Chem.* **1985**, *50*, 4490.
(7) Burnier, J. S.; Jorgensen, W. L. *J. Org. Chem.* **1983**, *48*, 3923.
(8) Burnier, J. S.; Jorgensen, W. L. *J. Org. Chem.* **1984**, *49*, 3001.

0022-3263/86/1951-3513$01.50/0

A new algorithm for pK_a determination has been developed that is both more accurate and general. The extensive experimental work by Bordwell and co-workers has provided a comprehensive data base for the algorithm which pertains to acidities in the common organic solvent Me_2SO.[9] Two key aspects of the new algorithm are that it operates on a condensed data base and that it can predict the acidities of a vast number of organic compounds to within 2 pK_a units. The implementation and scope of the new algorithm will be discussed following a brief presentation of acidifying effects and other predictive methods for pK_a values.

II. Background and Methods of pK_a Prediction

A. Effects of Structure on pK_a. Molecular features that modify pK_a are well-known.[10] They include mainly field-inductive and electron-delocalization effects, hydrogen bonding, hybridization, and steric factors. Field effects have inductive and electrostatic components. The inductive effect, in particular, is known to decrease with increasing distance from the acidic site. In most cases, the effect has little influence on a site that is four or more atoms away. Electron-delocalization or resonance effects stabilize systems in which an anionic center is α to an unsaturated bond. Additional stabilization occurs when the charge can be resonated to a heteroatom. It is well-known that the electronic character of a substituent can be transmitted through an unsaturated system as described by Fuson's principle of vinylogy.[11] Such delocalization effects fall off less rapidly than inductive effects.

The remaining factors are typically less common or less activating than field or electron-delocalization effects. Intramolecular hydrogen bonding is geometry-dependent. For example, *o*-hydroxyphenol is significantly more acidic than the para isomer; the pK_a difference in Me_2SO is 4.[9f] The hybridization effect follows from the amount of "s character" or intrinsic electrophilicity for the acidic atom. This is helpful in explaining the acidity of many hydrocarbons as well as molecules containing heteroatoms. Steric factors are much more diverse. They may play a role in the proton transfer (although seldom encountered), solvation, or resonance for a molecule. Steric inhibition of resonance is observed in many systems including 3,5-dimethyl-4-nitrophenol in which the methyl groups prevent the nitro group from being in the aryl plane thereby especially diminishing the resonance stabilization of the anion.

B. Methods of pK_a Prediction. There are several methods available for pK_a prediction that offer varying scope and accuracy.[12] Linear free energy relationships have been defined and employed in this way. They can be accurate to within a few tenths of a pK_a unit; however, they are specific to particular types of systems. The most widely used linear free energy relationships are the Hammett (1) and Taft (2) equations.[13,14] The Hammett equation is designed to handle substituted aromatic acids and bases, whereas the closely related Taft equation focuses on aliphatic systems. For particularly complex molecules, representative fragments can often be identified for treatment by these relationships. Of course, the σ and σ^* values for the substituents must be known.

$$pK_a = pK_a^\circ - \rho\Sigma\sigma \quad (1)$$

$$pK_a = pK_a^\circ - \rho^*\Sigma\sigma^* \quad (2)$$

Several other methods exist for pK_a prediction which are not derived from a linear free energy relationship. These alternatives include predictions for specific classes of compounds, predictions by analogy or extrapolations, and theoretical predictions. The acidities of aliphatic carboxylic acids and amines have been shown to be relatively insensitive to changes in the length or branching of the hydrocarbon chain.[12] Hence, simple relationships have been devised to predict acidities for these classes of compounds. Prediction by analogy is common and often functional group based. Of course, it requires knowledge of a reference set of experimental pK_a's.[12] Extrapolation allows for pK_a prediction of an organic acid in a particular solvent if its acidity is known in another solvent of similar solvating character.[15] Finally, pK_a predictions can be made by theoretical methods. In particular, gas-phase acidities have been estimated by molecular orbital calculations.[16,17]

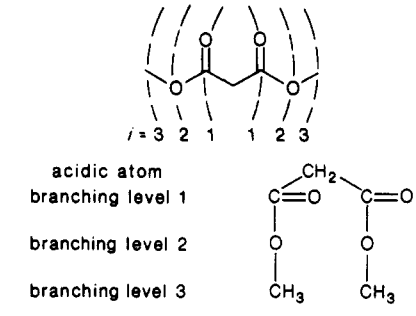
III. General Algorithm for pK_a Prediction in Me_2SO

A. pK_a Formulation. Analysis of Bordwell's data suggests that the effects of substituents on an acidic site can be treated as if they were approximately additive. This is the basis of the present algorithm which can be generalized in terms of the following empirically derived equation. In eq 3, pK_a° is the pK_a of a reference compound,

$$pK_a = pK_a^\circ - \sum_{i}^{\text{levels}} \sum_{j}^{\text{groups}} c_{ji}\alpha_{ji}/i^2 \quad (3)$$

i is the branching level index and corresponds to the distance from the acidic site measured in number of atoms, c is a leveling factor, and α is an acidfying constant that is characteristic of a given substituent. Details on these parameters are discussed below.

The standard procedure followed by the algorithm addresses the acidic site of interest and then branches out level by level as illustrated in the following structure. The



branching levels are designated by i. The factor $1/i^2$ is

(9) (a) Bordwell, F. G. *Pure Appl. Chem.* **1977**, *49*, 963. (b) Bordwell, F. G.; Fried, H. E. *J. Org. Chem.* **1981**, *46*, 4327. (c) Bordwell, F. G.; Algrim, D. *J. Org. Chem.* **1976**, *41*, 2507. (d) Olmstead, W. N.; Bordwell, F. G. *J. Org. Chem.* **1980**, *45*, 3299. (e) Olmstead, W. N.; Margolin, Z.; Bordwell, F. G. *J. Org. Chem.* **1979**, *45*, 3295. (f) Bordwell, F. G., personal communications.

(10) March, J. *Advanced Organic Chemistry*; McGraw Hill: New York, 1977; pp 238–245.

(11) Fuson, R. C. *Chem. Rev.* **1935**, *16*, 1.

(12) For a review, see: Perrin, D. D.; Dempsey, B.; Sergeant, E. P. *pK_a Prediction for Organic Acids and Bases*; Chapman and Hall: London, 1981.

(13) Hammett, L. P. *Chem. Rev.* **1935**, *17*, 125. Hammett, L. P. *Physical Organic Chemistry*; McGraw-Hill: New York, 1940; Chapter 7.

(14) Taft, R. W.; Lewis, I. C. *J. Am. Chem. Soc.* **1959**, *81*, 5343. For a recent review on acid-base equilibria, see: Taft, R. W. *Prog. Phys. Org. Chem.* **1983**, *14*, 247.

(15) Hall, N. S. *J. Am. Chem. Soc.* **1930**, *52*, 5115. Bordwell, F. G.; Branca, J. C.; Hughes, D. L.; Olmstead, W. N. *J. Org. Chem.* **1980**, *45*, 3305.

(16) Arnett, E. M.; Chawla, B.; Bell, L.; Taagepera, M.; Hehre, W. J.; Taft, R. W. *J. Am. Chem. Soc.* **1977**, *99*, 5729. Reynolds, W. F.; Mezey, P. G.; Hehre, W. J.; Topsom, R. D.; Taft, R. W. *J. Am. Chem. Soc.* **1977**, *99*, 5821.

(17) Chandrasekhar, J.; Andrade, J.; Schleyer, P. von R. *J. Am. Chem. Soc.* **1981**, *103*, 5609.

Table I. Values for the Leveling Constant *c*

substit no.	values of *c*		
	normal	nitro	nitrile
1	1	1	1
2	1	2/3	1
3	1/3	1/3	2/3
4	1/4	1/4	1/4

applied to substituents in branching level i to account for their diminishing contribution with increasing distance. For example, the malonate shown above has two carbonyls in branching level 1 that would contribute $1/1^2$ or 100% whereas the oxygens in branching level 2 would only contribute $1/2^2$ or 25%. Only 3 levels are considered, though an aromatic fragment can be included in 1 level (vide infra).

The factor c represents a leveling effect observed for a branching level. It accounts for diminishing activation with increasing number of substituents on the acidic site. The values of c are applied to substituents in a branching level according to their activating ability. Specifically, the most strongly activating substituents are leveled less than the more weakly activating ones. In a given branching level, the substituents are ordered from high to low and each substituent is leveled according to the corresponding c value given in Table I. As a rule, no significant leveling is observed until a third substituent is introduced. Two key exceptions are for activation by nitro and nitrile groups. Whenever a nitro group is α to the acidic site, leveling is observed as soon as a second substituent is introduced as illustrated below. In contrast, whenever a

	CH_3NO_2	vs. $PhSO_2CH_2NO_2$
exptl[9a]	17.2	7.1
calcd	17	6 (2 without leveling)

calculations:
45 - [1(1 x 28)] = 17
45 - [1(1 x 28 + 2/3 x 14) + 1/4(1 x 6)] = 6

nitrile group is α to the acidic site, diminished leveling is observed as witnessed below with the phenyl-substituted

	CH_3CN	vs. $PhCH_2CN$	vs. Ph_2CHCN
exptl[9b]	31.3	21.9	17.5
calcd	31	22	18

calculations:
45 - (-3) - [1(1 x 17)] = 31
45 - [1(1 x 17 + 1 x 6)] = 22
45 - [1(1 x 17 + 1 x 6 + 2/3 x 6)] = 18

nitrile alkanes. Values for c reflecting these observations are also given in Table I. (The calculations for the computed pK_a's are provided with most of the examples for illustration. The base pK_a for saturated carbon acids is 45. Any adjustments as with the −3 for acetonitrile are discussed below.)

The constant α is a measure of a substituent's activating ability. The algorithm currently uses a database of 29 representative atomic groupings with corresponding α's to determine acidities. The α value for any given unit was determined by taking an average of the unit's activating effects on numerous mono- and diactivated systems. A listing of these units and their α's is given in Table II. It is important to recognize that these units are typically fragments of functional groups. Thus, the algorithm does not require functional group perception. Restriction of predictions to only perceived functional groups is consequently avoided. The values for α are representative of the total activation (both inductive and resonant) of a substituent relative to hydrogen. Activation by a substituent is approximated to be the same through any molecular chain, i.e., activation is attenuated through -S-, -O-, -CO-, or any other unit in Table II to the same extent as through $-CH_2-$.

Table II. Substituent Activation Values on Saturated and Aromatic Sites[a]

atom or group	saturated	aromatic	
		OPACT	MACT
H	0.0	0.0	0.0
O	1.0	−1.0	0.0
N	3.0	−2.0	−1.0
C	−0.3	0.0	0.0
F	3.0	0.0	2.0
Cl	5.0	1.0	2.0
Br	2.0	2.0	2.0
I	2.0	2.0	2.0
Si	6.0	(2.0)	(2.0)
P	10.0	−1.0	−1.0
S	7.0	2.0	2.0
Se	6.0	(2.0)	(2.0)
aromatic	6.0	1.0	0.0
olefinic	6.0	1.0	0.0
C≡CH	7.0	1.0	0.0
C=N	10.0	(3.0)	(2.0)
N=C	8.0	(1.0)	(1.0)
N=N	(12.0)	(3.0)	(2.0)
CH=S or N=S	17.0	3.0	2.0
NO_2	28.0	7.0	4.0
R_2S^+	22.0	(4.0)	(3.0)
R_3N^+	10.0	3.0	2.0
R_3P^+	19.0	(4.0)	(3.0)
C≡N	17.0	5.0	3.0
C=O	18.0	3.0	2.0
P=O	12.0	(3.0)	(2.0)
N=O	(22.0)	9.0	4.0
SO_2	14.0	4.0	3.0
S=O	10.0	3.0	2.0

[a] The values in parentheses were approximated by relating those substituents' activations to similar substituents with known activations.

B. Generalized Treatment of Acidic Systems. The following sections deal with the development of key principles used by the algorithm to predict pK_a. The main subroutine in CAMEO for pK_a perception is called PKAVAL. Its organization is diagrammed in Figure 1.

1. Determination of $pK_a°$. A generic saturated carbon has been designated as the reference for $pK_a°$ from which the acidities of all other compounds are derived. An acidity of 45 was chosen empirically for this reference in order to give reasonable activation values for substituents. Presently, there are few experimental data available for compounds that are less acidic than Me_2SO. Thus, the treatment for such compounds is less refined, as discussed below.

Acidities for other parent systems are determined according to the nature of the acidic atom. Acidic atoms are classified by the program as either carbon acids, hetero acids, or special compounds. The treatment of carbon acids is the most general. The generic carbon provides the base pK_a for all saturated, acidic carbon sites. Thus, for acetone the carbonyl group is treated as a substituent directly activating the parent carbon. And, the predicted acidity is a reflection of the sum of substituent activations operating on the acidic site.

Adjustments are made to account for vinylogous and unsaturated acidic sites. The treatment for vinylogy applies to hetero acids as well as carbon acids. Before the parent designation is made, the program checks for heteroatoms or polarized multiple groups that are in resonant positions with the acidic atom through intervening multiple bonds. If such atoms or groups are found, they are treated as if they were directly activating the acidic atom and the

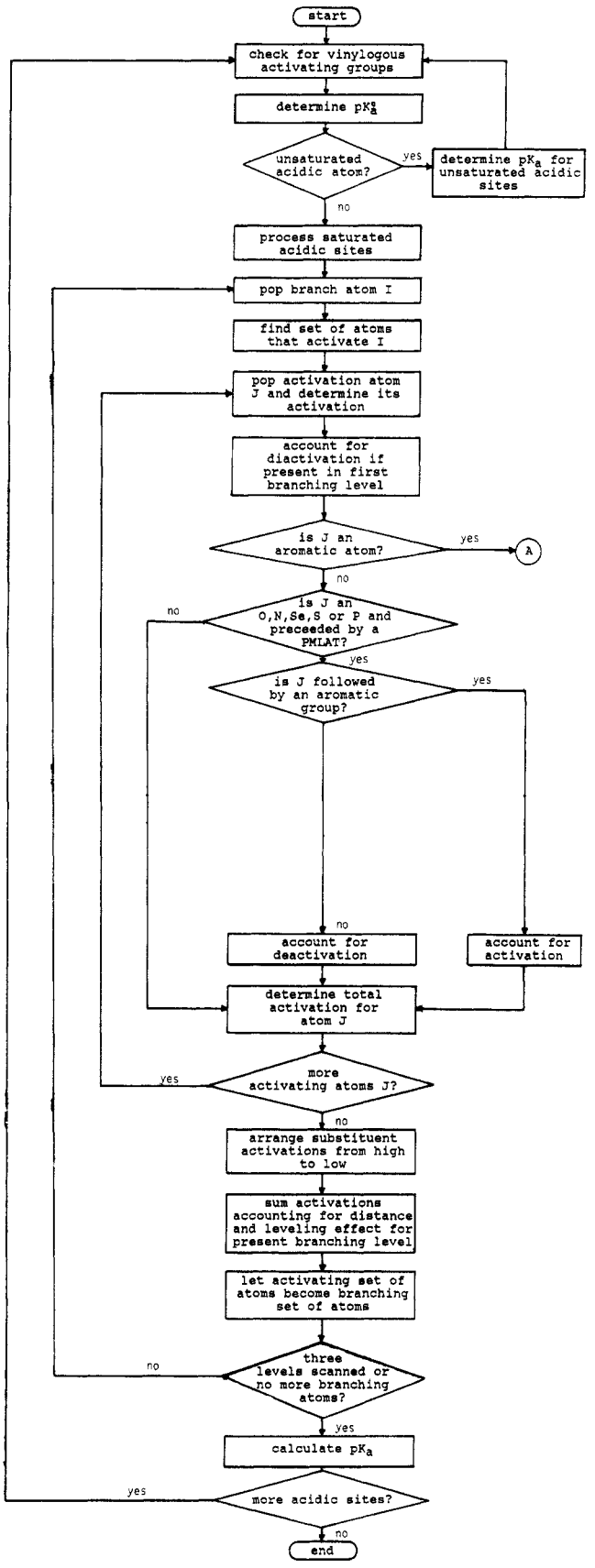


Figure 1. Flow chart for pK_a prediction. Point A branches to Figure 2.

parent acidity is lowered by 2 units to account for additional resonance stabilization. Unsaturated acidic atoms also require adjustment to account for the hybridization effect. pK_a° for vinylic sites is lowered by 2, to 43, whereas acetylenic atoms are acidified by 15 units to 30.

Table III. Parent Acidity Values for Heteroatomic Acids and Other Compounds[a]

parent	pK_a°	parent	pK_a°
R′R″R‴CH	45		44
ROH	30		
PhOH	24	C=CHR	43
RSH	17	C≡CH	30
RR′NH	38	cyclopropane	43
PhRNH	36	HX where X = halogen	1

[a] The parent acidity values are indicative only of the acidic atom and therefore do not include the acidifying effects of neighboring atoms or groups. Thus, the predicted acidity of phenol is actually 24 − 6 = 18.

Table IV. Activation Values for a Heteroatom α to an Acidic Heteroatom

acidic heteroatom	α heteroatom	activation[a]
N	N	6.0
O	N	6.0
all other combinations of O, N, S, Se, or P		11.0

[a] Values to be subtracted from pK_a°.

Hetero acids require the use of a reference other than the parent carbon. A small number of hetero acids and other parent compounds that must be found explicitly are listed in Table III with their pK_a° values. Since no data are available for phosphorus or selenium acids, their acidities are approximated to be similar to nitrogen and sulfur acids, respectively. Hetero acids not listed in the table are classified according to the acidic atom. For instance, the parent designation for carboxylic acids is alcohol. Hetero acids in which the acidic atom is either an oxygen, nitrogen, selenium, sulfur, or phosphorus atom and is directly attached to either an oxygen, nitrogen, selenium, sulfur, or phosphorus atom are activated beyond what is found for these atoms or groups activating carbon. As a result, hetero acids bearing this relationship receive an increment according to the generalized activations presented in Table IV. The acidity of the amide nitrogens in benzamide and benzhydroxamic acid given below illustrates the point. These rules for hetero–hetero acti-

vation coupled with the designation of hetero parents allow treatment of a large variety of hetero acids.

Adjustments are also needed for compounds that display steric inhibition to resonance and aromatic stabilization upon deprotonation. Steric inhibition of resonance occurs in compounds that have a bridgehead atom activated by a mulitple bond. Of course, the extent of deactivation is dependent on the size of the rings involved. However, since acidity data for such systems are limited, a generic adjustment of 6 for bridgehead atoms is applied. The compounds that were used to derive this value were bicyclo[3.3.1]nonan-2-one and 3,3-dimethylbicyclo[3.3.1]nonan-2-one shown below. It is interesting to note that diacti-

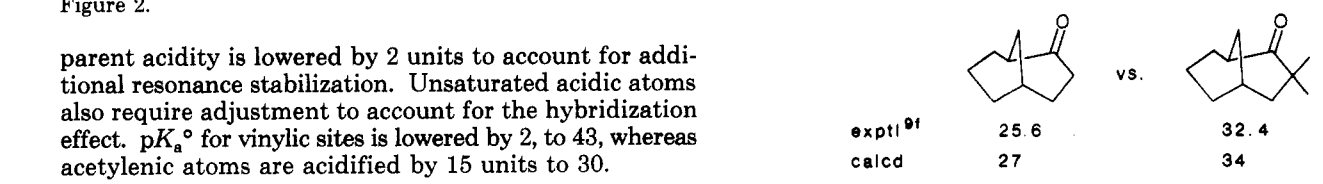
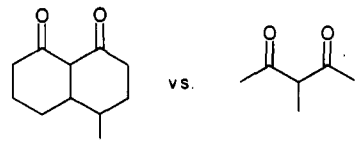


Table V. Miscellaneous Adjustments to Activation Parameters

condition	adjustment[a]
vinylogous activation	+2
resonance-activated bridgehead atom	−6
diresonance-activated fusion atom	−3
five-membered aromatic ring formation upon deprotonation	+9
R − CN where R = alkyl	−3
1,3-dicarbonyls	−4
1,3-dithiones	+5
1,3-disulfonyls	+2
aromatic rings	
ortho substituent	+1
ortho hydrogen bonding substituent	+3
ring follows carbonyl	2 × substituents
ring follows sulfoxide or sulfone	3 × substituents
heteroaromatic rings	
heteroatom in resonant position	+4
heteroatom in nonresonant position	+3
ester-like systems 	
G = O, S, Se or P	−8
when R′ = unsaturated group	+8
G = N	−16
when R′ = unsaturated group	−8
unsaturated acidic sites	
α activation for vinylic sites	0.3 × substituent
α heteroaromatic atom to aromatic site	+6

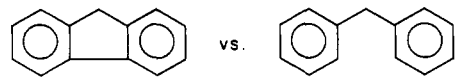
[a] Values to be subtracted from pK_a°.

vation at fusion atoms is also diminished, as observed with 4-methyl-1,8-decalindione. Such compounds are consequently deactivated by 3 units. Systems such as fluorene,



exptl[9d]	18.1	15.0
calcd	17	13

indene, and cyclopentadiene exhibit enhanced acidities due to aromatic stabilization upon deprotonation. The predicted pK_a's for these molecules are lowered by 9 units to account for the enhanced activation. Consequently fluorene is far more acidic than diphenylmethane. Pres-



exptl[9a]	22.6	32.2
calcd	22	33

calculations:
45 − 9 − [1(1 × 7 + 1 × 7)] = 22
45 − [1(1 × 6 + 1 × 6)] = 33

ently, there are no adjustments made for compounds that become antiaromatic upon deprotonation. Adjustments could be made when acidity data become available for such compounds. The adjustments just discussed are summarized in Table V.

2. Treatment of Saturated Acidic Sites. The processing of saturated acidic sites constitutes a major portion of the subroutine for pK_a perception. It is here that the actual branching is performed and substituent activations determined. The branching radiates out from the acidic site of interest in layers. Once an atom or group has been processed, it is placed in a set termed BRATMS which stands for branching atoms. Atoms or groups in BRATMS are not to be branched to again. This prevents the processing of an atom or group twice which could potentially

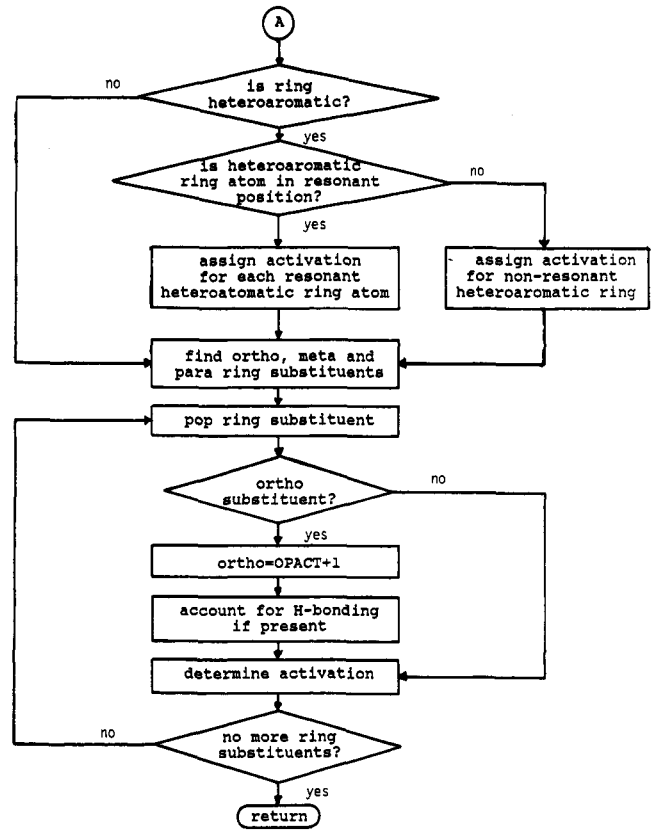


Figure 2. Flow chart for processing aromatic substituents.

occur in a cyclic molecule. The substituent values used are the ones given in Table II except for a few diactivated cases, simple nitriles, and aromatic and ester-like systems, as discussed below.

Substituent activations applied in the first branching level are the largest in magnitude and, therefore, could produce the greatest error. Consequently, fine tuning is required for this level. Molecules containing fragments corresponding to 1,3-dicarbonyls, 1,3-dithiones, and 1,3-disulfonyls presently require modifications. Additionally, monoactivated nitrile alkanes are adjusted to reflect their relative low acidity with respect to their di- and triactivated counterparts. The adjustments for these cases are listed in Table V. All other diactivated compounds are handled without adjustment.

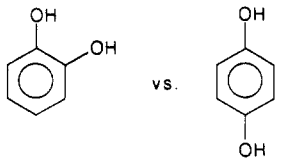
Aromatic substituents are treated as a single group. The subroutine ARMVAL is called whenever an aromatic atom is encountered. Its organization is diagrammed in Figure 2. The program assigns an initial activation of 6 for the aromatic substituent which corresponds to activation by an unsubstituted benzene ring. The purpose of ARMVAL is to adjust for heteroaromatic activation and determine the additional activation by any substituents on the ring. Upon exiting ARMVAL, a value indicative of the total activation by the aromatic group is returned.

The treatment of heteroaromatics considers whether the negative charge for the incipient anion can be resonated to a heteroatom in the aromatic ring. A generalization which gives good results is to lower the acidity by 4 pK_a units for each heteroatom in a resonant position and by 3 for each heteroatom in a nonresonant position. Relatively accurate predictions can then be made for derivatives of compounds such as pyrrole, imidazole, triazoles, and tetrazole. Predictions tend to be less satisfactory for heteroaromatics involving heteroatoms other than nitrogen.

The treatment of substituents on aromatic rings requires the use of two data sets characteristic of a substituent's

activating ability. These sets are OPACT and MACT for ortho/para and meta activation. The activation values were determined from data for a large number of substituted phenols and are given in Table II.[9f] As a rule, a group is more activating when it is ortho than when it is para by 1 pK_a unit. This generalization is true for all substituents except hydrocarbon groups which have roughly equal ortho and para activations.

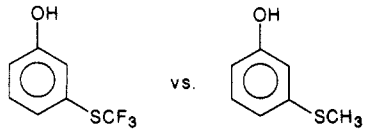
An additional rule is used to account for intramolecular hydrogen bonding. Anilines, phenols, and thiophenols typically have their acidities enhanced by an extra 3 units for hydrogen-bonding substituents in the ortho positions. Hence, catechol is more acidic than hydroquinone as indicated below. Finally, only activating groups directly



exptl[9f]	15.7	19.7
calcd	15	19

calculations:
24 - [1(1 x 9)] = 15
24 - [1(1 x 5)] = 19

attached to the ring are considered to have any effect on the total activation. Groups that are farther away tend to contribute only slightly and are therefore not considered.

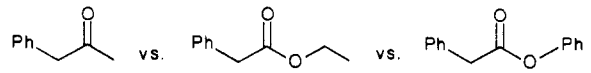


exptl[9f]	15.4	15.8
calcd	16	16

calculation:
24 - [1(1 x 8)] = 16

Aromatic ring substituents display enhanced activation when the ring directly follows a polarized, multiply bonded group, e.g., carbonyl, sulfonyl, etc. A factor is applied to the total ring substituents' activation when such a case is encountered. When the polarized multiple group is a sulfone or sulfoxide, a factor of 3 is used. Similarly, when the polarized multiple group is a carbonyl, a factor of 2 is appropriate. An illustration is given in the section below on sample pK_a predictions.

The last class of compounds that require special attention are ester-like systems. Polarized, multiply bonded groups directly followed by a heteroatom bearing a lone pair of electrons typically deactivate the polarized group unless the heteroatom is in turn followed by a multiple bond, as illustrated below. The deactivating effect of



exptl[9b,f]	19.8	22.7	18.7
calcd	21	23	18

calculations:
45 - [1(1 x 18 + 1 x 6) + ¼(1 x -0.3)] = 21
45 - [1(1 x 18 + 1 x 6) + ¼(1 x -8) + ⅛(1 x -0.3)] = 23
45 - [1(1 x 18 + 1 x 6) + ¼(1 x 8) + ⅛(1 x 6)] = 18

oxygen in the former situation is handled by assigning a base deactivation of 8 pK_a units, while for nitrogen it is 16 units. These base values are then leveled to net deactivations of 2 and 4 units when distance is considered in eq 3. Conversely, the presence of a multiply bonded group immediately following oxygen makes oxygen activating by a base value of 8 units. The effect is less pronounced for nitrogen and hence nitrogen remains deactivating by 8. There is limited data for sulfur, selenium, and phosphorus. At the present time, they are approximated to behave similarly to oxygen analogues.

3. Treatment of Unsaturated Acidic Sites. Acidic sites in which the proton is bonded to an unsaturated atom are handled separately from the main portion of the program (Figure 1). The acidities of many of these weak acids have not yet been experimentally determined. Hence, only a cursory treatment of these compounds can be provided at this time. For acidic sites in double bonds, only α activation is considered. The substituent activation values for vinylic sites are applied at 30% the activation shown in Table II in order to reproduce the relative acidity levels used in the program's old perception algorithm.[1] Aromatic sites are handled separately. For these sites, an activation of 6 is applied for each α heteroatom. This treatment

exptl[18]	37.1	28.8
calcd	37	31

calculations:
43 - [1(1 x 6)] = 37
43 - [1(1 x 6 + 1 x 6)] = 31

allows reasonable pK_a prediction for heteroaromatics such as furans, thiophenes, triazoles, and other related heterocycles. The acidities for thiophene and thiazole illustrate this rule.

C. Difficult Cases and Limitations. The nature of some compounds presents difficulties for predicting acidity in a generalized fashion. Nitrocyclopropane, for instance, is conisderably less acidic than nitromethane; the observed pK_a's are 27.0 and 17.2, respectively.[19] However, the acidity of nitrocyclobutane is 17.8.[19] Furthermore, related systems such as cyclopropyl phenyl ketone and acetophenone display a much diminished effect. The experimentally determined pK_a's for these ketones are 28.2 and 24.6, respectively. The anomalous acidity of nitrocyclopropane has been attributed to a different mechanism of deprotonation.[19] Thus, nitrocyclopropane can be treated as a special case or considerable error can be accepted by using a general algorithm.

Prediction for some aromatic systems is also problematic. Cyclopentadiene and indene show significant deactivation with alkylation. For example, the experimental pK_a of indene is 20.1, whereas the pK_a of 3-methylindene is 22.5.[20] This deactivation is unusual for the addition of only one methyl group. For example, benzenoid systems show very little change in activation with alkylation; the pK_a's for phenol and *o*-methylphenol are both 18.0.[9f] Heteroaromatic ring systems are also difficult to handle due to the unique activating abilities of specific heteroatoms in different sized rings. Generalization is elusive for such systems, so accurate prediction requires that they be handled explicitly.

Finally, as noted above, acidity predictions for compounds that are less acidic than Me_2SO present difficulties since there are few experimental data available for these

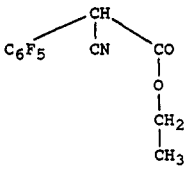
(18) These acidities were measured in cyclohexylamine and corrected to the Me_2SO scale by a statistical relationship correlating acidities in cyclohexylamine and Me_2SO: Buncel, E.; Durst, T. *Comprehensive Carbanion Chemistry*; Elsevier Scientific Publishing Company: Amsterdam, 1980; pp 355, 363.

(19) Bordwell, F. G.; Bartmess, J. E.; Hautala, J. A. *J. Org. Chem.* **1978**, *43*, 3113.

(20) Bordwell, F. G.; Bausch, M. J. *J. Am. Chem. Soc.* **1983**, *105*, 6188.

1. Representation

Acidic Atom
Branching Level 1
Branching Level 2
Branching Level 3



2. Calculation

pK_a^i = 45

Aromatic Activation:

Ring	=	6
Ortho = 1 + 1	=	2
Meta = 2 + 2	=	4
Para	=	0
Total Activation	=	12

Branching Level Calculations:

Level 1 = 1[1x18) + (1x17) + (0.67x12)] = 43.0

Level 2 = 0.25[1 x (-8)] = -2.0

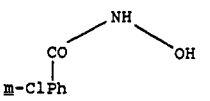
Level 3 = 0.11[1 x (-0.3)] = 0.0

pK_a = 45 - [43.0 - 2.0 + 0.0] = 4

Figure 3. Sample pK_a calculation for ethyl α-cyano-α-pentafluorophenyl acetate.

1. Representation

Acidic Atom
Branching Level 1
Branching Level 2



2. Calculation

pK_a^i = 44 - 11 = 33

Aromatic Activation:

Ring	=	6
Meta = 2x2	=	4
Total Activation	=	10

Branching Level Activations:

Level 1 = 1[(1x18) + (1x1)] = 19

Level 2 = 0.25 (1x10) = 2.5

pK_a = 33-[19 + 2.5] = 12

Figure 4. Sample pK_a calculation for ethyl *m*-chlorobenzohydroxamic acid.

compounds. For the purposes of the program, the acidities of these compounds have been extrapolated from existing aqueous acidity data.[1,2] Since the present reference value for saturated carbon is lower than reported acidities for alkanes in Me_2SO,[21,22] a factor is applied to expand the acidity scale for compounds less acidic than Me_2SO. This expansion is effected as follows: if the pK_a of a compound is calculated to be greater than 36, the difference from $pK_a°$ is determined and multiplied by 2.5. This amount is then added to 36 to give a rough pK_a for the compound. While the acidities predicted in this range may not be quantitatively accurate, the qualitative results should be reasonable. When more data become available, the algorithm can be improved.

D. Sample pK_a Predictions. Two examples are presented here which highlight key aspects of the algorithm. A full presentation on experimental vs. predicted acidities is provided in a subsequent section. It should be noted that the program rounds off the calculated acidities to the nearest integer.

The pK_a prediction for the highly functionalized ethyl α-cyano-α-pentafluorophenyl acetate covers many principles discussed so far. Figure 3 gives a representation of the molecule as perceived by the branching algorithm in addition to a calculation based on eq 3. The parent is designated saturated carbon. The groups that directly activate the acidic site are then determined and placed in the first branching level. The substituent activations for the carbonyl and nitrile groups are taken directly from Table II, whereas the pentafluorophenyl unit must be handled separately. The calculated substituent activation for the pentafluorophenyl unit is a sum of the base ring activation and ring substituent activations which are also taken from Table II. Note, even though OPACT for fluorine is 0 the ortho activation for each fluorine is 1, recalling that ortho groups are more activating than para groups (Table V). These substituents are now ordered from high to low according to their activations and the corresponding leveling values for nitrile systems are applied as given in Table I. Since these substituents are in the first branching level, no leveling factor is applied for distance. The second branching level only contains the ester oxygen. Its base deactivation of 8 is leveled to 2 by the distance factor. The algorithm processes the third branching level in a similar fashion and terminates branching. Finally, the activations determined for the three branching levels are subtracted from the parent acidity to give the predicted pK_a of 4. This is in close agreement with the experimental value of 5.1.[9f]

The pK_a prediction for *m*-chlorobenzohydroxamic acid in Figure 4 illustrates the treatment of hetero acids. The parent is recognized initially as an amide and assigned the $pK_a°$ given in Table III. Further evaluation realizes the enhanced activation due to the α oxygen; the parent acidity is adjusted further according to the value given in Table IV. Substituent activations for the carbonyl and oxygen are then determined, allowing calculation for the first branching level. The aromatic group found in the second branching level is once again treated as a single substituent. However, since it directly follows a carbonyl, a factor of 2 is applied to the ring substituent activations to account for their enhanced activation in such systems (Table V). The second branching level activation is calculated and the predicted acidity determined. The computed pK_a of 12 compares well with the experimental value of 12.7.[9f]

E. Utilization in CAMEO. The ability to predict relative acidities is vital to the CAMEO program. The knowledge of pK_a values is used to determine the feasibility of proton transfer as well as to help gauge the reactivity of nucleophiles and leaving groups in nucleophilic reactions.[1,2] A pK_a window is imposed, specifying the range for proton transfer. In general, the initial base generates the weakest base and any others within 4 pK_a units. This limit can be

(21) Bordwell, F. G.; Algrim, D.; Bares, J. E.; Branca, J. C. *J. Org. Chem.* **1978**, *43*, 5024.
(22) Breslow, R.; Grant, J. L. *J. Am. Chem. Soc.* **1977**, *99*, 7745.
(23) McEwen, W. K. *J. Am. Chem. Soc.* **1936**, *58*, 1124.
(24) Pearson, R. G.; Dillon, R. L. *J. Am. Chem. Soc.* **1953**, *75*, 2439.

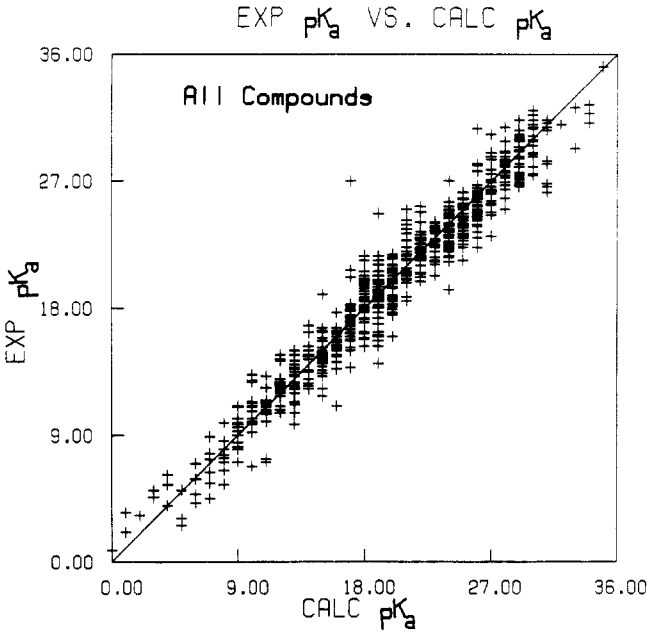


Figure 5. Experimental pK_a vs. calculated pK_a for all compounds addressed in the Me_2SO database (n = 763, σ = 1.49 and r = 0.979).

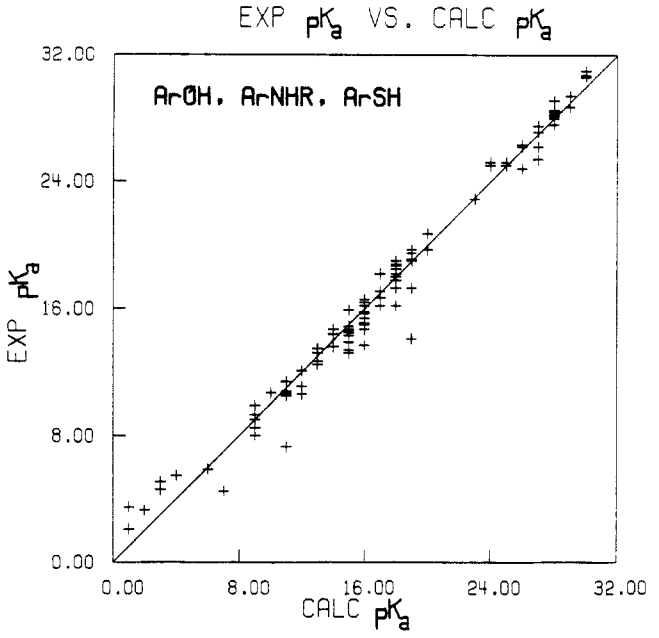


Figure 6. Experimental pK_a vs. calculated pK_a for phenols, anilines, benzenethiols, and related compounds (n = 103, σ = 1.12, and r = 0.981).

expanded to 10 pK_a units interactively by the user if it is deemed desirable.

The basicity of organolithium reagents requires special attention. Increasing alkyl substitution increases basicity. Thus, *t*-BuLi is a stronger base than *sec*-BuLi which is stronger than *n*-BuLi. This reactivity is handled as follows: tertiary, secondary, and primary alkyllithium reagents are allowed to deprotonate acids with pK_a's up to 54, 48, and 42, respectively. Grignard reagents behave as weaker bases than the corresponding organolithium compounds and are therefore only allowed to deprotonate compounds that have a pK_a of 42 or below.

IV. Discussion of Results

Predictions have been made using this approach for over 700 compounds with experimentally known acidities.

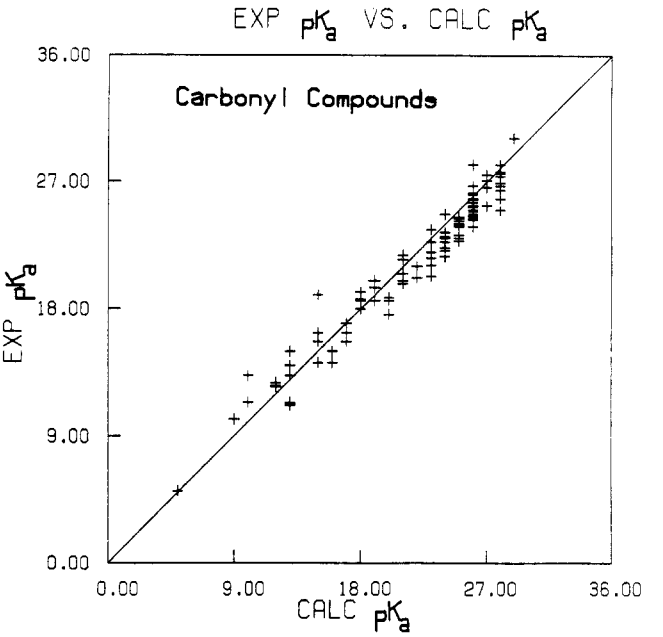


Figure 7. Experimental pK_a vs. calculated pK_a for ketones, esters, and related compounds (n = 97, σ = 1.35, and r = 0.975).

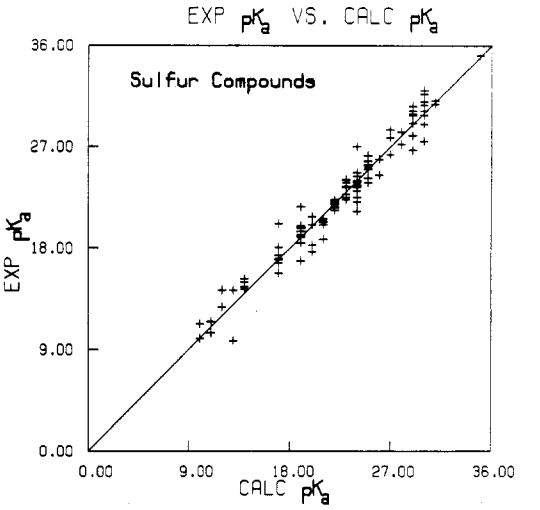


Figure 8. Experimental pK_a vs. calculated pK_a for sulfones, sulfoxides, sulfides, sulfonates, thiosulfonates, and related compounds (n = 115, σ = 1.13, and r = 0.977).

Figure 5 is a plot of the experimental vs. calculated pK_a's for all of the compounds. The standard deviation is 1.5 pK_a units. Figures 6–12 are plots showing the correlation for specific classes of acids.

A. Correlation for Different Classes of Compounds. The correlation for substituted phenols, anilines, and benzenethiols is quite good (Figure 6). Most of the acidity predictions are accurate to within 1 pK_a unit. However, the acidity tends to be overestimated when there are two or more strongly electron-withdrawing groups on the ring. Ketones, esters, and related compounds give similarly good correlations (Figure 7). Some of the overestimated pK_a's are for cyclic ketones, e.g., cyclopentanone, cyclohexanone, etc. An adjustment has not been made for the fact that these molecules are more acidic than their acylic counterparts.

There are no obvious trends for the sulfur compounds in Figure 8. Generally, the program does well for most sulfones, sulfoxides, sulfides, sulfonates, thiosulfonates, and related compounds. In Figure 9, the one particularly bad point is for nitrocyclopropane which was discussed above. Since other related compounds do not display such a

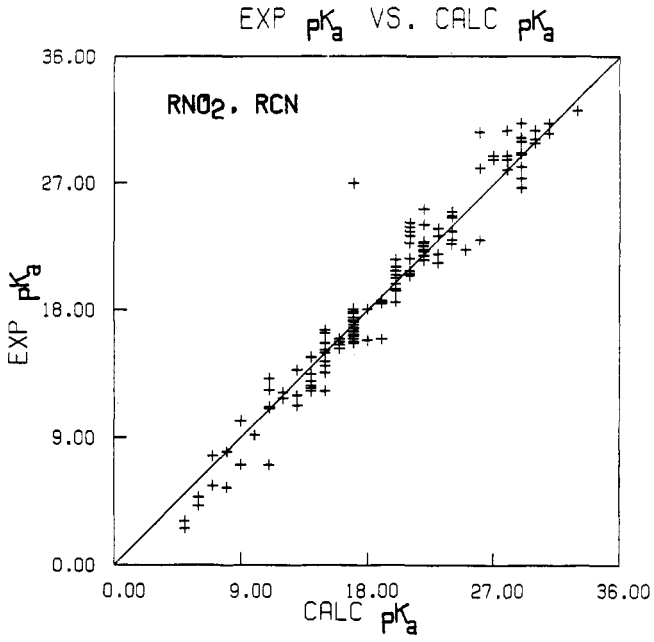


Figure 9. Experimental pK_a vs. calculated pK_a for aliphatic nitro, cyano, and related compounds (n = 130, σ = 1.64, and r = 0.972).

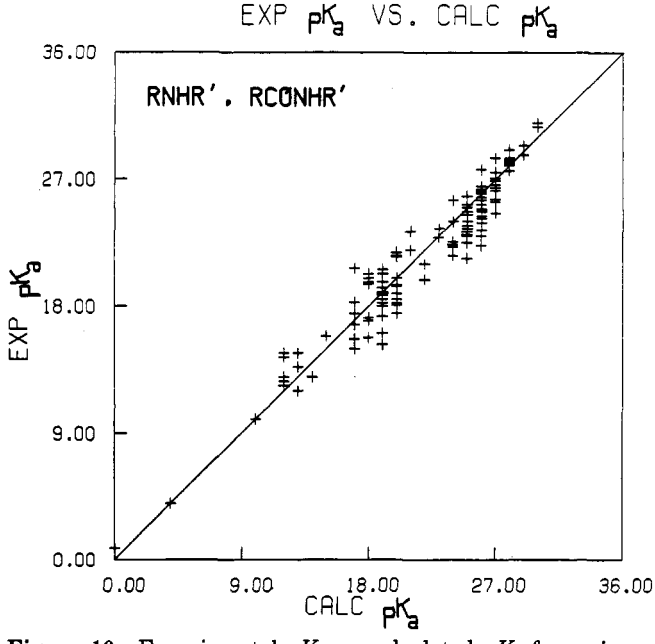


Figure 10. Experimental pK_a vs. calculated pK_a for amines, amides, and related compounds (n = 130, σ = 1.48, and r = 0.962).

marked change in acidity, no special accommodation is currently made in this case. Triactivated compounds which are activated by two nitrile substituents also tend to deviate. These are more acidic than what the algorithm predicts even when the third substituent is bulky such as a phenyl ring. Results of similar quality are obtained for amines, amides, and related compounds as shown in Figure 10.

The correlation for the unsaturated compounds in Figure 11 is less accurate. This reflects the unique activating ability of the heteroatoms in various sized heteroaromatic rings as mentioned above. The two largest errors are for *N*-methyl-2-pyrroleacetonitrile and tetraphenylmethylenediphosphine disulfide. Finally, the results for polycyclic aromatics in Figure 12 are reasonable. However, 9-(alkylthio)fluorenes are predicted to be significantly less acidic (3–5 pK_a units) than observed. All other substituted fluorenes give consistently accurate results.

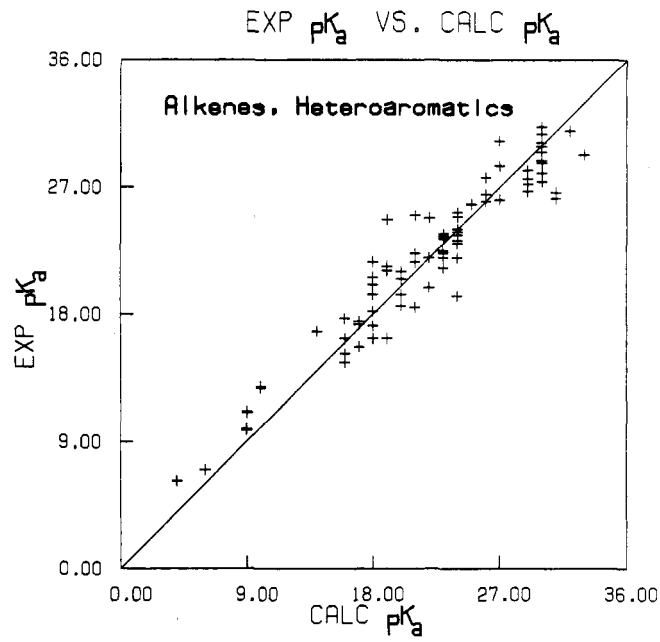


Figure 11. Experimental pK_a vs. calculated pK_a for aromatic heterocycles, alkenes, and related compounds (n = 83, σ = 1.93, and r = 0.956).

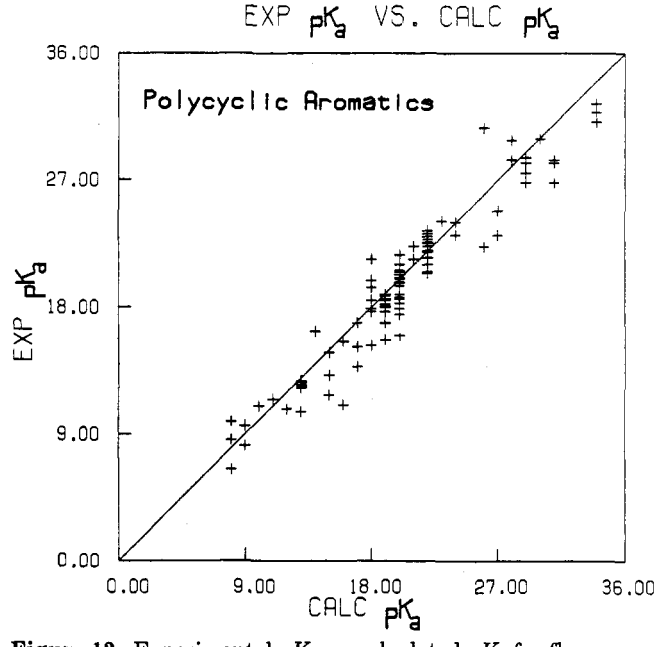


Figure 12. Experimental pK_a vs. calculated pK_a for fluorenes, anthracenes, xanthenes, and related compounds (n = 105, σ = 1.68, and r = 0.959).

B. Me_2SO vs. H_2O. There exist certain general relationships between Me_2SO and water acidities that are noteworthy.[9a] For many compounds that yield moderately charge-delocalized anions, e.g., carbonyl compounds and nitroalkanes, the pK_a in water is often about 7 units less than that in Me_2SO. However, due to its non-hydrogen-bond-donor nature, Me_2SO is less capable of stabilizing charge localized anions than water. As a result compounds such as alcohols are much more acidic in water than in Me_2SO. Similarly, compounds that yield highly charge

	CH_3CH_2OH	$CH_3COCH_2COCH_3$
$pK_a(Me_2SO)$	29.8[9e]	13.3[9d]
$pK_a(H_2O)$	18[23]	9[24]

delocalized anions that are less prone to participation in hydrogen bonding, e.g., acetylacetone, are relatively more

acidic in Me_2SO. Although these notions do not give quantitative acidity values, they may be of some use to the organic chemist.

V. Conclusion

In summary, a new algorithm for pK_a prediction has been implemented in the computer synthesis program CAMEO. Key organizing principles have been developed to account for substituent effects on acidity. The algorithm is general and capable of predicting pK_a's for a vast number of organic compounds with an average error of 1.5 pK_a units.

Acknowledgment. This work was supported by a grant from the National Science Foundation. We are also grateful to Professor Frederick G. Bordwell and his coworkers for helpful comments and for providing the experimental database that made this study possible.

Registry No. Me_2SO, 67-68-5.

Selective Fluorination of Steroids Using Elemental Fluorine

Shlomo Rozen* and Giora Ben-Shushan

School of Chemistry, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel

Received December 31, 1985

Elemental fluorine, when diluted with nitrogen, replaces tertiary deactivated hydrogens via the extremely rare electrophilic substitution on a saturated sp^3 carbon. This work describes such substitutions on steroids including the cases of the less common 3α-ol and 5β series. Depending on the location of the deactivating oxygenated functions in the steroid one can direct the substitution on any of the tertiary 5, 9, 14, 17, or 25 positions. A full retention of configuration is observed, in accordance with the proposed substitution mechanism. One of the major differences between the 3α- and the 3β-sterols is the fact that in the latter series we have never witnessed fluorination at the 5 position, while in the 3α-ol derivatives the formation of the 5C–F bond is quite common. A second difference between the 5α- and the 5β-steroids concerns the fluorination of the electronically favorable 9 position. In the 5β series the A/B cis arrangement sterically prevents the approach to this position so the 9C–H bond remains untouched. Some leads are also presented as to the question whether or not this reaction can be carried out efficiently at higher than the usual −75 °C. CNDO calculations, coupled with the MM1 program, offer good criteria as to which single tertiary site of the few present will be substituted by the fluorine atom.

It is quite likely that no other single chemical group of compounds has captured the chemist's attention so intensely as the steroids. Despite this, however, only extremely few works have been devoted to the study of activating sites on the four-ring skeleton, when these are remote from any conventional functional group such as a double bond, a carbonyl, or a heteroatom. Among these works one should mention the outstanding Breslow's "remote control" functionalization[1] and Mazur's radical chlorination[2] and silica supported ozone oxidation.[3] Most of these examples, and a few other attempts, are of a radical nature and have certain limitations. To perform reactions on deactivated sites in an ionic mode, a simple and yet unique reagent has been lacking.

In the past, one of us participated in the only published work showing that elemental fluorine can substitute a deactivated tertiary hydrogen in some common steroids.[4] We describe now the unusual reaction of F_2 with some less common 5β- and 3α-sterols which may help clarify the role of electronic as against stereochemical factors, concentrations, temperature, and solvent influences on the reaction course.[5] This approach can also be of biological importance since many selectively fluorinated compounds, especially among steroids, have potential biological activity either in pharmacology.[6] or in radiodiagnostic studies using ^{18}F derivatives with positron emitting tomography (PETT).[7]

As a part of our program in studying the unique substitution reaction at saturated centers, we have conducted research also on substrates other than steroids, showing that the mechanism of this substitution is of an ionic nature.[8] The reactive species is the electrophilic part of the F_2 dipole, which attacks the electrons of the corresponding C–H bond leading to a substitution with a full retention of the configuration.[8] The medium in which the reaction is performed is of crucial role. If a nonpolar solvent such as $CFCl_3$ is used, no encouragement for the F_2 polarization exists and the main course of the reaction is found to be an indiscriminate radical one. The same is true with hydrogen containing nonpolar solvents such as hexane or pentane. When, however, $CHCl_3$ is used as a cosolvent with $CFCl_3$[9] it serves as a radical scavenger and probably more important, it increases the medium polarity and acts

(1) Breslow, R. *Acc. Chem. Res.* **1980**, *13*, 170.
(2) Cohen, Z.; Mazur, Y. *Angew. Chem., Int. Ed. Engl.* **1978**, *17*, 281.
(3) Cohen, Z.; Mazur, Y. *J. Org. Chem.* **1979**, *44*, 2318.
(4) Alker, D.; Barton, D. H. R.; Hesse, R. H.; James, J. L.; Markwell, R. E.; Pechet, M. M.; Rozen, S.; Takeshita, T.; Toh, H. T. *Nouv. J. Chim.* **1980**, *4*, 239.
(5) For a preliminary communication on the fluorination of bile acids, see: Rozen, S.; Ben-Shushan, G. *Tetrahedron Lett.* **1984**, *25*, 1947.
(6) Filler, R.; Kobayashi, Y., Ed. *Biomedicinal Aspects of Fluorine Chemistry*; Elsevier Biomedical Press: Amsterdam, 1982.
(7) See for example: Dagani, R. *Chem. Eng. News* **1981**, *59(13)*, 30.
(8) See, for example: (a) Rozen, S.; Gal, C. *J. Fluorine Chem.* **1985**, *27*, 143. (b) Gal, C.; Rozen, S. *Tetrahedron Lett.* **1984**, *25*, 449. (c) Gal, C.; Rozen, S. *Tetrahedron Lett.* **1985**, *26*, 2793. (d) Rozen, S.; Gal, C.; Faust, Y. *J. Am. Chem. Soc.* **1980**, *102*, 6860.
(9) The $CHCl_3$ which usually serves as a cosolvent with $CFCl_3$ can be replaced with other polar solvents which have somewhat acidic hydrogen such as CH_3NO_2 or AcOH, with little change in the reaction outcome.

0022-3263/86/1951-3522$01.50/0