To the Graduate Council:

I am submitting herewith a dissertation written by Daniel Nathan Sorensen entitled "Three Studies in Ion Thermodynamics." I have examined the final copy of this dissertation for form and content and recommend that it will be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Chemistry.

Dr. John E. Bartmess, Major Professor

We have read this dissertation
and recommend its acceptance:

Accepted for the Council:

Associate Vice Chancellor
and Dean of the Graduate School

THREE STUDIES IN ION THERMODYNAMICS

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Daniel Nathan Sorensen

August 1994

DEDICATION

This dissertation is dedicated to my wife

Charlene Carpenter Sorensen

and to our son

Andrew Cornelius Sorensen

for their encouragement while completing my work.

ABSTRACT

This work is actually three different projects, each addressing a problem in ion thermodynamics. The first project addresses equilibrium affinity scales. These scales contain many Hess cycles which may not be numerically close due to experimental uncertainties. Experimenters needed an unbiased estimator to evaluate the affinity of one compound with respect to another. To rectify this problem, the programs LADDER and GRAPHLDR were developed. These programs take a set of data, analyze it under three different algorithms to ascertain the best fit, and then graph the scale. Partitioning of the data into subsets simplifies the analysis. Traditional least squares anlysis is considered the best unbiased estimate toward fitting the data. The iterative smoothing algorithm tends to reproduce the least squares solution. The least trimmed squares algorithm locates possible suspect data within the fusion set. The least trimmed squares algorithm works best for subsets of four to ten compounds.

The second project evaluated sulfonates and phenyl-dimethylsilyl enol ethers for use as precursors undergoing dissociative attachment in ion-molecule chemistry. Two silyl enol ethers were synthesized from isobutyraldehyde and cyclohexanone. Each compound was verified by its NMR spectra, but complete separation from solution was not

feasible. It was not possible to verify whether the observed enolate ions were from dissociative attachment or whether they were from secondary ion processes. Dimethyl sulfate yielded the methyl sulfate anion. Methyl methanethiosulfonate cleaved at the sulfur-sulfur bond to produce $MeSO_2^-$. The gas phase basicity for $MeSO_2^-$ was bracketed between $ClCH_2CO_2^-$ and $CF_3CO_2^-$.

The third project involved traditional solution calorimetry on selected anilines. Neutral solvation enthalpies were obtained from the heat of reaction of the neutral anilines into aqueous acid. The solvation enthalpies for five anilinium ions were calculated. The electrostatic term of the solvation enthalpy is highly correlated with the proton affinity. The effect of the electrostatic term toward solvation was greater in ring-substituted compounds than in successive alkylations of the nitrogen.

ACKNOWLEDGEMENTS

I would like to thank Charlene and Andrew Sorensen for their patience and encouragement while completing this work. I would also like to thank my major professor, Dr. John E. Bartmess, for his expertise, guidance, and patience. I would like to thank Dr. Terry Sumpter, Catherine Crowder, and Dr. Deborah Thomas Grimm for their invaluable assistance and comradery during the first few years that I was in Knoxville. I also want to thank Edwin W. Conover and Dr. Nathan Stump for their friendship during those first few years of classes. The families of Dr. Paul Kelley and Mr. Stan White have also been very generous in their hospitality this past year.

I would also like to thank Dr. John Bloor for allowing me to use his computer accounts to run the *ab initio* calculations. Special thanks also goes to Jack Dongarra, J.P. Chandler, Charles Lawson, and Richard Hanson for their permission to use their software.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

THE MATHEMATICS OF LADDER ALGORITHMS IN THE DETERMINATION OF ABSOLUTE THERMODYNAMIC AFFINITIES FROM EQUILIBRIUM VALUES

I. INTRODUCTION

The impetus for the study in this chapter arises from conceptually simple reactions for which the thermochemistry is empirically difficult or impossible to measure. These thermochemical measurements include gas phase acidities (see reaction 1.1),[1-5] gas phase basicities (reaction 1.2),[6-13] electron affinities (reaction 1.3),[14-17] ionization energies (reaction 1.4),[18] cation affinities (reaction 1.5),[19-27] anion affinities (reaction 1.6),[28-36] and their many solution phase analogues.[37,38]

(1.1) $HA \rightleftharpoons H^+ + A^-$

(1.2) $HA^+ \rightleftharpoons A + H^+$

(1.3) $A^- \rightleftharpoons A + e-$

(1.4) $A \rightleftharpoons A^+ + e^-$

(1.5) $MA^+ \rightleftharpoons A + M^+$

(1.6) $AX^- \rightleftharpoons A + X^-$

Each equation represents either a simple addition or simple dissociation. In each reaction, the value of the thermodynamic affinity for the reaction is a state function such as ΔG or ΔH. Most of the reactions (1.1)-(1.6) have been investigated in aqueous solution. For example, the aqueous acidity of HA is easily measured if the compound has an acidity between that of H_3O^+ and H_2O. The acidity of weak acid HA can be determined in water by titration with a strong base and a pH meter if the pK_a is less than 14.[39] In nonaqueous solutions, the acidity of HA is more difficult to measure and does not necessarily agree with the aqueous value.

The attractive forces surrounding both the neutral molecule and the dissociated ions are solvent dependent. Since the solution acidities are solvent dependent, it is important to separate the intrinsic acidity of the molecules from the solvent. Conceptually, the intrinsic acidity can be measured in the gas phase where the particles are much more isolated. However, the gas phase acidity (reaction 1.1) is highly endothermic and cannot be measured directly with existing instrumentation. Reactions (1.2) through (1.6) are also highly endothermic and most of these processes cannot be thermodynamically measured using the current mass spectrometric technology. Therefore, although the thermodynamic affinities of equations (1.1) - (1.6) are of fundamental importance, the affinities for most of the

thermodynamic processes, as written, cannot be reliably measured.

The thermochemistry of the reactions for electron affinity (1.3) and ionization energy (1.4) can be measured as written via a variety of spectroscopic techniques,[40,41] but often equilibrium measurements prove to be more reliable than direct determination. Electron photodetachment studies are the most reliable measure of electron affinities for small species, but for polyatomic asymmetric species the spectra are very difficult to interpret. High resolution photoelectron spectroscopy results in accurate vertical ionization energies. However, the adiabatic ionization energy is often lower than the vertical ionization energy. For both electron affinities and ionization energies, equilibrium measurements are preferred over direct measurements if the predicted structures of the ionic and neutral species are dissimilar.

In a strict sense, the reactions (1.1) to (1.6) all share two common characteristic: (1) each reaction is a simple association or dissociation reaction, and (2) each reaction has an associated numerical value. In mathematical terms, these similarities allow one to use a single reaction type to discuss mathematical methodologies that lead to the mathematical treatment of the other five reactions. Therefore, this discussion will focus on reaction (1.1), the gas phase acidity.

Since the thermochemical equation written in (1.1) is too endothermic to be reliably measured, it becomes necessary to look at another reaction (i.e the proton transfer equilibrium) to obtain a affinity value for the acidity reaction (1.1). Although the acidity of HA as written in (1.1) is too endothermic to be measurable, the proton transfer equilibrium (1.7) is measurable if the acidities of HA and HB are nearly the same.

(1.7) $HA + B^- \rightleftharpoons A^- + HB \quad \Delta G_{eq} = 0.3$ kJ/mol

The term ΔG_{eq} defines the equilibrium ΔG for a transfer reaction such as (1.7). The assigned ΔG_{eq} is arbitrary for this example.

Equilibrium (1.8) describes acidity of HA relative to HC.

(1.8) $HA + C^- \rightleftharpoons A^- + HC \quad \Delta G_{eq} = 0.7$ kJ/mol

The relative acidities of HB and HC can then be determined via Hess's Law. If the absolute acidity of HA is known from other types of experiments or theory, the absolute acidities of HB and HC can be calculated from (1.7) and (1.8). The measurement of a third equilibrium (1.9) will close a simple Hess cycle.

(1.9) $HB + C^- \rightleftharpoons B^- + HC \quad \Delta G_{eq} = 0.2$ kJ/mol

Adding the thermodynamic affinities of equations (1.7) and (1.9) (shown here as ΔG_{eq}) should agree with the ΔG_{eq} of

equation (1.8). In practice, apparent discrepancies for the thermodynamic values in the Hess cycles are often observed. (Here ΔG_{eq} for equation (1.8) is 0.7 kJ/mol while the sum of (1.7) and (1.9) is 0.5 kJ/mol).

The assignment of the acidities of HB and HC relative to HA (based on the ΔG_{eq} values of the reactions) often becomes an arbitrary process. To give an example: the experimenter might assign ΔG values of 0.4 kJ/mol and 0.7 kJ/mol to HB and HC relative to HA based only on the Hess cycle. By simple inspection, another scientist could argue that a second set of ΔG values of acidities for HB and HC relative to HA are 0.3 kJ/mol and 0.6 kJ/mol, in an attempt to find a median in the deviation. Thus the same three equilibrium measurements can lead to two equally plausible assignments as shown by Table 1-1.

Which assignment is correct? The three equilibria could be remeasured until a better numerical agreement was

Table 1-1: Assignments of Relative ΔG values for Three Equilibria.

	ΔG(kJ/mol), experimenter 1	ΔG(kJ/mol), experimenter 2	ΔG(kJ/mol), real value
HA	0.0	0.0	0.0
HB	0.3	0.4	?
HC	0.6	0.7	?

reached within the Hess law cycle. Re-investigation of larger Hess cycles in order to identify assignment errors of the relative ΔG's for individual compounds becomes increasingly difficult as the number of equilibria grows. For Hess law cycles involving more than about ten compounds, trying to find a single measurement by hand is difficult. For Hess cycles involving ten or more compounds, even the initial assignment of the relative ΔG values becomes tedious when done by hand. A consistent, unbiased estimation of the absolute affinity values is needed. This chapter discusses the development of programs LADDER and GRAPHLDR, which were developed to meet these challenges.

These programs were designed to read the equilibrium data, assign the best set of relative affinity values for the compounds (relative to one compound) according to a least squares fit, point out possible errors in the equilibrium measurements, and graph the data for a final visual inspection.

A ladder is defined as the connectivity graph of multiple, overlapping equilibrium affinity measurements. A scale is a ladder with the thermodynamic affinity assigned for each of its members. A solution vector is the set of thermodynamic affinity values resulting from one of the various algorithms used by the program. The assigned affinity values in a scale are all relative to a value for a single member within the scale, which is the anchor member.

An absolute scale is a scale where the absolute thermodynamic affinity is known for the anchor member, whether by experiment or by theory.

Many of these scales already appear in the literature.[1-36] The impetus for the development of programs LADDER and GRAPHLDR arose from the discrepancies in the reported scales. For example, the calculated difference in the proton affinities, ΔH_{PA} (from reaction (1.2)), between ammonia and water was found to be 33.0 kcal/mol by Wolf *et. al.*,[6] but only 31.8 kcal/mol by Yamdagni and Kebarle.[7] Experimenter bias in the assignment of the relative affinities needed to be resolved.

Several algorithms are employed in accomplishing the goals of the programs. The least squares algorithm originates from the LINPACK subroutines SQRDC and SQRSL.[42] An iterative smoothing algorithm is also employed to suggest values for equilibrium measurements which have not been made. The simplex algorithm from QCPE is used to suggest possible errors in data collection.[43] Several of the algorithms used by the program require a sorting of the data. The sorting algorithm used is an adaptation of the Heapsort algorithm.[44] Vector manipulations throughout the program utilize the Basic Linear Algorithms, BLAS.[45,46]

The first report of the program was at the 1989 Annual Conference of the American Society for Mass Spectrometry,[47]

and significant improvements were reported at the 1991 conference.[48]

II. MATRIX AND VECTOR NOTATION

The mathematics discussed in this section require some notational preliminaries. For the purposes of this chapter a vector is defined as a rectangular collection of real numbers aligned in either a row or a column. If there are n elements to the vector, the vector is said to be a real n-tuple, i.e, the vector defines a single point in n-space. Vectors are denoted by small letters in bold face type, for example **a**, **x**, **v**. Components of a row vector are subscripted, for example, the n^{th} component of **a** is denoted a_n. The transpose of the row vector **v** is given by column vector $\mathbf{v}^t$. Three vectors will be of particular importance. The zero vector, **0**, is a vector of zeroes. The **1** vector is an n-tuple of ones. The row identity vector is given by definition (1.10).

(1.10) Definition: row identity vector, $\mathbf{e}_n \equiv$ a real vector of size k where n represents the nonzero column of **e**.

$$\mathbf{e}_n = [\ 0,\ 0,\ 0\ \ldots\ \underset{\substack{\uparrow \\ n^{th}\ \text{position}}}{1},\ 0,\ 0\ \ldots\ ]$$

For the purposes of this work, a matrix is defined as a rectangular collection of real numbers in m rows and n columns. The previous definition can be shortened to a matrix of real (m * n). Matrices are denoted by capital letters in bold face type, for example, **A**, **M**, **U**. The matrices are denoted in bold face type with a subscript when the number of columns should be distinguished, for example $\mathbf{A}_3$, $\mathbf{M}_n$. The notation for parts of a matrix are given by several forms. The individual elements of **A** are denoted by the lower case with subscripts, for example, a_{11}, a_{12}, a_{1n}. In each case, the first subscript denotes the row while the second denotes the column. There will be a comma separating the row and column subscripts only if either exceeds nine, for example, a_{19} but $a_{12,9}$. Since the matrix **A** can be thought of as a collection of m rows, the rows of **A** are denoted by $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, etc. Likewise, **A** can be partitioned into n columns, denoted by $\mathbf{a}^1$, $\mathbf{a}^2$, $\mathbf{a}^3$ etc., where superscripting distinguishes columns of **A** from the subscripted rows. The transpose of matrix **A** is given by $\mathbf{A}^t$. The identity matrix, an (n * n) matrix with 1's on the diagonal and 0's off the diagonal is given by the matrix **I**. The inverse of the square matrix **A** is $\mathbf{A}^{-1}$.

III. DATA ENTRY

The pre-analysis features of the program are designed to simplify the analysis. While the discussion in this subsection is not strictly mathematical, it is relevant to the mathematical algorithms which follow. The pre-analysis consists of four parts (1) data entry, (2) data verification, (3) data partitioning and (4) sorting. Each of the four parts involve two different stages.

The initial data entry process is fairly simple. First, the program reads the data file for bibliographical information regarding the experimental conditions (methodology, temperature, etc.). The program's output file then repeats the bibliographic information so the user can distinguish between two or more separate sets of equilibrium data. Next, the program reads the control parameters which govern the analysis, controlling certain program parameters such as algorithmic constants and the amount of output that will be generated to the output files during the course of the program run. The control parameters are also repeated within the program's output file. In the final stage of data entry, the program reads each recorded equilibrium. Each record of an individual equilibrium consists of the two compound names of the equilibrium pair as character strings, an affinity value as a real number, and possibly a weighting option as a real value.

While processing the entire list of recorded equilibria, the data entry algorithm creates several internal lists which will be manipulated in turn by later algorithms. For example, whenever the algorithm encounters a new name, it is added to the list of possible compounds and is given a reference number corresponding to this list. The compounds making up each equilibrium pair are cross-referenced against the list of all possible compound names. The algorithm counts both the number of compounds in the potential ladder as well as the number of individual measurements. In order to comply with the Fortran fixed format constraints, the format of the data within each equilibrium record has a few constraints placed upon it. First, the format of the record is represented as two character strings of fifteen spaces followed by the affinity as a real value in f10.5 format. If weighting information is to be entered, then this real f10.5 formatted value follows the affinity. Second, each chemical name is uniquely defined by a character string: all other references to the same compound must have the identical character string. For example, compound "c" is always "c" and cannot be called "cd" or even "C" in one equilibrium. The order of the acids listed for a given equilibrium should be such that equilibria values are listed as all positive. Thus, an experimental value of e, c, -0.2 should be given as: c, e, 0.2 in the equilibria data file, since all of the

equilibria are listed as positive values. The equilibrium affinities are best accepted as linear relationships (ΔG_{eq}, ΔS_{eq}, ΔH_{eq}, etc.). Equilibrium pairs listed with K_{eq} will be transformed into their ΔG_{eq} values at 300 K. If the affinities are presented in the format of equilibrium constants, then the listing should be such that all K_{eq}'s are less than one so that when the K_{eq}'s are converted to ΔG_{eq}'s, all of the ΔG_{eq}'s are positive. Some of the algorithms are not completely robust: confusion results whenever negative ΔG_{eq}'s are encountered. Therefore, to simplify the ordering processes of these algorithms, the K_{eq}'s should all be listed as less than one and the ΔG_{eq}'s should be positive throughout the data.

The next algorithm assigns the weighting to the data if a nonuniform weighting scheme is desired for the least squares algorithm. Ordinarily a uniform weighting scheme is used by the program. The weighting scheme to be used can be entered as either individual weights or the program can assign one of six specific weighting schemes. The details for the weighting schemes will be discussed later in this chapter.

By re-examining equation (1.7), one will find that each equilibrium measurement within the data file can be defined in terms of a vector, b_i.

(1.7) $\quad HA + B^- \rightleftharpoons A^- + HB \qquad \Delta G_{eq}=b_i$

By the conventions of this chapter, b_i is assumed to be positive, and equation (1.7) is written for the endoergonic reaction. Therefore b_i is seen as the difference between HB and HA. This difference will be defined by (1.11), the difference vector.

(1.11) Definition: Difference vector ≡ a real vector of length n, $\mathbf{d}_n$, that has only two nonzero values, a real positive value w and its additive inverse, -w. The negative value usually has the lower index of the two nonzero values.

In terms of this discussion, the difference vector is of length n since that will be the total number of compounds to be considered by later algorithms. Definition (1.11) has been generalized to include the possibility of nonuniform weighting. Nonuniform weighting can either be entered into the data file or the experimenter can choose from a variety of weighting techniques. For purposes of the LADDER program, the weights, w_i, are always positive. In the absence of any weighting constraints, the difference vector is said to be normalized, and the nonzero elements of the difference vector are 1 and -1. When cross-referenced against the list of possible compounds, each of the equilibria measurements can thus be portrayed as a difference vector. For example, if the entire list of possible compounds was HA, HB, HC and HD, the difference vector representing the measurement between HA and HB as in

equation (1.7) would be given as -1, 1, 0, 0. If a weight of 0.9 were to be given to the HA ⇔ HB equilibrium, then the difference vector would be -0.9, 0.9, 0, 0.

If all of the possible difference vectors of the n compounds are collected into a matrix, the resulting matrix is the complete positional matrix, $\mathbf{C}_n$ given by definition (1.12).

(1.12) Definition: Complete positional matrix ≡ the real k * n matrix made up of individual rows $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3 \ldots \mathbf{c}_k\}$ where each of the k rows is a difference vector described by definition (1.11). For every column in **C**, there are exactly n-1 nonzero elements. Of the k rows, no row can be transformed into any other row by scalar multiplication. In addition no row or column of **C** contains only zeros.

From the definition (1.12), an example of three compounds is given by equation (1.13).

$$\mathbf{C}_3 = \begin{bmatrix} -w_1 & w_1 & 0 \\ -w_2 & 0 & w_2 \\ 0 & -w_3 & w_3 \end{bmatrix} \tag{1.13}$$

Similarly, the complete positional matrix for four compounds would be given by equation (1.14).

$$\mathbf{C}_4 = \begin{bmatrix} -w_1 & w_1 & 0 & 0 \\ -w_2 & 0 & w_2 & 0 \\ -w_3 & 0 & 0 & w_3 \\ 0 & -w_4 & w_4 & 0 \\ 0 & -w_5 & 0 & w_5 \\ 0 & 0 & -w_6 & w_6 \end{bmatrix} \tag{1.14}$$

By inspection of equations (1.13) and (1.14), it can be shown that the number of rows, k, in the complete positional matrix is given by equation (1.15).

(1.15) $$k = n! / 2!(n-2)! = \binom{n}{2}$$

Unfortunately, chemical problems often preclude the generation of the complete positional matrix. Such problems would include the investigation of transfer reactions between indistinguishable isomers, reactions of abnormally slow kinetics, reactions too endothermic to measure precisely, and the presence of reactions competitive with the desired transfer process. Of course, most often the experimenter simply has not completed the positional matrix. Instead of matrix **C**, the program must analyze **M**, the measurement positional matrix (1.16).

(1.16) Definition: Measurement positional matrix ≡ any real (m * n) matrix (m ≥ (n-1)) constructed of m row difference vectors of length n. None of the m row vectors can be transformed into another by only scalar multiplication. In addition, each of the rows and each of the columns of the measurement positional matrix has at least one nonzero element.

Thus the measurement positional matrix is relatively sparse, and the nonzero elements of the matrix **M** can be completely random. For example, 24 different measurement

positional matrices can be derived from (1.14). Two examples are shown in (1.17) and (1.18).

$$(1.17)\quad \mathbf{M} = \begin{bmatrix} -w_1 & w_1 & 0 & 0 \\ -w_3 & 0 & 0 & w_3 \\ 0 & -w_4 & w_4 & 0 \\ 0 & -w_5 & 0 & w_5 \end{bmatrix}$$

[elimination of rows 2 and 6 from **C** in (1.14)]

$$(1.18)\quad \mathbf{M} = \begin{bmatrix} -w_1 & w_1 & 0 & 0 \\ -w_2 & 0 & w_2 & 0 \\ 0 & -w_4 & w_4 & 0 \\ 0 & 0 & -w_6 & w_6 \end{bmatrix}$$

[elimination of rows 3 and 5 from **C** in (1.14)]

Both equations (1.17) and (1.18) are typical examples of matrices which could be found in the literature.

IV. DATA VERIFICATION

Once the initial data are read, the data input file is closed and the data verification process begins. The initial check is for duplicate data entry. If the equilibrium pair appears in the list twice (for example the experimenter types in two entries for the reaction between methanol and ethanol), the affinity values are compared. If the two measurements are the same, the second appearance is deleted from the data set. If the values differ, the program will abort with an error message telling the user where within the data file the difference lies. This verification step is important because the optimization algorithms are designed to accommodate only one affinity

value per equilibrium pair. For example, in equation (1.19), the last row is a scalar multiple of the previous row.

$$(1.19) \quad \mathbf{M} = \begin{bmatrix} -w_1 & w_1 & 0 & 0 \\ -w_2 & 0 & w_2 & 0 \\ 0 & -w_4 & w_4 & 0 \\ 0 & 0 & -w_6 & w_6 \\ 0 & 0 & -w_a & w_a \end{bmatrix} \quad \begin{array}{l} \text{[addition of a row to} \\ \quad \mathbf{C} \text{ in(1.18)]} \end{array}$$

Once the program is aborted, the experimenter has to decide which of the two measurements to keep. If both measurements appear equally valid to the researcher, the first occurring affinity value should be replaced with the mean of the two affinity values. The second recorded entry of this equilibrium pair should then be deleted from the data file. If the experimenter wishes, the standard deviation of the two experiments may be placed in the data file under the column of weights.

The fourth operation is a check for data continuity. When a real scale is first developed, the experimenters may have trouble ordering the initial scale. In the 1970's, many ladders were just beginning to be developed. Experimenters would often obtain measurements for the compounds in question with respect to one of several reference compounds. These references would have an affinity difference larger than the experimenter could measure precisely. For example, the practical measurement limit for room temperature Ion Cyclotron Resonance (ICR) mass

spectrometry is 3-4 kcal/mol.[3] For High Pressure Mass Spectrometry (HPMS), the practical measurement limit is slightly larger, around 10 kcal/mol due to a larger dynamic range for ion determination.[15,18] The experimenter would then have many equilibria measured to a single reference compound before finding one compound which would tie two reference compounds together. The program can now verify whether all of the compounds are related to one another by at least one pathway through the recorded equilibria. The program checks to see if the measurement positional matrix **M** is intraconnected according to definition (1.20).

(1.20) Definition: Intraconnected ≡ The (m*n) measurement positional matrix **M** is intraconnected if for all nonzero elements m_{ij} of **M**, two conditions exist.
(1) In either the i^{th} column or j^{th} column of **M**, at least one other nonzero element exists. Pairs of such elements between columns are called connection paths.
(2) Every nonzero element of **M** is related to every other nonzero element by one or more connection paths.

If the ladder given by **M** is not intraconnected, the Hess Law cycle cannot be calculated for all of the compounds. Often **M** is not intraconnected due to a single isolated measurement between two compounds which are not otherwise part of the ladder as shown by the expanded measurement positional matrix in (1.21).

$$(1.21) \qquad \mathbf{M} = \begin{bmatrix} -w_1 & w_1 & 0 & 0 & 0 & 0 \\ -w_2 & 0 & 0 & w_2 & 0 & 0 \\ 0 & -w_3 & 0 & w_3 & 0 & 0 \\ 0 & 0 & -w_4 & 0 & w_4 & 0 \\ 0 & 0 & 0 & -w_5 & 0 & w_5 \\ 0 & -w_6 & 0 & 0 & 0 & w_6 \end{bmatrix}$$

Although difficult to visually recognize at first, the fourth row is completely isolated from the rest of **M**. One of the benefits of using the program is the comparatively rapid identification of such an isolated measurement, particularly if the number of recorded equilibria exceeds twenty.

Another problem could arise when a smaller ladder is isolated from the larger ladder. The program checks for intraconnectedness, and if **M** is not intraconnected, the program exits after listing which acids are connected in each of the two (or more) ladders. For example, equation (1.22) shows two disjointed ladders.

$$(1.22) \qquad \mathbf{M} = \begin{bmatrix} -w_1 & w_1 & 0 & 0 & 0 & 0 \\ -w_2 & 0 & 0 & w_2 & 0 & 0 \\ 0 & -w_3 & 0 & w_3 & 0 & 0 \\ 0 & 0 & -w_4 & 0 & w_4 & 0 \\ 0 & 0 & -w_5 & 0 & 0 & w_5 \\ 0 & 0 & 0 & 0 & -w_6 & w_6 \end{bmatrix}$$

One ladder consists of columns 1,2 and 4 while the other consists of columns 3,5 and 6.

Using the program, scales such as the phenol acidity scale[1] could easily be intraconnected before publication.

The phenol acidity scale of McMahon and Kebarle actually contains two isolated scales: a portion of higher acidity anchored to CF_2HCO_2H and a portion of lower acidity anchored to propionic acid. The two separate scales could have been connected by as little as one measurement between catechol and p-chlorophenol. This measurement should have been performed. The researchers also investigated a single isolated acidity measurement between m-chlorophenol and formic acid. The isolated measurement could also have been easily connected to the rest of the scale by a single measurement. Any of the three monofluorophenols could have been measured against formic acid to tie the isolated measurement to the lower acidity portion of the scale.

V. DATA PARTITIONING

The first stage of data partitioning temporarily removes all pendant acids from the data set. A pendant acid is defined as one which is connected to the rest of the ladder by a single measurement. The single measurement connecting the pendant acid to the rest of the scale is called a pendant measurement. Because only one measurement is known to the pendant acid, it must be assumed to be correct and thus can do nothing for the optimization of the errors in the rest of the scale. The last row of equation

(1.18) is a pendant row since it is the row difference vector describing the pendant measurement.

$$(1.18) \quad \mathbf{M} = \begin{bmatrix} -w_1 & w_1 & 0 & 0 \\ -w_2 & 0 & w_2 & 0 \\ 0 & -w_4 & w_4 & 0 \\ 0 & 0 & -w_6 & w_6 \end{bmatrix} \quad \text{[elimination of rows 3 and 5 from } \mathbf{C} \text{ in (1.14)]}$$

When a column has only one nonzero value, the pendant row is one which contains the only non-zero value for that column. In this case, the fourth row is the pendant row since the fourth column only contains one nonzero value. Inclusion of the pendant measurement(s) will increase the computational time and may lead to errors in the accuracy of the final solution. Removal of the pendants is recursive: removal of pendant acids in one cycle may result in the isolation of pendant acids in the next cycle. Only after a cycle is completed in which no pendant acids are found is pendant recognition/removal portion of the program completed.

The second stage of the data partitioning breaks down the data into smaller subsets if possible. Partitioning is done for two reasons. The first consideration is computational time. The fastest of the algorithms in the analysis reaches a solution in a time proportional to the cube of the number of acids. Thus it is computationally more efficient to break a set of 50 acids into 2 subsets of 25 (if possible) to do the analysis. The second reason for partitioning the scales pertains to the detailed error

analysis. The algorithms can more accurately identify measurement errors in smaller subsets due to computational reasons.

The program partitions the subsets by analyzing the pairs of ordered data. For example, the expanded form of the measurement positional matrix in (1.23) can be broken into two smaller subsets shown by (1.24) and (1.25).

$$(1.23) \quad \mathbf{M} = \begin{bmatrix} -w_1 & w_1 & 0 & 0 & 0 & 0 \\ -w_2 & 0 & w_2 & 0 & 0 & 0 \\ 0 & -w_3 & w_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & -w_4 & w_4 & 0 \\ 0 & 0 & -w_5 & 0 & 0 & w_5 \\ 0 & 0 & 0 & 0 & -w_6 & w_6 \\ 0 & 0 & -w_7 & w_7 & 0 & 0 \end{bmatrix}$$

$$(1.24) \quad \mathbf{M_1} = \begin{bmatrix} -w_1 & w_1 & 0 \\ -w_2 & 0 & w_2 \\ 0 & -w_3 & w_3 \end{bmatrix}$$

$$(1.25) \quad \mathbf{M_2} = \begin{bmatrix} 0 & -w_4 & w_4 & 0 \\ -w_5 & 0 & 0 & w_5 \\ 0 & 0 & -w_6 & w_6 \\ -w_7 & w_7 & 0 & 0 \end{bmatrix}$$

The third column of matrix (1.23) is defined as a <u>fusion point</u>, and each of the matrices (1.24) and (1.25) is a <u>fusion set</u>. <u>Fusion points</u> are where the total measurement positional matrix can be separated into nonzero intraconnected submatrices. <u>Fusion sets</u> are the two subsets associated with the fusion point.

In terms of the actual ladder of compounds, a fusion point is the only compound where the two fusion sets are connected. The example of a slightly more complicated total measurement matrix (1.26) into the submatrices (1.27) and (1.28) was recently suggested.

$$(1.26) \qquad \mathbf{M'} = \begin{bmatrix} -w_1 & w_1 & 0 & 0 & 0 & 0 \\ -w_2 & 0 & 0 & 0 & w_2 & 0 \\ 0 & -w_3 & 0 & 0 & w_3 & 0 \\ 0 & 0 & -w_4 & w_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & -w_5 & w_5 \\ 0 & 0 & -w_6 & 0 & 0 & w_6 \\ 0 & 0 & 0 & -w_7 & w_7 & 0 \end{bmatrix}$$

$$(1.27) \qquad \mathbf{M'}_1 = \begin{bmatrix} -w_1 & w_1 & 0 \\ -w_2 & 0 & w_2 \\ 0 & -w_3 & w_3 \end{bmatrix}$$

$$(1.28) \qquad \mathbf{M'}_2 = \begin{bmatrix} -w_4 & w_4 & 0 & 0 \\ 0 & 0 & -w_5 & w_5 \\ -w_6 & 0 & 0 & w_6 \\ 0 & -w_7 & w_7 & 0 \end{bmatrix}$$

Matrix (1.26) is slightly more complicated than (1.23) in that the resulting submatrices are intertwined. Submatrix (1.27) is constructed from columns one, two and five from matrix (1.26), while (1.28) is constructed from columns three through six of the larger matrix. The second example of fusion set splitting given by (1.26) through (1.28) has been complicated by a <u>riser</u> in the total measurement positional matrix, **M'**. A <u>riser</u> is defined as a compound

which has only measurements leading away from it to compounds of higher index, that is, it has no direct path connecting it to compounds of lower index. In the case of (1.26), column three depicts a riser in the data.

The program LADDER in its current version cannot account for the splitting of the fusion sets as shown by matrices (1.26) through (1.28). Matrix (1.26) is therefore analyzed as a total measurement positional matrix. The decision to postpone the revision of the fusion set recognition algorithm only affects the size of the fusion sets that will be entered into the detailed analysis sections of the program. The only expected effects of this decision are to increase computational time and to decrease the efficiency of locating possible errors in the affinity measurement data.

VI. ORDERING SCHEMES

Three different ordering schemes take place in the pre-analysis. The original equilibrium data file may not have a particular order. Some logical ordering is necessary for both the recognition of fusion sets and the two algorithms used in the detailed analysis. For example, if the positions of the rows of **M** from (1.23) are altered as in (1.29), it becomes difficult, if not impossible, to detect fusion sets.

$$(1.29) \qquad \mathbf{M} = \begin{bmatrix} -w_1 & w_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -w_4 & w_4 & 0 \\ 0 & 0 & -w_5 & 0 & 0 & w_5 \\ -w_2 & 0 & w_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -w_6 & w_6 \\ 0 & 0 & -w_7 & w_7 & 0 & 0 \\ 0 & -w_3 & w_3 & 0 & 0 & 0 \end{bmatrix}$$

$$(1.23) \qquad \mathbf{M} = \begin{bmatrix} -w_1 & w_1 & 0 & 0 & 0 & 0 \\ -w_2 & 0 & w_2 & 0 & 0 & 0 \\ 0 & -w_3 & w_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & -w_4 & w_4 & 0 \\ 0 & 0 & -w_5 & 0 & 0 & w_5 \\ 0 & 0 & 0 & 0 & -w_6 & w_6 \\ 0 & 0 & -w_7 & w_7 & 0 & 0 \end{bmatrix}$$

As previously discussed matrix (1.23) is in near block form. This form makes it easier to see the possible division into two fusion sets. Matrix (1.29) must be premultiplied by a permutation matrix in order to form (1.23). Rather than developing an algorithm to create a permutation matrix necessary to convert (1.29) into (1.23), the program uses the least squares algorithm to give a preliminary sorting of the data in order of increasing relative affinities from the initial least squares solution.

Once the fusion sets have been recognized, each of the three sortings is repeated for each fusion set. The first sorting is based on a strict least squares algorithm, described below in section F. The algorithm for least squares values is considered to give the best unbiased solution of the experimental ΔG_{eq}'s. The least squares algorithm returns a <u>solution vector</u> which is defined as the listing of initial relative affinities based on the

algorithm. The acids are then sorted with reference to the solution obtained from this initial pass. The second ordering scheme depicts the minimum absolute affinity value each acid can have, and sorts the ladder into a minimum vector. Similarly, the maximum vector resulting from the third scheme depicts the maximum value the acid can have. For example, given the data from equations (1.7) through (1.9):

(1.7) $HA + B^- \rightleftharpoons A^- + HB \quad \Delta G_{eq} = 0.3$ kJ/mol

(1.8) $HA + C^- \rightleftharpoons A^- + HC \quad \Delta G_{eq} = 0.7$ kJ/mol

(1.9) $HB + C^- \rightleftharpoons B^- + HC \quad \Delta G_{eq} = 0.2$ kJ/mol

the minimum value for acid HC relative to HA is 0.5 kJ/mol based on the addition of reaction equilibria (1.7) and (1.9). The maximum value of 0.7 kJ/mol for HC relative to HA is given directly by equilibrium (1.8). Further discussion of the minimum and maximum vector algorithms occurs in section VIII. Bounding the Solution.

VII. LEAST SQUARES

A. General Methodology.

The principles behind the solution to the generalized least squares problem are generally attributed to Gauss from

his work in the late 1800's.[49] Gauss also developed methods of elimination to solve linear equations such as equation (1.30), a method which (with some modifications) is still widely used.[50,51]

(1.30) Let **A** be a nonsingular (n*n) matrix, and **x** and **b** be column vectors of length n. Then the equation **A x** = **b** is exactly solvable.

The least squares problem considered here is almost always overdetermined, that is, there are more equations than variables as shown in (1.31).

(1.31) Let **A** be a nondegenerate (m*n) matrix, and **x** and **b** be column vectors of length n and m, respectively. The equation **A x** = **b** is not usually exactly solvable.

The only time (1.31) is exactly solvable is when **b** - **A x** results in the zero vector. The least squares algorithm attempts to solve problem (1.32) for the more generalized case.

(1.32) Given a real nondegenerate matrix **A** and a real vector **b**, find the solution vector **x** of the equation **A x** = **b** such that the value of

$$\{n^{-1}\Sigma(b_i - \mathbf{A}_i\mathbf{x})^2\}^{\frac{1}{2}} = \| \textbf{residuals} \| = \| \mathbf{b} - \mathbf{A}\,\mathbf{x} \|$$

is a minimum.

The $\|$ symbol in problem (1.32) represents the L_2 norm of the residuals, that is, the square root of the summation of the squares of the individual residual values. The normal equations are often used to solve problem (1.32).[49]

The normal equations are obtained by first multiplying both sides of the equation $\mathbf{A}\,\mathbf{x} = \mathbf{b}$ by the transpose of $\mathbf{A}$ to get equation (1.33).

(1.33) $$\mathbf{A}^t\mathbf{A}\mathbf{x} = \mathbf{A}^t\mathbf{y}$$

As long as $\mathbf{A}$ is nondegenerate, the solution $\mathbf{x}$ can be obtained from equation (1.34).

(1.34) $$\mathbf{x} = [\,\mathbf{A}^t\mathbf{A}]^{-1}\ [\,\mathbf{A}^t\mathbf{y}]$$

In terms of the measurement positional matrix, problem (1.32) can be restated as problem (1.35)

(1.35) Let $\mathbf{M}$ be a (k*n) intraconnected measurement positional matrix, and $\mathbf{x}$ and $\mathbf{g}$ be column vectors of length n and k, respectively. Find the solution vector $\mathbf{x}$ given in equation $\mathbf{M}\,\mathbf{x} = \mathbf{g}$ such that the value of

$$\{n^{-1}\Sigma(g_i - \mathbf{M}_i\mathbf{x})^2\}^{\frac{1}{2}} = \|\ \mathbf{residuals}\ \| = \|\ \mathbf{g} - \mathbf{M}\,\mathbf{x}\ \|$$

is a minimum.

Within the context of (1.31), $\mathbf{g}$ is a vector of ΔG_{eq}'s.

B. Rank Deficiencies.

The principal difference between problems (1.32) and (1.35) is that the matrix **M** has less than full rank as defined by (1.36).

(1.36) Definition - Rank of a matrix ≡ The rank, k, of a (m*n) matrix **A** is the number of linearly independent rows or linearly independent columns within **A**.

Any intraconnected measurement matrix **M** of n compounds has a rank of (n-1). While the mathematical proof which follows is lengthy, a chemical example of that proof is rather simple. Consider the thermodynamic equations (1.7)-(1.9) relating HB and HC to HA: the knowledge of the affinities for two of the three equations should, according to Hess's law, reveal the affinity of the third.

There are a number of ways to find the rank of a given matrix **A**. Perhaps the easiest way is to reduce **A** to its row-echelon form.[52] The measurement positional matrix given by equations (1.7)-(1.9) is restated by matrix (1.13).

$$(1.13) \quad \mathbf{M} = \mathbf{C} = \begin{bmatrix} -w_1 & w_1 & 0 \\ -w_2 & 0 & w_2 \\ 0 & -w_3 & w_3 \end{bmatrix}$$

Scalar multiplication of each row i by $-w_i^{-1}$ yields (1.37).

$$(1.37)\quad \mathbf{M'} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}$$

The interchange of rows two and three yields (1.38).

$$(1.38)\quad \mathbf{M'} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$

The subtraction of row one from row three yields equation (1.39).

$$(1.39)\quad \mathbf{M'} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & -1 \end{bmatrix}$$

The rank degeneracy can easily be seen from equation (1.39) since now rows two and three are identical. The addition of row two to row one now yields (1.40).

$$(1.40)\quad \mathbf{M'} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 1 & -1 \end{bmatrix}$$

The subtraction of row two from row three yields equation (1.41), the row echelon form with a rank of two.

$$(1.41)\quad \mathbf{M'} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

To extend the proof to show that any intraconnected measurement matrix of n columns has a rank of (n-1), the basis of $\mathbf{M}_n$ must first be defined as shown by equation (1.42), where the subscript n denotes an **M** of n columns.

(1.42) Definition: Basis of **M** ≡ The basis **B** of all intraconnected measurement positional matrices of n columns is a matrix constructed of an {(n-1)*(n-1)} identity matrix augmented by a column of -1's.

$$\mathbf{B}_n = \begin{bmatrix} 1 & 0 & 0 & . & . & . & -1 \\ 0 & 1 & 0 & . & . & . & -1 \\ 0 & 0 & 1 & . & . & . & -1 \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & & & 1 & -1 \end{bmatrix}$$

It is easily seen that $\mathbf{B}_n$ forms a basis for any matrix $\mathbf{M}_n$. Every row within the matrix $\mathbf{M}_n$ can be formed by the linear combinations of the (n-1) rows in $\mathbf{B}_n$. For example, choose an arbitrary row, k, of $\mathbf{M}_n$ with the nonzero values of $-w_k$ in the i^{th} column and w_k in the j^{th} column, according to equation (1.43).

(1.43) row k of **M** = $[\ .\ .\ .\ -w_k\ .\ .\ .\ \ w_k\ .\ .\ .]$

From definition (1.42) and equation (1.43), row k of **M** is given by equation (1.44) from the i^{th} and j^{th} rows of $\mathbf{B}_n$.

(1.44) $[\ .\ .\ .\ -w_k\ .\ .\ .\ .\ w_k\ .\ .\ .] = -w_k\,[\,b_i - b_j\,]$.

An important corollary to equation (1.42) is that any subset of the complete positional matrix, $\mathbf{C}_n$, containing at least (n-1) rows and intraconnected according to equation (1.20) is row equivalent to the matrix $\mathbf{B}_n$ given by definition (1.42). Note that by definition every intraconnected measurement positional matrix $\mathbf{M}_n$ fits this criterion.

According to the row-echelon form, $\mathbf{B}_n$ is of rank (n-1). Therefore any intraconnected measurement positional matrix $\mathbf{M}_n$ is also of rank (n-1) since $\mathbf{M}_n$ is row equivalent to $\mathbf{B}_n$.

C. Row Augmentation.

In order to solve the problem given by (1.35), the degeneracy must be removed. The degeneracy can be removed simply by adding a single row identity vector to the intraconnected measurement positional matrix of n compounds, $\mathbf{M}_n$, as shown by theorem (1.45).

(1.45) Theorem: If $\mathbf{A}_n$ is a row augmented matrix defined by adding a single row identity vector to the intraconnected measurement positional matrix $\mathbf{M}_n$: that is,

$$\mathbf{A}_n = \begin{bmatrix} \mathbf{M}_n \\ e_n \end{bmatrix},$$

then the new matrix $\mathbf{A}$ will be of full rank n.

Proof of theorem (1.45) can be shown by substituting the matrix $\mathbf{B}_n$ for the matrix $\mathbf{M}_n$ into (1.45) as equation (1.46).

$$(1.46)\quad \mathbf{A_n}' = \begin{bmatrix} \mathbf{B_n} \\ e_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & . & . & . & -1 \\ 0 & 1 & 0 & . & . & . & -1 \\ 0 & 0 & 1 & . & . & . & -1 \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & 1 & -1 \\ 0 & 0 & 0 & . & . & 0 & 1 \end{bmatrix}$$

Note the matrix $\mathbf{B_n}$ will have the fewest rows of any intraconnected subset, $\mathbf{M_n}$, of the complete measurement matrix $\mathbf{C_n}$. To determine the rank of $\mathbf{A_n}'$, simply add the last row of $\mathbf{A_n}'$ to each of the other rows of $\mathbf{A_n}'$ in equation (1.46) to yield the identity matrix proving that $\mathbf{A_n}'$ is full rank. Since all intraconnected subsets of $\mathbf{C_n}$ containing more than (n-1) rows can be row reduced to the matrix $\mathbf{B_n}$, every intraconnected $\mathbf{M_n}$ will also be full rank once it has been row augmented with the row identity.

Since $\mathbf{M_n}$ has been augmented by a row forming $\mathbf{A_n}$ to remove the degeneracy in $\mathbf{M_n}$, the vector $\mathbf{g}$ must also be augmented by a row as shown by equation (1.47) in order to balance both sides of problem (1.32).

(1.47) Let $\mathbf{y}$ be a real (m+1)-tuple defined by

$\mathbf{y}^t = [\ \mathbf{g}^t \mid v\ ]$, where v is a real value.

Given the augmentations of $\mathbf{M_n}$ and $\mathbf{g}$, the least squares problem is now restated as (1.48).

(1.48) Given a real nondegenerate matrix **A** and a real vector **y**, find the solution vector **x** of the equation **A x** = **y** such that the value of

$$\|\mathbf{residuals}\| = \|\mathbf{y} - \mathbf{A}\mathbf{x}\| = \left\| \begin{bmatrix} \mathbf{g} \\ \mathbf{v} \end{bmatrix} - \begin{bmatrix} \mathbf{M}_n \\ \mathbf{e}_n \end{bmatrix} \mathbf{x} \right\|$$

is a minimum.

Within the context of the chemistry involved, equation (1.48) is almost always overdetermined, that is, there are more measurements between compounds than there are compounds. The augmented, intraconnected matrix **A** can never be underdetermined because of the definition (1.20) of **M** and the definition of intraconnectedness. The only time **A** is invertible would be if **M** were constructed entirely of pendant measurements (then there would be n-1 rows). The data partitioning algorithm has already excluded most of the pendant measurements, but the algorithm cannot recognize pendant measurements which connect two otherwise disjointed fusion sets. The most often observed case from the literature is when only one measurement connects the two disjointed fusion sets. Large fusion sets constructed only of pendant measurements are possible and have been observed.[21] When the matrix **A** in problem (1.48) is invertible, no statistical assessment of measurement errors can be performed.

In every case where the intraconnected measurement positional matrix $\mathbf{M}_n$ is overdetermined, there exists a unique solution to the least squares problem. For the

overdetermined set of equations given in problem (1.48), two different types of methods can be used to solve the linear least squares problem. The first method is to use the normal equations, while the second prefactors the **A** matrix. In theory, both methods result in the same solution. The method of the normal equations has already been described by (1.32) through (1.34). Since discussion regarding the validity of adding the row augmentation to form **A** from **M** is more easily followed if the normal equations are used, the rest of the discussion of this section will use the normal equations. Discussion of the methods which prefactor the **A** matrix will be presented in section VII. E., Improvements to the LU Decomposition.

The first consideration will be the effect of adding the additional row, $\mathbf{e}_n$, to **M** to form **A**. The block form which shows the effect of the row augmentation in the product of the matrices $\mathbf{A}^t\mathbf{A}$, denoted by **P**, is given by equation (1.49).

$$(1.49) \qquad \mathbf{P} = \mathbf{A}^t\mathbf{A} = \left[\mathbf{M}_n^t \mid \mathbf{e}_n^t \right] \begin{bmatrix} \mathbf{M}_n \\ \hline \mathbf{e}_n \end{bmatrix}$$

The only effect of adding the extra row $\mathbf{e}_n$ to $\mathbf{M}_n$ in terms of the normal equations is to add an extra value of one to element $P_{n,n}$. This addition removes the degeneracy from **P**, that is **P** is invertible. In terms of the right-hand side of

equation (1.33), the only difference between $\mathbf{A}^t\mathbf{y}$ and $\mathbf{M}_n\mathbf{g}$ is the inclusion of the term v in the last row of the product as seen by equation (1.50).

$$(1.50) \quad \mathbf{A}^t\mathbf{y} = [\, \mathbf{M}_n^t \mid \mathbf{e}_n^t \,] \begin{bmatrix} \mathbf{g} \\ \mathbf{v} \end{bmatrix} = [\, \mathbf{M}_n^t\mathbf{g} + v\mathbf{e}_n^t \,]$$

Therefore the effects of the addition of the extra row on $\mathbf{A}^t\mathbf{y}$ becomes entirely dependent upon the value of v in problem (1.48). Before the inclusion of the extra row, the linearly dependent columns in $\mathbf{M}_n$ could be written as a family of equations all dependent on the single parameter v. The extra row now gives the mechanism for incorporating the value of the parameter. The numerical value of v, while it could have any real number, has two reasonable choices. First, v could be the absolute affinity value of the n^{th} compound obtained from other methods. The choice of absolute affinity could then automatically anchor the solution vector to the absolute affinity value of that n^{th} compound. The deciding factor against this choice is that the $\mathbf{y}$ vector would easily become skewed in favor of the large anchoring affinity value.[53] For measurements such as the gas phase acidities, the anchoring affinity would be several orders of magnitude larger than any of the other measurements in the $\mathbf{y}$ vector. Typical gas phase acidities are between 300 to 400 kcal/mol, while the typical equilibrium affinities in ICR work are often less than 3.0

kcal/mol. Some loss of accuracy in the final solution vector would be obtained. The best choice is then to set the value of v to some small value. For purposes of the algorithm, v is constrained to be zero. Since v is declared to be zero, the left hand side of the augmented vector in (1.33) is the same as the nonaugmented form in (1.50) as shown by equation (1.51).

$$(1.51) \quad \mathbf{A}^t\mathbf{y} = [\ \mathbf{M}_n{}^t\mathbf{g} + 0\mathbf{e}_n{}^t\ ] = \mathbf{M}_n{}^t\mathbf{g}$$

The dimensions of the far left and the far right sides of equation (1.51) are both (n*1).

If any of the values of the least squares solution are then less than zero, it is probable that a riser exists within the data set. A riser can be defined as a compound which has only measurements leading away from it, that is, it has no direct path connecting it to compounds of lower index. Here the riser has an affinity index less than the first compound of the scale, but is only directly connected to compounds with a larger relative affinity index than the first compound.

The three compound case of HA, HB, and HC given by equations (1.7) through equation (1.9) is now used to illustrate the normal equations. The **M** matrix was given by (1.13), and adding the extra row yields equation (1.52).

$$(1.52) \quad \mathbf{A}\,\mathbf{x} = \begin{bmatrix} -w_1 & w_1 & 0 \\ -w_2 & 0 & w_2 \\ 0 & -w_3 & w_3 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Adding the extra zero row to **g** to form **y** is depicted in (1.53).

$$(1.53) \quad \mathbf{y} = \begin{bmatrix} 0.3 \\ 0.7 \\ 0.2 \\ 0.0 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ 0.0 \end{bmatrix}$$

The far right-hand side of equation (1.53) has been written in terms of its row elements. To simplify further discussion, the use of uniform weights of the **A** matrix in equation (1.52) results in equation (1.54).

$$(1.54) \quad \mathbf{A}\,\mathbf{x} = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

The vector $\mathbf{A}^t\mathbf{y}$ from equation (1.51) is then given by equation (1.55).

$$(1.55) \quad \mathbf{A}^t\mathbf{y} = \begin{bmatrix} -1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ 0.0 \end{bmatrix} = \begin{bmatrix} -y_1 - y_2 \\ y_1 - y_3 \\ y_2 + y_3 + 0.0 \end{bmatrix}$$

Since the sum of the elements in the last column vector is zero, $\mathbf{A}^t\mathbf{y}$ is said to be orthogonal to **1**. By similar

argument, any $\mathbf{A}^t\mathbf{y}$ with the value of v defined to be zero can be shown to be orthogonal to **1** because in each of the first m columns of $\mathbf{A}^t$, the two nonzero values are equal but opposite in sign.

Now returning to the left-hand side of equation (1.33), consider the three compound example. The $\mathbf{A}^t\mathbf{A}$ matrix is given by equation (1.56).

$$(1.56) \quad \mathbf{P} = \mathbf{A}^t\mathbf{A} = \begin{bmatrix} -w_1 & -w_2 & 0 & 0 \\ w_1 & 0 & -w_3 & 0 \\ 0 & w_2 & w_3 & 1 \end{bmatrix} \cdot \begin{bmatrix} -w_1 & w_1 & 0 \\ -w_2 & 0 & w_2 \\ 0 & -w_3 & w_3 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} = \begin{bmatrix} w_1^2+w_2^2 & -w_1^2 & -w_2^2 \\ -w_1^2 & w_1^2+w_3^2 & -w_3^2 \\ -w_2^2 & -w_3^2 & w_2^2+w_3^2+1 \end{bmatrix}$$

D. LU Decomposition.

Since the matrix $\mathbf{A}^t\mathbf{A}$ is square, the solution **x** can now be obtained using Gaussian elimination.[50,51] The effects of Gaussian elimination on **P** can be described by the LU decomposition,[50] where **L** is a lower triangular matrix and **U** is an upper triangular matrix. The ranks of **L** and **U** are the same as **A**, that is n.[54] Substituting the **LU** decomposition product for $\mathbf{A}^t\mathbf{A}$ in equation (1.34) yields equation (1.57).

$$(1.57) \quad \mathbf{x} = [\,\mathbf{A}^t\mathbf{A}\,]^{-1} \cdot [\,\mathbf{A}^t\mathbf{y}\,] = [\,\mathbf{LU}\,]^{-1} \cdot [\,\mathbf{A}^t\mathbf{y}\,]$$
$$= [\,\mathbf{U}\,]^{-1} \cdot [\,\mathbf{L}\,]^{-1} \cdot \mathbf{A}^t\mathbf{y}$$

Assuming uniform weights are used to simplify discussion, equation (1.56) results in the expanded equation (1.58).

$$(1.58)\quad \mathbf{A^tA} = \begin{bmatrix} -1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 3 \end{bmatrix}$$

Using the LU decomposition, the matrix **P** in equation (1.58) becomes equation (1.59).[51]

$$(1.59)\quad \mathbf{P = LU} = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ -.5 & 1.0 & 0.0 \\ -.5 & -1.0 & 1.0 \end{bmatrix} \cdot \begin{bmatrix} 2.0 & -1.0 & -1.0 \\ 0.0 & 1.5 & -1.5 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}$$

The inverse of **L** and **U** can then be substituted into equation (1.57) as shown by equation (1.60).

$$(1.60)\quad \mathbf{x} = [\,\mathbf{U}\,]^{-1} \cdot [\,\mathbf{L}\,]^{-1} \cdot \mathbf{A^t y}$$

$$= \begin{bmatrix} .500 & .333 & 1.00 \\ .000 & .667 & 1.00 \\ .000 & .000 & 1.00 \end{bmatrix} \cdot \begin{bmatrix} 1.00 & .00 & .00 \\ .50 & 1.00 & .00 \\ 1.00 & 1.00 & 1.00 \end{bmatrix} \cdot \begin{bmatrix} -y_1 - y_2 \\ y_1 - y_3 \\ y_2 + y_3 + v \end{bmatrix}$$

Substituting the numerical affinities given in equations (1.7) through (1.9) yields equation (1.61).

$$(1.61)\quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} .50 & .33 & 1.00 \\ .00 & .67 & 1.00 \\ .00 & .00 & 1.00 \end{bmatrix} \cdot \begin{bmatrix} 1.00 & 0.00 & .00 \\ .50 & 1.00 & .00 \\ 1.00 & 1.00 & 1.00 \end{bmatrix} \cdot \begin{bmatrix} -1.0 \\ 0.1 \\ 0.9+v \end{bmatrix}$$

Performing the multiplication on the second two matrices yields equation (1.62).

$$(1.62)\quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} .50 & .33 & 1.00 \\ .00 & .67 & 1.00 \\ .00 & .00 & 1.00 \end{bmatrix} \cdot \begin{bmatrix} -1.0 \\ -0.4 \\ 0.0 + v \end{bmatrix} = \begin{bmatrix} -0.633 + v \\ -0.267 + v \\ v \end{bmatrix}$$

Thus the resulting solution becomes entirely dependent on the value of the parameter v. As previously discussed, the value of v should be numerically close to the values for the individual elements of the vector of measured affinities (usually less than 3.0 kcal/mol for ICR work) and therefore has been defined as zero.

If the numerical value of v were changed in equation (1.62) so that x_1 was zero instead of x_3, the new solution would then be given by equation (1.63).

$$(1.63)\quad \mathbf{x}^t = [\, v,\ .3667 + v,\ .6333 + v \,]$$

The position of the nonzero element within the augmented row in the original problem (1.52) must be discussed. To simplify the discussion, the row identity $\mathbf{e}_n$ was used to augment the original measurement positional matrix, but any row identity of length n could have been used to give equivalent results. For example, if $\mathbf{e}_1$ of length n had been used, the solution obtained would be the same as in equation (1.63). Therefore it does not matter

where the nonzero column of the row identity is; the algorithm constrains the solution of that particular column to be zero.

E. Algorithmic improvements to LU decomposition.

The basic problem of using the LU decomposition to accomplish the least squares analysis has to do with the error obtained in the calculated solution. The algorithm usually will not calculate the exact solution, but instead it will generate an exact solution to a nearby problem. Computer processing units manipulate and store numbers in binary arithmetic.[55,56] Binary arithmetic has two limitations. First, it cannot reproduce all numbers exactly, just as there is no exact decimal equivalent to an irrational number. Second, the computer processing unit has a limitation with respect to the relative size of two numbers that it can accurately add. For example, on any computer there is a sufficiently small number that when that number is added to one the result is still equal to one. For the DEC VAX and IBM PC systems that value is about 2^{-24}. The next larger power of two, 2^{-23}, is called the machine epsilon.[55]

When the limitations of binary arithmetic are imposed on matrices, the error of the generated solution becomes a factor of three variables: the size of each element in the

solution, the machine epsilon, and the condition number of the matrix. The condition number is a measure of how close the matrix is to being degenerate. Degenerate matrices do not guarantee a least squares solution, since too many parameters are sought in solving problem (1.32). Matrices with large condition numbers result in less accurate solutions than ones with small condition numbers.[50] The greater the condition number, the less accuracy can be obtained from the least squares solution. The size of the condition number depends not only on the original **A** matrix but also on the algorithm used. Fortunately the condition number of most intraconnected measurement matrices is small. The uniformly weighted complete positional matrix $\mathbf{C}_3$ has a condition number of 3.7, while $\mathbf{C}_{10}$ has a condition number of 10.9. The condition numbers for less than complete positional matrices are slightly larger: the nine-measurement, ten compound case of equation (1.46) has a condition number of 13.9. All of these condition numbers are relatively small. For terms of this work, concern over condition numbers occurs primarily when (1) the **A** matrix is large, and (2) nonuniform weighting is used.

In general, the condition number of a matrix **A** will be the square root of the condition number of the matrix $\mathbf{A}^t\mathbf{A}$. Thus if **A** is nearly degenerate, then the solution to the least squares problem might have very few significant figures. Because the LU decomposition algorithm works

initially on $\mathbf{A}^t\mathbf{A}$, its solution will have even fewer significant figures than the solution of an algorithm that prefactors **A**.

Several algorithms exist which solve the least squares problem by prefactoring **A**.[49,57] Of these methods, the Householder orthogonalization algorithm[58] associated with the LINPACK routine SQRDC[42] was chosen for use in the program. The Householder orthogonalization algorithm is the fastest of the major algorithms, according to Golub and Van Loan.[57] Given the initial conditions imposed on intra-connected measurement matrices, no matrix was experimentally observed to be row or column degenerate. Therefore use of the computationally more expensive Single Value Decomposition algorithm was deemed unnecessary.

A second reason for using the Householder orthogonalization (available as SQRDC and SQRSL from LINPACK)[42] was that these programs were encoded with the Basic Linear Algebra Subroutines, BLAS.[45,46] The BLAS subroutines perform basic vector manipulations and operations such as scalar multiplication, copying one vector into another, and dot products. The BLAS subroutines are encoded in FORTRAN language as *Transactions on Mathematical Software Algorithm 539*.[45] Using the BLAS library will increase the portability of the programs developed in this chapter. A further incentive toward using these routines is that the ASSEMBLY language subroutines are available through

Transactions on Mathematical Software Algorithm 653.[46] The ASSEMBLY language version of these routines proved not only to be computationally more efficient but also more accurate than either the FORTRAN BLAS[45] algorithms or personal attempts to encode the same vector processes. The assembly language BLAS subroutine SCOPY was so fast that it could copy one vector to another faster than the Fortran code could fill that vector with zeros.

The third reason the LINPACK subroutines SQRDC and SQRSL were chosen for use by this program was the ease of iterative refinement to improve solution accuracy.[59-65] The use of iterative refinement in effect lowers the condition number of the **A** matrix even further by accumulating updates in double precision arithmetic.[55]

F. The QR Algorithm.

The least squares algorithm the program utilizes will now be discussed. The initial part of the algorithm prefactors **A** into two orthogonal matrices **Q** and $\mathbf{K}^t$ and an upper triangular (or diagonal) matrix **R** as in equation (1.64).[66]

$$\mathbf{A} = \mathbf{QRK}^t \tag{1.64}$$

Within the context of the algorithm, the matrix **K** is a permutation matrix which keeps track of any pivoting performed by the algorithm. Pivoting improves the accuracy of solution, but the Householder algorithm[58] can be performed without pivoting.[42] When no pivoting occurs, $\mathbf{K}^t$ is the identity matrix, **I**. To simplify the presentation, $\mathbf{K}^t$ will be omitted in further discussion.

Equation (1.34) can now be solved by substituting in equation (1.64) as shown by equation (1.65).

$$(1.65) \qquad \mathbf{x} = [\,(\mathbf{QR})^t\mathbf{QR}]^{-1}\ [\,(\mathbf{QR})^t\mathbf{y}] = [\,\mathbf{R}^t\mathbf{Q}^t\mathbf{QR}]^{-1}\ [\,\mathbf{R}^t\mathbf{Q}^t\mathbf{y}]$$

Since **Q** is orthogonal, $\mathbf{Q}^t\mathbf{Q} = \mathbf{I}$ and the far right side of (1.65) reduces to (1.66).

$$(1.66) \qquad \mathbf{x} = [\,\mathbf{R}^t\mathbf{R}]^{-1}\ [\,\mathbf{R}^t\mathbf{Q}^t\mathbf{y}] = [\,(\mathbf{R})^{-1}(\mathbf{R}^t)^{-1}]\ [\,\mathbf{R}^t\mathbf{Q}^t\mathbf{y}]$$

Since **R** as an upper triangular matrix is invertible, **x** can be obtained by equation (1.67).

$$(1.67) \qquad \mathbf{x} = (\mathbf{R})^{-1}\ [(\mathbf{R}^t)^{-1}\mathbf{R}^t]\ Q^t y = (\mathbf{R})^{-1}\mathbf{Q}^t\mathbf{y}$$

As no algorithmic changes have been made from the original LINPACK decomposition algorithm SQRDC,[42] no further discussion of the Householder decomposition is necessary here.

Initial attempts to form the solution to the least squares problem gave some inaccurate results. For example, rounding errors associated with the algorithm would alter the values obtained from the least squares solution so the solution would only be good to five significant figures. It was then realized that the least squares algorithm can be performed with iterative improvement of the initial solution, thus improving the accuracy of the least squares solution. The first iterative improvement scheme was done by Bjorck,[59,60] followed by subsequent papers by Wampler[61,62] and Golub.[63-65] The iterative scheme is based upon equation (1.68).[57,59,61]

$$\begin{bmatrix} \mathbf{I} & \mathbf{A} \\ \hline \mathbf{A}^t & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{r} \\ \hline \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ \hline \mathbf{0} \end{bmatrix} \tag{1.68}$$

where **A**, **x**, and **y** are the matrices described by equation (1.48) and **r** is the vector of residuals.

The iterative process for the k^{th} step begins by solving the supermatrix (1.69) as given in Golub and Van Loan.[57]

$$\begin{bmatrix} \mathbf{f}^{(k)} \\ \hline \mathbf{g}^{(k)} \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ \hline \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{I} & \mathbf{A} \\ \hline \mathbf{A}^t & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{r}^{(k)} \\ \hline \mathbf{x}^{(k)} \end{bmatrix} \tag{1.69}$$

where $\mathbf{f}^{(k)}$ and $\mathbf{g}^{(k)}$ are the iterative updates and **A**, **r**, **x** and **y** are already defined by equation (1.68). Here is one area

where the unique architecture of **A** allows an efficient shortcut. Most of the computational time in solving (1.69) is spent finding the products of $\mathbf{Ax}^{(k)}$ and $\mathbf{A}^t\mathbf{r}^{(k)}$. Looking at the expanded product in equation (1.70),

$$(1.70) \qquad \mathbf{Ax}^{(k)} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \ldots a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \ldots a_{2n}x_n \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x3 \ldots a_{mn}x_n \end{bmatrix}$$

it can be seen the product of the matrix $\mathbf{Ax}^{(k)}$ requires m*n flops, where one flop is one multiplication and one addition.[56] However, only two elements in each row of **A** are nonzero. If only the nonzero elements of **A** are stored (and they have been in vectors), then this product involves only 2*n flops as shown by equation (1.71).

$$(1.71) \qquad \mathbf{Ax}^{(k)} = \begin{bmatrix} a_{1,i}x_i + a_{1,j}x_j \\ a_{2,i}x_i + a_{2,j}x_j \\ \ldots\ldots\ldots \\ a_{m,i}x_i + a_{m,j}x_j \end{bmatrix}$$

where i and j represent the two nonzero elements of each row. Similarly, equation (1.72) shows that the product of $\mathbf{A}^t\mathbf{r}^{(k)}$ can also be reduced to a 2*n process.

$$(1.72) \qquad \mathbf{A}^t\mathbf{r}^{(k)} = \begin{bmatrix} a^{1,i}r_i + a^{1,j}r_j \\ a^{2,i}r_i + a^{2,j}r_j \\ \ldots\ldots\ldots \\ a^{m,i}r_i + a^{m,j}r_j \end{bmatrix}$$

where i and j represent the two nonzero elements of each row of $\mathbf{A}^t$. The rest of the algorithm proceeds as according to the treatise of Golub and Van Loan.[57]

G. Weighting.

Weighting of the initial data can be performed by two different methodologies.[67-70] The first method relies on the fact that certain measurements are not as precise as others and therefore weights the data before the least squares algorithm is performed. The second approach is performed after the least squares algorithm and makes attempts at evaluating the validity of the solution attained by least squares. Both the preweighting and the post-weighting schemes can be done with the program LADDER. The preweighted least squares is discussed in this section while the post least squares weighted algorithms will be discussed in section X. Least Trimmed Squares.

There are several different ways to preweight the data. The simplest weighting scheme is to have a uniform weight throughout the entire data set, as has been used for illustration earlier. The uniform weighting scheme then becomes the normalized measurement positional matrix such as was demonstrated in equations (1.54) and (1.55). In the absence of any further knowledge as to how the experimental data was generated, the uniform weighting scheme is not only

the simplest but the best weighting scheme (by default) to use. Since most of the equilibrium data within the literature is not reported with error estimates on a measurement-by-measurement basis, uniform weighted least squares also makes intuitive sense: each measurement can initially be treated as equally valid as any other measurement. Uniformly weighted least squares is the default algorithm of the program.

A second weighting scheme can be derived from the precision of the initial measurement data, providing it is known. Each equilibrium measurement is usually performed several times in order to obtain one thermodynamic relative affinity. For example, when a gas phase acidity is measured, the experimenter conducts the experiment at several different pressure ratios of HA to HB. The experimenter can then average the data to obtain the mean and standard deviation for the measured affinity, as demonstrated by equations (1.73) and (1.74).

$$x_{mean} = 1/n \sum x_i \tag{1.73}$$

$$\sigma^2 = 1/n \sum (x_i - x_{mean})^2 \tag{1.74}$$

When the mean, x_{mean}, and standard deviation, σ, are known for each individual measurement, then the weighted least squares algorithm can be evaluated as problem (1.75) instead of problem (1.32).

(1.75) For the real nondegenerate matrix **A** and the real vector **y**, find the solution vector **x** for the equation $\mathbf{A}\cdot\mathbf{x} = \mathbf{y}$ such that $\{n^{-1}\Sigma\ w_i(y_i - \mathbf{A}_i^t\mathbf{x})^2\}^{\frac{1}{2}}$ is a minimum.

In problem (1.75) w_i is the reciprocal of the standard deviation of the i^{th} measurement. The least squares algorithm can readily solve for **x** when **A** has nonuniform weights by altering the **y** vector. Each element in **y**, y_i is replaced by $w_i y_i$, where w_i is the weight of the i^{th} row of **A**.[57] The augmented rows in **A** and **y** that were necessary to give full rank to **A** are treated separately in each of the possible weighting schemes: the augmented rows remain e_1 for the last row in **A** and zero for the last row of **y**.

Few cases exist within the literature where the standard deviation of the measured affinity has been recorded for every measurement. When the standard deviation for each individual measurement is not known or has not been recorded, there are still several ways by which to weight the data based upon assumptions regarding the measured affinities.

The third weighting scheme uses the reciprocal of the individual relative affinity values, g_i, for problem (1.75) as shown by equation (1.76).

(1.76) $$w_i = (g_i)^{-1}$$

If the measured affinity is zero, a suitable finite weight must be substituted for that particular value of w_i. The ladder program currently makes the assumption that the principal users are chemists studying gas phase equilibria with a precision limited to ±0.1 kcal/mol. Therefore, any affinity less than 0.1 kcal/mol will be given the same weighting as an affinity of 0.1 kcal/mol. Therefore a measurement of any affinity less than 0.1 kcal/mol in equation (1.76) carries a weight of 10. The weighting scheme portion of the program makes no distinction between the units of kcal/mol and kJ/mol, assuming that the affinities are listed in kcal/mol. Most of the scales that were studied were in units of kcal/mol.

The fourth weighting scheme employed by the program squares the reciprocal of the measured affinity for the weight in problem (1.75) as shown by equation (1.77).

$$w_i = (g_i)^{-2} \tag{1.77}$$

Measurements of less than 0.1 kcal/mol (or kJ/mol) carry a weight of 100.

The fifth weighting scheme is actually two-fold and is based on a further analysis of the limitations of the instruments used in evaluating gas phase equilibria. The overall effect of the scheme is designed to be a descending scale estimate[67,69] such that when the numerical affinity is

large, the weight given to the measurement is small. The result is a Gaussian style weighting function as described by equation (1.78).

$$(1.78) \qquad \Sigma\, w_i(\mathbf{y} - \mathbf{A}^t\mathbf{x})^2 = \Sigma\, \exp[-\alpha * \mathbf{y}_i^2] * (\mathbf{y} - \mathbf{A}^t\mathbf{x})^2$$

where α is a constant dependent on which gas phase method is used to obtain the equilibria measurements. In order to understand the basis for this weighting scheme, a discussion of how gas phase equilibria are measured follows.

The equilibrium affinities defined by reactions (1.1) to (1.6) are logarithmic functions of the concentrations of the reactants and products. In gas phase equilibria studies the concentration of a neutral species is given by the pressure of the gas. The concentration of the ionic species is determined by the signal intensity of that ion. Thus in the gas phase, the equilibrium constant for reaction (1.7) is given by equation (1.79).

$$(1.79) \qquad K_{HA-HB} = \frac{[(\text{Pressure of HB})\ (\text{Intensity of } A^-)]}{[(\text{Pressure of HA})\ (\text{Intensity of } B^-)]}$$

The limitations on obtaining K_{HA-HB} (as well as other equilibrium constants) are based upon the measured precisions of both pressures and both intensities.

The errors associated with intensities are often less than a few percent. Often the peak observed at a

mass/charge ratio corresponding to either A^- or B^- is the base peak of the spectrum. In this case the relative height of the less intense peak is more responsible for the limitations due to intensity on the observable affinity. In order to obtain accurate intensities using current technology, the ratio of intensities should not exceed about 100:1 for ICR using the capacitance bridge detector. For HPMS work which uses an electron multiplier for detection of ion signals, the ratio of ion intensities should not exceed 1000:1.

For ICR measurements, the errors associated with the pressure measurement are considered to be larger than errors in intensity. In ICR and Fourier Transform Mass Spectrometry (FTMS), pressures are indirectly measured, and the accuracy is often no better than about ten percent.[3] The error associated with pressure measurement can be as high as ±30% in some ICR work.[21] The errors in pressure measurement for HPMS are smaller than in ICR because the ratio of reagent pressures is based on the stoichiometry of the condensed phase mixture[15] which can be directly measured.

To understand how individual errors accumulate, let the relative errors in the pressures of HA and HB be denoted as $\sigma(HA)$ and $\sigma(HB)$. Similarly, let the relative errors in the intensities of A^- and B^- be represented as $\sigma(A^-)$ and $\sigma(B^-)$.

The most probable error in K_{eq} is then described by equation (1.80).[71,72]

$$(1.80) \quad \sigma(K_{eq}) = [\ \{\sigma(HA)\}^2 + \{\sigma(HB)\}^2 + \{\sigma(A^-)\}^2 + \{\sigma(B^-)\}^2\ ]^{\frac{1}{2}}$$

Examination of equation (1.80) shows that much of the most probable error in the K_{eq} will be associated with the largest relative error. When pressure values are much less exact than the intensities (as in ICR measurements), the most probable error is primarily a function of the errors in the pressure measurements. Due to the limitations on obtaining accurate measurements of both pressures and intensities, any measured ΔG_{eq} larger than approximately 3.0 kcal/mol in ICR equilibrium chemistry should be regarded with caution. In order to obtain a ΔG of 3.0 kcal/mol, the K_{eq} must be approximately 0.006. To obtain this equilibrium constant the ratio of intensities from (1.79) nears the limit of $100(B^-):1(A^-)$. In addition the pressure of HB much be nearly twice (or more) that of the pressure of HA. Therefore, while an ICR ΔG_{eq} measurement of 1.0 kcal/mol is generally known to ±0.1 kcal/mol, a measurement of 3.0 kcal/mol is known with much less precision.

For ICR measurements, the scaling constant α in equation (1.78) is 0.1733. Therefore the weight given to an affinity of 1.0 kcal/mol is 0.841, while the weight for an affinity of 3.0 kcal/mol is only 0.595. A measured affinity

of 0.0 kcal/mol (or kJ/mol) is given the highest weight of 1.000.

The precision in obtaining pressure information directly from the liquid volumes and the use of electron multipliers for signal detection in HPMS[15] work allows the scaling constant in equation (1.78) to be smaller than for ICR work. HPMS experiments are also usually obtained at higher temperatures, which would increase the size of the ΔG_{eq} measurements relative to the same nominal equilibrium constant. Therefore the accepted precision of a measured value of 3.0 kcal/mol is higher for HPMS work than for ICR work. The scaling constant adopted for HPMS work is 0.04233. A measured HPMS affinity of 1.0 kcal carries a weight of 0.959 while the a 3.0 kcal/mol measurement has a slightly smaller weight of 0.881.

The Gaussian style weight presented as (1.79) is an example of a redescending estimator: the larger the measurement is the smaller the weight will be. Several other redescending scale estimators have appeared in the literature[67,69,70] but were not studied in the course of this work. The Gaussian scheme presented in this work seems logical when considering that most of the thermodynamic affinities are themselves logarithmic functions. Using an exponential function to weight the data in essence performs a first approximation of a weighting scheme on the original equilibrium constants.

The final weighting scheme attempts to reduce the influence of individual equilibria within the ladder. The more frequently that a particular compound appears in the list of recorded equilibria, the less influence the absolute affinity of that compound has on the entire scale. Occasionally, the equilibria leading to a compound will not show agreement to the affinity determinations by other measurements. For example, the gas phase acidity of HF is significantly weaker than expected from the corresponding dissociation-electron affinity calculation.[3] The influence of HF to the rest of the equilibrium scale would be even greater if more measurements involved this compound. To reduce the influence of an individual measurement, the weight of that measurement is calculated according to equation (1.81).

(1.81) w_i = (0.02*the total number of measurements)/($f_1 f_2$)

where f_1 and f_2 correspond to the frequency with which compounds one and two appear within the list of recorded equilibria.

H. Errors Within the Least Squares Solution.

As already suggested, errors exist in the least squares solution obtained from the analysis. Errors can occur in

four different stages of data collection and processing: (1) errors within the equipment resulting in an improper measurement of the equilibrium, such as too short delay times or transient signals; (2) errors associated with unique and undetected ion-molecule chemistry for specific compounds; (4) errors in data recording and data entry; and (4) algorithmic/machine errors inherent in performing the least squares algorithm on computers using a finite number system.

Errors in the fourth stage of data collection and manipulation are actually the easiest to avoid. Sections F and G hinted at a few of the errors associated with the algorithm. If each of the initial three stages of data collection are correct, then only three possible problems could occur from a computational perspective. The first of the computational problems is that the measurement matrix will still be degenerate after the inclusion of the augmented row. Every augmented measurement matrix is nondegenerate in exact mathematics according to theorems developed in section F, but the use of the floating point binary arithmetic could incorrectly assess a nearly-degenerate matrix as degenerate.[51] So far, every augmented matrix observed by this work has been nondegenerate. The second algorithmic error might occur if any matrices force an illogical operation, such as division by zero. No errors of this sort have been observed with the existing program on

the tested scales. The third error is that the computer almost never calculates the exact solution to the given problem, but rather the exact solution to a nearby problem that the computer can interpret within the limitations of the floating point binary arithmetic.[55] Fortunately the errors due to machine interpretation of the least squares problem are several orders of magnitude smaller than the errors associated with the measurement, unique chemistry, or data entry.

The errors resulting from human interpretation of the experiments are more difficult to find. If a systematic error were to occur over an entire equilibrium data set, then it could go undetected until a separate laboratory reproduced the experiments. Fortunately, errors tend to occur in isolated equilibria rather than over the entire work. Here the difficulty is in determining which equilibrium affinities are correct and which should be treated as suspect. In a large ladder, isolating the few truly suspicious measurements by hand is difficult.

A principal problem with the least squares algorithm occurs while trying to minimize the residuals. One or more suspect measurements can skew the solution obtained away from reality.[69,70,73] Two types of measurement can result in skewed solutions. The first measurement type has to do with influential measurement points, while the second type is

that of true outliers, which are values that are significantly different than expected.

In terms of the chemistry involved, influential measurements result whenever very few pathways connect the compounds in the scale. The more pathways from one compound in the ladder to the rest, the lower the influence of each measurement. For example, if the pendant measurements were not removed prior to least squares analysis, each pendant measurement would be considered influential. Relatively influential affinities occur whenever only two or three paths connect a compound to the rest of the compounds in the ladder. Every measurement within those paths is then relatively influential. In terms of the mathematics, the relative influence of a single point can be determined through the use of the projection (or Hat) matrix defined by equation (1.82).[74]

(1.82) Definition: Projection matrix ≡ the projection matrix, **H**, of the augmented row matrix **A** is given by:

$$\mathbf{H} = \mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{A}^t$$

The diagonal elements of **H** give the relative influence of the individual measurements. For example the value of h_{kk} and h_{ll} denote the influence of the k^{th} and l^{th} measurements, respectively. The off diagonal elements of **H** give a measure

of the influence one measurement has on another. The off diagonal elements should be nearly zero.

The program can calculate the projection matrix after first calculating the variance-covariance matrix as defined by equation (1.83).[74]

(1.83) Definition: Variance-covariance matrix ≡ the matrix, **V**, of the row augmented matrix **A** is given by

$$\mathbf{V} = (\mathbf{A}^t\mathbf{A})^{-1}.$$

The variance-covariance matrix can be rewritten in terms of the QR algorithm used to solve the initial least squares solution. The simplification is given by equation (1.84).

$$\mathbf{V} = (\mathbf{A}^t\mathbf{A})^{-1} = ((\mathbf{QR})^t\mathbf{QR})^{-1} = (\mathbf{R}^t\mathbf{Q}^t\mathbf{QR})^{-1} = (\mathbf{R}^t\mathbf{R})^{-1} \tag{1.84}$$

Substitution of (1.84) into definition (1.82) yields the first step toward calculating the projection matrix as shown in terms of **Q** and **R** by equation (1.85).

$$\mathbf{H} = \mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{A}^t = \mathbf{A}(\mathbf{R}^t\mathbf{R})^{-1}\mathbf{A}^t = \mathbf{AVA}^t \tag{1.85}$$

To find the individual element h_{kl} of **H**, the algorithm again takes advantage of the relative sparsity (lack of nonzero terms) in the original **A** matrix. Due to the nature of **A**, the product element of $\mathbf{VA}^t$ results in the addition of only

two rows of **V**. The product elements of **AV** result in the addition of only two columns of **V**. When these two multiplications are intersected, only four elements of **V** are required to form h_{kl} as shown by equation (1.86).

$$(1.86) \qquad h_{kl} = a_{li,lj} - a_{li,kj} - a_{ki,lj} + a_{ki,kj}$$

where k and l denote the k^{th} and l^{th} measurements. The subscripts i and j denote the i^{th} and j^{th} nonzero elements of rows k and l in the augmented measurement positional matrix, **A**.

For elements on the diagonal of **H**, equation (1.86) simplifies to equation (1.87).

$$(1.87) \qquad h_{kk} = a_{jj} + a_{ii} - 2a_{ij}$$

where i and j denote the nonzero elements of the k^{th} row in the augmented measurement positional matrix, **A**.

Most of the literature scales are made primarily from very few measurements. Therefore investigation of the projection matrix for these matrices of very few measurements led to the conclusion that most of the measurements show a strong relative influence on the final solution. The lack of measurements between compounds makes it harder to separate and recognize true outliers, as will now be discussed.

Outliers are measurements which are appreciably different from the value expected. There is no clear, exact definition, primarily due to disagreements on the exact limits of acceptable measurement.[74,75]

In a one parameter case, repeated measurements of a property should result in a Gaussian style distribution about a single value, the mean, as described in equation (1.73).[71] The standard deviation from the mean can be used as a measure of precision for that mean value. According to the Gaussian distribution, 67 percent of measurements fall within two standard deviations of the mean, while 93 percent fall within three standard deviations.[71] Some statisticians regard as an outlier any measured value which falls outside the neighborhood of three standard deviations. The basic limitation of using equations (1.73) and (1.74) for outlier identification is that both of these values are implicit functions of the outlier.[69,70,73] If the outlier is also influential, the mean and standard deviation will be skewed in the direction of the outlier.[70,73] What is true in the one parameter case is also true for the multiple variable functions. For functions of more than one parameter, it is assumed that the residuals from the calculated function will be normally distributed about the mean. In the two parameter case (linear least squares), it is simple to show graphically whether a particular measurement is an outlier. To look for the presence of outliers, one first standardizes

the residuals by dividing by the root mean square of the total as shown by equation (1.88).[73]

$$(1.88)\ \text{standardized res}_i = \text{res}_i \,/\, (1/n\ \Sigma\ \text{res}_i^2)^{1/2}$$

Here any standardized residual greater than approximately 2.5 to 3.0 is considered to be an outlier. The standardized residuals are plotted against the calculated value of the function using the best fitting slope and intercept. A second residual plot can be constructed using the individual values of x as the ordinate and the standardized residuals as the abscissa. In both plots the residuals should show no pattern but should be randomly distributed about the zero line.[73]

Finding outliers for multivariate functions is much more difficult.[70,76] Residual plots are limited to only one variable. Furthermore, a one variable residual plot may appear skewed when in fact it is not. A plot which appears to fit may actually be skewed.[70] There are several statistical ways to analyze the problem in an attempt to isolate the truly outlying value. For the classical least squares algorithm, the method chosen in this research for testing the solution is to analyze the residual versus the influence based on the projection matrix.[76] Each residual is plotted against the corresponding trace of the projection matrix as shown by Figures 1-1 and 1-2. The best equilibria

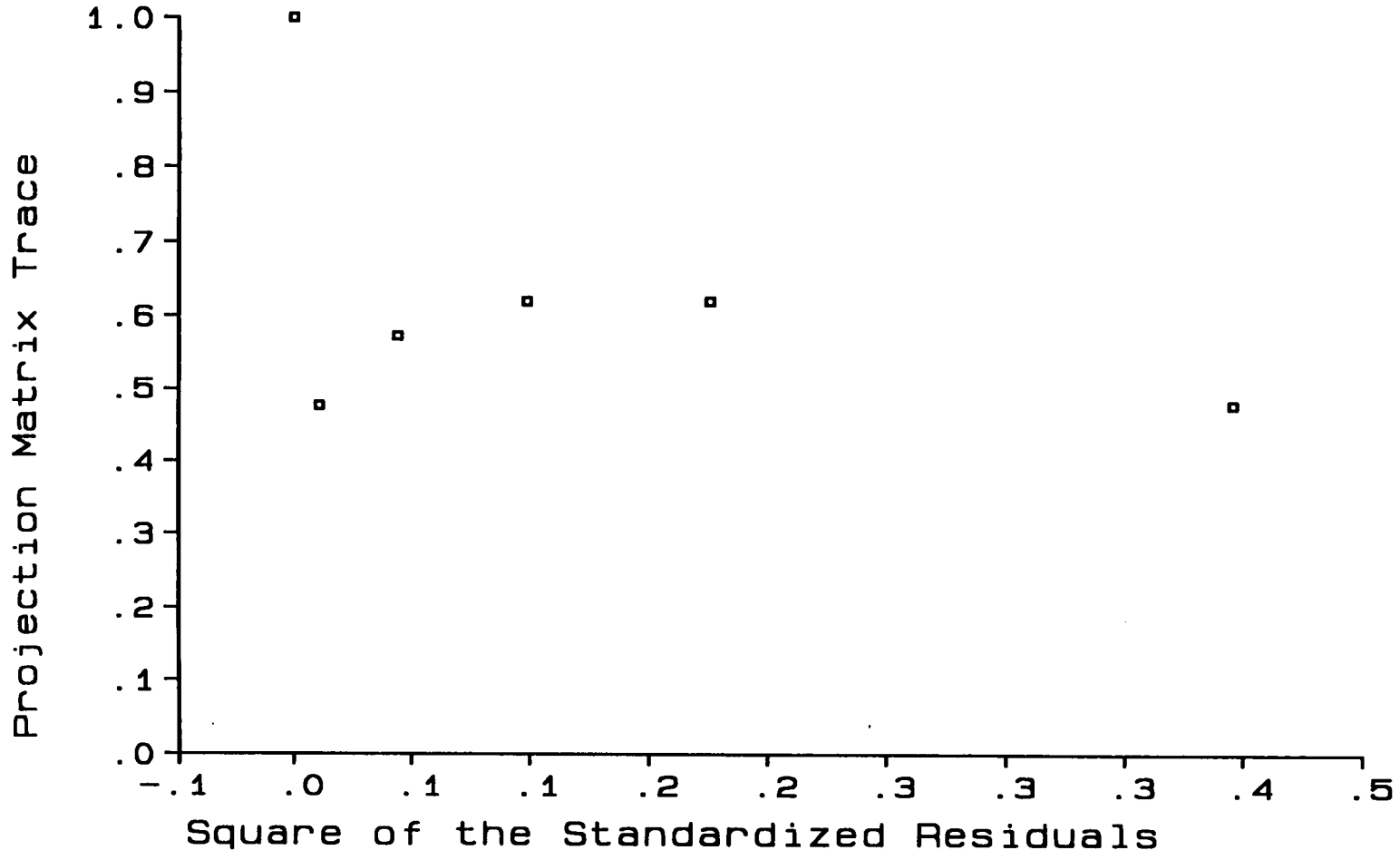


Figure 1-1: Influence Graph in MeOH to EtOH Fusion Set.
Source: Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979**, *101*, 6046-56.

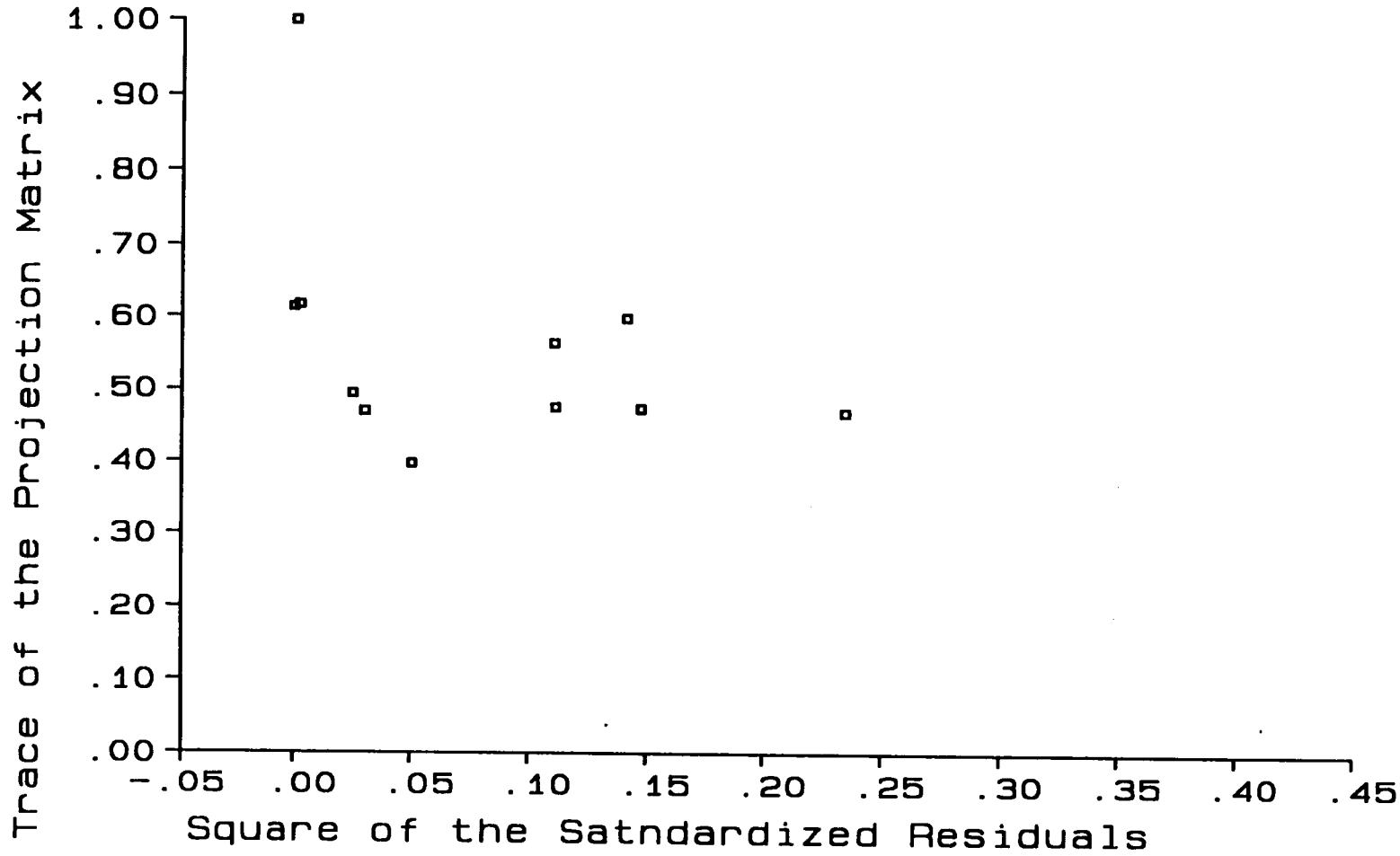


Figure 1-2: Influence Graph in EtOH to n-BuOH Fusion Set.
Source: Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979**, *101*, 6046-56.

are those where both the ordinate and the abscissa are close to zero. True outliers have a relatively small abscissa showing little influence but large residual value. Points that are only influential depict little residual value but an abscissa close to one.[76] In Figures 1-1 and 1-2 the point at (0,1) depicts the value ascribed to the row augmentation. This value is highly influential since it the value that keeps the **A** matrix full rank in the least squares problem.

Testing a variety of scales shows that most of the scales within the literature have so few measured equilibria that each equilibrium must be regarded as an influential point. The greater the influence, the more accurate the measurement must be in order to validate the entire ladder. Unfortunately, with little or no verification of the accuracy of these influential measurements, the scales tend to fall into the realm of guess-work. For this reason, methods other than least squares have proven to be necessary to give a thorough investigation of ladders.

VIII. BOUNDING THE SOLUTION

The program described in this chapter is not limited to the least squares algorithm but instead uses several different algorithms in the analysis of the data. The second major algorithm (after least squares) used by the

program sorts the absolute affinities and forms a minimal and maximal boundary. For example, given the chemical equations (1.7) through (1.9) and constraining the value of HA to be zero results in two pathways to both HB and HC as can be seen by Figure 1-3.

The algorithm begins by assigning the compound with the lowest affinity according to the least squares algorithm to have a relative affinity of zero, in this case HA. From the initial vantage point, the algorithm tentatively assigns the minimum values of HB and HC to be 0.3 kJ/mol and 0.7 kJ/mol. The algorithm then proceeds to the next lowest minimum affinity, in this case HB. From the vantage point of HB the affinity value of HC is reassigned as 0.5 kJ/mol from the sum of 0.2 kJ/mol and 0.3 kJ/mol. After assigning affinities from the next to highest member of the ladder, the algorithm terminates the minimum vector search.

The maximum vector algorithm proceeds the same way as the minimum vector algorithm does. It begins with the lowest member of the ladder and seeks the initial values based on the members that have equilibria tying them to the

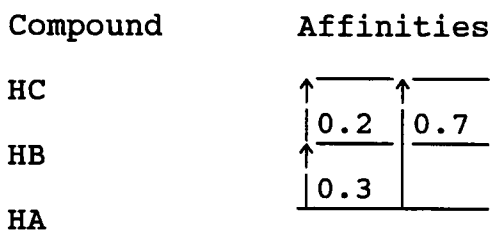


Figure 1-3: Graph of equations (1.7) to (1.9).

lowest member. The first pass is now complete. It then proceeds with the second lowest member, the third lowest, the fourth and so on until all but the highest member of the ladder has been processed. The only difference between the two algorithms is that the minimum vector keeps the value obtained from the minimum pathway while the maximum vector keeps the values pertaining to the largest pathways. Therefore the maximum vector would have an affinity of 0.2 kJ/mol for HB and an affinity value of 0.7 kJ/mol for HC.

While either algorithm is in progress, the order of the measured affinities may change. Ordering changes occur in scales primarily for two reasons: the existence of risers within the ladder as depicted by Figure 1-4 and inconsistent data.

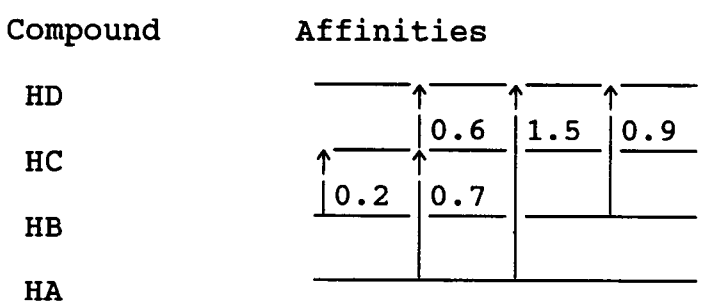


Figure 1-4: Graph Showing a Riser in the Data

In Figure 1-4 compound HB is considered to be a riser. Both the minimum and the maximum algorithms begin by examining HA. The next step in each algorithm is to examine

the rest of the scale beginning with the acid with the next lowest affinity associated with HA, namely HC. However, HB has a lower affinity index than HC, so the algorithm must re-examine the order.

The least squares algorithm is unable to point out an obvious inconsistency in the small ladder shown by Figure 1-5.

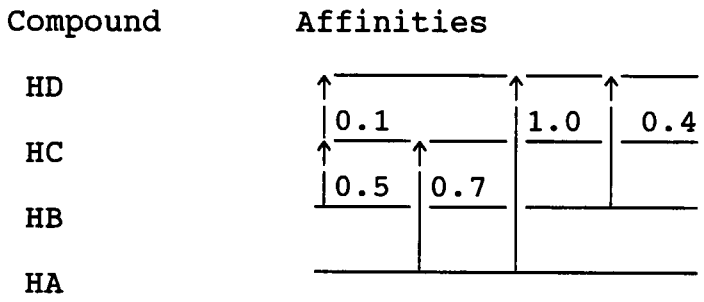


Figure 1-5: Inconsistent Data Set.

Here the algorithm starts with HA as before. From HC, it realizes that HB is lower in affinity than HC and switches the order as described above. Therefore the order is HA < HB < HC < HD. Analysis from the HB - HD pathway shows HD to be less than HC. This order is in disagreement with the two other pathways HA - HD and HA - HC - HD. Trying to reconcile the disagreement could throw the algorithm into an endless loop, but an alternate termination route exists. From experimental observation, the algorithm returns any reasonable solution within a number of passes. When the

number of passes (evaluations leading from a single compound) exceeds twice the number of measurements, the algorithm is given an abnormal termination status. The solution vector returned will be mathematically skewed low (or high in the case of the maximum vector), and the algorithm warns the user that the solution will be skewed.

A second error which is easily ignored by a blind least squares approach has a strong probability of being noticed by the minimum and maximum sorting routines. Consider the affinities in equilibria (1.7) through (1.9). Figure 1-6 gives the small ladder showing what happens if an error in data entry exists such that the affinity order were reversed for one measurement. The affinity of HA is less than HC according to the two step process in Figure 1-6, but the affinity of HA is greater than HC when measured directly. The algorithm as described until now would keep attempting to switch the compounds HA and HC, resulting in an endless loop. To avoid endless cycles between the two compounds, a counter was placed into the algorithm. When the loop count exceeds twice the number of compounds the program aborts the

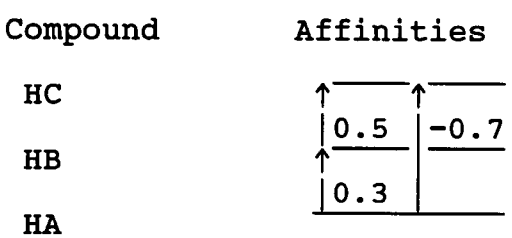


Figure 1-6: Reversal of Affinity Order in a Fusion Set.

algorithm and gives warning that it has terminated the algorithm. Errors ofthis type result in minimum vectors that are ridiculously low. In some examples, the solution obtained after the abnormal termination is a minimum vector where each element of the solution is negative. There are a few logical reasons for errors of this type. First, the experimenters may have omitted a sign when constructing the original data table. Second, the affinity may have been reversed by an editor compiling a large data set from many different sources and/or laboratories, such as in the compilation of the gas phase acidity scale or the fluoride affinity scale. Both of these errors should be obvious, but the least squares answer is often skewed toward the incorrect affinity. A third kind of error occurs when the affinities tying the three compounds together approach the experimental precision. For ICR data, the precision is usually assumed not to be greater than 0.1 kcal/mole. If the affinity for HA relative to HB is zero and HB relative to HC is 0.1 kcal/mol, then it is entirely feasible for the affinity of HA relative to HC to be -0.1 kcal/mol and still be within the experimental precision. Here the experimental errors interfere with what would normally be considered mathematically logical.

IX. ITERATIVE SMOOTHING

Once the minimum and maximum boundaries have been calculated for the fusion set, the next algorithm the program investigates is iterative smoothing. This algorithm is primarily an artifact of the early stages of ladder research, but it remains as it carries a few unique properties not found in the other algorithms.

The individual affinities can be projected onto an affinity surface as shown by Figures 1-7 and 1-8. On these surfaces the x and y coordinates represent the absolute affinities of individual compounds, with x being identical to y. The z coordinate represents the affinity between two compounds. The surface is approximately planar and contains the line of zero affinity. Deviations from the plane surface are due to either the imprecision of the individual measurements or due to outlying measurements. The steeper the hills and valleys, the more likely it is to find one (or more) errors.

In Figure 1-7, the planar part of the surface has been constrained to zero affinity, while Figure 1-8 shows the surface where a plane which minimizes the hills and valleys has been calculated.

In terms of the mathematics, the transition between Figures 1-7 and 1-8 begins by rewriting a portion of the gas phase acidity scale[3,77] shown by Figure 1-9 as the n by n

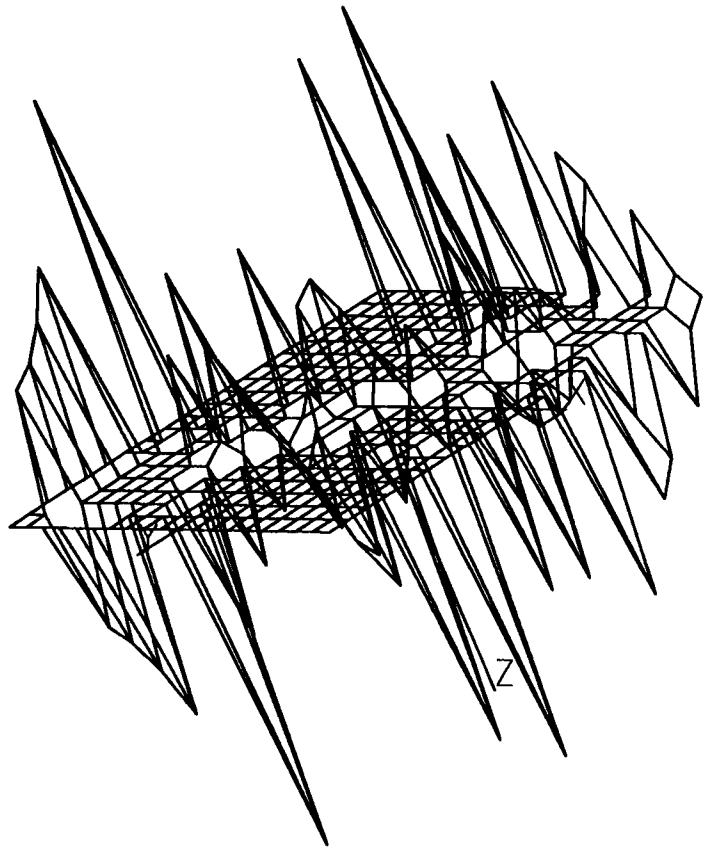


Figure 1-7: Unsmoothed data prior to iterative procedure.
Source: Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979**, *101*, 6046-56.

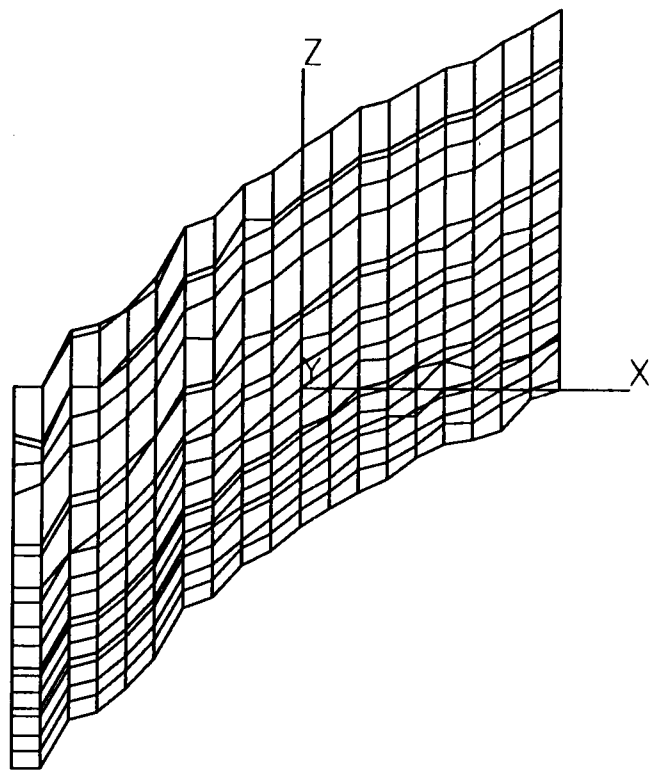


Figure 1-8: Smoothed data After the iterative procedure.
Source: Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979**, *101*, 6046-56.

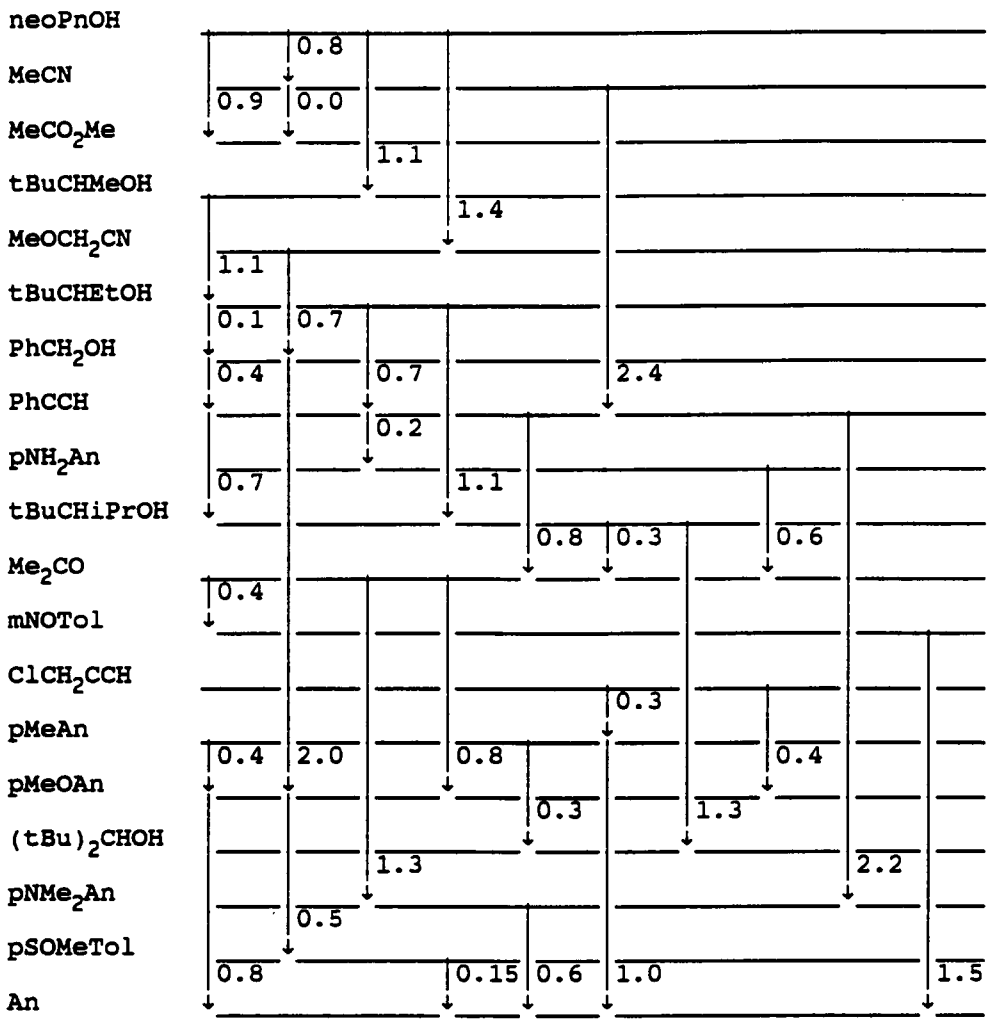


Figure 1-9. Graph of 74BAR\SCO used to Demonstrate the Iterative Algorithm.
Source: Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979** 101, 6046-56.; Lias, S.G.; Bartmess, J.E.; Liebman, J.F.; Holmes, J.L.; Levin, R.D.; Mallard, W.G.; *J. Phys. Chem. Ref. Data*, **1988**, *17*, Suppl. 1.

affinity matrix, **F**, shown in Figures 1-10 and 1-11. Each row or column of **F** depicts one compound in the ladder. The diagonal of **F** is defined as zero since the compound must have zero affinity with itself as shown by equation (1.89).

(1.89) $HA + A^- \rightleftharpoons HA + A^-$ $\Delta G = 0.0$ kcal/mol

Off diagonal numerical components represent a measured affinity between two compounds. For example row two and column four show the ΔG_{eq} measurement between neoPnOH and $MeCO_2Me$.[3]

The algorithm discussed in this section does not provide the actual smooth plot shown in Figure 1-8, but rather it generates the interpolated and extrapolated points used in constructing the curve. Several algorithms exist which interpolate data,[78-81] but it is more difficult to reasonably and accurately extrapolate data points successfully. The algorithm described herein does a fair job of extrapolating data.

The algorithm is based on the engineering metal template problem.[81] In that problem, two or more edges of the metal surface are considered to be at known (and fixed) but differing temperatures. The purpose is to discover the temperature at each measurable point on the interior of the surface. One rational approach to solving the problem is to set up a two dimensional matrix array of known and unknown

0.0	0.0	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
0.0	0.0	0.8	0.9	1.1	1.4	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
xxx	-0.8	0.0	xxx	xxx	xxx	xxx	xxx	2.4	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
xxx	-0.9	xxx	0.0	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
xxx	-1.1	xxx	xxx	0.0	xxx	1.1	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
xxx	-1.4	xxx	xxx	xxx	0.0	xxx	0.7	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
xxx	xxx	xxx	xxx	xxx	-0.7	-0.1	0.0	0.4	0.0	xxx	xxx	xxx	xxx	xxx	2.0	xxx	xxx	xxx	xxx
xxx	xxx	-2.4	xxx	xxx	0.0	-0.7	-0.4	0.0	0.2	0.7	0.8	xxx	xxx	xxx	xxx	xxx	2.2	xxx	xxx
xxx	xxx	xxx	xxx	-1.1	0.0	0.0	0.1	0.7	xxx	1.1	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
xxx	xxx	xxx	xxx	xxx	xxx	0.0	0.0	-0.2	0.0	xxx	0.6	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
xxx	xxx	xxx	xxx	xxx	xxx	-1.1	0.0	-0.7	xxx	0.0	0.3	xxx	xxx	xxx	xxx	1.3	xxx	xxx	xxx
xxx	xxx	xxx	xxx	xxx	xxx	xxx	0.0	-0.8	-0.6	-0.3	0.0	0.4	xxx	xxx	0.8	xxx	1.3	xxx	xxx
xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	-0.4	0.0	xxx	xxx	xxx	xxx	xxx	xxx	1.5
xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	0.0	0.3	0.4	xxx	xxx	xxx	xxx
0.0	0.0	0.0	0.0	0.0	0.0	xxx	xxx	xxx	xxx	xxx	xxx	xxx	-0.3	0.0	0.4	0.3	0.0	xxx	1.0
xxx	xxx	xxx	xxx	xxx	xxx	xxx	-2.0	xxx	xxx	xxx	-0.8	xxx	-0.4	-0.4	0.0	xxx	xxx	0.5	0.8
xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	-1.3	xxx	xxx	xxx	-0.3	xxx	0.0	xxx	xxx	0.0
xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	-2.2	xxx	xxx	-1.3	xxx	xxx	xxx	xxx	xxx	0.0	xxx	0.6
xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	-0.5	xxx	xxx	0.0	0.15
xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	-1.5	xxx	-1.0	-0.8	xxx	-0.6	-0.15	0.0

Figure 1-10: Portion of 79BAR/SCO Used to Demonstrate Iterative Algorithm: Before Iterative Solution.
Source: Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979** 101, 6046-56.; Lias, S.G.; Bartmess, J.E.; Liebman, J.F.; Holmes, J.L.; Levin, R.D.; Mallard, W.G.; *J. Phys. Chem. Ref. Data*, **1988**, *17*, Suppl. 1.

0.0	0.0	0.6	0.7	1.1	1.5	2.2	2.4	2.8	3.1	3.4	3.7	4.0	4.1	4.3	4.5	4.7	4.9	5.1	5.4
0.0	0.0	0.8	0.9	1.1	1.4	2.2	2.4	2.8	3.1	3.4	3.7	4.0	4.1	4.3	4.5	4.7	4.9	5.1	5.4
-0.6	-0.8	0.0	0.0	0.4	0.8	1.5	1.7	2.4	2.4	2.7	3.0	3.3	3.4	3.6	3.9	4.0	4.2	4.5	4.7
-0.7	-0.9	0.0	0.0	0.3	0.7	1.4	1.6	2.0	2.3	2.6	2.9	3.2	3.3	3.5	3.8	3.9	4.1	4.3	4.6
-1.1	-1.1	-0.4	-0.3	0.0	0.4	1.1	1.2	1.7	1.9	2.3	2.5	2.8	2.9	3.2	3.4	3.5	3.8	4.0	4.3
-1.5	-1.4	-0.8	-0.7	-0.4	0.0	0.7	0.7	1.3	1.5	1.8	2.1	2.4	2.5	2.8	3.0	3.1	3.4	3.6	3.8
-2.2	-2.2	-1.5	-1.4	-1.1	-0.7	0.0	0.1	0.7	0.8	1.1	1.4	1.7	1.8	2.0	2.3	2.4	2.7	2.9	3.1
-2.4	-2.4	-1.7	-1.6	-1.2	-0.7	-0.1	0.0	0.4	0.6	1.0	1.2	1.5	1.7	1.9	2.0	2.2	2.5	2.7	3.0
-2.8	-2.8	-2.4	-2.0	-1.7	-1.3	-0.7	-0.4	0.0	0.2	0.7	0.8	1.1	1.2	1.4	1.7	1.8	2.2	2.3	2.5
-3.1	-3.1	-2.4	-2.3	-1.9	-1.5	-0.8	-0.6	-0.2	0.0	0.3	0.6	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.3
-3.4	-3.4	-2.7	-2.6	-2.3	-1.8	-1.1	-1.0	-0.7	-0.3	0.0	0.3	0.5	0.6	0.9	1.1	1.3	1.5	1.7	1.9
-3.7	-3.7	-3.0	-2.9	-2.5	-2.1	-1.4	-1.2	-0.8	-0.6	-0.3	0.0	0.4	0.4	0.6	0.8	1.0	1.3	1.4	1.7
-4.0	-4.0	-3.3	-3.2	-2.8	-2.4	-1.7	-1.5	-1.1	-0.9	-0.5	-0.4	0.0	0.1	0.3	0.5	0.6	0.9	1.1	1.5
-4.1	-4.1	-3.4	-3.3	-2.9	-2.5	-1.8	-1.7	-1.2	-1.0	-0.6	-0.4	-0.1	0.0	0.3	0.4	0.5	0.8	1.0	1.3
-4.3	-4.3	-3.6	-3.5	-3.2	-2.8	-2.0	-1.9	-1.4	-1.2	-0.9	-0.6	-0.3	-0.3	0.0	0.4	0.3	0.6	0.8	1.0
-4.5	-4.5	-3.9	-3.8	-3.4	-3.0	-2.3	-2.0	-1.7	-1.4	-1.1	-0.8	-0.5	-0.4	-0.4	0.0	0.1	0.3	0.5	0.8
-4.7	-4.7	-4.0	-3.9	-3.5	-3.1	-2.4	-2.2	-1.8	-1.6	-1.3	-1.0	-0.6	-0.5	-0.3	-0.1	0.0	0.2	0.4	0.7
-4.9	-4.9	-4.2	-4.1	-3.8	-3.4	-2.7	-2.5	-2.2	-1.8	-1.5	-1.3	-0.9	-0.8	-0.6	-0.3	-0.2	0.0	0.2	0.6
-5.1	-5.1	-4.5	-4.3	-4.0	-3.6	-2.9	-2.7	-2.3	-2.0	-1.7	-1.4	-1.1	-1.0	-0.8	-0.5	-0.4	-0.2	0.0	0.1
-5.4	-5.4	-4.7	-4.6	-4.3	-3.8	-3.1	-3.0	-2.5	-2.3	-1.9	-1.7	-1.5	-1.3	-1.0	-0.8	-0.7	-0.6	-0.15	0.0

Figure 1-11: Portion of 79BAR/SCO Used to Demonstrate Iterative Algorithm: After Iterative Solution.
Source: Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979** 101, 6046-56.; Lias, S.G.; Bartmess, J.E.; Liebman, J.F.; Holmes, J.L.; Levin, R.D.; Mallard, W.G.; *J. Phys. Chem. Ref. Data*, **1988**, *17*, Suppl. 1.

values demonstrated by Figure 1-12.

The algorithm then begins with the first array element. If the array element has a known measured temperature, then that element is left alone and the algorithm moves to the next element in the first row. For the problem given in Figure 1-12, the algorithm would continue its search for unknown values until it reach the second row and second column. Unknown values are interpolated from the four nearest neighbors providing these values are known or have already been interpolated. Therefore, the unmeasured second row-second column element of the matrix will be the average of the first row-second column and the second row first column elements. The process continues until an initial value has been assigned to each of the unmeasured interpolants. The algorithm then cycles back to the first interpolant and begins updating the individual values. When the interpolants reach consistency on successive iterations

34.0	34.0	34.0	34.1	34.1	34.2	34.1	34.1	34.0	34.0	34.0
34.0	xx	xx	xx	xx	xx	xx	xx	xx	xx	34.1
33.9	xx	xx	xx	xx	xx	xx	xx	xx	xx	34.2
33.7	xx	xx	xx	xx	xx	xx	xx	xx	xx	34.4
33.4	xx	xx	xx	xx	xx	xx	xx	xx	xx	34.7
33.0	xx	xx	xx	xx	xx	xx	xx	xx	xx	35.0
32.6	xx	xx	xx	xx	xx	xx	xx	xx	xx	35.5
33.0	xx	xx	xx	xx	xx	xx	xx	xx	xx	35.0
33.7	xx	xx	xx	xx	xx	xx	xx	xx	xx	34.4
33.9	xx	xx	xx	xx	xx	xx	xx	xx	xx	34.2
34.0	xx	xx	xx	xx	xx	xx	xx	xx	xx	34.1
34.0	34.0	34.0	34.1	34.1	34.2	34.1	34.1	34.0	34.0	34.0

Figure 1-12: Engineering Template Problem Surface.

below some threshold value, the iterative process is terminated.

There are several algorithmic changes from this basic idea. First, one extra row and column have been added to the initial affinity matrix. Thus in the first row, the second column through the $(n+1)^{th}$ column becomes thesolution to the iterative smoothing algorithm. To demonstrate the difference, the scale depicted by Figure 1-13 will be shown as equation (1.90).

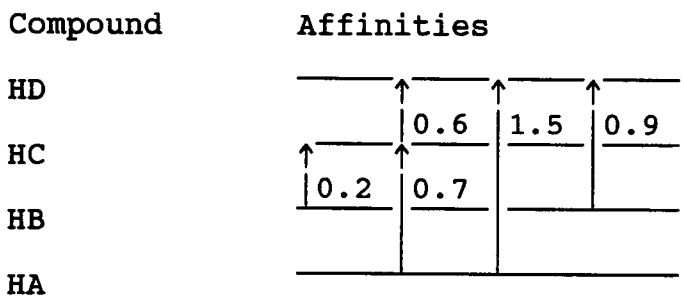


Figure 1-13: Scale Used to Demonstrate Iterative Solution *vs.* Engineering Template Problem.

$$\text{(1.90)} \qquad \mathbf{A} = \begin{array}{ccccc} \text{HA} & \text{HB} & \text{HC} & \text{HD} & \\ 0.00 & \text{XXXX} & \text{XXXX} & \text{XXXX} & \\ 0.00 & \text{XXXX} & 0.70 & 1.50 & \text{HA} \\ 0.00 & 0.00 & 0.20 & 0.90 & \text{HB} \\ -0.70 & -0.20 & 0.00 & 0.60 & \text{HC} \\ -1.50 & -0.90 & -0.60 & 0.00 & \text{HD} \end{array}$$

Without the additional row and column, it would be more difficult to assign the best fit to the data. The measured affinities for HA relative to HC and HA relative to HD do not depict the best fit values. Without the extra row,

these values would have to be masked off to obtain the best fit and then measurements re-inserted in order to obtain values for the rest of the row. Computationally, adding the extra row and column is more efficient.

The next difference between the iterative smoothing algorithm used by program LADDER and the engineering template problem is the order in which the interpolants are calculated. In the template problem, the order for calculations is done row by row, that is, all interpolants are calculated for the second row before the third row is calculated. In the LADDER algorithm, the procedure begins by assigning values to the nearest off-diagonal elements. Thus in equation (1.90) the value for $a_{2,3}$ is assigned before $a_{1,3}$ because the former is closer to the diagonal of zero affinities. The rationale here is that the interpolants furthest from the diagonal are often the ones known with the least amount of accuracy. Calculating interpolants on a row by row basis rather than on the off diagonal basis also resulted in abnormally low solution values.

The third difference is in the way the interpolants are averaged. The template method is an average over at most the four nearest neighbors. Use of the template algorithm's averaging method results in an unreasonable solution. The LADDER algorithm calculates an interplant by averaging over an entire row and column. For example, to calculate the

value of $a_{2,3}$, the sum of the entire second row and third column are considered. When updates to $a_{2,3}$ are considered, twice the value of the old $a_{2,3}$ is removed from the sum of the second row and third column. There is not much difference between the least squares algorithms and the iterative smoothing algorithm in terms of computational speed. The slow portion of the iterative smoothing algorithm is the initial filling of the **F** matrix. Once an initial value has been interpolated for each unmeasured value in the affinity matrix, the LADDER algorithm adds each column of the affinity matrix **F** in double precision arithmetic. This allows the algorithm to take advantage of the skew symmetric properties of the affinity matrix: the sum of the second row is equal but opposite to the sum of the second column. Therefore updating $a_{2,3}$ can be performed by subtracting the second column's double precision sum from the third column's sum. After each interplant is updated, the corresponding columns are also updated in double precision arithmetic. With this fast update procedure, the number of additions is reduced considerably, and solution time decreases.

The algorithm could be terminated in a number of ways. In the ideal **F** matrix, the changes made to each interpolated value become infinitesimal. In this ideal case, the various interpolants remain static over several iterations. The tolerance value for terminating the program could then be set to zero. In practice, zero tolerance is frequently

unattainable. Minor changes in the interpolants cause the first row, also called the edge, of the matrix to either shrink or increase with each additional iteration. The exact reasoning for the increasing/decreasing of the edge is not presently known (some conclusions are drawn in the results section). To compensate for a matrix which will not ease into stasis, the tolerance test checks the edge sum and the range to determine whether these values are steadily increasing or decreasing. The check is made both on the edge sum and also the last column of the edge row, called the range. Values for the edge sum and the range are kept for three iterative passes. Rather than comparing the values for the range over the course of three values to see that it has not changed, the relative differences in the range values for consecutive passes are compared. When the change in the differences between the range values over consecutive passes is less than the set tolerance, one criterion for algorithmic termination has been met. The other criterion is that the change in the differences for the edge vector sums is also either static, linearly increasing, or linearly decreasing within the set tolerance limits. When both criteria have been met at the same time, then the algorithm terminates. Most of the time the solution obtained from the iterative smoothing algorithm will be close to that obtained from the least squares algorithm already described.

There can be differences between the two algorithms, however. The least squares algorithm is more global in its scope[57] while the interpolants generated from the iterative smoothing process are more dependent on neighboring values. For an unmeasured affinity in the interior of the affinity matrix, the dependencies can cause changes between the interpolated value and the value calculated from the least squares solution. Therefore, while the least squares solution is often accepted as the best solution for the relative absolute affinities, the best guide in making new measurements is the iterative smoothing algorithm.

The second way by which differences exist between the least squares solution and the iterative smoothing edge is a result of repositioning of the affinity matrix. In the ideal case, each row of the affinity matrix should be monotonically increasing while each column is decreasing due to the skew symmetry of the matrix. In some real cases, the lack of monotonicity within the known values of the affinity matrix affects the monotonicity of the edge row. After the tenth iteration, the algorithm checks the edge to verify that it is monotonic. If the check warrants a switch between the position of two (or more) compounds, the algorithm then swaps the appropriate rows and columns. The check and swap process can occur via two pathways: either only the swap occurs or the interpolants of the matrix can be reset to their initial value after the swap. A user

supplied parameter controls how many times a complete re-initialization of the matrix can occur. The parameter was deemed necessary since, in the largest of scales the re-initialization/swap check would result in an infinite loop, between several different edge orderings.

X. LEAST TRIMMED SQUARES

The final analysis algorithm makes attempts at pointing out discrepancies within the data for further investigation by the researcher. The principal algorithm used here is the simplex algorithm following Nelder and Mead[82] and adapted by Chandler.[43] Beginning with an initial guess solution, the simplex procedure creates a n-dimensioned hyperspace, where n is the number of compounds.[82] The algorithm proceeds to minimize the error observed from the worst point in that hyperspace. The hyperspace can expand or contract as necessary in seeking the solution[82] until one of three termination criteria is met. The simplex optimization will terminate if 1) the error function returns the same value for each of the hyperspace solutions, 2) the amount that the hyperspace is expanding or contracting is less than some tolerance value, or 3) the number of calls to the error function exceeds a preset limit.[43]

Perhaps the greatest advantage of the simplex procedure is that it will minimize any error function. If the error

function is least squares, then the final solution vector will be very similar to the least squares solution vector (by very similar we mean that values in the two vectors are the same for the first 4 to 5 significant figures). Several error functions were tried. Prior to the discovery that the least squares solution could be obtained directly, the error function used was that of the least squares problem.

The second error function tried was that of least sums,[83-87] as the least sums algorithm has been used by the chemical community to isolate outliers.[88] The least sums error function used gave comparable results to the least sums algorithm of *Transactions on Mathematical Software #551*.[87] The latter algorithm depends on the simplex tableau used in linear programming. The principal disadvantage toward using the linear programming algorithm is that it requires at least n-measurements to be fit exactly, where n is again the number of compounds within the fusion set. Often this perfect-n-measurement fitting did not leave many measurements that the algorithm could average in and therefore make predictions regarding outliers. Thus the returned least sums solutions were often arbitrary with the warning that the absolute affinities for the compounds in the scale could be arranged in other ways and still arrive at the same error minima. The prospect of reporting several equivalent solutions is not desirable, but the fault lies with the limited amount of data available rather than with

the algorithm itself. Since most of the fusion sets are built from almost as many compounds as there are measurements connecting them, the least sums approach was finally abandoned. Other approaches toward finding outliers in the data exist, but many of these were not tried. The infinity norm approach[89] was not seriously considered as these algorithms tend to minimize outliers.[70]
The minimax approach was not attempted,[90] primarily due to failed attempts to incorporate the minimax algorithm available through the International Mathematic and Scientific Library of subroutines (IMSL)[91] into the program. Attempts to use the algorithm for ladders where the are more than 50 compounds resulted in runtime errors within the IMSL subroutines.

One approach tried briefly was to use fractional norms as according to equation (1.91).

$$(1.91) \quad N\text{-}norm = LN = [\Sigma\ \text{residual}_i^N]^{1/N}$$

This approach was abandoned in favor of the least trimmed squares procedure, after the pattern of Rousseeuw.[92] The least trimmed squares is called such because the largest of the residuals are actually removed from the error function. The process is as follows. Initially, the residual rms (root mean square of entire sum) is calculated as in equation (1.92).

(1.92) $rms = [\Sigma \text{residual}_i^2]^{0.5} = \|\text{residuals}\|_2 = \|\mathbf{b}-\mathbf{Ax}\|_2$.

One (or more) of the residuals which contribute to the *rms* may be much larger than the rest. The least trimmed squares algorithm begins with the assumption that the residuals should follow the standard Gaussian "bell curve" error function distribution. Since one particular measurement can skew the results, several methods have been proposed to point out these observations.[67,69,70] The least trimmed squares algorithm scale estimator σ is prepared based on equation (1.93).[93]

(1.93) σ = C*rms/(square root of number of measurements)

where C is a normalization constant. The individual residuals are then compared with σ, and if any residual is 2.5 times as large as σ then the residual can be temporarily discarded from consideration with 90% confidence as depicted by equation (1.94).[94]

(1.94) $$\begin{bmatrix} \text{residual}_i = \text{residual}_i & \text{if} & |\text{residual}_i/\sigma| < 2.5 \\ \text{residual}_i = 0 & \text{if} & |\text{residual}_i/\sigma| \geq 2.5 \end{bmatrix}$$

If the residual remains discarded from consideration, the measurement associated with that residual is a candidate for re-evaluation or further experiments as a poor measurement. The error function is then calculated as the trimmed rms, *trms*, of the new set of residuals as shown by equation (1.95),

$$(1.95)\quad trms = \left[\sum_{1}^{h} \text{residual}_i^2 \right]^{0.5} = \quad residual_2.$$

where h is the number of residual values remaining after application of (1.94). The value *trms* is the value returned from the simplex error function. If none of the residuals have been discarded from the least trimmed squares analysis, then the solution vector obtained should be very similar to that of the least squares procedure.

In practice, the least trimmed squares algorithm allows the largest of the residuals (the probable outliers) associated with a data set to be large without penalizing the overall fit. Therefore the existence of an outlier (either a poorly measured equilibrium measurement or a typographical data entry error) will not skew the fit.

CHAPTER 2

APPLICATIONS OF LADDER ALGORITHMS IN THE DETERMINATION OF ABSOLUTE THERMODYNAMIC AFFINITIES FROM EQUILIBRIUM VALUES

I. INTRODUCTION

The mathematical algorithms described in chapter one were written as two separate programs: LADDER and GRAPHLDR. The first program, LADDER, analyzes the data according the least squares (hereafter LS), iterative smoothing and least trimmed squares (hereafter LTS) algorithms. GRAPHLDR is used to construct a visual representation of the scale such as was seen in Figure 1-9.

The programs were tested on IBM-compatible personal computers and DEC VAX systems. The source codes for LADDER and GRAPHLDR primarily utilized an IBM PC with an Intel 80287 math coprocessor and the subset of the ANSI Fortran 77 code used by MICROSOFT FORTRAN 3.2.[93] The source code is not limited to the math coprocessor, although the PC analysis is more efficient when the coprocessor can be used.

Many of the scales used in testing various program attributes are compilations from the literature or single literature references. When the literature scale is unique to a paper, it will be coded by the year of publication

followed by the first three letters of the last names of the first two authors, for example, 92MIS/KAN.[12] Scales compiled from several papers will be coded by the type of affinity, for example, FDATA is the compiled data from several papers written during the course of Larson and McMahon's studies of fluoride affinities.[29,30,36]

II. CONVERSION OF FORTRAN CODES

Conversion of the source code from MICROSOFT FORTRAN 3.2 to DEC VAX FORTRAN[94] is relatively simple. The primary restriction of the PC version is due to the amount of memory used by the program. The program requires between 415 and 490 kbytes from the lower 640 kbytes of free memory to operate, as currently dimensioned for a scale containing 150 compounds and 300 equilibria. A minimum of three files are active during each program run, so it is recommended that at least ten files should be specified in the CONFIG.SYS when turning on the personal computer.

A second restriction toward using the PC version of the programs is the time it takes the central processing unit, (cpu), to process the data. Each of the analysis techniques has a solution speed which is proportional to a power function of the number of compounds within the fusion set.[57,82] LS is the fastest of the three algorithms,[57] while the solution speed of Simplex increases greatly with each

compound added to the scale.[82] When the scale either becomes larger than the currently allotted dimensions or the user finds it more convenient to use a faster cpu, then the conversion to the DEC VAX Fortran can performed using a five step process.

The first step of the conversion is to link the eleven different files of the main program and subroutines to a single file. The twelfth Microsoft file can be omitted entirely when upgrading to the DEC VAX as most main frame systems hold the Basic Linear Algorithms, BLAS, in memory for library compilations.[91] While not absolutely necessary, rewriting the eleven files as one unit allows further conversion to go more smoothly. The second step of the conversion process is to delete or to comment out the $metacommands used by the Microsoft compiler.[93] The $metacommands appear in the first few lines of each of the initial eleven files and handle special commands unique to the compiler such as storing integers as 2 bit words to save memory. The third conversion step is the expanding of the initial arrays if necessary. Array expansion tends to slow the solution time of even the main frame cpu. Therefore the array should remain untouched unless the user has a huge scale to analyze. In addition to the actual changing of the array size, the initialization of the arrays within individual subroutines will need appropriate expansion. The fourth step involves allowing the DEC VAX Fortran compiler[94]

to use a subroutine from the International Mathematic and Scientific Library, IMSL.[91] The subroutine that the program may call if the array size is limited to less than 50 compounds is LSBRR.[91] Subroutine LSBRR performs the iterative LS solution similar in nature to the one the program generates. The use of internal assembly language calls within LSBRR allows the library subroutine to obtain a slightly more exact solution than the one obtained from the iterative LS procedure encoded within the program. The trade off is that real success using LSBRR is limited to small scales. Attempts to use LSBRR to process large scales caused errors within the executable code, even when the allocated work space memory was explicitly declared.[91] The use of the LSBRR subroutine can be unlocked by removing comment statements from a few lines of the PC code. The final stage of the PC to VAX conversion is to rewrite the timing algorithm, Etime. For the PC, the timing algorithm first makes calls to the internal clock and then compares the returned time to a previously recorded time. Because the DEC VAX is not a dedicated machine, the use of real time to evaluate cpu speed is inaccurate. Instead, another IMSL subroutine CTIME is used to obtain the record of the cpu time.[91]

For comparative purposes, the Fortran BLAS[45] subroutines used for the personal computer version of the source code were also tested. The assembly language version

of the BLAS code[46] performs not only more efficiently than its Fortran counterpart, but also with higher precision in executing many of the floating point operations. Therefore, it is recommended to use the Assembly language version of these subroutines when the PC has the necessary coprocessor.

The following sections list the results of many tests of the program. Except where noted, the default parameters for the various algorithms were used. Section III gives examples about relative solution speeds, while later sections devote attention to fusion sets of various sizes.

III. ANALYSIS TIMES

Since the program requires a large number of mathematical operations to conduct the detailed analysis of equilibrium scales, some researchers would be interested in the amount of time the program takes when processing the data. For comparative purposes, the analysis times of four literature scales are found in Table 2-1. The central processing unit used to construct the Table 2-1 was an IBM PS/2 Model 50Z with an 8087 Intel math coprocessor at 12MHz speed. The LADDER program used for the timing process incorporated the assembly language codes for Basic Linear Algorithms. Solution times for each of the three main algorithms are reported. The number of cycles considered by each algorithm is also in Table 2-1. For LS, the time it

Table 2-1: Literature Scales Timed on the IBM PS/2 Model 50.

Scale			Lst Squares[a]		Iterative		Simplex		Total Time[c]
Fusion Set	No. Cpd.	No. Meas.	No. Pert.	Time[b]	Pass	Time[b]	Calls	Time[c]	
76AUE/WEB[d]									0:53.39
aziridine - $EtNH_2$	3	3	0	.11	3	0.00	113	0:00.39	
$nPrNH_2$ - $tamylNH_2$	9	14	0	.22	66	1.16	390	0:04.89	
$tamylNH_2$ - nPr_2NH	9	12	0	.16	50	.94	887	0:09.39	
iPr_2NH - sBu_2NH	3	3	0	.06	12	.06	91	0:00.33	
Et_3N - nBu_3N	6	7	0	.17	23	.17	682	0:03.84	
76YAM/KEB[e]									1:53.13
H_2O - H_2S	3	3	0	.11	12	.00	91	0:00.33	
HCO_2H - $mC_6H_4F_2$	3	3	6	.27	12	.05	82	0:00.22	
$mC_6H_4F_2$ - MeCHO	8	14	8	.50	40	.60	928	0:08.19	
EtOH - NH_3	16	31	14	1.82	129	6.31	2405	1:05.14	

Notes: [a] No. Pert. refers to the number of perturbations calls made by the algorithm.
[b] in seconds | [c] in minutes:seconds

Table 2-1: (continued)

Scale			Lst Squares[a]		Iterative		Simplex		Total Time[c]
Fusion Set	No. Cpd.	No. Meas.	No. Pert.	Time[b]	Pass	Time[b]	Calls	Time[c]	
89CHO/KIS[f]									4:19.91
4-CNNB - 3-MeNB	20	29	0	.61	56	5.16	6101	3:58.76	
89CAL/REN[g]									6:46.72
$MeCO_2H$ - $MeOCH_2CO_2H$	19	35	12	2.19	90	6.59	5567	3:58.76	
$MeOCH_2CO_2H$ - F_2CHCO_2H	18	35	12	2.03	28	3.62	3402	1:56.44	
F_2CHCO_2H - CF_3CO_2H	3	3	0	.11	12	.05	87	0:00.33	

Notes: [a] [b] No. Pert. refers to the number of perturbations calls made by the algorithm. in seconds ¦ [c] in minutes:seconds

Sources: [d] Aue, D.H.; Webb, H.M.; Bowers, M.T. *J. Am. Chem. Soc.* **1976**, *98* 311. ¦[e] Yamdagi, R.; Kebarle, P. *J. Am. Chem. Soc.* **1976**, *98*, 1320. ¦[f] Chowdhury, S.; Kishi, H.; Dillow, G.W.; Kebarle, P. *Can. J. Chem.* **1989**, *67*, 603. ¦[g] Caldwell, G.; Renneboog, R.; Kebarle, P.; *Can. J. Chem.* **1989**, *67*, 611.

takes the computer to solve the problem is often based upon the number of perturbations of the initial problem the algorithm solves. Perturbations occur when one or more residuals obtained from the initial solution is greater than the value of 0.1 kcal/mol. It takes less time to solve the fusion set of 20 compounds when no perturbation requests are made than to solve the fusion set of 16 compounds when 14 perturbations occurred. The calculation time performing the iterative smoothing procedure is based on the number of iterative passes, but it can be complicated by any number of re-initialization sweeps that the algorithm calculates. No re-initialization sweeps are necessary for the fusion sets listed from these four scales, so the analysis time is purely a reflection of the number of iterative passes. The Simplex solution's time is a reflection of the number of function calls the algorithm makes while calculating the LTS solution. The total amount of analysis time the LADDER program spent on each scale is also reported in Table 2-1. As can be seen by the final two scales, running the simplex algorithm on a large fusion set begins to take substantially longer times to process the data.

Table 2-2 was the result of several artificial scales which were generated for timing information. The artificial scales had 5, 10, 15, 20, 25 and 50 compounds each. In every synthetic scale each compound had a "measured" affinity of 0.1 to its nearest neighbor, a "measured"

Table 2-2: Synthetic Scales Timed on the IBM PS/2 Model 50.

Scale	Least Squares[a]		Iterative		Simplex		Total
	No. Pert.	time[b]	Pass	time[b]	Calls	time[c]	Time[c]
5 compounds 15 measure.	10	.44	25	.11	268	0:01.43	0:10.71
10 compounds 30 measure.	8	.76	60	1.10	1284	0:18.90	0:31.97
15 compounds 45 measure.	8	1.48	103	4.06	11523	5:24.34	5:44.44
20 compounds 60 measure.	8	2.47	135	9.56	4860	3:34.54	4:04.20
25 compounds	8	3.79	135	15.27	9639	10:30.87	11:11.24
50 compounds	8	16.20	51	42.02	13488	51:23.96	53:07.60

Notes: [a] No. Pert. refer to the perturbation calls by the algorithm.
[b] in seconds | [c] in minutes:seconds

affinity of 0.2 to the compound two acids away, and a "measured" affinity of 0.3 to the third compound above it. One error existed in each of these otherwise sets of perfectly closed Hess cycles. The third measurement from each of the prepared synthetic scales had an affinity exactly opposite of the one expected based on the remaining measurements. Reversing the direction of equilibria was observed to be a common mistake in the preparation of an initial data file. The data listing for the five compound case is given in Table 2-3 and the graph of this scale given by Figure 2-1.

IV. RESULTS BASED ON FUSION SETS

A. Pendant scales.

From the definition of pendants given in Chapter 1, one would think that by the time the program reaches the analysis stage all problems pertaining to pendant scales would have been removed. The smallest fusion set isolated would be of a closed Hess cycle between three compounds as given by equations (1.7) through (1.9). Analysis of literature scales found two additional problems. The scale of $FeBr^+$ affinities based on the work of Kappes and Staley[25] presented a new problem to the program. The ladder consists of an elaborate network of pendant measurements, without

Table 2-3: Data for the Synthetic Scale of Five Compounds.

Compound 1	Compound 2	ΔG_{eq} (kcal/mol)
a1	a2	0.1
a1	a3	0.2
a4	a1	0.3
a2	a3	0.1
a2	a4	0.2
a2	a5	0.3
a3	a4	0.1
a3	a5	0.2
a4	a5	0.1

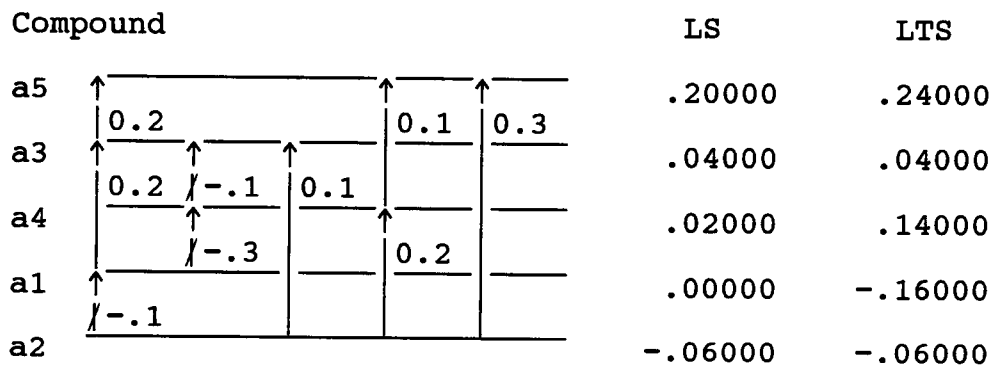


Figure 2-1: Synthetic Scale of 5 Compounds.

Initial data in disagreement with Least Squares solution
a1 is numerically larger than a2, by .06000
a4 is numerically larger than a1, by .02000
a3 is numerically larger than a4, by .02000
changing the way these measurements are graphed
/ symbol flags these changes in the graph.

closing (however imperfectly) any Hess cycles. Therefore, the pendant stripping algorithm originally described in Chapter 1 completely removes every single measurement. The pendant stripping algorithm had to be rewritten so it could process a completely pendant scale such as 82KAP/STA2.[25] Entirely pendant fusion sets or ladders are perhaps the most questionable in terms of reliability. If a single measurement is incorrect, then only the ladder between the anchor member and the incorrect measurement is valid. Relative affinities beyond the suspect measurement are incorrect, because these affinities are based on a poor affinity value. Thus in 82KAP/STA2, if the ΔG_{eq} between dimethyl ether and methyl formate is incorrect, every ΔG beyond methyl formate is also incorrect. Further, there is no way to statistically analyze the data to verify whether the dimethyl ether/methyl formate equilibrium is correct. Statistical analysis becomes important because the precision of pressure measurements is only 20%. The statistical error due to the equilibrium pressures of a binary system then leads to an overall error of at least 28% in the measurement. Therefore, increasing the number of measurements between compounds such that only a few, if any, pendants remain within the ladder becomes vital.

Pendant ladders can also arise when the program isolates fusion sets. For example, the program isolates a literature scale into three fusion sets X, Y and Z, where Y

connects X to Z. By itself, Y is a ladder of only pendant measurements, but the pendants were not removed since the lowest index in Y is tied to fusion set X while the last affinity index in Y is tied to fusion set Z. The absolute size of fusion set Y can vary. Most of these pendant sets are simply between the two extremes of more complicated ladders.[7,21] Most of the appearances of pendant fusion sets within the literature also have two or more connected pendant fusion sets in between more detailed fusion sets.[21-24] For example, the scale by Jones and Staley connects the fusion set ending with 2-bromopropane to the one beginning with 1-propanol.[21] The connection involves four connected pendant sets, as shown by Figure 2-2. Although Figure 2-2 shows six compounds between 2-bromopropane and 1-propanol, ethanol and propylene are pendant measurements that the program isolated and removed before the fusion subset search. One measurement connects 2-bromopropane with 2-bromobutane, which is then connected to dimethyl ether, which is connected to methyl formate, which is finally connected to 1-propanol.

In terms of the analysis, the program can only do two things to a pendant ladder. It can check the ladder for continuity, and it can construct the files necessary for a graphical display. For the detailed analysis, no statistical information can be obtained from LS, iterative smoothing, or LTS: the analysis will result in zero

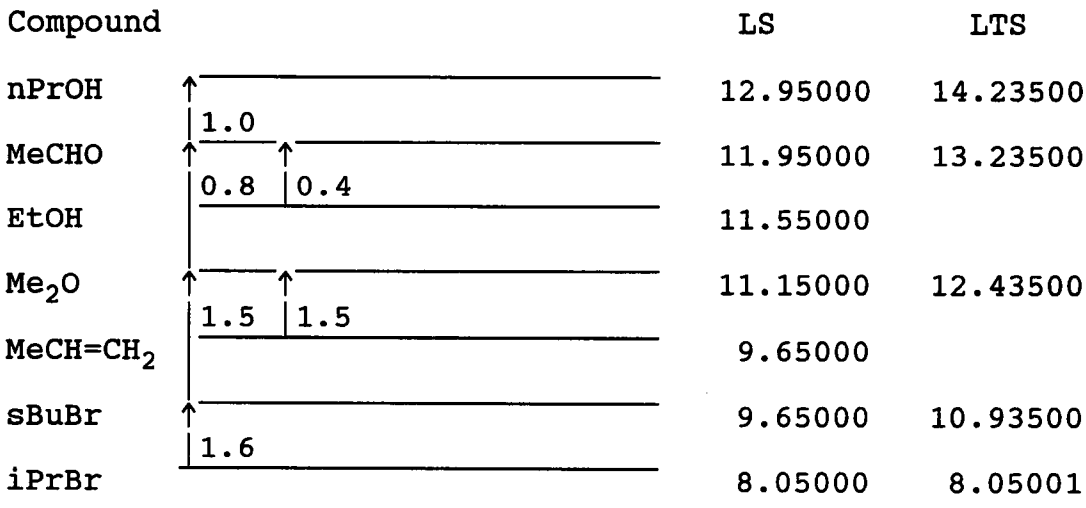


Figure 2-2: Portion of 82JON/STA: Extended Pendant System Source: Jones, R.L.; Staley, R.H. *J. Am. Chem. Soc.* **1982** 104, 2296.

residuals. Within the literature references, pendant fusion sets usually contain only one compound. For the sake of efficiency in cases where only one measurement ties two otherwise disjoint fusion sets, the program simply stores the values and ignores the analysis stage of the two-compound ladder.

B. Three-Membered Scales.

The simplest fusion set which is marginally acceptable is a closed Hess cycle between three compounds, such as given by equations (1.7) through (1.9). Many of these fusion sets exist within the literature as shown in Table 2-4, which presents a short summary of the results from the LS algorithm. Within Table 2-4 and subsequent tables, the

Table 2-4: Least Squares Summary of Fusion Sets with Three Compounds.

Scale	Set	# Pert	L2 norm	Range
82JON/STA[a]	MeCOEt - Et_2CO	0	.0577350	1.5333
82KAP/STA[b]	Me_2CO - $EtCO_2Et$	0	.01154705	2.1233
82KAP/STA	MeCN - EtCN	0	.1674316	2.1467
82UPP/ST2[c]	MeSH - EtSH	3	.1905256	1.3800
82UPP/STA[d]	MeCOEt - Et_2CO	0	.04618803	1.3167
88OPE/TEW[e]	EtOH - nPrOH	0	.01732052	1.1800
89CAL/REN[f]	F_2CHCO_2H - CF_3CO_2H	0	.1154700	6.4333
76YAM/KEB[g]	H_2O - H_2S	0	.05773504	3.0333
76YAM/KEB	HCO_2H - $mC_6H_4F_2$	3	.2309402	3.6333
77WOL/STA[h]	H_2O - H_2S	0	.05773511	3.6667
77WOL/STA	$CF_3CH_2NH_2$ - NH_3	0	.1154700	1.6667
76AUE/WEB[i]	aziridine - $EtNH_2$	0	.91694E-22	1.1000
76AUE/WEB	iPr_2NH - sBu_2NH	0	.1154701	1.9667
FDATA[j]	tBuF - tBuCHO	0	.34413E-07	1.5000
87KEB/CHO[k]	4-CNNB - 4-NO_2NB	0	0.1732051	6.4000
87KEB/CHO	4-NO_2NB - 2,3-Cl_2NpQ	0	.05773504	4.6333
91MEO/SIE[l]	MeOH - C_6H_6	3	.2309401	1.4333
94KOP/TAF[m]	Tf_2NH -(CF_3CO)TfNH	0	.0577349	4.0667
94KOP/TAF	CF_3COSH-$PhCH(CN)_2$	1	.1732051	1.6000

Notes: Affinity values are in kcal/mol. # Pert. refers to the perturbation calls made by the least squares algorithm.

Sources:[a] Jones, R.L.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 2296. [b] Kappes, M.M.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 1813. [c] Uppal, J.S.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 1238. [d] Uppal, J.S.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 1235. [e] Operti, L.; Tews, E.C.; Freiser, B.S. *J. Am. Chem. Soc.* **1988** *110*, 3847. [f] Caldwell, G.; Renneboog, R.; Kebarle, P.; *Can. J. Chem.* **1989** *67*, 611. [g] Yamdagi, R.; Kebarle, P. *J. Am. Chem. Soc.* **1976** *98*, 1320.

Table 2-4: (continued)

[h] Wolf, J.A.; Staley, R.H.; Koppel, I.; Taagepera, M.; McIver, R.T., Jr.; Beachamp, J.L.; Taft, R.W. *J. Am. Chem. Soc.* **1977** *99*, 5417. [i] Aue, D.H.; Webb, H.M.; Bowers, M.T. *J. Am. Chem. Soc.* **1976** *98*, 311. [j] Larson, J.W.; McMahon, T.B. *J. Am. Chem. Soc.* **1983**, *105*, 2944.; *J. Phys. Chem.* **1984** *88*, 1083.; *J. Am. Chem. Soc.* **1985** *107*, 766. [k] Kebarle, P., Chowdhury *Chem. Rev.* **1987** *87*, 513. [l] Meotner, M.; Siek, L.W. *J. Am. Chem. Soc.* **1991** *113*, 4448. Affinities are ΔG's [m] Koppel, I.A.; Taft, R.W.; Anvia, F.; Zhu, S.-Z.; Hu, L.-Q.; Sung, K.-S.; DesMarteau, D.D.; Yagupolskii, L.M.; Yagupolskii, Y.L.; Ignat'ev, N.V.; Kondratenko, N.; Volkonskii, A.Yu.; Vlasov, V.M.; Notario, R.; Maria, P.C.; *J. Am. Chem. Soc.* **1994**, *116*, 3047.

term #Pert. refers to the perturbation calls requested by the LS algorithm whenever residuals greater than 0.1 kcal/mol are encountered. The range is the relative affinity difference (usually ΔG) between the two compounds listed. Unless specified, the value of the range is given in kcal/mol.

The analysis portions of the program are somewhat limited due to the lack of measurements in the original measurement matrix. The maximum number of possible measurements between three compounds is only three. Multiple measurements of each of the equilibria can be performed, but the multiple tests on any equilibrium lead to a statistical evaluation of a single measurement, not to the entire ensemble of ΔG_{eq}'s.

If any residual is found from an imperfectly closed Hess cycle in the three compound case, the residual is then evenly split between the three measurements. The normalized residuals for the three measurements all have exactly the same magnitude: the square root of one third.

Only four of the scales in Table 2-4 contained a measurement where there was more than 0.3 kcal/mol disagreement among the measured ΔG_{eq}'s. These four scales are readily identified in that the perturbation portion of the LS algorithm was called at least once by the program. As each measurement has exactly the same influence on the LS solution, testing for large residuals gives no useful

information as to which of the measurements should be re-investigated.

Since the measurement matrix is completely filled for the iterative smoothing technique, the resulting solution for this algorithm is identical to that obtained from LS to at least five significant figures, far greater than the precision of the measurements. Termination of the iterative algorithm occurs when the values in the edge row do not change after several cycles, resulting in a static matrix.

The three compound case also appears to be immune to LTS, that is, no measurement was temporarily discarded in favor of the other two. The breakdown point of the algorithm has been exceeded. The returned solution for each of the three-compound scales resembles its respective LS solution to at least four significant figures, which again far exceeds the precision of the measurement. The observed difference seen in some of the three compound scales is probably due to a localized minima near the global minimum.

Three compound scales, while very popular within the literature and useful in demonstrating the benefits of the programs LADDER and GRAPHLDR, should be avoided in general practice. The principal problem lies in the relative influence of each component within the ladder. If one of the experiments is suspect, then each of the three equilibria must be re-investigated. The program is unable to rationalize which of the three measurements causes the

appearance of large residual values in the LS solution. Where possible, three member ladders should be extended to larger fusion sets.

C. Four-Membered Scales.

Four-membered fusion sets are also common in the literature. The number of measurements, whether it be three, four, five or six, determine the information which can be obtained. The results from the LS algorithm on each of these fusion sets appears as Table 2-5, while graphs of each of the fusion sets appear in Figures 2-3 to 2-11.

The first case is the three measurement case found connecting propanal to 2-methylpropanal for Ni^+ complexes studied by Kappes and Staley.[24] This three measurement case describes the only pendant fusion set larger than a single measurement found within the literature scales. A graph of this small fusion set appears as Figure 2-3. The main distinguishing feature of the larger pendant fusion set is that there is exactly one less measurement than the number of compounds within the fusion set. The mathematical definition of an entirely pendant fusion set would be the minimal basis set or linear combination of the minimal basis set as described by equation (1.42) in Chapter One. The program can do absolutely nothing for the statistical evaluation of such data. The program analysis simply

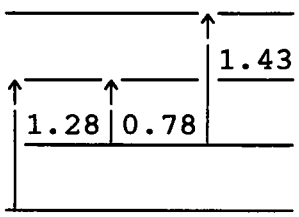
Table 2-5: Least Squares Summary of Fusion Sets with Four Compounds.

Scale	No. Meas.	Set	# pert	RMS	Range1
82KAP/STA[a]	3	EtCHO - iPrCHO	0	.0000000	1.9300
82JON/STA[b]	4	EtCHO - iPrOH	0	.9999999E-01	1.6000
82KAP/STA	4	MeCO2Me - Me2CO	0	.3464098E-01	2.1300
82KAP/STA	4	EtCO2Et - Et2CO	0	.2999997E-01	2.2750
82UPP/ST2[c]	4	MeCHO - EtCHO	0	.1500000E-01	1.3350
82UPP/STA[d]	4	HCO2nPr - Me2CO	0	.3499999E-01	.8850
82UPP/STA	4	EtCN - iPrCHO	0	.7000000E-01	.8950
77WOL/STA[e]	4	CF3CO2Et - MeCHO	0	.9999996E-01	1.7500
94KOP/TAF[f]	4	Tf2CH2 - m-CNPhCH(CN)2	4	.3500000	2.7500
Document.[g]	4	d - h	0	.2485076E-15	.7000
8UPP/STA	5	Me2CO - MeCOEt	0	.8660250E-01	1.5300
92MIS/KAN[h]	5	pCF3 - C6H5CHO	2	.1732051	3.9000
93DEC/GAL[i]	5	MeSH - EtSH	4	.9658414	7.6825
94KOP/TAF[f]	5	CF3CO2H - HBr	3	.2915476	1.8500
Document.[g]	5	a - d	0	.6123724E-01	.4875
77WOL/STA[e]	6	H2S - H2CO	1	.2121320	1.1750

Table 2-5: (continued)

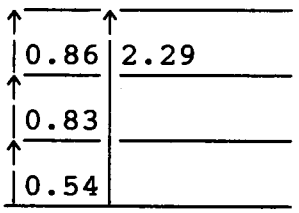
Notes: Affinity values are in kcal/mol except for 93DEC/GAL which is in kJ/mol.
Pert. refers to the perturbation calls made by the least squares algorithm.
Sources: [a] Kappes, M.M.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 1813. [b] Jones, R.L.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 2296. [c] Uppal, J.S.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 1238. [d] Uppal, J.S.; Staley, R.H. *J. Am. Chem. Soc.* **1982**, *104*, 1235. [e] Wolf, J.A.; Staley, R.H.; Koppel, I.; Taagepera, M.; McIver, R.T., Jr.; Beachamp, J.L.; Taft, R.W. *J. Am. Chem. Soc.* **1977** *99*, 5417. [f] Koppel, I.A.; Taft, R.W.; Anvia, F.; Zhu, S.-Z.; Hu, L.-Q.; Sung, K.-S.; DesMarteau, D.D.; Yagupolskii, L.M.; Yagupolskii, Y.L.; Ignat'ev, N.V.; Kondratenko, N.; Volkonskii, A.Yu.; Vlasov, V.M.; Notario, R.; Maria, P.-C.; *J. Am. Chem. Soc.* **1994**, *116*, 3047. [g] Synthetic scale used in the documentation of program LADDER. [h] Mishima, M.; Kang, C.H.; Fujio, M.; Tsuno, Y. *Chem. Lett.* **1992**, 493. [i] Decouzon, M.; Gal, J.; Gayraud, J.; Maria, P.; Viaglo G.; Volpe P. *J. Am. Soc. Mass Spectrom.* **1993** *4*, 54.

Compound		LS	LTS
iPrCHO		9.27333	8.76260
nPrCHO		8.62333	8.11260
iPrOH		7.84333	7.33260
EtCHO		7.34333	6.83260



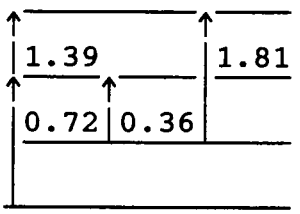
(a)

Compound		LS	LTS
Et_2CO		17.27167	16.76094
Me_2S		16.39667	15.88593
MeCOEt		15.55167	15.04094
$EtCO_2Et$		14.99667	14.48594



(b)

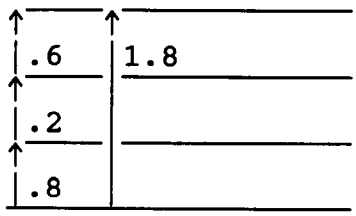
Compound		LS	LTS
Me_2CO		12.87333	12.36261
tBuCHO		11.46333	10.95261
Et_2O		11.08333	10.57261
$MeCO_2Me$		10.74333	10.23260



(c)

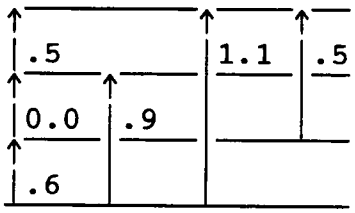
Figure 2-3: Portions of 82KAP/STA: Four Compound Scales.
Source: Kappes, M.M.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 1813.
Relative bond dissociation energies for two-ligand complexes of Ni^+ with organic molecules in the gas phase.
(a) Three equilibria, (b) and (c) four equilibria.

Compound	LS	LTS
MeCHO	16.12893	16.16393
CF_3CO_2nBu	16.77893	16.81392
CF_3CO_2nPr	17.02893	17.06392
CF_3CO_2Et	17.87893	17.91392



(a)

H_2CO	25.94893	26.02392
H_2Se	26.37394	26.52392
HCN	26.44893	26.52392
H_2S	27.12394	27.12392



(b)

Figure 2-4: Portions of 77WOL/STA: Four Compound Scales. Source: Wolf, J.A.; Staley, R.H.; Koppel, I.; Taagepera, M.; McIver, R.T., Jr.; Beachamp, J.L.; Taft, R.W. *J. Am. Chem. Soc.* **1977** *99*, 5417.

Gas phase basicities. (a) four equilibria (b) six equilibria.

Compound	LS	LTS
iPrOH	15.55000	16.83503
nPrCHO	15.30000	16.58502
nBuOH	14.50000	15.78502
EtCHO	13.95000	15.23500

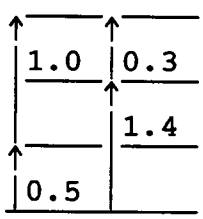


Figure 2-5: Portions of 82JON/STA: Four Compound Scales. Source: Jones, R.L.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 2296.

Relative bond dissociation energies for two-ligand complexes of Cu^+ with organic molecules in the gas phase.

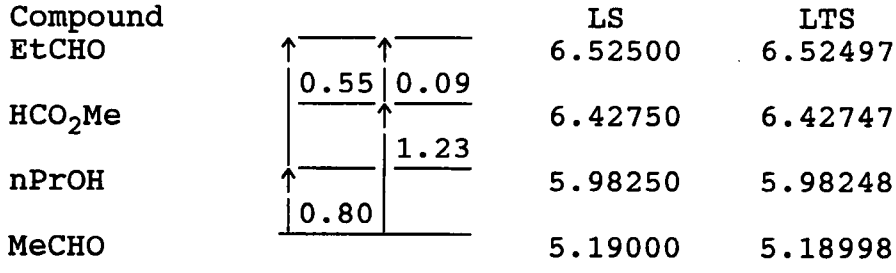


Figure 2-6: Portion of 82UPP/STA2: Four Compound Scale. Source: Uppal, J.S.; Staley, R.H. *J. Am. Chem. Soc.* **1982**, *104*, 1238.

Relative binding energies of Mn^+ with organic molecules in the gas phase.

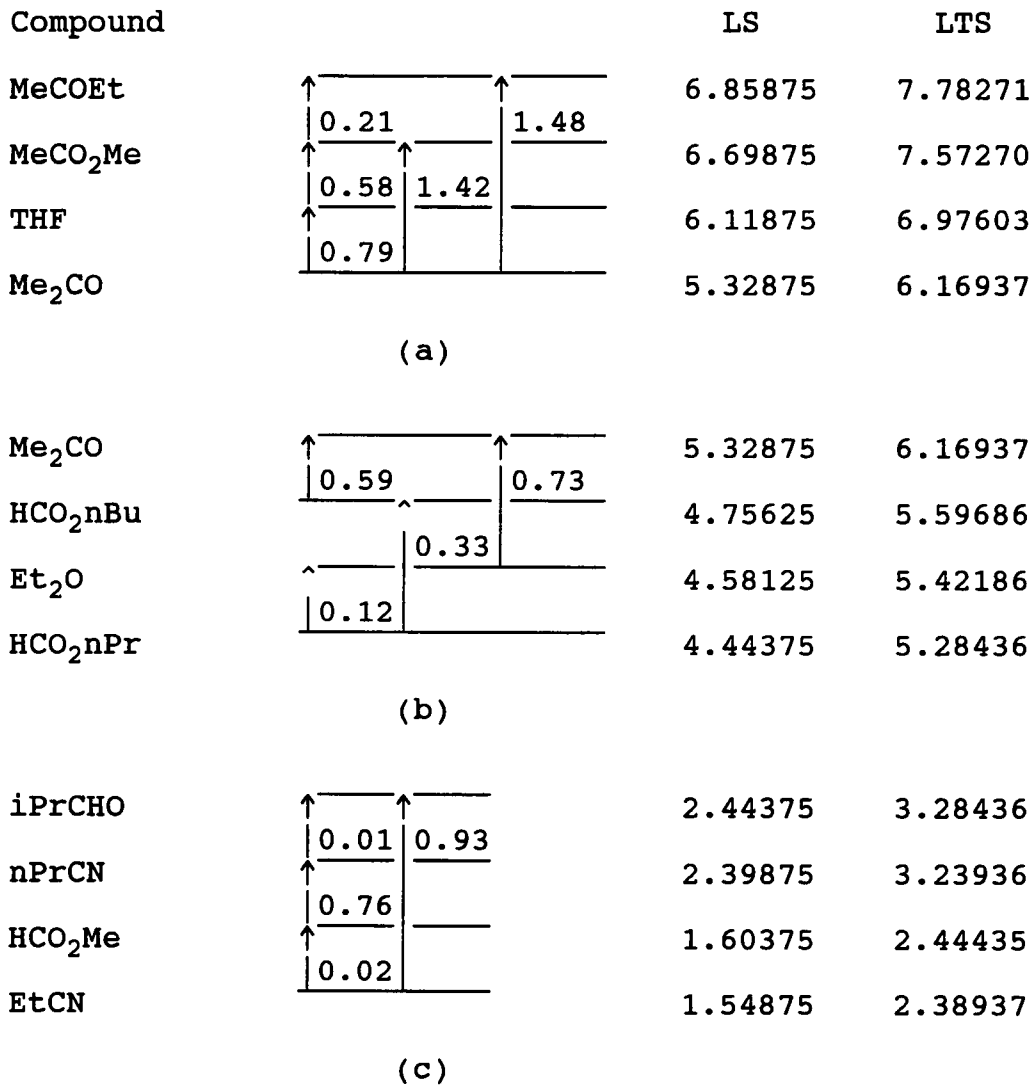


Figure 2-7: Portions of 82UPP/STA: Four Compound Scales Source: Uppal, J.S.; Staley, R.H. *J. Am. Chem. Soc.* **1982**, *104*, 1235.

Relative binding energies of Al+ with organic molecules in the gas phase. (a) five equilibria, (b) and (c) four equilibria.

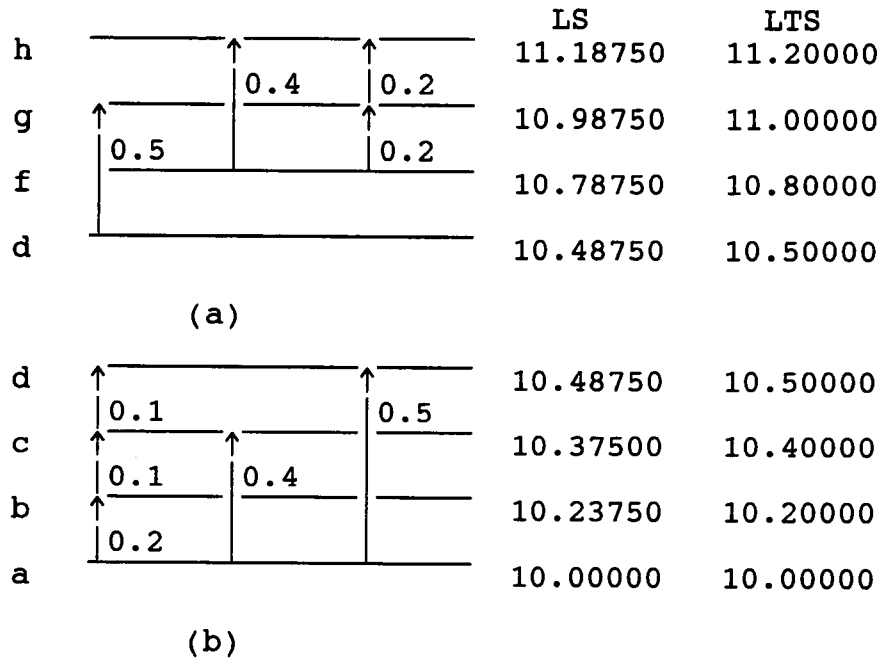
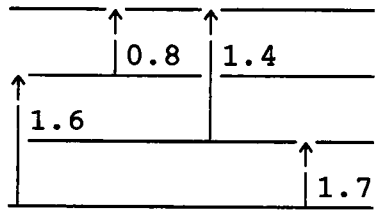
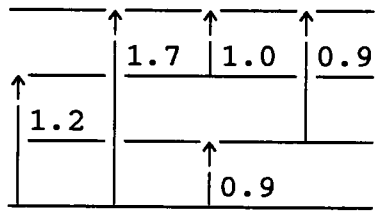


Figure 2-8: Portions of the Synthetic Scale Used in the Documentation of LADDER: Four Compound Scales.
(a) four equilibria (b) five equilibria

Compound		LS	LTS
m-$CNPhCH(CN)_2$		304.30020	304.26200
$(C_3F_7CO)_2CHCF_3$		303.32520	303.28700
m-$NO_2PhCH(CN)_2$		303.07520	303.03710
Tf_2CH_2		301.55020	301.51200

(a)

HBr		318.20000	320.35220
$3,5(CF_3)_2$pyrazo		317.37500	319.35220
8Ffluorene		317.27500	319.48560
CF_3CO_2H		316.35000	318.61890

(b)

Figure 2-9: Portions of 94KOP/TAF: Four Compound Scales. Source: Koppel, I.A.; Taft, R.W.; Anvia, F.; Zhu, S.-Z.; Hu, L.-Q.; Sung, K.-S.; DesMarteau, D.D.; Yagupolskii, L.M.; Yagupolskii, Y.L.; Ignat'ev, N.V.; Kondratenko, N.; Volkonskii, A.Yu.; Vlasov, V.M.; Notario, R.; Maria, P.-C.; *J. Am. Chem. Soc.* **1994**, *116*, 3047.

Lower end of acidity scale from HBr to $(C_4F_9SO_2)_2NH$.
(a) four equilibria (b) five equilibria

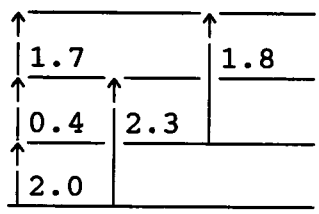
Compound		LS	LTS
C_6H_5CHO		-1.56667	-1.43334
mCl		-3.16667	-3.13333
mF		-3.46667	-3.49999
pCF_3		-5.46667	-5.46667

Figure 2-10: Portion of 92MIS/KAN: Four Compound Scale. Source: Mishima, M.; Kang, C.H.; Fujio M.; and Tsuno, Y. *Chem. Lett.* **1992**, 493.

Gas phase basicities of Acetophenones toward TMS cation. C_6H_5CHO is used as reference base.

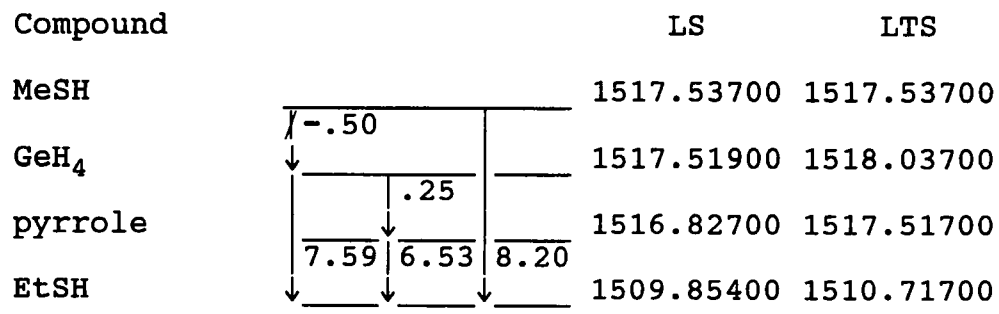


Figure 2-11: Portion of 93DEC/GAL: Four Compound Scale. Source: Decouzon, M.; Gal, J.; Gayraud, J.; Maria, P.; Viaglo G.; Volpe P. *J. Am. Soc. Mass Spectrom.* **1993** *4*, 54.

Gas phase acidities of germane.
Initial data in disagreement with Least Squares solution
GeH_4 is numerically smaller than MeSH by .01794
changing the way this measurement is graphed
/ marks the change within the graph.

evaluates and stores the fusion set with absolutely no residual. The simplex subroutine proves to be the least accurate of the three major algorithms, although the lack of accuracy is trivial. The minimum returned by the simplex algorithm for the root mean square (L2 norm) is the machine epsilon instead of zero.

Nine four-compound four-measurement fusion sets were found. The diethyl ketone - ethyl acetate fusion set of 82KAP/STA[24] is similar to the pendant fusion sets except that a single measurement connects the extreme compounds, thereby initiating statistical analysis. A similar four-equilibria fusion set in Table 2-5 comes from the 1977 Wolf *et. al.* paper on gas phase basicities.[6] Four measurements are found between ethyl trifluoroacetate and acetaldehyde. One measurement connects the two extreme compounds, while the other three measurements connect the two extremes by a separate path. Without the connection between the first and fourth measurement, the fusion set would again be three connected pendant fusion sets as described in section C. A third such scale is found in the EtCN - iPrCHO fusion set of 82UPP/STA.[23]

The fusion set between methyl acetate and acetone in 82KAP/STA[24] when examined proves to be nothing more than a three compound closed Hess cycle with an extra measurement attached from methyl acetate to the middle member of the smaller cycle as shown by Figure 2-3. A similar pattern was

used in the small scale to document of the LADDER program (Figure 2-8). In terms of analysis, both of these fusion sets are actually three compound scales with a single but nonpendant measurement fused to it. Hence the relative affinity values for the second through fourth compounds within the scale are completely dependent on the accuracy of the measurement between the first and the third compounds.

The remaining fusion sets all show a pattern of no measurements connecting the second and third compound or first and fourth compounds. Only the extreme values have two independent pathways connecting them. Therefore no statistical predictions can be made on the relative affinities of the second or the third compounds in the scale.

The results from iterative smoothing and LTS prove to be of little value for the four measurement case. There is simply not enough information to give any result other than the original LS result. Therefore, these results agree with the LS analysis to at least four significant figures, which again exceeds the original experimental precision. Perhaps the only important item of note in the supplemental algorithms is that the iterative smoothing algorithm occasionally detects an error in the row monotonicity in the measurement matrix as ordered by the LS solution. The error within the matrix is not apparent in the final solution edge, but the iterative smoothing acknowledges the difference.

The LS data from the four-compound five-measurement case is listed for four literature scales and for one synthetic one.

The iterative smoothing routine is still insensitive to alterations from the LS algorithm, although it does list problems dealing with row monotonicity. The big difference between the four-measurement and the five-measurement case is in the ability of the LTS routine to suggest suspect equilibria. In each scale the LTS algorithm selects one measurement as suspect. It then tries to minimize the residuals from the remaining measurements while allowing the suspect measurement to grow as large as possible. The suspect value suggested by the LTS program cannot be pinpointed as the worst measurement with complete certainty. In fact, a different set of simplex control parameters including the initial step size and the initial guess solution may affect which measurement the program suggests as the suspect value. There are still too few measurements for the algorithm to make an exact judgement in every case. It should be noted, however, that the LTS algorithm does point out significant errors in five-measurement scales. For the article concerning the gas phase acidities of germanes,[5] the calculated LS and LTS solutions differ greatly as shown by Figure 2-11. The initial data is in disagreement with the LS solution because the latter is skewed in favor of reducing the total overall residual. The

actual measured affinities agree better with the LTS solution. In light of the experimental uncertainties given by Decouzon *et. al.*,[5] a thorough re-investigation of the chemistry in this fusion set is in order. The probe could begin by measuring the comparative acidity between MeSH and pyrrole. This equilibrium has been previously measured,[3] but the repetition might be useful. If, once this equilibrium is remeasured, the affinity for the pathway going from MeSH to pyrrole to EtSH agrees with the direct measurement between these values, then the suspect measurement will be the MeSH - GeH_4 equilibrium.

Only one reference, 77WOL/STA,[6] is given in the literature for a ladder of four compounds and six measurements. The ladder is shown by Figure 2-4. This fusion set between H_2S and H_2CO is perhaps the best statistically designed in all of the literature. The well designed scale has a fair closure, but the LTS correctly points out that all of the error can be attributed to a single measurement of 0.9 kcal/mol between H_2S and H_2Se. According to the affinities from the other pathways, this measurement is expected to be 0.6 kcal/mol. Every compound within the scale has three separate pathways connecting it to the first compound within the scale. As previously mentioned, all of the error can be attributed to one equilibrium. The perturbation method of the LS algorithm also points out this solitary suspect measurement. Because

the scale is well-ordered, the iterative smoothing algorithm does not deviate from the LS solution.

D. Five-membered Scales.

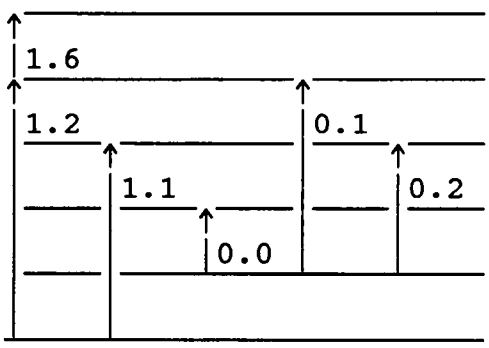
Fewer fusion sets exist within the literature references which contain five compounds. A summary of the LS results for nine five-compound sets appears in Table 2-6. Graphs of these fusion sets appear as Figure 2-1 for the synthetic scale and Figures 2-12 to 2-18 for the remaining scales.

The first of the five-compound scales to be investigated is the set of Cu^+ affinities between isobutyl bromide and sec-butyl bromide from 82JON/STA.[21] This five-compound, five-measurement ladder is actually a four-compound, four-measurement fusion set which is connected to a single measurement fusion set as shown in Figure 2-12. The reason the scale is observed as a five-membered ladder as opposed to two separate ladders is due to the relative affinities of oxirane and isopropyl bromide equilibrium. Here the ΔG is zero between the species according to the calculated LS values. The fusion set recognition algorithm is currently unable to discern the numerical values pertaining to the compounds in a ladder. Instead, it makes the assumption that the order given to it is one where the listed compounds are monotonically increasing (or

Table 2-6: Least Squares Summary of Fusion Sets with Five Compounds.

Scale	# Eq	Fusion Set	# Pert.	Rms	Range
82JON/STA[a]	5	iBuBr - sBuBr	0	.1000000	2.7500
82UPP/STA[b]	6	EtOH - EtCN	0	.1289865	1.5488
87KEB/CHO[c]	7	C_6F_6 - NB	5	.5329166	10.950
79BAR/SCO[d]	7	MeOH - EtOH	0	.1362770	3.0857
77WOL/STA[e]	7	nPrCHO - HCO_2Et	1	.1767767	1.4875
85MCM/KEB[f]	7	CF_3CN - H_2S	0	.1234427	5.2905
94KOP/TAF[g]	7	$(C_4F_9SO_2)$TfNH - Tf_2NH	0	.1447494	4.5048
94KOP/TAF	7	(CF_3CO)TfNH - Tf_2CH_2	1	.1851641	3.0857
SYNTHETIC	9	a2 - a5	5	.4381781	.2600

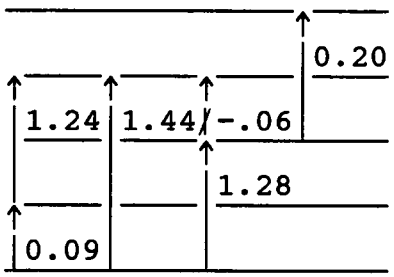
Notes: Affinity values are in kcal/mol. # Pert. refers to the perturbation calls made by the least squares algorithm.
Sources:[a] Jones, R.L.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 2296. [b] Uppal, J.S.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 1235. [c] Kebarle, P., Chowdhury *Chem. Rev.* **1987** *87*, 513. [d] Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979** *101*, 6046. [e] Wolf, J.A.; Staley, R.H.; Koppel, I.; Taagepera, M.; McIver, R.T., Jr.; Beachamp, J.L.; Taft, R.W. *J. Am. Chem. Soc.* **1977** *99*, 5417. [f] McMahon, T.B.; Kebarle, P. *J. Am. Chem. Soc.* **1985** 107, 2612. [g] Koppel, I.A.; Taft, R.W.; Anvia, F.; Zhu, S.-Z.; Hu, L.-Q.; Sung, K.-S.; DesMarteau, D.D.; Yagupolskii, L.M.; Yagupolskii, Y.L.; Ignat'ev, N.V.; Kondratenko, N.; Volkonskii, A.Yu.; Vlasov, V.M.; Notario, R.; Maria, P.-C.; *J. Am. Chem. Soc.* **1994**, *116*, 3047.



Compound	LS	LTS
sBuBr	9.65000	10.93500
iPrBr	8.05000	8.05001
Oxirane	8.05000	8.04998
$MeNO_2$	7.90000	
MeOH	7.90000	7.89999
iBuBr	6.90000	6.90000

Figure 2-12: Portion of 82JON/STA: Five Compound Scale Source: Jones, R.L.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 2296.

Relative bond dissociation energies for two-ligand complexes of Cu+ with organic molecules in the gas phase.



Compound	LS	LTS
EtCN	1.54875	2.38937
nPrOH	1.35750	1.35747
EtCHO	1.34875	1.34874
C_2H_3CN	0.10375	0.10375
EtOH	0.00000	0.00000

Figure 2-13: Portion of 82UPP/STA: Five Compound Scale. Source: Uppal, J.S.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 1235.

Relative binding energies of Al^+ with organic molecules in the gas phase.

/ Marks the exoergonic reaction.

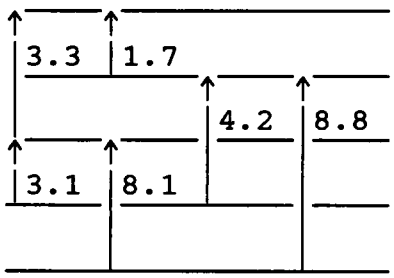
Compound		LS	LTS
NB		-2.20909	-2.57171
$2CH_3NB$		-4.06909	-4.35918
$2,3Me_2NB$		-5.34909	-5.78417
Az		-8.35909	-8.72169
C_6F_6		-13.15909	-13.15909

Figure 2-14: Portion of 87KEB/CHO: Five Compound Scale. Source: Kebarle, P., Chowdhury *Chem. Rev.* **1987** *87*, 513.

Electron affinities of nitrobenzenes.

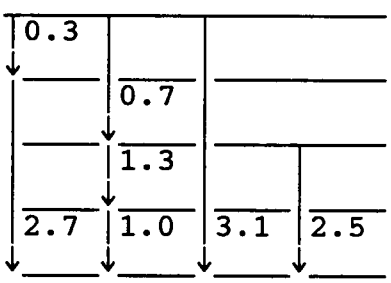
Compound		LS	LTS
MeOH		374.00000	374.00000
PhMe		373.65710	373.65000
MeCCH		373.32860	373.35000
nPrCCH		371.97140	372.05000
EtOH		370.91430	370.90000

Figure 2-15: Portion of 79BAR/SCO: Five Compound Scale. Source: Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979** *101*, 6046.

Gas Phase Acidities

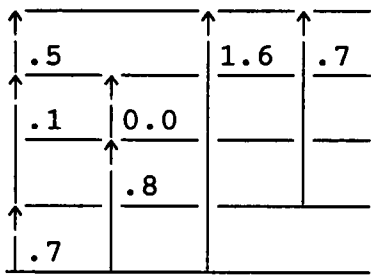
Compound		LS	LTS
HCO_2Et		10.02803	10.17534
iPrCHO		10.61553	10.71170
nBuCHO		10.66553	10.72078
Me_2O		10.75303	10.83897
nPrCHO		11.51553	11.52988

Figure 2-16: Portion of 77WOL/STA: Five Compound Scale. Source: Wolf, J.A.; Staley, R.H.; Koppel, I.; Taagepera, M.; McIver, R.T., Jr.; Beachamp, J.L.; Taft, R.W. *J. Am. Chem. Soc.* **1977** *99*, 5417.

Gas phase basicities.

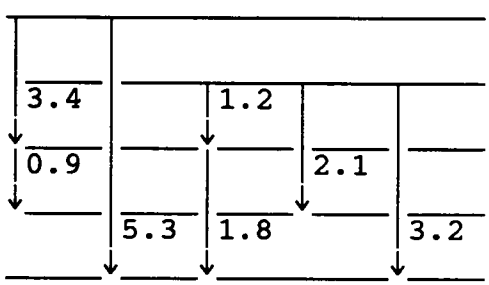
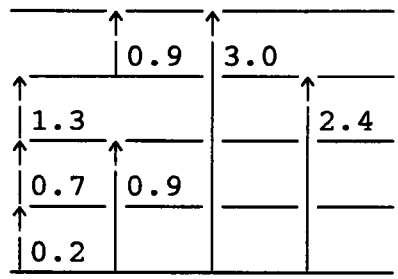
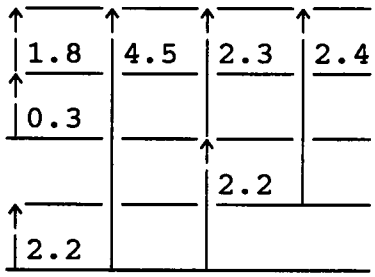
Compound		LS	LTS
CF_3CN		-2.12381	-1.61716
H_2O		-4.28572	-3.85049
CS_2		-5.53333	-5.05050
CF_3CH_2OH		-6.40952	-5.95051
H_2S		-7.41429	-6.88383

Figure 2-17: Portion of 85MCM/KEB: Five Compound Scale. Source: McMahon, T.B.; Kebarle, P. *J. Am. Chem. Soc.* **1985** 107, 2612.

Weaker end of the Basicity scale



Compound	LS	LTS
Tf_2CH_2	301.55020	301.51200
$(CF_3CO)_3CH$	300.73600	300.63930
$(C_3F_7CO)_3CH$	299.39310	299.36660
Tf_2CHPh	298.67880	298.67570
(CF3CO)TfNH	298.46450	298.48480



Compound	LS	LTS
Tf_2NH	294.39790	294.41810
Tf_3CH	292.51690	292.61810
$(C_4F_9SO_2)_2CH_2$	292.13600	292.10560
$CF_2(CF_2SO_2)_2NH$	292.04550	292.05560
$(C_4F_9SO_2)TfNH$	289.89310	289.89310

Figure 2-18: Portion of 94KOP/TAF: Five Compound Scales. Source: Koppel, I.A.; Taft, R.W.; Anvia, F.; Zhu, S.-Z.; Hu, L.-Q.; Sung, K.-S.; DesMarteau, D.D.; Yagupolskii, L.M.; Yagupolskii, Y.L.; Ignat'ev, N.V.; Kondratenko, N.; Volkonskii, A.Yu.; Vlasov, V.M.; Notario, R.; Maria, P.-C.; *J. Am. Chem. Soc.* **1994**, *116*, 3047.

Lower end of acidity scale from HBr to $(C_4F_9SO_2)_2NH$.

decreasing) in affinities. From the ordering scheme which occurred after the first LS pass over the entire data set, the algorithm placed oxirane after isopropyl bromide even though these compounds have the exact same numerical affinity to at least six figures according to the LS algorithm. The agreement between these two relative affinities far exceeds the reported experimental error of ±0.2 kcal/mol. The fusion set recognition algorithm first examines the equilibria leading away from isopropyl bromide and incorrectly concludes that isopropylbromide - sec-butyl bromide is part of the fusion set. What is perhaps more interesting is the role of the extra compound in determining the most suspicious measurement in LTS. Here the algorithm determined the most suspicious measurement to be the pendant value between iPrBr and sBuBr. With this occurrence, one sees the LTS algorithm cannot be the absolute judge of suspicious equilibria: it merely suggests possibilities for re-investigation. The researcher must then decide whether the equilibria suggested by the program should be remeasured.

The second five-compound case is constructed from six measurements. The Al^+ binding energy scale of Uppal and Staley[22] has significant problems in the fusion set between ethanol and propionitrile. There are several sizeable residuals to the LS solution although no single residual is greater than 0.1 kcal/mol. The iterative solution repeats

the LS analysis. The LTS declares the measurement having the lowest residual value from LS to be suspicious. However, the solution obtained using the default settings for the simplex algorithm is one which cannot be trusted. The omitted value from LTS is actually a pendant measurement connecting this ladder with the next ladder.

The other five-compound cases found in the literature each contained seven measurements. In the perfluorobenzene to nitrobenzene fusion set of 87KEB/CHO,[15] two measurements are treated with equal suspicion by the LS perturbation algorithm. Together these two suspicious measurements are the only pathways away from C_6F_6 to the rest of the ladder. The LTS algorithm treats C_6F_6 - 2,3-Me_2NB as the more inaccurate measurement, but there are problems in agreement elsewhere in the fusion set. The entire fusion set should probably be restudied. The only other noticeable detail is that iterative smoothing and the LS algorithm differ slightly (by 0.08 cal/mol) in their respective solutions. Both return the same residual root-mean-square value. The difference results from the iterative scale being terminated before the solution wandered down to the LS values.

The gas phase acidity scale between methanol and ethanol[3] is an example showing four different pathways connecting the two extremes, which would seem to give a fair estimate of the equilibria between the two compounds. The minimum and maximum vectors show that the differences in the

affinities for the four paths disagree by only 0.2 kcal/mol. The iterative smoothing algorithm agrees with the LS solution after reaching its static convergence. The LTS algorithm suggests the 1-pentyne to ethanol measurement at 1.0 kcal/mol to be the worst measurement, while the normal LS shows the biggest problem is between propyne and ethanol. Each of the compounds between the two extremes is a carbon acid, and carbon acid measurements tend to be less precise than the acids of other functional groups for kinetic reasons.[3] Therefore it is difficult to predict from only seven measurements exactly where the disagreement lies. Each of the equilibria needs to be retested. The other alternative would be to increase the number of measurements to at least nine.

The third five-compound, seven-measurement case is the butanal to ethyl formate portion of 77WOL/STA shown by Figure 2-16.[6] The LS algorithm places most of the residual error in the measurement between the two extremes, and the LTS algorithm agrees with this evaluation. Here the single measurement value is larger than the affinity sums of the other pathways connecting the compounds by at least 0.2 kcal/mol. The LTS solution is probably the better result for this fusion set. The iterative smoothing algorithm again provides no new information: the iterative solution and the LS solution agree.

The fourth five-compound, seven-measurement work depicts a basicity ladder between CF_3CN and H_2S.[13] The LS solution finds fairly large residuals between both CS_2 to H_2S and H_2O to H_2S. The LTS suggests that the problem is primarily due to the latter measurement. The H_2O - H_2S equilibrium should be remeasured. To make further suggestions, measurements between CF_3CH_2OH - H_2S and $CF_3C{\equiv}N$ - H_2O would improve the statistical relevancy of the ladder.

The fifth and sixth fusion sets come from Taft's work on the lower end of the acidity scale.[95] In both cases the iterative smoothing algorithm supplies the identical solution as the LS solution. The LTS analysis on the first of Taft's fusion sets, the one between $(C_4F_9SO_2)TfNH$ and Tf_2NH, declares the measurement between $(C_4F_9SO_2)_2CH_2$ and tri-trifyl methane should be 0.2 kcal/mol larger than measured. The second fusion set is between $(CF_3C{=}O)TfNH$ and di-trifyl methane. Here, the LTS algorithm picks the $(CF_3C{=}O)TfNH$ - $(CF_3C{=}O)_3CH$ equilibria as the most suspicious. In both fusion sets no useful information beyond that found by LS can be obtained from the iterative solution.

The final five-membered fusion set studied is the nine member component of the set of synthetic scales as described in section III., ANALYSIS TIMES. The gross sign error for the a1 - a4 equilibrium skews the entire LS solution as can

be seen by Figure 2-1 (page 101). The LS algorithm does assign the largest residual to the erroneous value, but the resulting solution vector is not similar to the "correct" solution. The order is so bad that when the less robust minimum vector algorithm attempts its solution, it is forced to terminate with an error status. The iterative smoothing algorithm does not find any problems with the improper order as given by the LS algorithm, although it finds ten errors in the row monotonicity check. The presumed reason for not trying to switch the compound order is that the measurement surface for the smoothing algorithm is mostly filled for this small scale of synthetic data. The iterative solution and LS solutions are identical. The LTS has no problems in locating the error. LTS reduces all of the error to the expected erroneous measurement.

E. Six-Membered Scales.

Six different six-membered scales were found in the literature, and each had relatively few measurements in its construction. The LS information for all six is given in Table 2-7. Graphs of these scales appear as Figures 2-19 through 2-23.

The six-membered scales are the smallest fusion sets where a wide variety of agreement occurs between the equilibria making the fusion sets. One method used to

Table 2-7: Least Squares and Boundary Solution Summary of Fusion Sets with Six Compounds.

Scale	# Eq	Fusion Set	# Pert	Rms	Range:LS	Minimum	Maximum
82KAP/STA[a]	7	Me_2O - EtCHO	4	.5615218	3.2733	2.6600	3.8227
76AUE/WEB[b]	7	Et_3N - nBu_3N	0	.0500000	3.6500	3.6000	3.7000
93DEC/GAL[c]	8	EtSH - H_2S	4	.4051379	16.5543	16.3600	17.2400
82JON/STA[d]	8	HCO_2nBu - Me_2CO	5	.3202678	3.6543	3.2000	3.9000
94KOP/TAF[e]	8	m-CF_3PhCH$(CN)_2$ - p-ClPhCH$(CN)_2$	0	.1290994	2.7667	2.7000	2.8000
94KOP/TAF	9	PhCH$(CN)_2$ - CF_3CO_2H	4	.4092121	1.9545	1.7000	2.3000

Notes: Affinity values are in kcal/mol except for 93DEC/GAL which is in kJ/mol. # Pert. refers to the perturbation calls made by the least squares algorithm.

Sources: [a] Kappes, M.M.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 1813. [b] Aue, D.H.; Webb, H.M.; Bowers, M.T. *J. Am. Chem. Soc.* **1976** *98*, 311. [c] Decouzon, M.; Gal, J.; Gayraud, J.; Maria, P.; Viaglo G.; Volpe P. *J. Am. Soc. Mass Spectrom.* **1993** *4*, 54. [d] Jones, R.L.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 2296. [e] Koppel, I.A.; Taft, R.W.; Anvia, F.; Zhu, S.-Z.; Hu, L.-Q.; Sung, K.-S.; DesMarteau, D.D.; Yagupolskii, L.M.; Yagupolskii, Y.L.; Ignat'ev, N.V.; Kondratenko, N.; Volkonskii, A.Yu.; Vlasov, V.M.; Notario, R.; Maria, P.-C.; *J. Am. Chem. Soc.* **1994**, *116*, 3047.

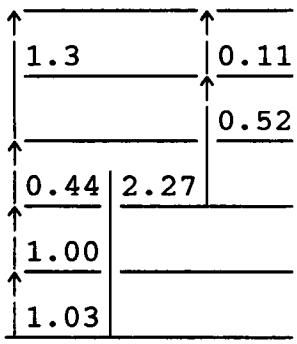
Compound		LS	LTS
EtCHO		7.34333	6.83260
MeSH		6.94800	6.38453
nPrOH		6.31867	5.96012
Me_2CCH_2		6.14267	5.51783
1,4dioxane		5.12133	5.51004
Me_2O		4.07000	4.07000

Figure 2-19: Portion of 82KAP/STA: Six Compound Scale. Source: Kappes, M.M.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 1813.

Relative bond dissociation energies for two-ligand complexes of Ni+ with organic molecules in the gas phase.

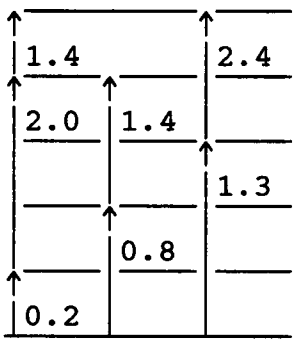
Compound		LS	LTS
nBu_3N		19.33044	19.26432
nPr_3N		17.90544	17.86432
tBuNHtamyl		16.95544	16.96432
$nPrNEt_2$		16.49294	16.46432
quinuclidine		15.89294	15.86432
Et_3N		15.68044	15.66432

Figure 2-20: Portion of 76AUE/WEB: Six Compound Scale. Source: Aue, D.H.; Webb, H.M.; Bowers, M.T. *J. Am. Chem. Soc.* **1976** *98*, 311.

Basicity of Alkyl Amines.

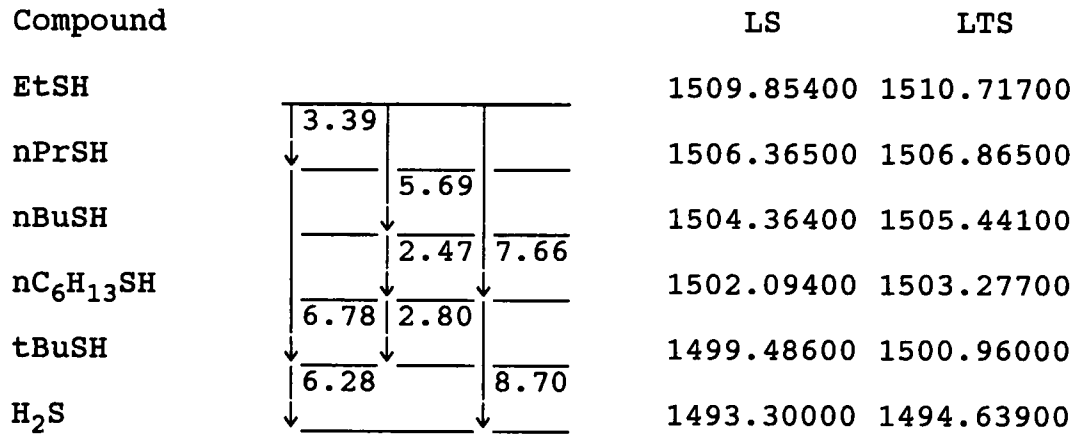


Figure 2-21: Portion of 93DEC/GAL: Six Compound Scale. Source: Decouzon, M.; Gal, J.; Gayraud, J.; Maria, P.; Viaglo, G.; Volpe, P. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 54.

Gas phase acidity of germane.

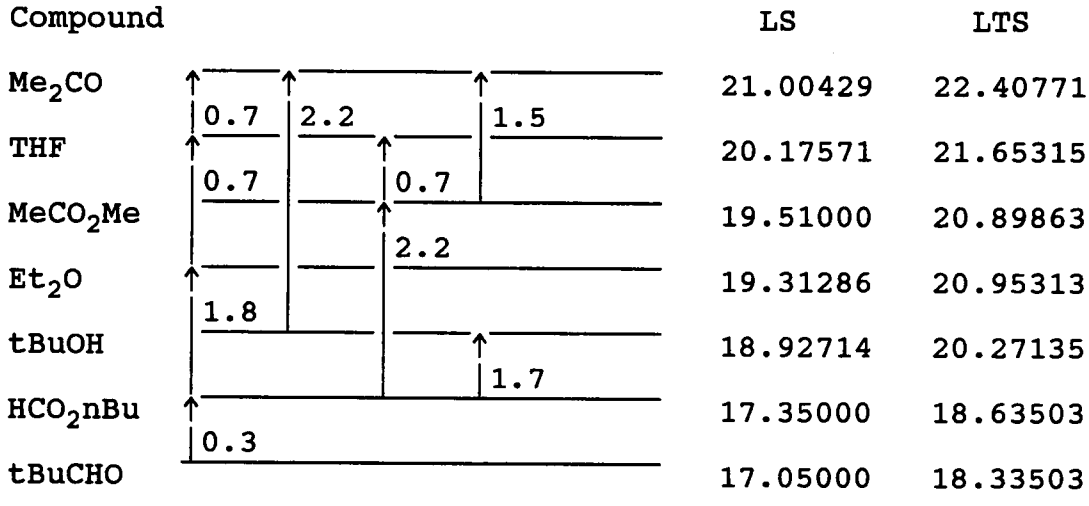


Figure 2-22: Portion of 82JON/STA: Six Compound Scale. Source: Jones, R.L.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 2296.

Relative bond dissociation energies for two-ligand complexes of Cu+ with organic molecules in the gas phase.

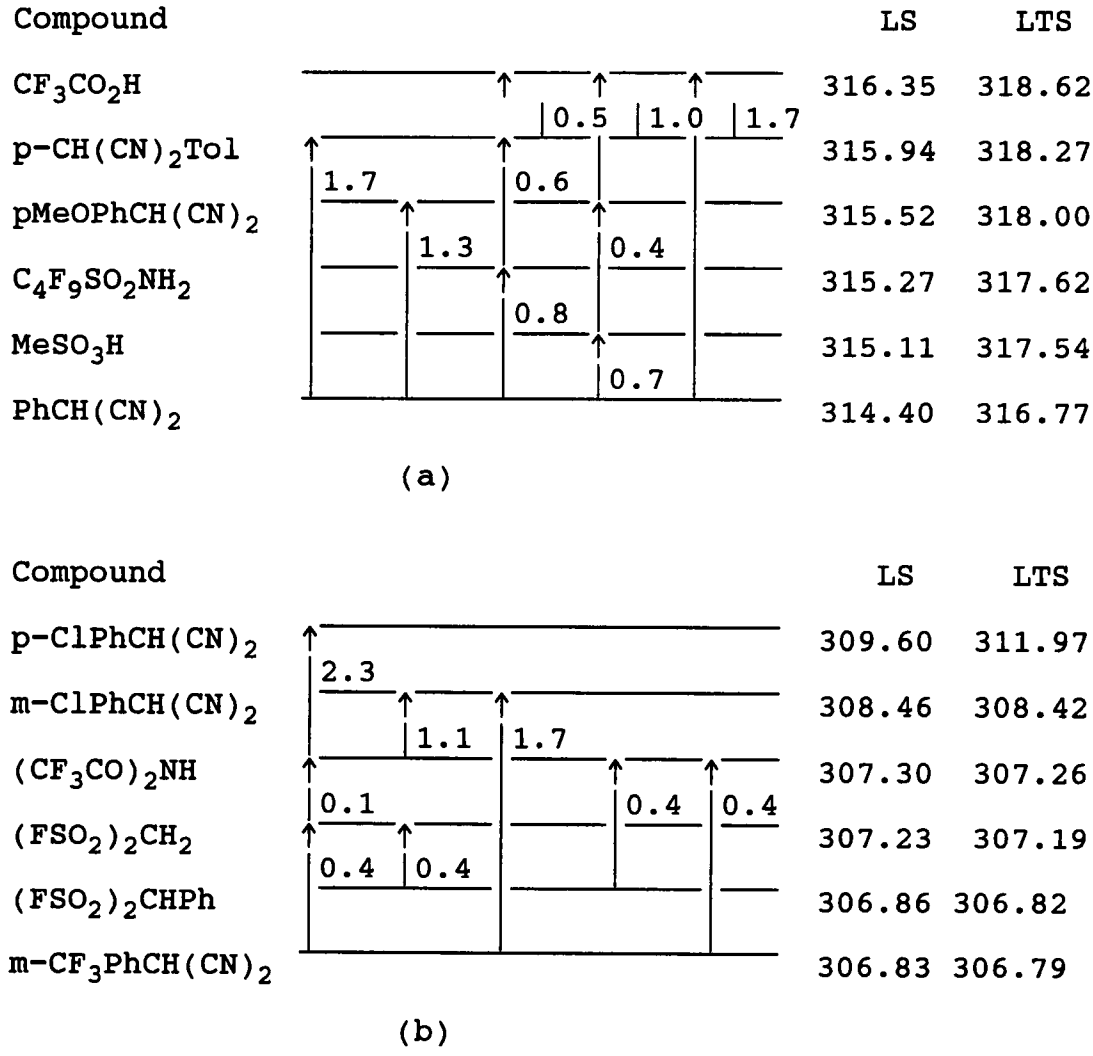


Figure 2-23: Portion of 94KOP/TAF: Six Compound Scale. Source: Koppel, I.A.; Taft, R.W.; Anvia, F.; Zhu, S.-Z.; Hu, L.-Q.; Sung, K.-S.; DesMarteau, D.D.; Yagupolskii, L.M.; Yagupolskii, Y.L.; Ignat'ev, N.V.; Kondratenko, N.; Volkonskii, A.Yu.; Vlasov, V.M.; Notario, R.; Maria, P.-C.; *J. Am. Chem. Soc.* **1994**, *116*, 3047.

Lower end of acidity scale from HBr to $(C_4F_9SO_2)_2NH$.
(a) eight equilibria (b) nine equilibria

determine the amount of agreement is to compare the values obtained for the range, or the last compound within the fusion set. This test is called the boundary solution test. The value obtained from the minimum vector is subtracted from the range value of the maximum vector. The difference is then divided by the number of compounds within the scale. If the quotient is greater than 0.05 kcal/mol, then the boundary solution test has shown poor agreement.

The lack of measurements to enable critical judgements becomes increasingly apparent as the size of the fusion set increases. For absolute judgement, the scale should have a minimum of three separate, distinct paths connecting each compound to the first compound within the ladder. The required number of equilibria is three times the number of compounds internal to the scale. For four compounds, six equilibria are required to pass judgement on the scale. For eight compounds, eighteen measurements are necessary to make completely pure judgements on any one equilibrium. The limited observable affinity range precludes the affinities being greater than about 3.0 kcal/mol in ICR. Therefore any compound must be measured relative to as many nearest neighbors as possible in order to be statistically accurate. Unfortunately, most experimenters make only a few more measurements than they believe they need to in order to establish their scales. The largest number of six compound ladders only contains eight measurements, which is still too

few to make absolute predictions as to quality of the solution. Thus the evaluation made by the LTS becomes only a tool for re-investigative predictions and not the absolute authority to be blindly followed.

Of the six-compound, seven-measurement scales, the first one covers the Ni^+ affinities between dimethyl ether and propanal from the work of Kappes and Staley.[24] The minimum/maximum vectors again show relatively large differences in their range values (1.2 kcal/mol) for the last compound in the fusion set. No extra information can be obtained from the iterative smoothing procedure. Four of the seven measurements in this fusion set have sizable residuals resulting from the LS solution, with three of the four having the exact same size. The problem with this ladder is the upper four compounds are connected by four measurements, creating two separate paths between isobutylene and propanal which have no other overlaps. Furthermore, the two affinities for the paths disagree by 1.12 kcal/mol, which is almost the entire range noticed by the minimum/maximum vectors. The lower subladder between dimethyl ether and 1-propanol does not have a sizeable error. Here the LTS algorithm is fooled by the ladder, choosing to fix the problem by creating more error in the lower subladder. Again, one can see the LTS solution is not entirely foolproof.

The other seven-measurement fusion set studies amine basicity from Et_3N to n-Bu_3N.[13] This ladder is fairly closed. No extra information is obtained from the iterative smoothing or the minimum/maximum vector searches. The LTS algorithm concentrated the entire error into one measurement: the one between tBuNHtAm and nBu_3N. Again, this suspicion is only a suggestion, and further measurements should probably be made within the ladder before any new calculations are performed.

There are three six-compound, eight-measurement fusion sets. The article concerning the gas phase acidity of germane, 93DEC/GAL,[5] is the first of the three. There is poor agreement among the values in this fusion set with several of the LS residuals above 0.1. However, the scale has been constructed in terms of kJ/mol, not kcal/mol, so the errors are less alarming. The LTS solution suggests the measurement of nPrSH - tBuSH for re-investigation, but further measurements are needed to ascertain whether there is statistical cause for suspicion. The most interesting part of the iterative algorithm's treatment of this scale is that the scale reaches a static convergence not often observed in other scales.

The Cu^+ affinity ladder between acetone and n-butyl formate of 82JON/STA[21] shows poor agreement among the equilibria. The iterative procedure does not exactly reproduce the LS solution, but the differences between the

solutions (0.01 cal/mol) are much smaller than the precision of the experiment (0.1 kcal/mol). For the interpolated values within the measurement plane, the interpolants agree with the iterative smoothing algorithm. The minimum/maximum vector search give a range of 0.7 kcal/mol for the ladder, which is relatively large considering the maximum range is 3.9 kcal/mol. There are five sizeable residuals within the LS solution, and the LTS algorithm concentrates the error into the largest residual seen by the LS algorithm. As previously stated, the re-investigation of the n-butyl formate - diethyl ether equilibrium pair can be taken only as a suggestion. The truest statement regarding the program's analysis is that there is a significant problem with at least one measurement.

The final eight-measurement fusion set is found in the 94KOP/TAF work on acidities.[95] The agreement within the eight-measurement set is such that no residual is larger than 0.1 kcal/mol. No further information is obtained from the boundary sorting algorithms or the iterative smoothing technique as the scale has relatively good agreement among the measurements. The LTS solution makes a mistake when analyzing the ladder. The principal defect to the ladder as seen by visual inspection is in a subladder between the lowest four compounds as shown by Figure 2-23. Instead of manipulating the affinity values within the lower fusion set, the LTS algorithm dumps a sizeable error into another

measurement. By visually inspecting the scale, it is found that the one measurement the LTS routine has isolated is actually a pendant. The measurement remained in the scale as this value connects this fusion set to the next one above it in 94KOP/TAF. This problem could possibly be avoided if the amount that the range is allowed to deviate from the maximum vector is lowered. The default value is 1.9 times of the maximum vector, allowing the simplex solution freedom to seek true outliers but also a chance to come up with abnormal solutions.

The six-compound, nine-measurement fusion set found in 94KOP/TAF[95] between $PhCH(CN)_2$ and CF_3CO_2H does not have good agreement among the equilibria, as can be seen by the boundary solution test. The iterative solution again repeats the LS values. Although the simplex solution isolates one measurement as suspicious, the size of the residuals on the remaining measurements demand that more work be studied in this area of the scale, either by re-investigating the measured equilibria or by making more measurements. Although not chosen by the LTS procedure, both measurement affinities of 1.7 kcal/mol appear suspicious by visual inspection.

F. Seven-Membered Scales.

As can be seen by the case studies in the previous section, the simplex/LTS procedure starts to break down beginning with the six component scales. Further degeneration of the LTS algorithm is observed in seven compound scales.

A summary of the LS results appears for five seven-compound scales in Table 2-8. Graphs of these fusion sets appear as Figures 2-24 to 2-27.

Perhaps the most noticeable difference between the seven-compound fusion sets and those already discussed is the seven compound cases tend to have a higher measurement to compound ratio. The literature fusion set with the fewest measurements is in the compiled references of fluoride affinities.[29,30,36] The ladder between F_2CO and AsF_3 contains nine measurements. Here the boundary solution test suggests only fair agreement between the equilibria. The iterative solution is similar to the LS solution fed into the algorithm, but the vectors do not agree completely. The interpolated values within the affinity matrix agree better with the values calculated from the iterative smoothing algorithm than they do with the values calculated from the LS algorithm. The largest residuals within the solution to the LS algorithm are the two measurements which lead to formic acid. The LTS solution allows the error in the

Table 2-8: Least Squares and Boundary Solution Summary of Fusion Sets with Seven Compounds.

Scale	# Eq	Fusion set	# Pert	Rms	Range:LS	Minimum	Maximum
FDATA[a]	9	F_2CO - AsF_3	5	.4286608	6.7875	6.6875	7.1000
79BAR/SCO[b]	11	tBuOH -neoPnOH	0	.0870260	1.4721	1.4000	1.5000
77WOL/STA[c]	11	MeOAc -$CF_3CH_2NH_2$	0	.1309686	4.3132	4.2000	4.4049
77WOL/STA	13	MeCHO - nPrCHO	1	.2259534	4.6134	4.4000	4.8332
88DIL/KEB[d]	14	C_6F_6 - $C_6F_5NO_2$	2	.2921593	12.5893	12.3411	13.0000
SYNTHETIC	21	A4-A7	8	.9751068	.3750	-7.5000	.6042

Notes: Affinity values are in kcal/mol. # Pert. refers to the perturbation calls made by the least squares algorithm.

Sources: [a] Larson, J.W.; McMahon, T.B. *J. Am. Chem. Soc.* **1983** *105*, 2944.; *J. Phys. Chem.* **1984** *88*, 1083.; *J. Am. Chem. Soc.* **1985** *107*, 766. [b] Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979** *101*, 6046. [c] Wolf, J.A.; Staley, R.H.; Koppel, I.; Taagepera, M.; McIver, R.T., Jr.; Beachamp, J.L.; Taft, R.W. *J. Am. Chem. Soc.* **1977** *99*, 5417. [d] Dillow, G.W.; Kebarle, P. *J. Am. Chem. Soc.* **1988** *110*, 4877.

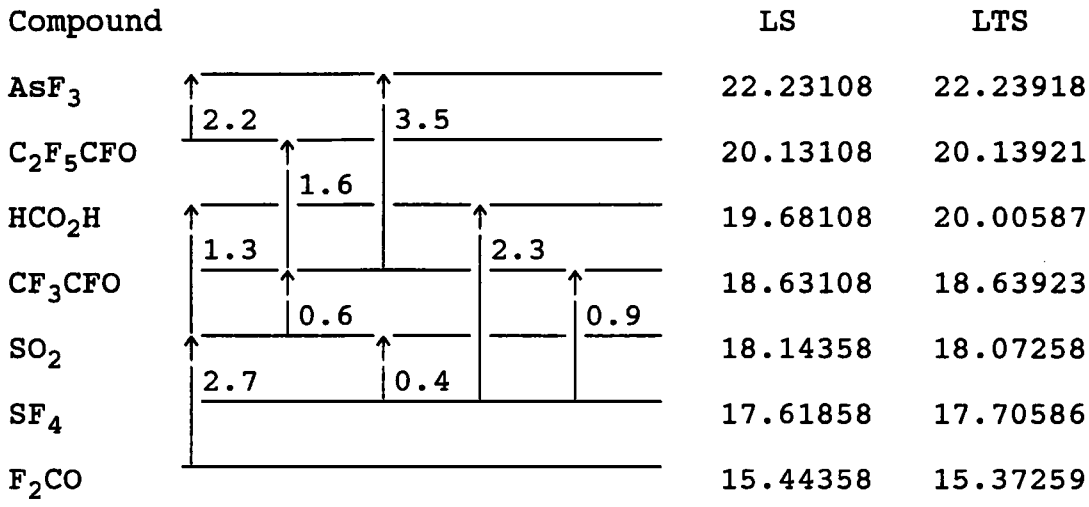


Figure 2-24: Portion of F_DATA: Seven Compound Scale. Source:Larson and McMahon, *J. Am. Chem. Soc.* **1983**, *105*, 2944; *J. Phys. Chem.* **1984** *88*, 10; *J. Am. Chem. Soc.* **1985** *107*, 766.

Fluoride Affinity Scale

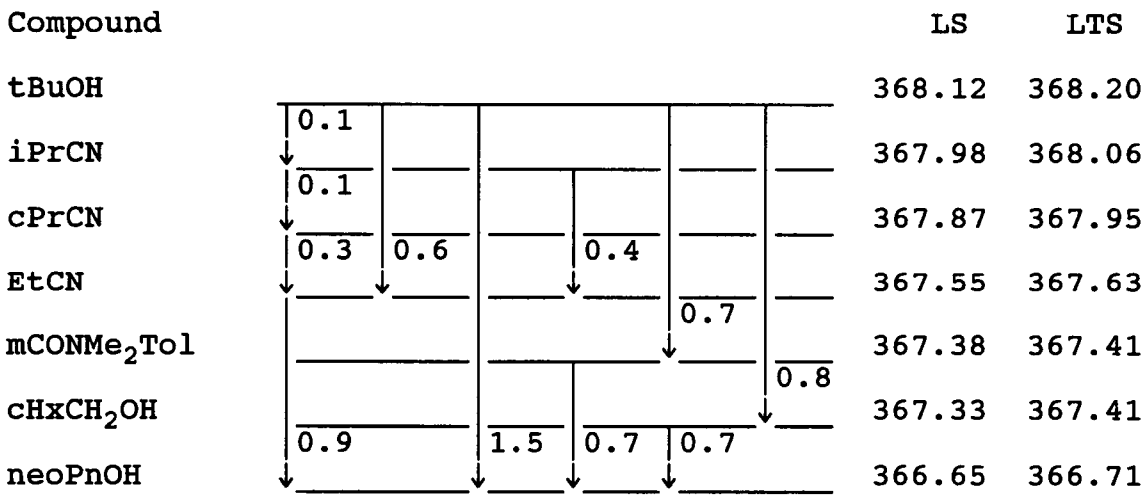


Figure 2-25: Portion of 79BAR/SCO: Seven Compound Scale. Source:Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979** *101*, 6046.

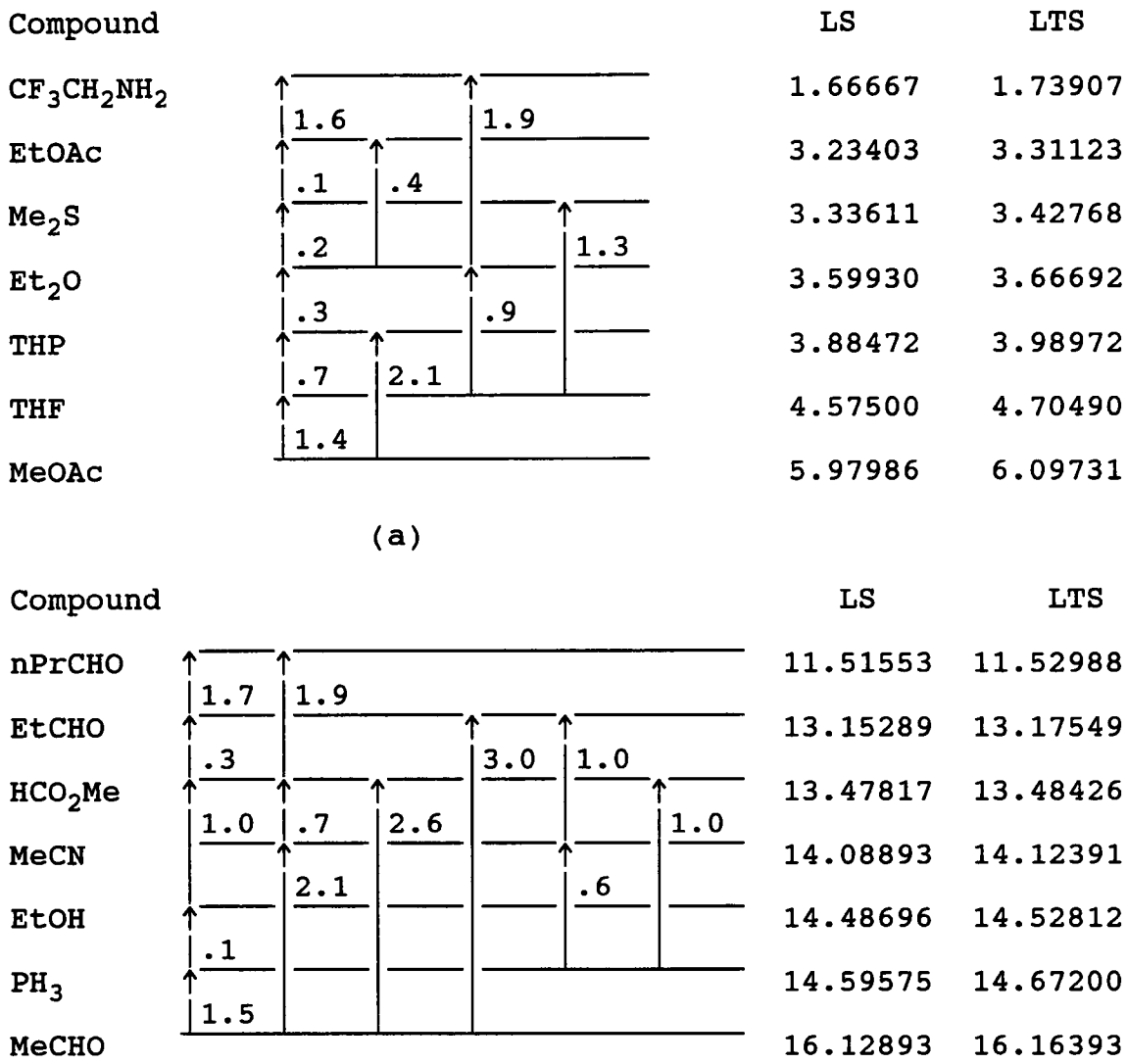


Figure 2-26: Portions of 77WOL/STA: Seven Compound Scales. Source: Wolf, J.A.; Staley, R.H.; Koppel, I.; Taagepera, M.; McIver, R.T., Jr.; Beachamp, J.L.; Taft, R.W. *J. Am. Chem. Soc.* **1977** *99*, 5417.

Gas Phase Basicities; (a) 11 equilibria (b) 13 equilibria

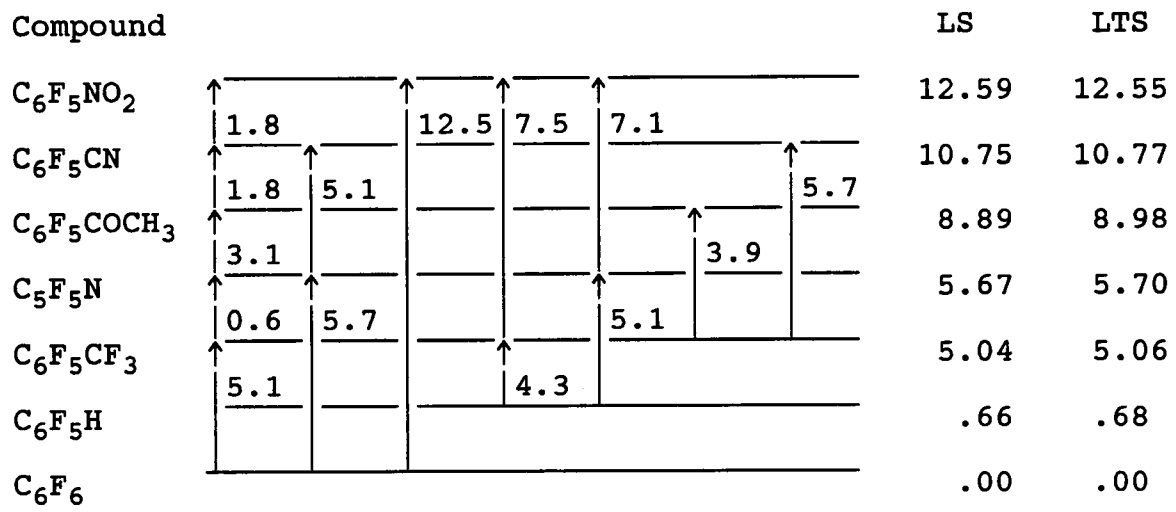


Figure 2-27: Graph from 88DIL/KEB: Seven Compound Scale. Source: Dillow, G.W.; Kebarle, P. *J. Am. Chem. Soc.* **1988** *110*, 4877.

Fluoride Meisenheimer Complexes.

sulfur dioxide - formic acid equilibrium to grow and accepts the sulfur tetrafluoride - formic acid as valid. The temporary deletion of either of these measurements would improve the LTS L_2 residual norm. Visual inspection of this fusion set region in the full fluoride affinity scale is cluttered by the addition of five pendant measurements. With these five pendants removed, the fusion set (Figure 2-24) is easier to read.

The next block of seven compound ladders to discuss is the eleven measurement case. One case is a partial ladder of the gas phase acidity scale,[3] containing compounds from t-butyl alcohol to neopentyl alcohol. The difference in the values obtained from the minimum and maximum vectors for neopentyl alcohol is only 0.1 kcal/mol. The iterative and LS solutions completely agree. Because the largest residual obtained by the LS algorithm was less than 0.04 kcal/mol, the LTS solution is actually unnecessary. The LTS algorithm suggests as its worst measurement one of the top three candidates from the LS algorithm.

The other eleven measurement case found in the literature is the $MeOCH_2CO_2H$ to $CF_3CH_2NH_2$ fusion set found in 77WOL/STA.[6] The agreement between minimum and maximum solution vectors suggest fairly good agreement among the eleven compounds. The iterative solution repeats the LS values exactly. The largest residual is the same for both

LS solutions and is only 0.14 kcal/mol for the LTS algorithm.

Another seven compound case may be found in the 77WOL/STA study between ethanal and butanal,[6] only here there are thirteen measurements. The perturbation algorithm contained in the initial LS finds one residual larger than 0.1 kcal/mol: the measurement between PH_3 and methyl formate. The same measurement is chosen as a suspected measurement by the LTS algorithm. Even with this one measurement removed, there are still several residuals larger than 0.05 kcal/mol in the LTS solution. The magnitude of these residual errors are also seen by the sizable differences in both the minimum and the maximum vector solutions. The iterative smoothing algorithm is of little use as it exactly duplicates the LS solution.

The final literature case contains fourteen measurements. This literature scale is the first of the ladders to be studied which is the entire scale. The electron affinities of the aryl fluorides[31] show reasonable agreement between the equilibria. While several of the equilibria have residuals greater than 0.5 kcal/mol, this High Pressure Mass Spectrometry work includes measurements of up to 12.5 kcal/mol. Therefore the 0.7 kcal/mol difference in the range of the minimum/maximum boundary solutions for $C_6F_5NO_2$ is trivialized by the size of the maximum value for this compound (13 kcal/mol). The differences between the

iterative and LS solutions can be attributed to the gradual decreasing of the iterative solution with each pass. If the current termination scheme used by the iterative algorithm were to be abandoned and the first thousand passes were calculated, the scale would shrink by over 6 cal/mol. This relatively small effect is what determines the differences between the solutions. This scale is also unique in that it is the first scale where two measurements are allowed to be dropped by the LTS algorithm. The removal of both measurements does not remove any compound from the scale, a problem which is observed in larger scales. The temporary removal of these two equilibria values does not significantly change the LTS solution from the LS solution except for $C_6F_5COCH_3$, as can be seen by Figure 2-27.

The final seven-compound scale in the study is the seven-membered synthetic scale as described in the timing section. The a1 - a4 pair again skews the LS solution. The problem is then continued as both the minimum and the maximum vectors have to be terminated with an error status warning that the solutions obtained from these algorithms can not be trusted. Even if the error flag were not there, these solutions are abnormal enough to indicate a problem in the algorithm. In fact, one of the reasons the synthetic scales were constructed was to evaluate the robustness of these algorithms. Even though the row monotonicity check finds eight errors, the iterative solution exactly repeats

the LS solution. Both the LS and the LTS solutions correctly list the a1 - a4 equilibria to be the worst of the fifteen. Once the erroneous measurement is removed, the resulting solution is perfectly closed.

G. Eight-Membered Scales.

Rather than emphasizing each point of the analysis for each of the individual ladders as has been done in the previous sections, the discussion for eight compound scales is a little more generalized. Only the aberrations from the previous trends will be noted. For any well behaved fusion set, little mention will be made other than the listing of the LS results as shown in Table 2-9. Graphs of the various scales appear as Figures 2-28 through 2-37.

The primary difference noticed in generating the scales was in the ability of the simplex solution to identify a rational solution based on the initial guess solution. When only six compounds were in the fusion set, one of the best ways to identify a particular erroneous value was to have the initial guess solution for the simplex procedure to be the zero vector. For seven compounds, some of the scales gave reasonable solutions with the zero initial vector, but others did not. For eight or more compounds the zero solution becomes impractical: the resulting solution is unreasonable as the step size diminishes before the simplex

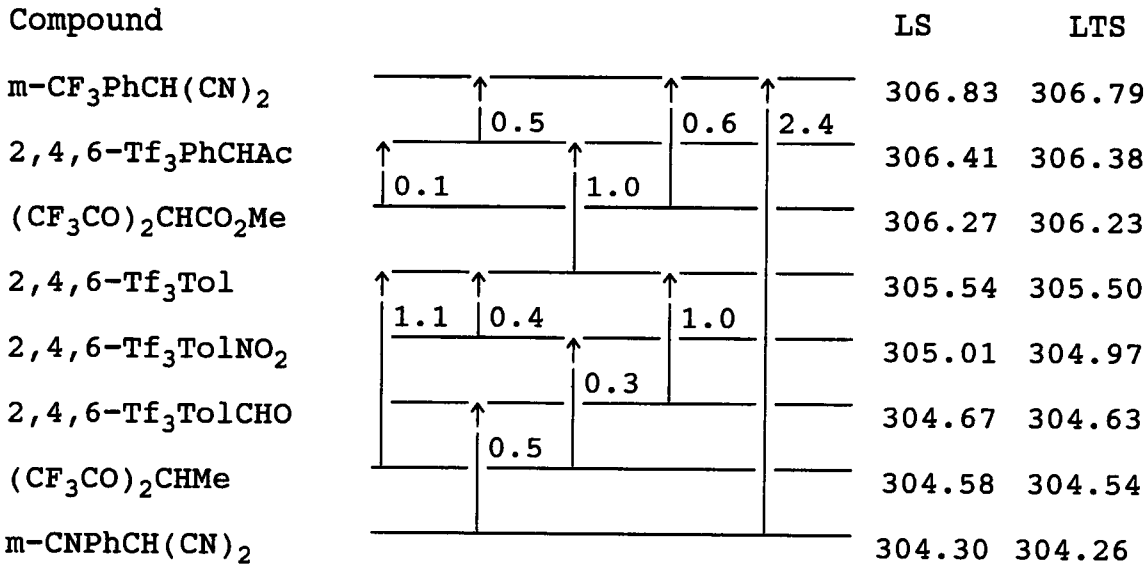
Table 2-9: Least Squares and Boundary Solution Summary of Fusion Sets with Eight Compounds.

Scale	# Eq	Fusion Set	# Pert	Rms	Range:LS	Minimum	Maximum
Taft[a]	10	m-CNPhCH(CN)$_2$ - m-CF$_3$PhCH(CN)$_2$	7	.3612148	2.5286	2.3429	3.0000
94KOP/TAF[b]	11	m-CNPhCH(CN)$_2$ - m-CF$_3$PhCH(CN)$_2$	7	.3642915	2.5375	2.2250	3.0000
88OPE/TEW[c]	11	nPrOH - Me$_2$CO	0	.1757839	3.4700	3.2800	3.5900
77WOL/STA[d]	12	HCO$_2$Et - MeOAc	3	.2580636	4.0482	3.8000	4.3000
77MCM/KEB[e]	12	CF$_2$HCO$_2$H - CH$_2$FCO$_2$H	3	.3147419	7.1396	6.7000	7.4000
79BAR/SCO[f]	13	EtOH - tBuOH	6	.4092604	2.7938	2.5000	3.3000
Cl data[g]	13	SO2 - HCO$_2$H	8	.5347134	3.4898	2.8000	3.8000
94KOP/TAF[b]	13	CF$_3$COSH - CF$_3$CO$_2$H	8	.5135304	3.6836	3.1000	4.1000
76YAM/KEB[h]	14	mC$_6$H$_4$F$_2$ - MeCHO	4	.3430431	3.8089	3.6000	4.3000
85MCM/KEB[i]	14	SO$_2$F$_2$ - CF$_3$CN	5	.9519404	14.1905	13.7000	15.5000
91MEO/SIE[j]	14	3F pyr - Me$_3$N	7	.8910059	11.8931	11.2000	13.0000
SYNTHETIC	24	a2 - a8	4	.4477910	0.5421	0.5421	0.6000

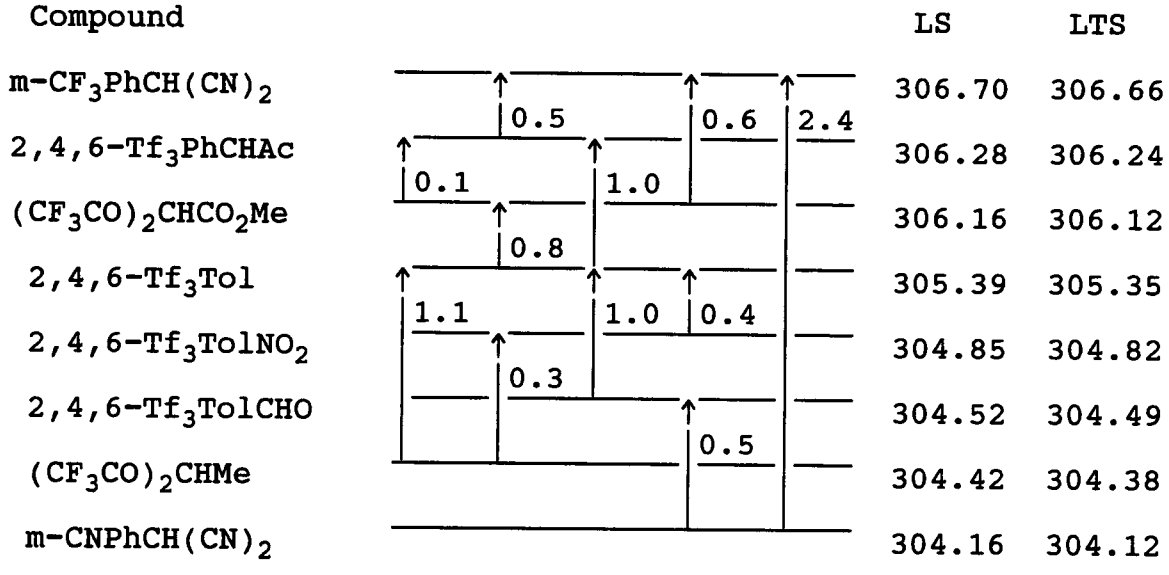
Notes: Affinity values are in kcal/mol.# Pert. refers to the perturbation calls made by the least squares algorithm.
Sources: [a] Taft, R.W., personal communication. [b] Koppel, I.A.; Taft, R.W.; Anvia, F.; Zhu, S.-Z.; Hu, L.-Q.; Sung, K.-S.; DesMarteau, D.D.; Yagupolskii, L.M.; Yagupolskii, Y.L.; Ignat'ev, N.V.; Kondratenko, N.; Volkonskii, A.Yu.; Vlasov, V.M.; Notario, R.; Maria, P.-C.; *J. Am. Chem. Soc.* **1994**, *116*, 3047.

Table 2-9: (continued)

[c] Operti, L.; Tews, E.C.; Freiser, B.S. *J. Am. Chem. Soc.* **1988** *110*, 3847. [d] Wolf, J.A.; Staley, R.H.; Koppel, I.; Taagepera, M.; McIver, R.T., Jr.; Beachamp, J.L.; Taft, R.W. *J. Am. Chem. Soc.* **1977** *99*, 5417. [e] McMahon, T.B.; Kebarle, P.; *J. Am. Chem. Soc.* **1977** *99*, 2222. Lower portion of phenol scale. [f] Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979** *101*, 6046. [g] Larson, J.W.; McMahon, T.B. J. Am. Chem. Soc. **1985** *107*, 766.; *J. Am. Chem. Soc.* **1984** *106*, 517.; *Can. J. Chem.* **1984** *62*, 675. *J. Phys. Chem.* **1984** *88*, 1083. [h] Yamdagi, R.; Kebarle, P. *J. Am. Chem. Soc.* **1976** *98*, 1320. [i] McMahon, T.B.; Kebarle, P. *J. Am. Chem. Soc.* **1985** *107*, 2612. [j] Meotner, M.; Siek, L.W. *J. Am. Chem. Soc.* **1991** *113*, 4448. Affinities here are ΔG's.



(a)



(b)

Figure 2-28: Portion of 94KOP/TAF: Eight Compound Scales. Source: Koppel, I.A.; Taft, R.W.; Anvia, F.; Zhu, S.-Z.; Hu, L.-Q.; Sung, K.-S.; DesMarteau, D.D.; Yagupolskii, L.M.; Yagupolskii, Y.L.; Ignat'ev, N.V.; Kondratenko, N.; Volkonskii, A.Yu.; Vlasov, V.M.; Notario, R.; Maria, P.-C.; *J. Am. Chem. Soc.* **1994**, *116*, 3047. (a) original communication. (b) second communication. Lower end of acidity scale.

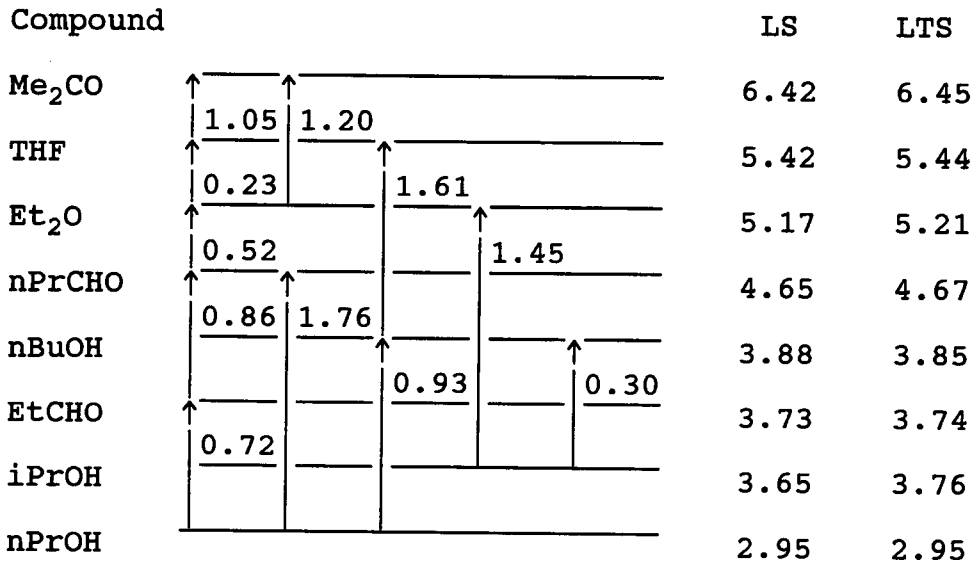


Figure 2-29: Portion of 88OPE/TEW: Eight Compound Scale. Source: Operti, L.; Tews, E.C.; Freiser, B.S. *J. Am. Chem. Soc.* **1988** *110*, 3847.

Gas phase ligand binding energies to Mg+

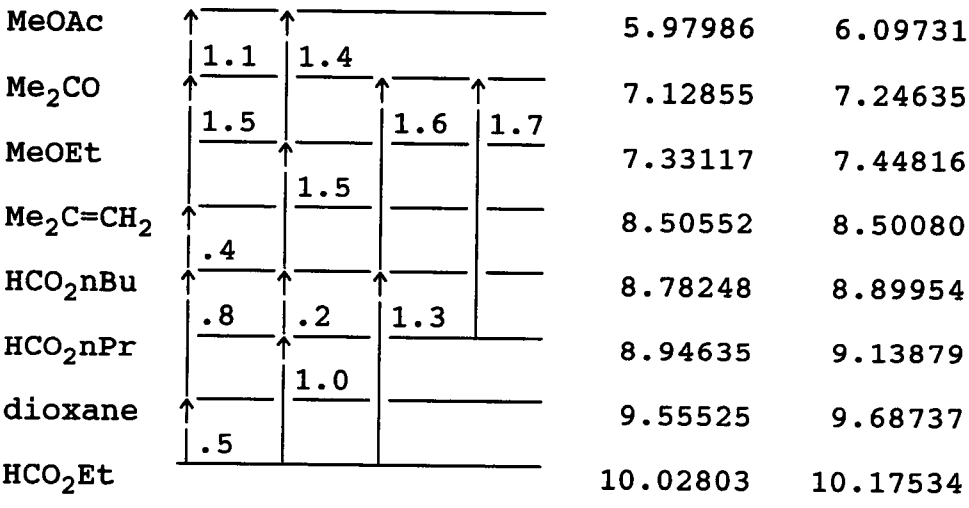


Figure 2-30: Portion of 77WOL/STA: Eight Compound Scale. Source: Wolf, J.A.; Staley, R.H.; Koppel, I.; Taagepera, M.; McIver, R.T., Jr.; Beachamp, J.L.; Taft, R.W. *J. Am. Chem. Soc.* **1977** *99*, 5417.
Gas phase basicities.

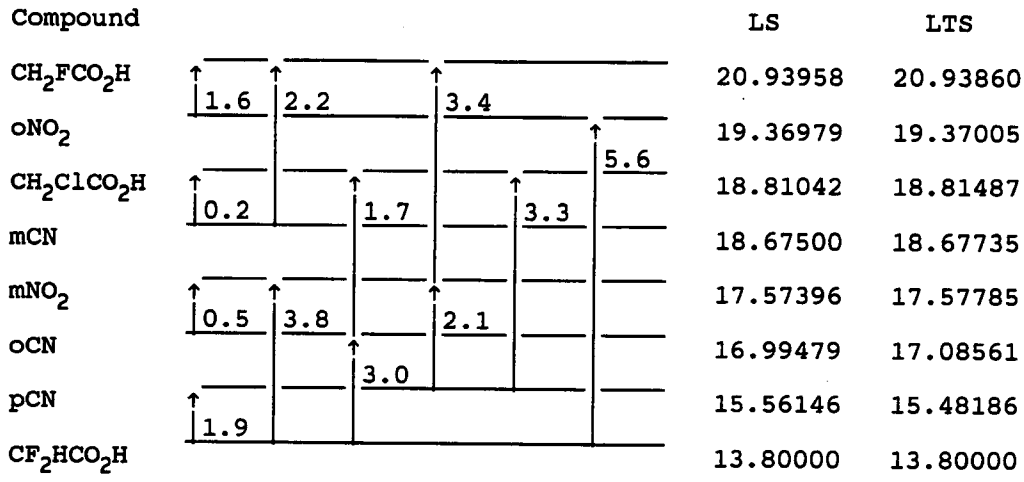


Figure 2-31: Portion of 77MCM/KEB: Eight Compound Scale. Source: McMahon, T.B.; Kebarle, P.; *J. Am. Chem. Soc.* **1977** *99*, 2222.

PARTIAL PHENOL SCALE- LOWER PORTION.

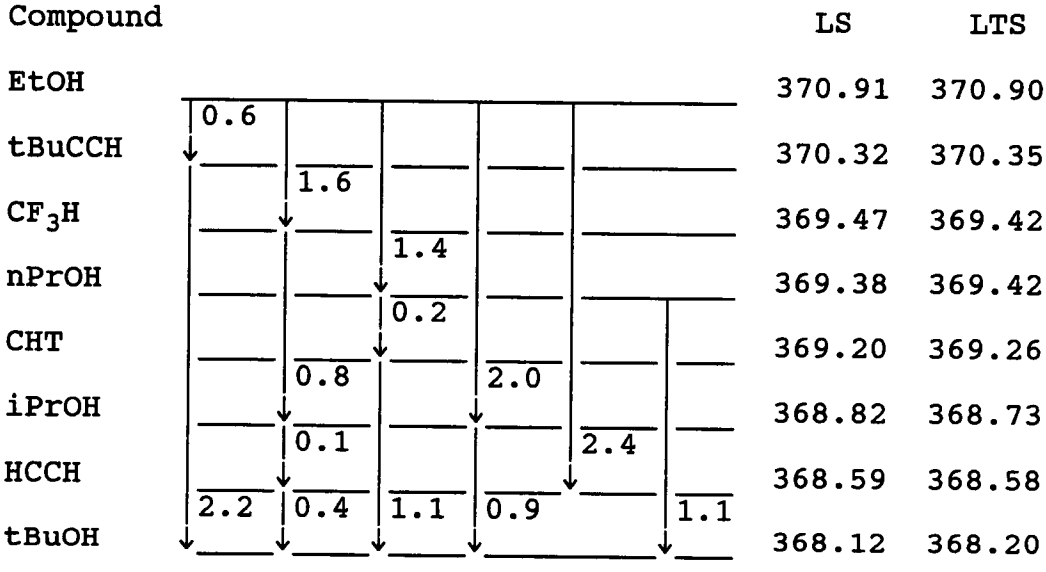


Figure 2-32: Portion of 79BAR/SCO: Eight Compound Scale. Source: Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979** *101*, 6046.

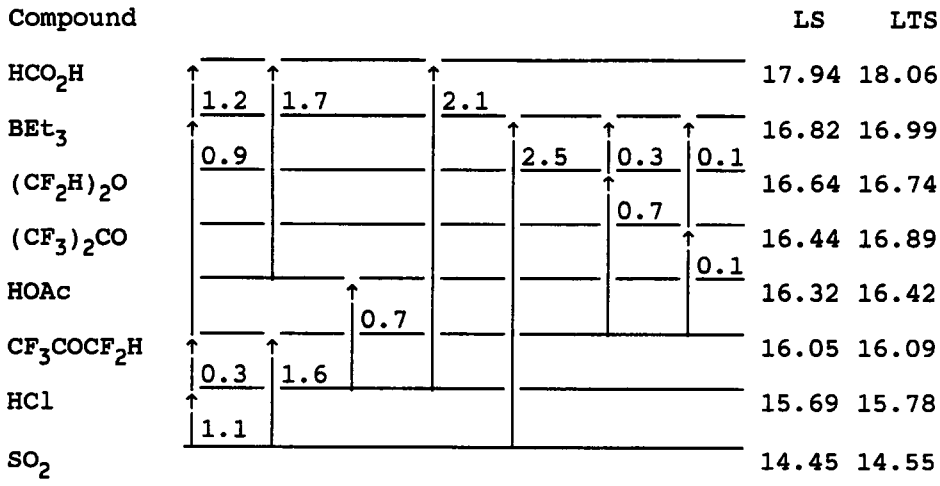


Figure 2-33: Portion of Cl Affinities: Eight Compound Scale. Source: Larson, J.W.; McMahon, T.B. *J. Am. Chem. Soc.* **1985** *107*, 766.; *J. Am. Chem. Soc.* **1984** *106*, 517.; *Can. J. Chem.* **1984** *62*, 675. *J. Phys. Chem.* **1984** *88*, 1083.

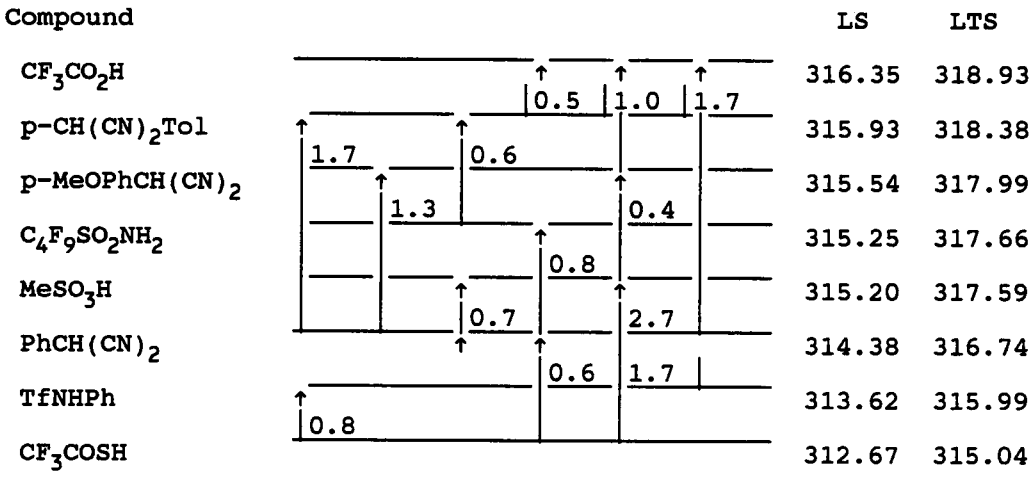


Figure 2-34: Portion of 94KOP/TAF: Eight Compound Scale. Source: Koppel, I.A.; Taft, R.W.; Anvia, F.; Zhu, S.-Z.; Hu, L.-Q.; Sung, K.-S.; DesMarteau, D.D.; Yagupolskii, L.M.; Yagupolskii, Y.L.; Ignat'ev, N.V.; Kondratenko, N.; Volkonskii, A.Yu.; Vlasov, V.M.; Notario, R.; Maria, P.-C.; *J. Am. Chem. Soc.* **1994**, *116*, 3047.

Lower end of acidity scale from HBr to $(C_4F_9SO_2)_2NH$.

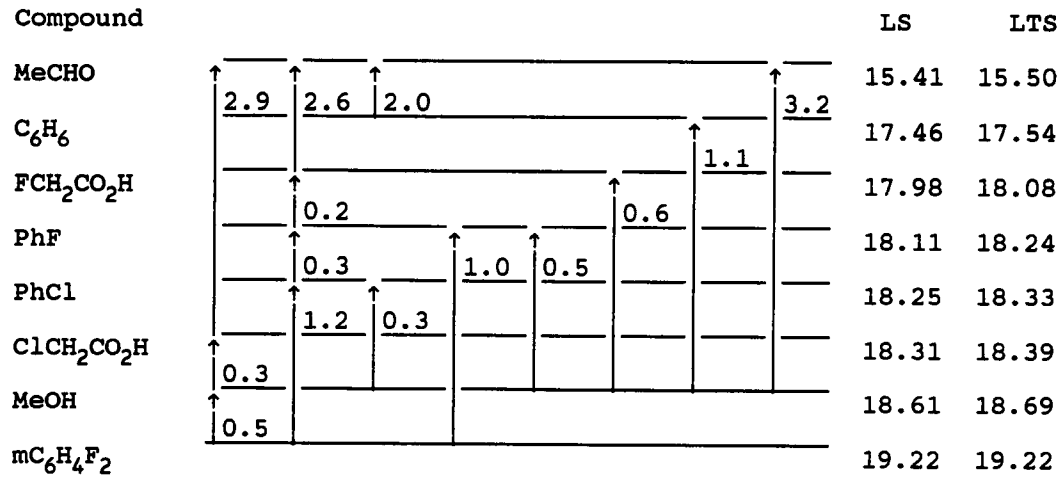


Figure 2-35: Portion of 76YAM/KEB: Eight Compound Scale. Source: Yamdagi, R.; Kebarle, P. *J. Am. Chem. Soc.* **1976** *98*, 1320.

Gas phase basicities

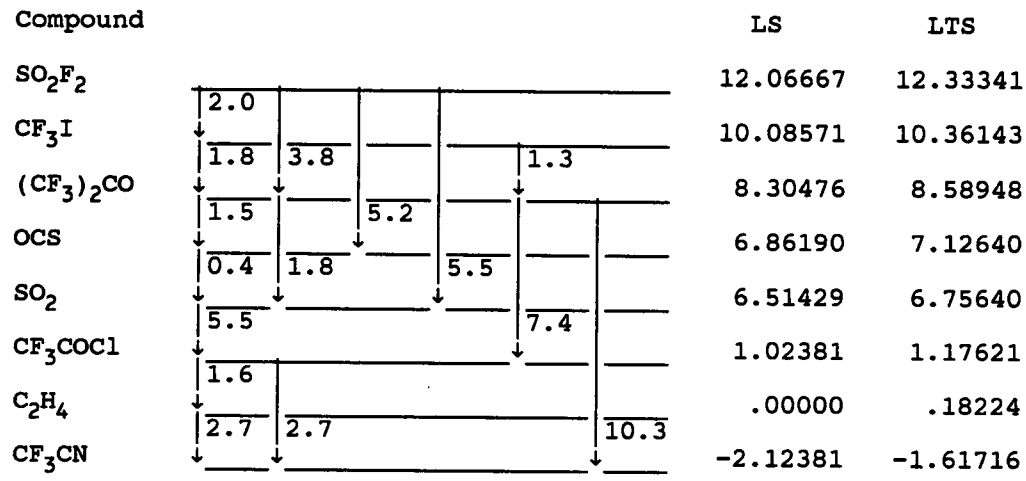


Figure 2-36: Portion of 85MCM/KEB: Eight Compound Scale. Source: McMahon, T.B.; Kebarle, P. *J. Am. Chem. Soc.* **1985** *107*, 2612.

Weaker end of the Basicity scale

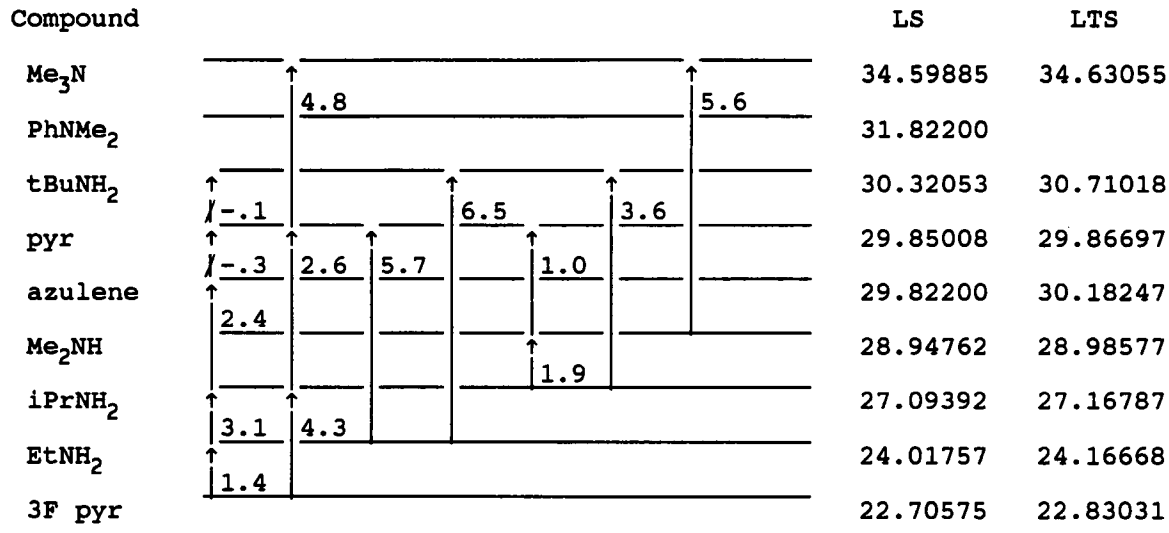


Figure 2-37: Portion of 91MEO/SIE: Eight Compound Scale. Source: Meotner, M.; Siek, L.W. *J. Am. Chem. Soc.* **1991** *113*, 4448.
Affinities here are ΔG's. Proton affinity -variable temperature Equilibrum measurements, Upper Proton Affinity Range

Initial data in disagreement with Least Squares solution
pyr is numerically larger than azulene, by .02808
tBuNH2 is numerically larger than pyr, by .47045
changing the way these measurements are graphed.
/ Marks the exoergonic reactions graphed differently.

reaches a point where all of the values are within a reasonable range. Therefore it was decided that the default methodology would be to use the zero initial vector for fusion sets of less than seven compounds, but to use the LS initially for fusion sets of seven or more compounds.

Most of the eight-membered fusion sets can be characterized as having relatively poor agreement among the individual equilibria. The best judge of the agreement is to look at the differences between the minimum and the maximum boundary vectors. If the values differ by more than 0.1 kcal/compound, then at least one measurement is usually in serious disagreement with the others in the fusion set.

Critical analysis begins with the lower portion of the gas phase acidity scale compiled as 94KOP/TAF.[95] Perhaps the most surprising detail from the analysis of the fusion set between the m-$CNC_6H_4CH(CN_2)_2$ and m-$CF_3C_6H_4CH(CN_2)_2$ is that the LTS algorithm makes no suggestions as to questionable equilibria. A visual inspection of the scale shows at least one problem within the small three compound cycle $(CF_3CO)_2CHMe$ - 2,4,6-Tf_3TolNO_2 - 2,4,6-Tf_3Tol as shown by Figure 2-28. The ΔG_{eq} directly connecting the first and third compounds of the cycle is 1.1 kcal/mol, while the sum of the two other equilibria totals only 0.7 kcal/mol.

After the results of the analysis of the scale was sent to Taft, he responded with corrections to the earlier version of the scale. The work has now been published.[96] In

the updated version of the gas phase scale, there is one extra measurement within this eight compound fusion set not apparent in the earlier tabulation. The additional equilibrium does not greatly affect the treatment by any of the analysis algorithms: the iterative smoothing and LTS algorithms have comparable solutions to the LS solution.

The Mg^+ affinity scale from the work of Operti, Tews, and Freiser between n-propyl alcohol and acetone is fairly well ordered. Only one compound is listed as suspicious by the LTS algorithm: isopropyl alcohol to n-butyl alcohol.

The basicity scale from ethyl formate to methyl acetate from the work of 77WOL/STA[6] shows only fair agreement, with the difference between the minimum and maximum boundary vectors for MeOAc equal to 0.5 kcal/mol. The LTS algorithm temporarily removes two measurements from consideration. Removal of the isobutylene - acetone pair by LTS increases the affinity of the isobutylene by 0.15 kcal/mol. Here the remaining tie of isobutylene to the rest of the scale shows very little residual error. The removal of the propyl formate - acetone pair results in a minor perturbation of the ester's ΔG because it is tied to the rest of the scale by two other paths.

The first of the eight-compound, twelve-measurement scales is the lower portion of the phenol acidity scale. Based on the results from the LS values, there are significant disagreements among the measurements leading

away from CF_2HCO_2H. The boundary test solution results agree with the existence of at least one poorly measured value. The LTS solution rejects both of these measurements within the ladder.

The partial fusion set from the gas phase acidity scale between ethanol and t-butyl alcohol[3] shows a fair amount of disorder. There are a number of residuals from the LS algorithm which are sizeable: that is, greater than 0.1 kcal/mol. The LTS algorithm flags the isopropanol - t-butyl alcohol measurement as being the worst, yet there are several other large residuals even after this measurement has been removed. The overall effect of removing the isopropanol - t-butyl alcohol equilibrium is to drop the t-butyl alcohol ΔG by 0.1 kcal.

For the SO_2 - HCO_2H fusion set of the chloride affinity compilation,[29,34-36] the disagreement among the equilibria is enough to flag a warning statement when the minimum boundary algorithm is processed. The problem is the resulting affinity sum from one of the three paths connecting $(CF_2H)_2CO$ and BEt_3 does not agree with the affinity sums from the other two paths, as seen by Figure 2-33. The smallest affinity sum is only 0.2 kcal/mol, while the other two paths have affinity sums of 0.9 and 1.0 kcal/mol. The probable cause here is that one of the 0.1 kcal/mol equilibria[36] is actually 1.0 kcal/mol. The simplex

technique does notice and correct for the error of the small value, but it cannot remove all of the error.

The updated version of the Taft scale depicts a larger fusion set between CF_3COSH and CF_3CO_2H than was in the first communication.[95,96] There is not good agreement among the thirteen equilibria as depicted by the boundary solution criterion. The LTS algorithm keeps as its worst measurement the worst one from LS solution: the $PhCH(CN)_2$ - CF_3CO_2H pair. Even with this value removed from consideration, the remaining residuals are sizeable.

Yamdagni and Kebarle's gas phase basicity scale[7] between $mC_6H_4F_2$ and MeCHO is also fairly disordered as attested to by four of the fourteen measurements having residuals larger than 0.1 kcal/mol. The largest residual value is associated with the m-$C_6H_4F_2$ to C_6H_5Cl measurement. Visual inspection of Figure 2-35 reveals that this 1.2 kcal/mol measurement disagrees with the values calculated from two other pathways. The m-$C_6H_4F_2$ → MeOH → C_6H_5Cl pathway has a calculated affinity of 0.8 kcal/mol while the m-$C_6H_4F_2$ → C_6H_5F ← C_6H_5Cl pathway is only calculated to be 0.7 kcal/mol. The differences between the boundary solutions is 0.5 kcal/mol or more for the latter portion of the ladder which also shows disagreement among the equilibria. While the iterative solution notes one error in its row monotonicity check, the iterative solution agrees with the LS solution to 0.01 cal/mol. Here the unmeasured

interpolated values exactly agree with the LS interpolation of these measurements, showing that the differences between the two scales is due to a gradual shrinking of the iterative solution as the number of iterative passes increase. The LTS solution removes as suspicious the m-$C_6H_4F_2$ - C_6H_5Cl measurement, but it also removes the C_6H_5Cl - C_6H_5F measurement. The temporary removal of these two values does improve the agreement of the measured values within the scale. Removal of the second measurement as suspicious is probably unnecessary. The difficulty with both of these measurements being removed is the estimation of the affinity for C_6H_5Cl. The only measurement connecting C_6H_5Cl to the rest of the scale is the one from methanol. Although this value should now be pendant and therefore fit exactly, it still has a sizeable error. A thorough re-investigation of the data might include the running of the entire equilibrium data file with the most offensive equilibrium removed by hand.

The weaker end of the basicity sales was studied by McMahon and Kebarle.[10] There is disagreement among the fourteen HPMS measurements as indicated by the differences in the boundary solution test. The biggest differences are for the last two compounds, C_2H_4 and CF_3CN. The LTS algorithm tries to reduce the residual rms by removing two measurements. The effect of the removal here is to invalidate the LTS solution, even though the values in

Figure 2-36 appear to be reasonable. The LTS algorithm has removed all of the references to C_2H_4 from consideration. Therefore, while the affinity given for ethylene seems valid, it could actually be any value, making the LTS solution incorrect. While incorrect, the algorithm has accomplished its purpose of identifying problem equilibria: the C_2H_4 - CF_3CN pair should be re-investigated.

The final eight-compound literature scale is also constructed from fourteen equilibria. The fusion set between 3-fluoropyridene and trimethyl amine[17] shows poor agreement among the equilibria, as can be attested to by several criteria. The LS L_2 norm is comparatively high. The boundary solution test shows differences in the range of over 0.2 kcal/mol. Several residuals are larger than 0.1 kcal/mol. Finally, the graph of this scale had to be modified to allow the equilibria to be drawn in: the LS solution has inverted the affinity order of pyridine in relation to t-butyl amine and in relation to azulene. The LTS solution attempts to correct the problems in the scale by removing two equilibrium pairs: $tBuNH_2$ - pyridine and $iPrNH_2$ - azulene. Removal of these values by LTS significantly raises the affinity value of $tBuNH_2$.

The eight-compound synthetic scale behaves similarly to its lower-membered analogues: the LS solution is skewed, causing problems in the boundary selection algorithms. The iterative algorithm repeats the LS solution. The LTS

algorithm correctly isolates the suspect equilibrium, and the resulting solution is perfectly closed without the error.

H. Nine-Membered Scales.

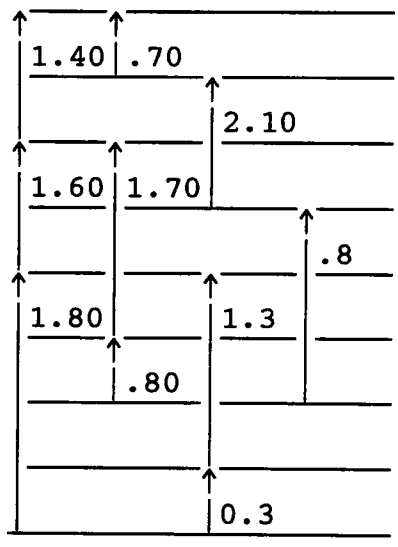
As the size of the fusion sets increases beyond eight compounds, the number of representative samples decreases. The LS results for the nine-compound cases are given in Table 2-10, with the respective graphs following the table as Figures 2-38 to 2-40.

The first of the ladders demonstrates the need for a program such as LADDER. Several problems existed in the published scale of cyanide affinities.[32] First the scale was primarily constructed from an elaborate network of pendant measurements which is not acceptable in terms of any statistical verification of the data. Second, there were compounds such as pyridine listed within the published work which did not have any graphical connection with the ladder. Both of these conditions could be corrected by the addition of just a few measurements. The final problem was initially harder to observe as there were errors in the graph of the published reference.[32] As there are only ten measurements between the nine compounds, the agreement among the equilibria is primarily based on the agreement within two semi-independent subladders, the smaller cycle being

Table 2-10: Least Squares and Boundary Solution Summary for Fusion Sets with Nine Compounds.

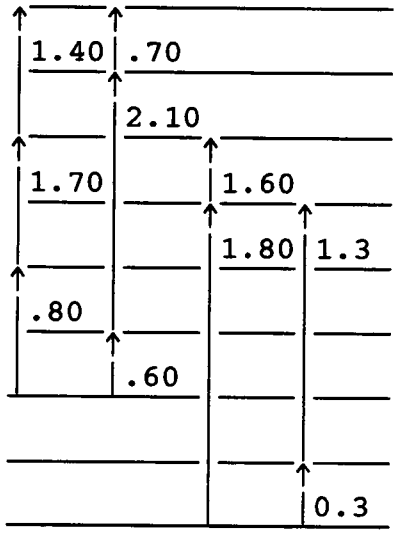
Scale	# Eq	Fusion Set	# Pert	Rms	Range:LS	Minimum	Maximum
87LAR/MCM[a]	10	CF_2HCFH_2 - CF_3CH_2OH	0	.1683251	4.6833	4.6000	4.8000
CN Data[b]	10	CF_2HCFH_2 - CF_3CH_2OH	0	.2345208	4.6500	4.1000	4.8000
76AUE/WEB[c]	12	tamylNH$_2$ - nPr$_2$NH	0	.0586302	4.9969	4.9000	5.0156
76AUE/WEB	14	nPrNH$_2$ - tamylNH$_2$	0	.1302861	3.7169	3.6000	4.0000
85MCM/KEB[d]	15	CH_4 - SO_2F_2	15	3.194657	14.2749	13.000	17.8562
SYNTHETIC	27	a4 - a9	8	.9703311	.5923	.5923	.7031

Notes: Affinity values are in kcal/mol. # Pert. refers to the perturbation calls made by the least squares algorithm.
Sources: [a] Larson, J.W.; McMahon, T.B. *J. Am. Chem. Soc.* **1987** *109*, 6230. [b] McMahon, T.B., personal communication regarding previous reference. [c] Aue, D.H.; Webb, H.M.; Bowers, M.T. *J. Am. Chem. Soc.* **1976** *98*, 311. [d] McMahon, T.B.; Kebarle, P. *J. Am. Chem. Soc.* **1985** *107*, 2612.



Compound	LS	LTS
CF_3CH_2OH	17.25000	17.31671
$(CFH_2)_2CHOH$	16.50000	16.56670
CF_2HOCF_2H	15.90000	15.96667
HCN	14.35000	14.41673
CF_3COMe	14.30000	14.36664
$ClCH_2CH_2OH$	14.25000	14.31669
FCH_2CH_2OH	13.50000	13.56674
CF_3OCF_2H	12.93333	13.06661
CF_2HCFH_2	12.56667	12.56667

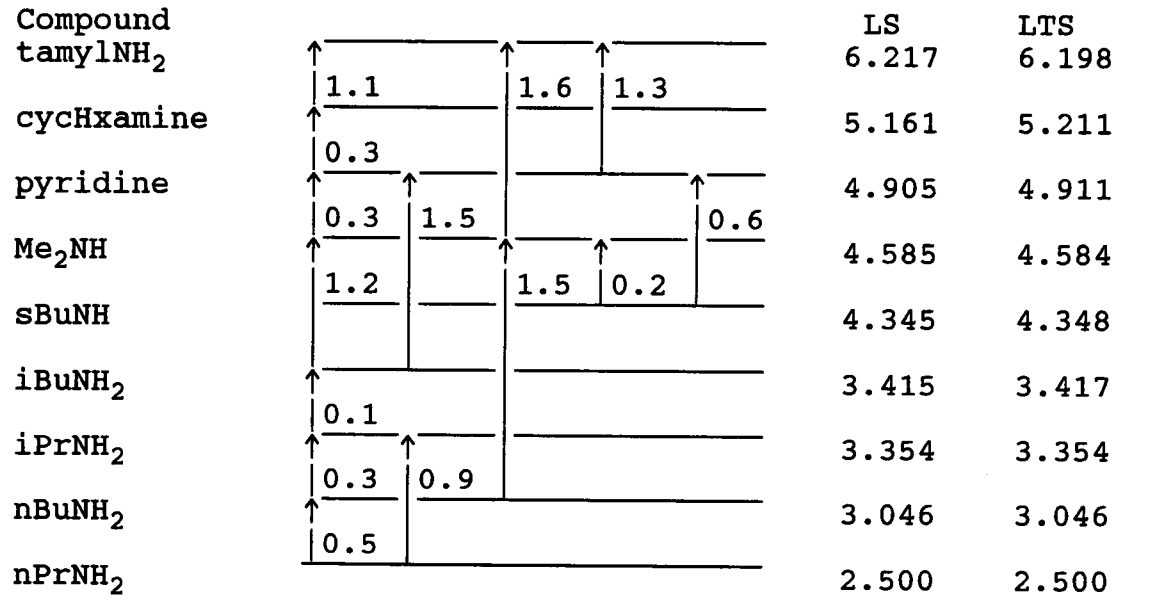
(a)



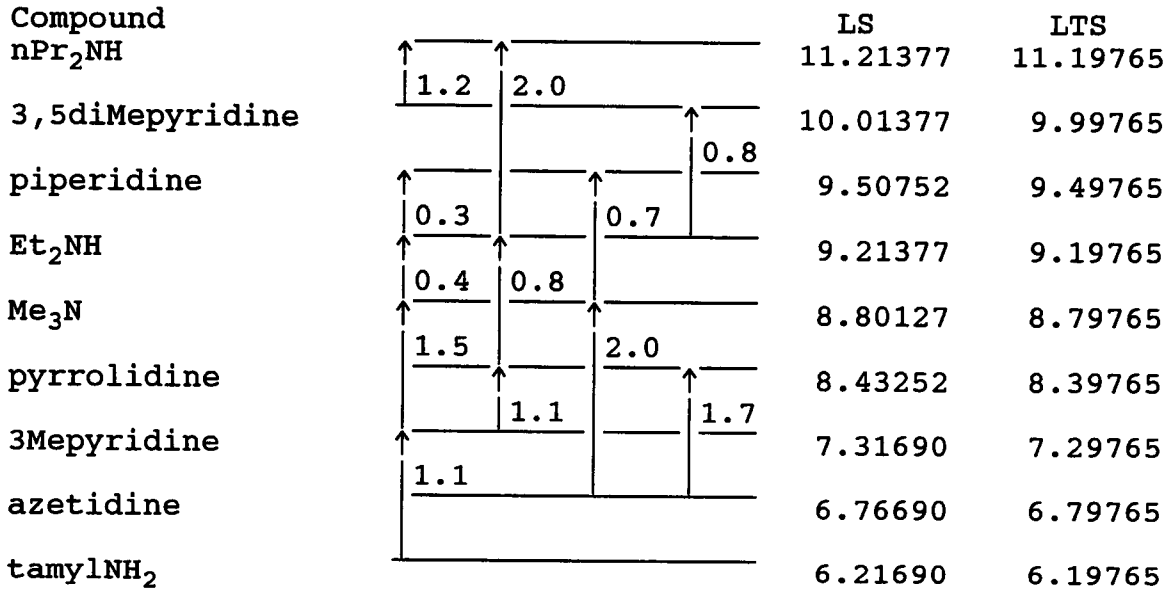
Compound	LS	LTS
CF_3CH_2OH	17.15000	17.15000
$(CFH_2)_2CHOH$	16.36667	16.36667
CF_2HOCF_2H	15.83333	15.83333
CF_3COMe	14.23333	14.23333
HCN	14.21667	14.21667
$ClCH_2CH_2OH$	14.18333	14.18333
FCH_2CH_2OH	13.50000	13.50000
CF_3OCF_2H	12.86667	12.86667
CF_2HCFH_2	12.50000	12.50000

(b)

Figure 2-38: Portion of 87LAR/MCM: Nine-compound Scale. Source: Larson, J.W.; McMahon, T.B. *J. Am. Chem. Soc.* **1987** *109*, 6230.
(a) Published scale of CN Affinities. (b) Same scale with corrections from T.B. McMahon.



(a)



(b)

Figure 2-39: Portion of 76AUE/WEB: Nine-compound Scale. Source: Aue, D.H.; Webb, H.M.; Bowers, M.T. *J. Am. Chem. Soc.* **1976** *98*, 311.

Basicity of alkyl amines. (a) 12 equilibria (b) 14 equilibria

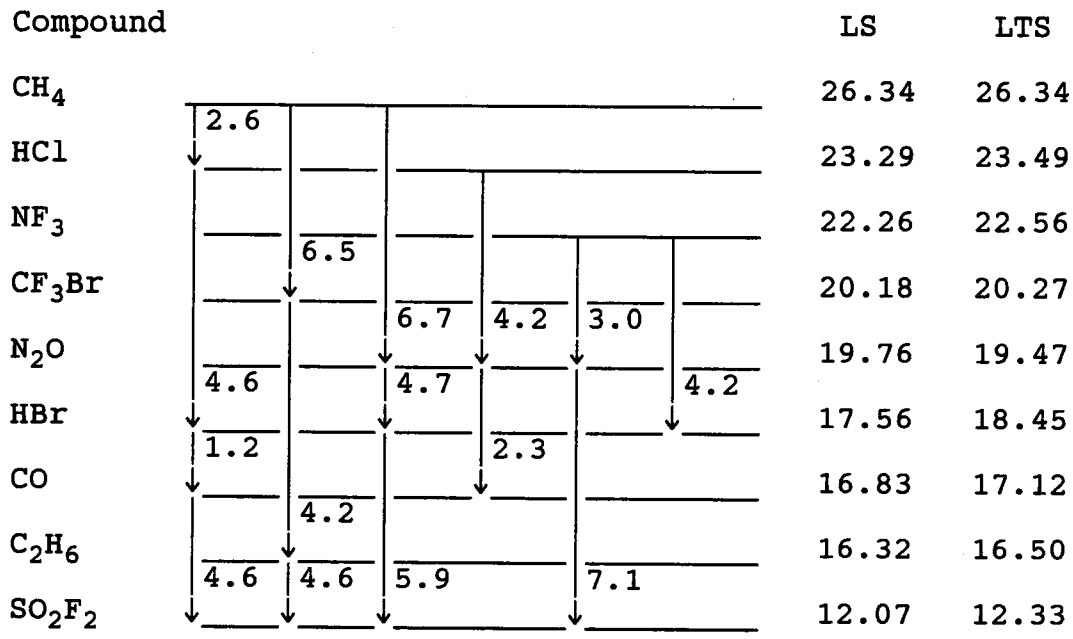


Figure 2-40 Portion of 85MCM/KEB: Nine-compound Scale. Source: McMahon, T.B.; Kebarle, P. *J. Am. Chem. Soc.* **1985** *107*, 2612.

Weaker end of the Basicity scale

$CF_2HCFH_2 \rightarrow CF_3OCF_2H \rightarrow CF_3C{=}OCH_3$. The small subladder is connected via 1,1,1-trifluoroacetone to the larger subladder. The boundary solution test shows decent agreement among the equilibria. The iterative smoothing solution obtains values slightly different than the those calculated by LS, but the differences can be attributed to rounding errors in the iterative solution. The LTS algorithm chooses the pair CF_3OCF_2H - CF_2HCH_2F as the statistically worst value among the ten measurements. However, this scale does not have enough measurements to make any real statistical evaluation of the data. Results

of the analysis were presented to McMahon, who responded with corrections to the original data used in constructing the scale.[97] The corrected fusion set appears as Figure 2-38 (b). The analysis of the corrected scale follows the analysis of the published scale with the exception that in the corrected scale the LTS algorithm exactly matches the LS solution. There is less agreement among the measurements in the corrected version of the scale. Again, with only ten measurements in a nine compound scale, no real statistical evaluations can be performed on the data.

The amine study between n-propyl amine and t-amyl amine shows fairly good agreement.[13] All of the residuals from the initial LS solution are less than 0.5 kcal/mol. The number of iterative passes causes the iterative solution to decrease slightly from its LS counterpart. The LTS solution picks one of the larger errors as the most suspicious, but the error is small as well as evenly spread. Any of the six worst measurements were likely candidates to erase only 0.1 kcal/mol worth of total residual. The fusion set between t-amyl amine and di-n-propyl amine within the same study shows even better agreement. In the latter case, all of the error can be attributed to one measurement by the LTS algorithm.

By contrast, the basicity fusion set from CH_4 to SO_2F_2 from the work of McMahon and Kebarle shows remarkably poor agreement. The difference between the boundary solutions is

4.8 kcal/mol, which averages to more than 0.5 kcal/mol difference per compound. The most suspicious measurement according to LTS is between N_2O and HBr. Visual inspection of Figure 2-40 shows that there is a large discrepancy between the measurements that make up the HCl → N_2O → HBr subladder. By direct measurement, ΔG_{eq} is 4.6 kcal/mol but the path summing two affinities is 8.9 kcal/mol. Clearly, there is an error here in one or more of the measured or reported values. The effect of removing the N_2O - HBr equilibrium by LTS lowers the affinity of HBr relative to CH_4 by almost 0.9 kcal/mol.

The nine-membered synthetic scale behaved exactly as the previous analogous synthetic scales.

I. Mid-Sized Scales.

As the number of compounds increases, the number of representative fusion sets will decrease accordingly. Therefore, fusion sets containing ten to twenty compounds will be grouped together. The LS results for these fusion sets are listed in Table 2-11. Figures 2-41 to 2-55 give graphical representation of these scales.

The electron affinity scale between nitrobenzene and p-cyanonitrobenzene[15] shows only fair agreement among the fifteen equilibria as judged by the boundary solution test. Much of the disagreement is between the affinities for the

Table 2-11: Least Squares and Boundary Solution Summaries for Fusion Sets Containing From Ten to Twenty Compounds.

Scale	# Cpd	# Eq	Fusion Set	# Pert	Rms	Range: LS	Minimum	Maximum
87KEB/CHO[a]	10	15	NB - 4CNNB	4	.3923510	16.1061	15.7000	16.6092
78WOO/BEA[b]	10	17	Me_2NH - H_2O	0	.1848942	6.7361	6.4800	7.0500
SYNTHETIC	10	30	a4 - a10	8	.9703910	.6897	.6897	.8030
77WOL/STA[c]	11	14	CF_3CO_2Et	0	.1522059	8.0700	7.9000	8.2000
77MCM/KEB[d] benzoic acd	11	17	mCH_3PhCO_2H - $mClPhCO_2H$	4	.3039424	5.5429	5.4000	5.6095
SYNTHETIC	11	33	a4 - a11	8	.9703991	.7925	.7925	.9030
91MEO/SIE[e]	12	24	2Fpyr - Me_3N	15	1.349114	17.3057	15.7000	18.9376
SYNTHETIC	12	36	a4 - a12	8	.9704000	.8915	.8915	1.0030
78CUM/KEB[f]	13	22	Pyrrole - Acetone	17	.598453	10.0793	9.6000	10.8000
SYNTHETIC	13	39	a4 - a13	8	.9704006	.9914	.9914	1.1030
92MIS/KAN[g]	14	23	C_6H_5CHO - pOMe	2	.3174748	8.5665	8.1000	8.8000
F DATA[h]	14	20	tBuOH - F_2CO	13	.6057705	8.1396	7.8169	9.3000
SYNTHETIC	14	42	a4 - a14	8	.9704008	1.0917	1.0917	1.2030

Notes: Affinity values are in kcal/mol. # Pert. refers to the perturbation calls made by the least squares algorithm.

Table 2-11: (continued)

Scale	# Cpd	# Eq	Fusion Set	# Pert	Rms	Range: LS	Minimum	Maximum
77MCM/KEB[d] phenols	15	27	pClPhOH - pNH_2PhOH	8	.5410760	11.0566	16.4000	11.4000
SYNTHETIC	15	45	a4 - a15	8	.9704008	1.1916	1.1916	1.3030
76YAM/KEB[i]	16	31	EtOH - NH_3	7	.9475662	14.0133	13.5000	15.8000
Cl data[j]	16	22	MeCl - SO_2	5	.4399168	8.3584	7.7000	8.5000
89CAL/REN[k]	18	35	$MeOCH_2CO_2H$ - F_2CHCO_2H	6	.4114173	11.2002	10.8000	11.8000
FDATA[h]	18	26	tBuCHO - tBuOH	5	.3406845	8.6028	7.9000	8.9000
89CAL/REN[k]	19	35	$MeCO_2H$ - $MeOCH_2CO_2H$	6	.4700878	6.1153	5.6000	6.6000
79BAR/SCO[l]	19	33	neoPnOH - An	11	.5669531	5.4441	4.7500	6.1000
89CHO/KIS[m]	20	29	4-CNNB - 3-MeNB	0	.1562133	16.5668	16.3000	16.7305
SYNTHETIC	20	60	a4 - a20	4	.4478773	1.7389	1.7389	1.7389

Notes: Affinity values are in kcal/mol. # Pert. refers to the perturbation calls made by the least squares algorithm.

Sources: [a] Woodin, R.L.; Beachamp, J.L. *J. Am. Chem. Soc.* **1977** *100*, 501. [b] Kebarle, P., Chowdhury *Chem. Rev.* **1987** *87*, 513. [c] Wolf, J.A.; Staley, R.H.; Koppel, I.; Taagepera, M.; McIver, R.T., Jr.; Beachamp, J.L.; Taft, R.W. *J. Am. Chem. Soc.* **1977** *99*, 5417. [d] McMahon, T.B.; Kebarle, P.; *J. Am. Chem. Soc.* **1977** *99*, 2222. [e] Meotner, M.; Sieck, L.W. *J. Am. Chem. Soc.* **1991** *113* 4448. Affinities here are ΔH's [f] Cumming, J.B.; Kebarle, P.; *Can. J. Chem.* **1978** *56*, 1. [g] Mishima, M.; Kang, C.H.; Fujio, M.; Tsuno, Y. *Chem. Lett.* **1992**, 493.

Table 2-11: (continued)

[h] Larson, J.W.; McMahon, T.B. *J. Am. Chem. Soc.* **1983** *105*, 2944.; *J. Phys. Chem.* **1984** *88*, 1083.; *J. Am. Chem. Soc.* **1985** *107*, 766. [i] Yamdagi, R.; Kebarle, P. *J. Am. Chem. Soc.* **1976** *98*, 1320. [j] Larson, J.W.; McMahon, T.B. *J. Am. Chem. Soc.* **1985** *107*, 766.; *J. Am. Chem. Soc.* **1984** *106*, 517.; *Can. J. Chem.* **1984** *62*, 675. *J. Phys. Chem.* **1984** *88*, 1083. [k] Caldwell, G.; Renneboog, R.; Kebarle, P.; *Can. J. Chem.* **1989** *67*, 611. [l] Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979** *101*, 6046.; Lias, S.G.; Bartmess, J.E.; Liebman, J.F.; Holmes, J.L.; Levin, R.D.; Mallard, W.G.; *J. Phys. Chem. Ref. Data*, **1988**, *17*, Suppl. 1. [m] Chowdhury, S.; Kishi, H.; Dillow, G.W.; Kebarle, P. *Can. J. Chem.* **1989** *67*, 603.

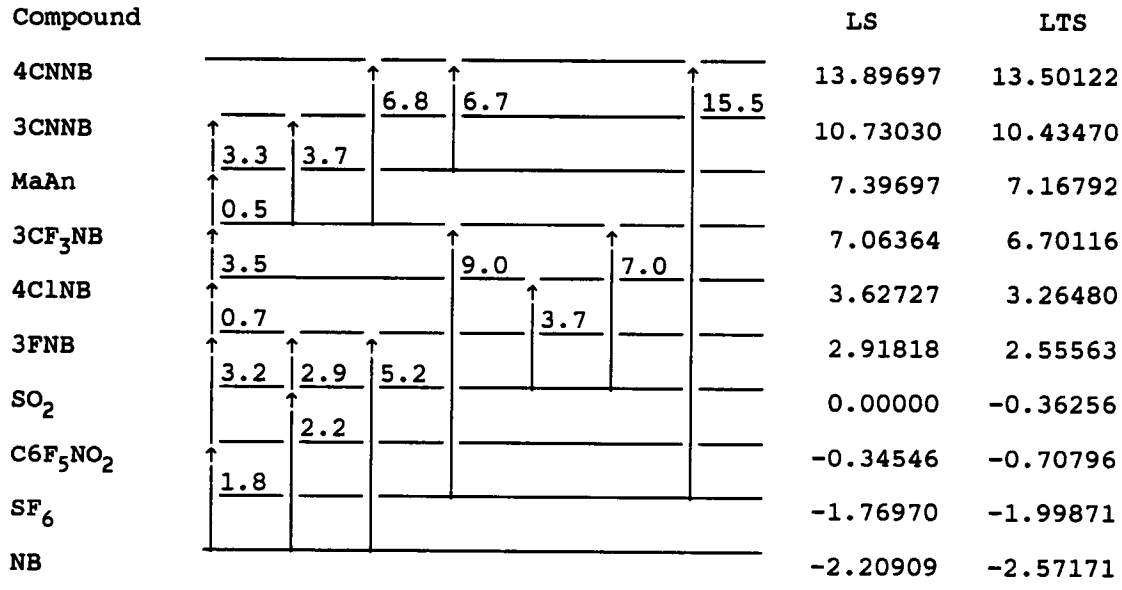


Figure 2-41: Portion of 87KEB/CHO: Ten Compound Scale: Source: Kebarle, P., Chowdhury *Chem. Rev.* **1987** *87*, 513. Electron affinities.

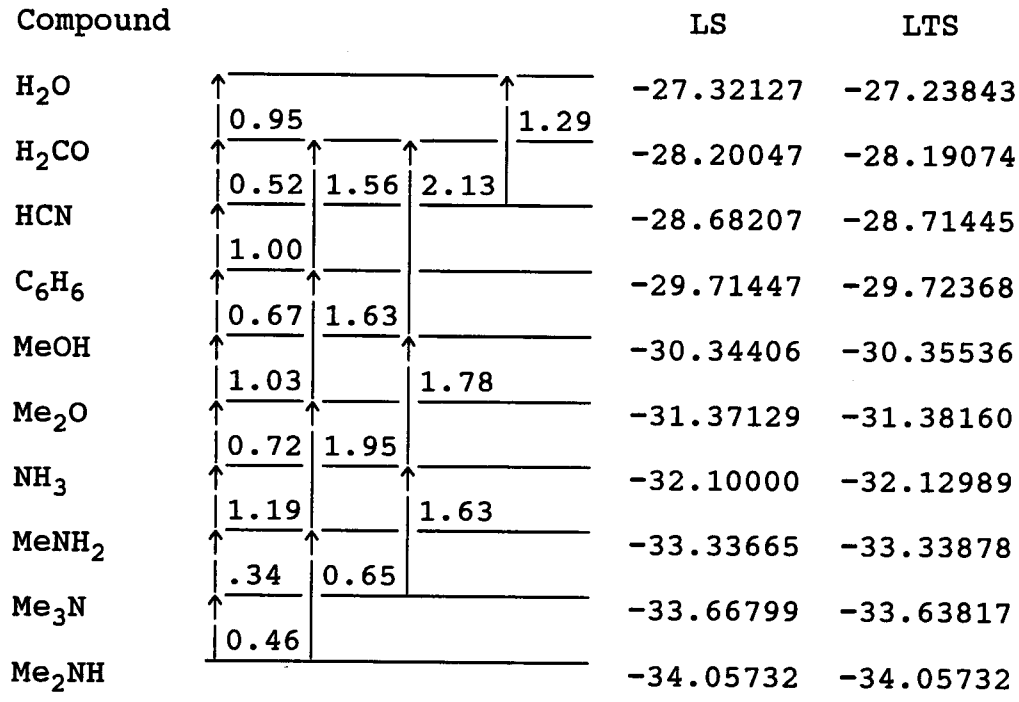


Figure 2-42: Portion of 78WOO/BEA: Ten Compound Scale
Source: Woodin, R.L.; Beachamp, J.L. *J. Am. Chem. Soc.* **1978** *100*, 501.

Binding of Li^+ to Lewis Bases in the Gas phase. ICR work.

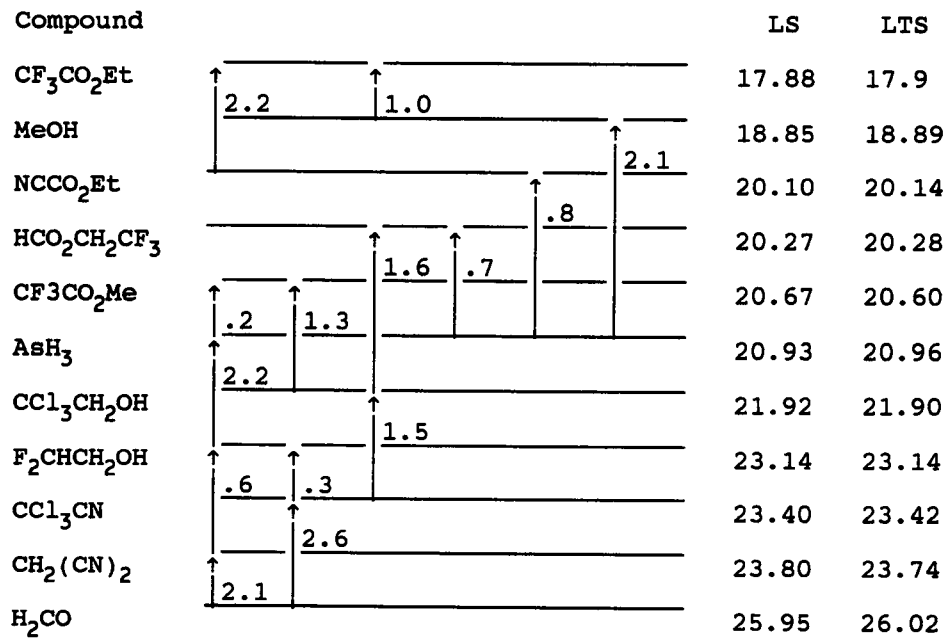


Figure 2-43. Portion of 77WOL/STA: Eleven Compound Scale Source: Wolf, J.A.; Staley, R.H.; Koppel, I.; Taagepera, M.; McIver, R.T., Jr.; Beachamp, J.L.; Taft, R.W. *J. Am. Chem. Soc.* **1977** *99*, 5417.

Gas Phase Basicities - ICR study.

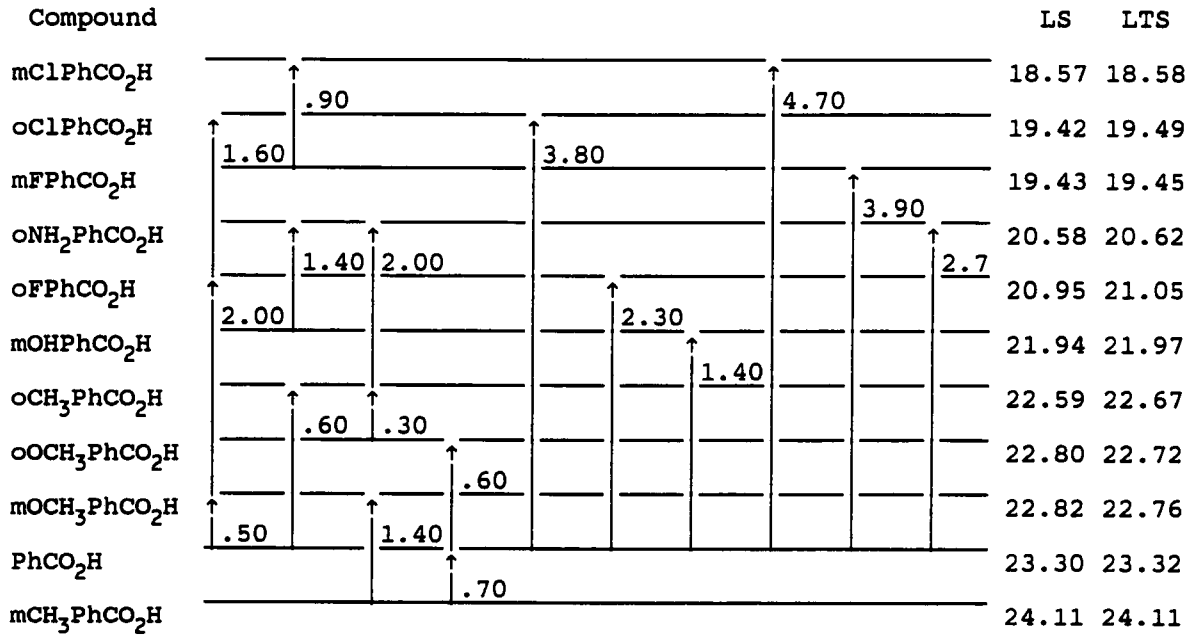


Figure 2-44: Portion of 77MCM/KEB: Eleven Compound Scale: Source: McMahon, T.B.; Kebarle, P.; *J. Am. Chem. Soc.* **1977** *99*, 2222.

HPMS acidities for substituted benzoic acids.

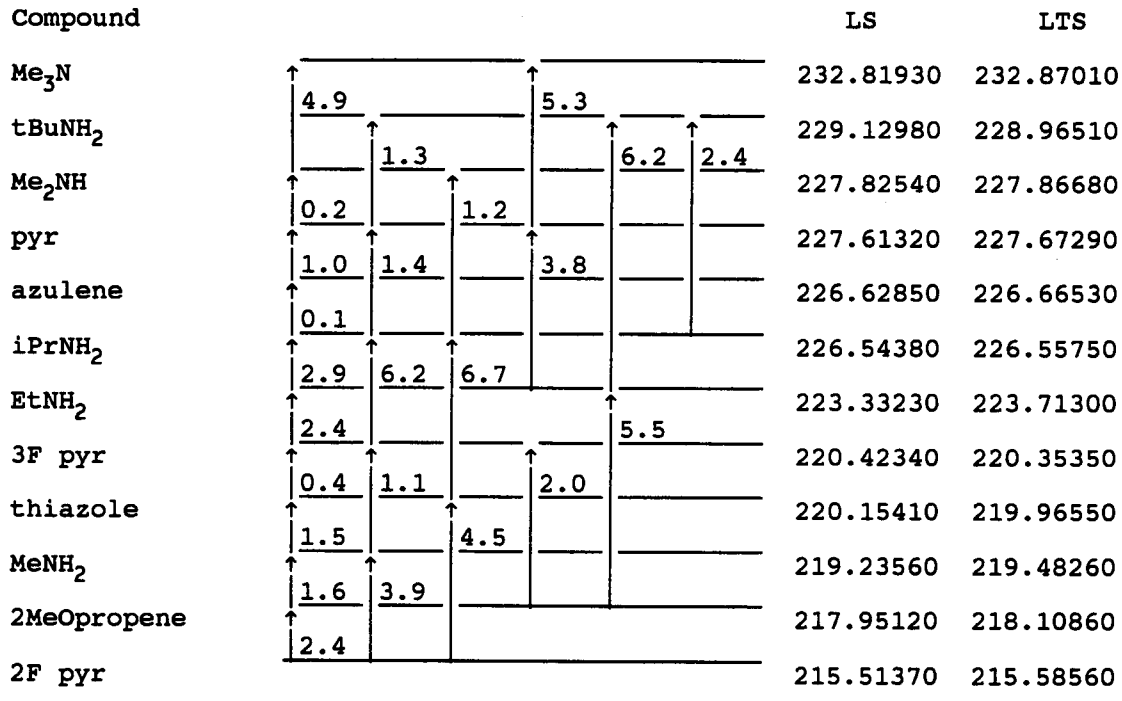


Figure 2-45: Portion of 91MEO/SIE: Twelve Compound Scale. Source: Meotner, M.; Siek, L.W. *J. Am. Chem. Soc.* **1991** *113*, 4448. Affinities here are ΔH's

Proton affinity -variable temperature equilibrium measurements, Upper Proton Affinity Range.

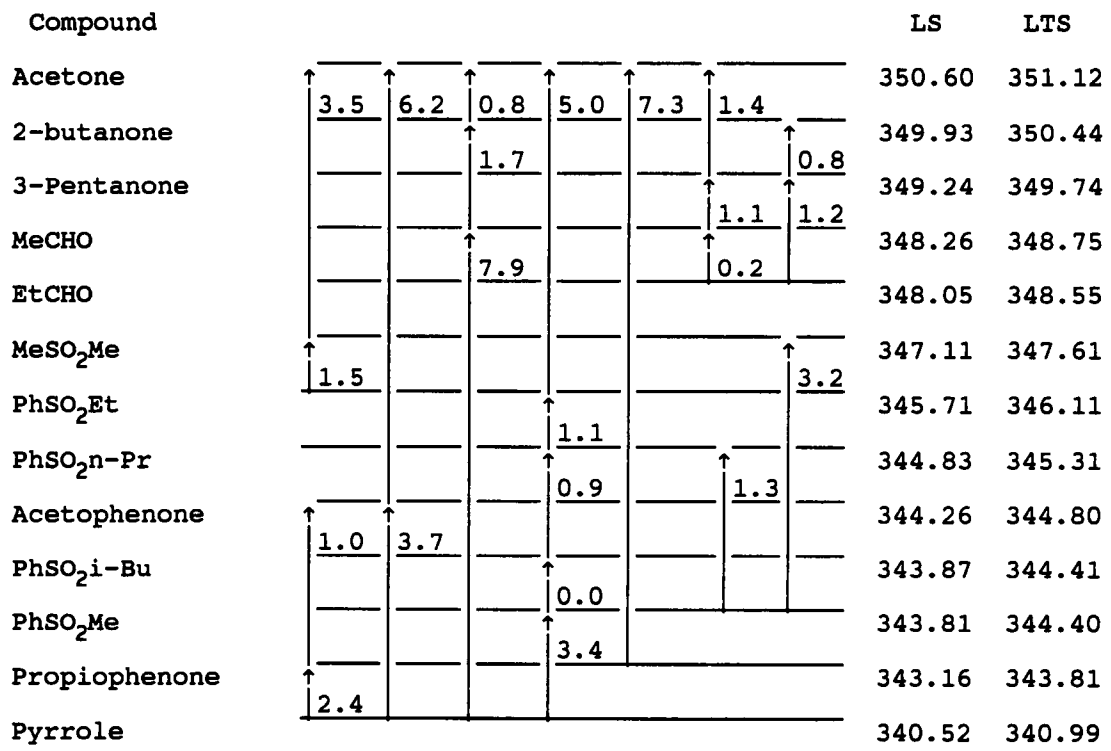


Figure 2-46. Portion of 88CUM/KEB: Thirteen Compound Scale. Source: Cumming, J.B.; Kebarle, P.; *Can. J. Chem.* **1978** 56, 1.

Gas phase acidities.

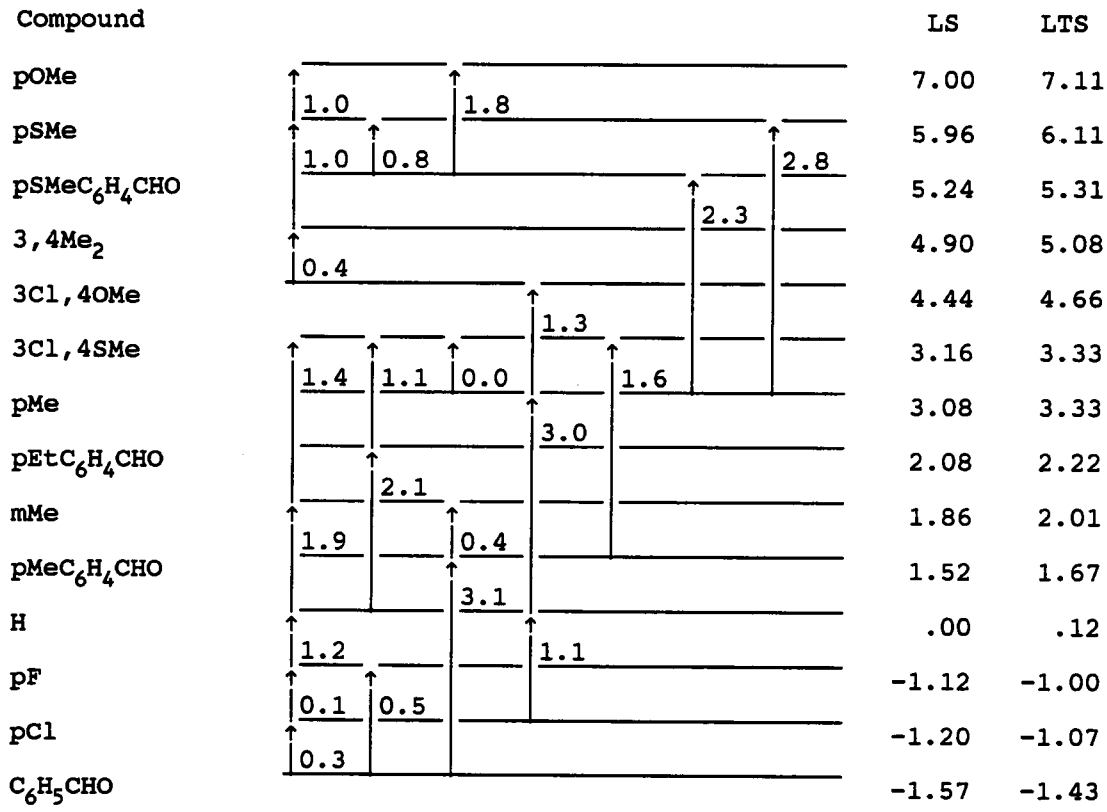


Figure 2-47. Portion of 92MIS/KAN: Fourteen Compound Scale. Source: Mishima, M.; Kang, C.H.; Fujio, M.; Tsuno, Y. *Chem. Lett.* **1992** 493.

Gas phase basicities of Acetophenones toward TMS cation

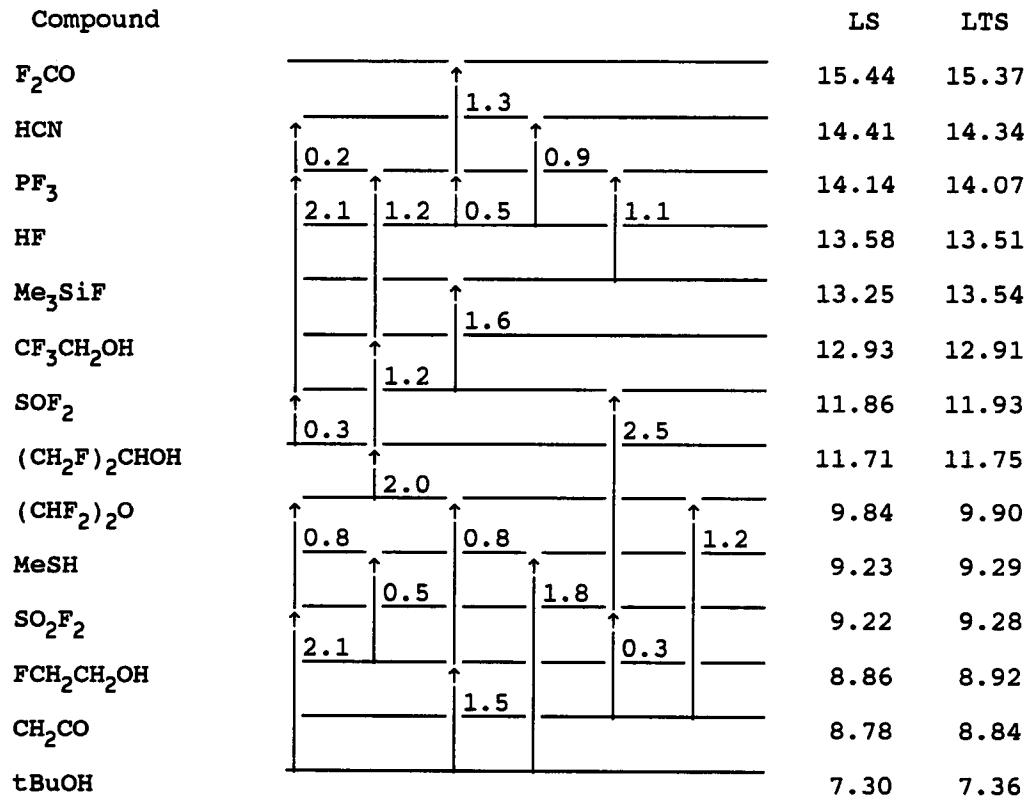


Figure 2-48. Portion of F^- DATA: Fourteen Compound Scale. Source: Larson, J.W.; McMahon, T.B. *J. Am. Chem. Soc.* **1983** 105, 2944.; *J. Phys. Chem.* **1984** 88, 1083.; *J. Am. Chem. Soc.* **1985** 107, 766.

Compilation of the Fluoride Affinity Scale.

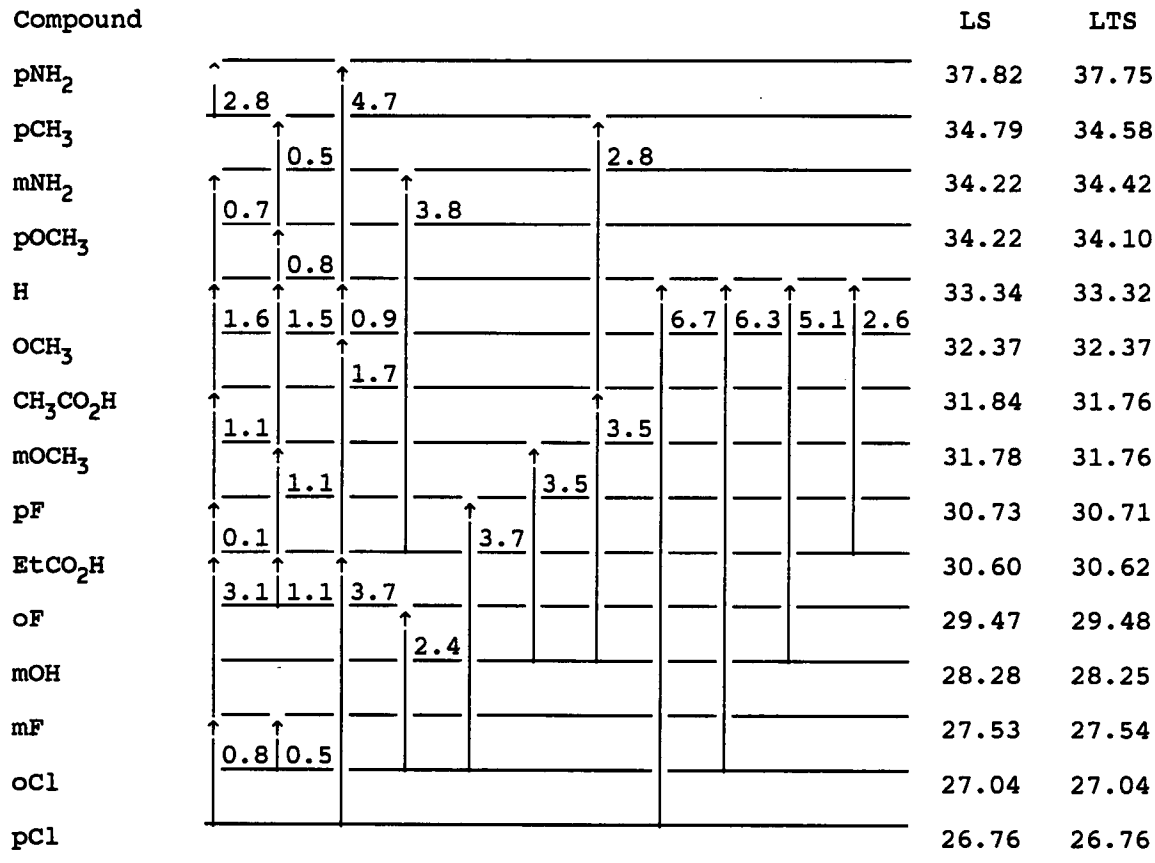


Figure 2-49: Graph of 77MCM/KEB: Fifteen Compound Scale. Source: McMahon, T.B.; Kebarle, P.; *J. Am. Chem. Soc.* **1977** *99*, 2222.

PARTIAL PHENOL SCALE- UPPER PORTION.

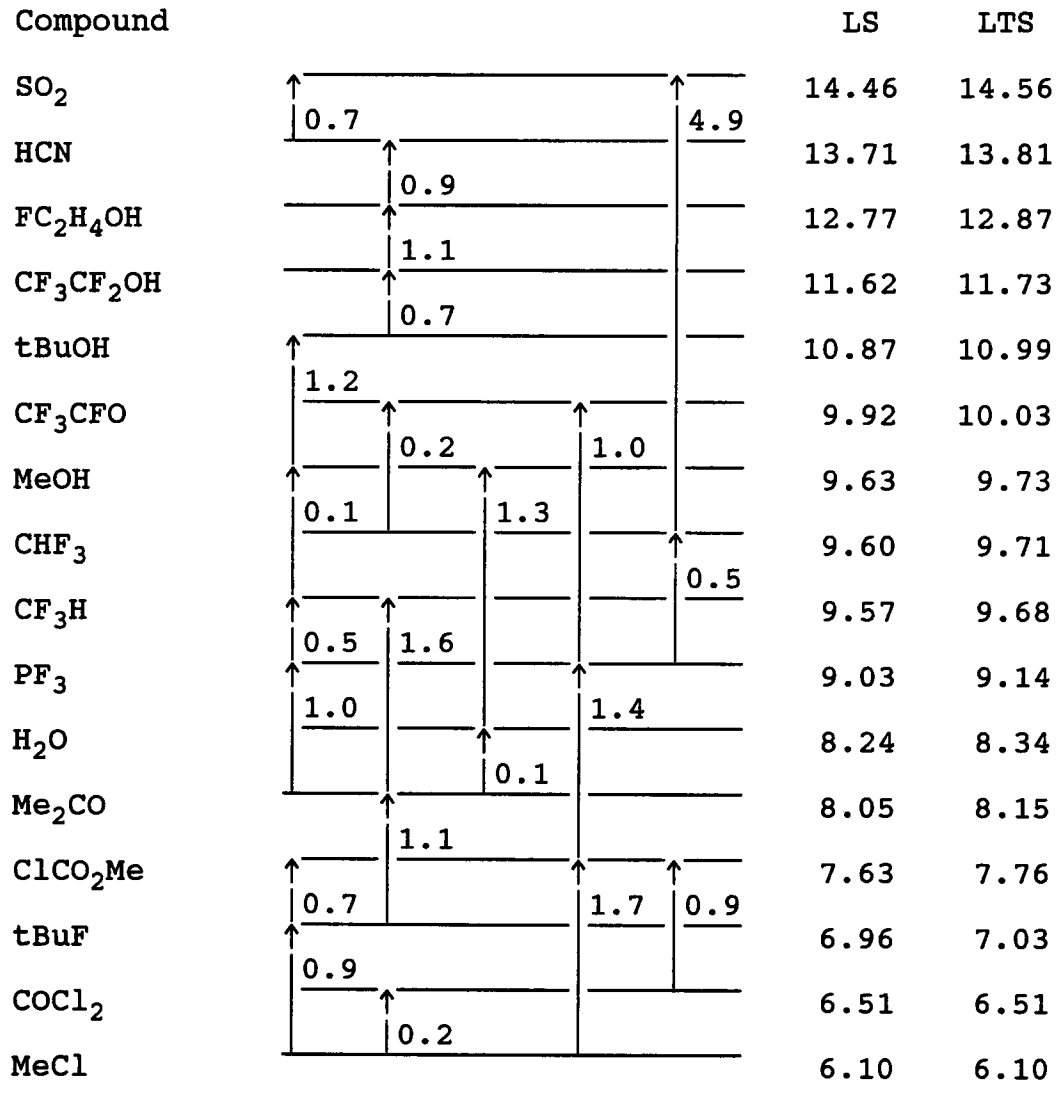


Figure 2-50. Portion of Cl^- data: Sixteen Compound Scale. Source: Larson, J.W.; McMahon, T.B. *J. Am. Chem. Soc.* **1985** 107, 766.; *J. Am. Chem. Soc.* **1984** 106, 517.; *Can. J. Chem.* **1984** 62, 675. *J. Phys. Chem.* **1984** 88, 1083.

Compilation of Chloride Affinities

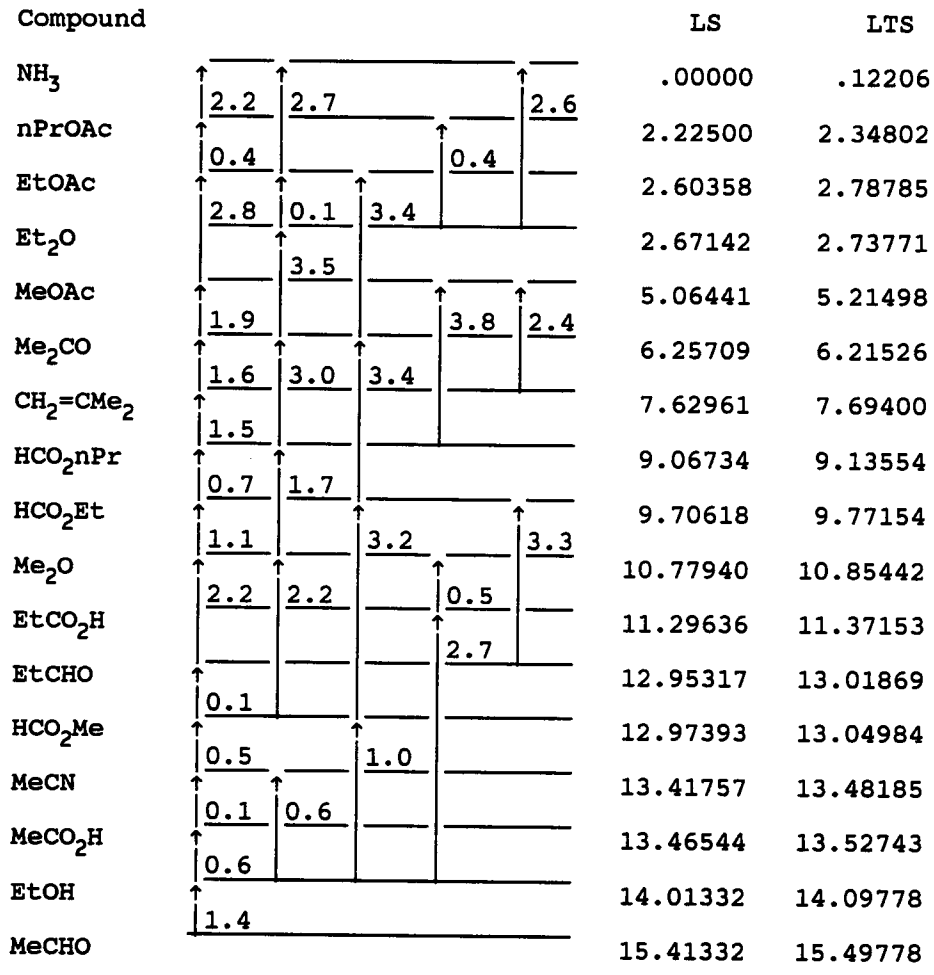


Figure 2-51. Portion of 76YAM/KEB: Seventeen Compound Scale. Source: Yamdagi, R.; Kebarle, P. *J. Am. Chem. Soc.* **1976** 98, 1320.

Gas phase basicities

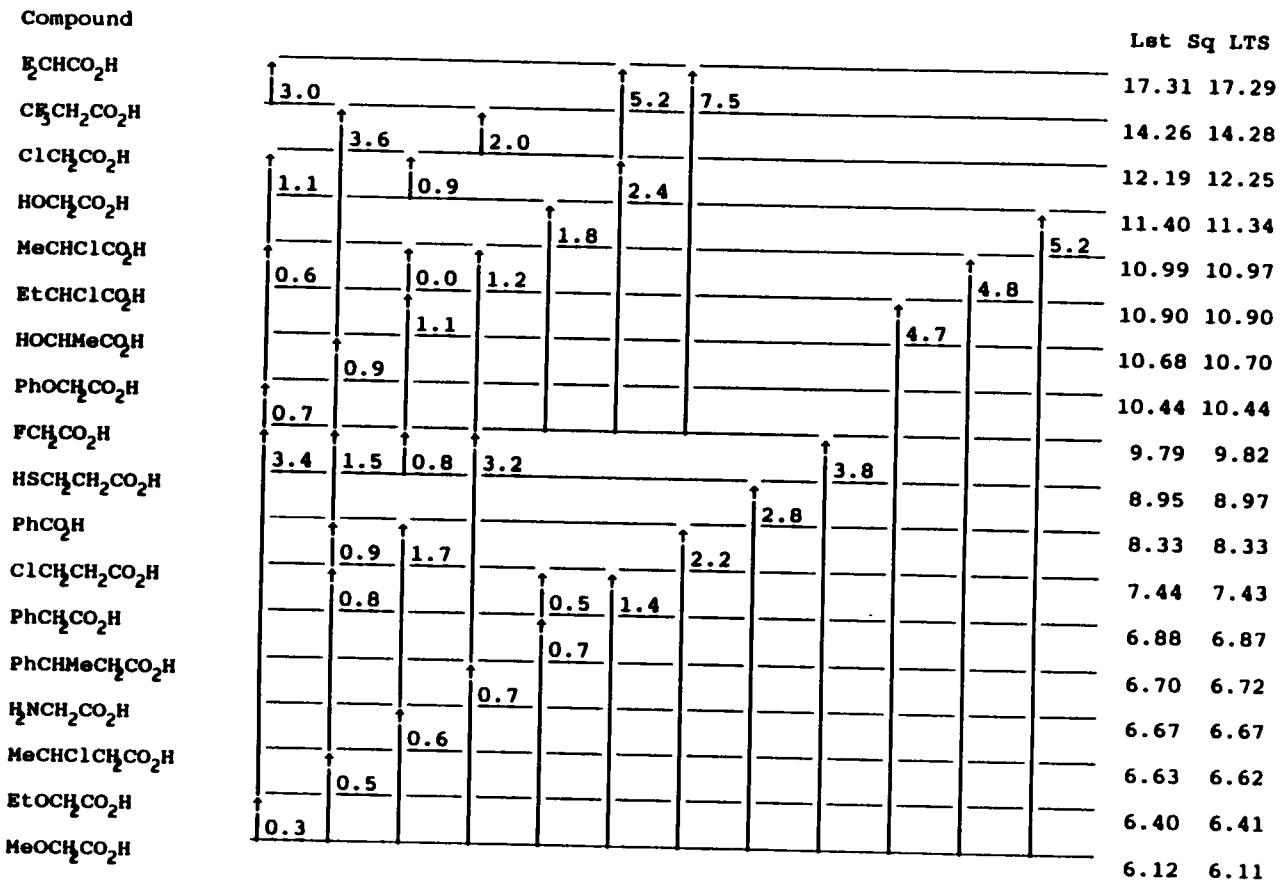


Figure 2-52. Portion of 89CAL\REN: Eighteen Compound Scale. Source: Caldwell, G.; Renneboog, R.; Kebarle, P.; *Can. J. Chem.* **1989** 67, 611-18.

Aliphatic carboxylic acid acidities, based on H^+ transfer.

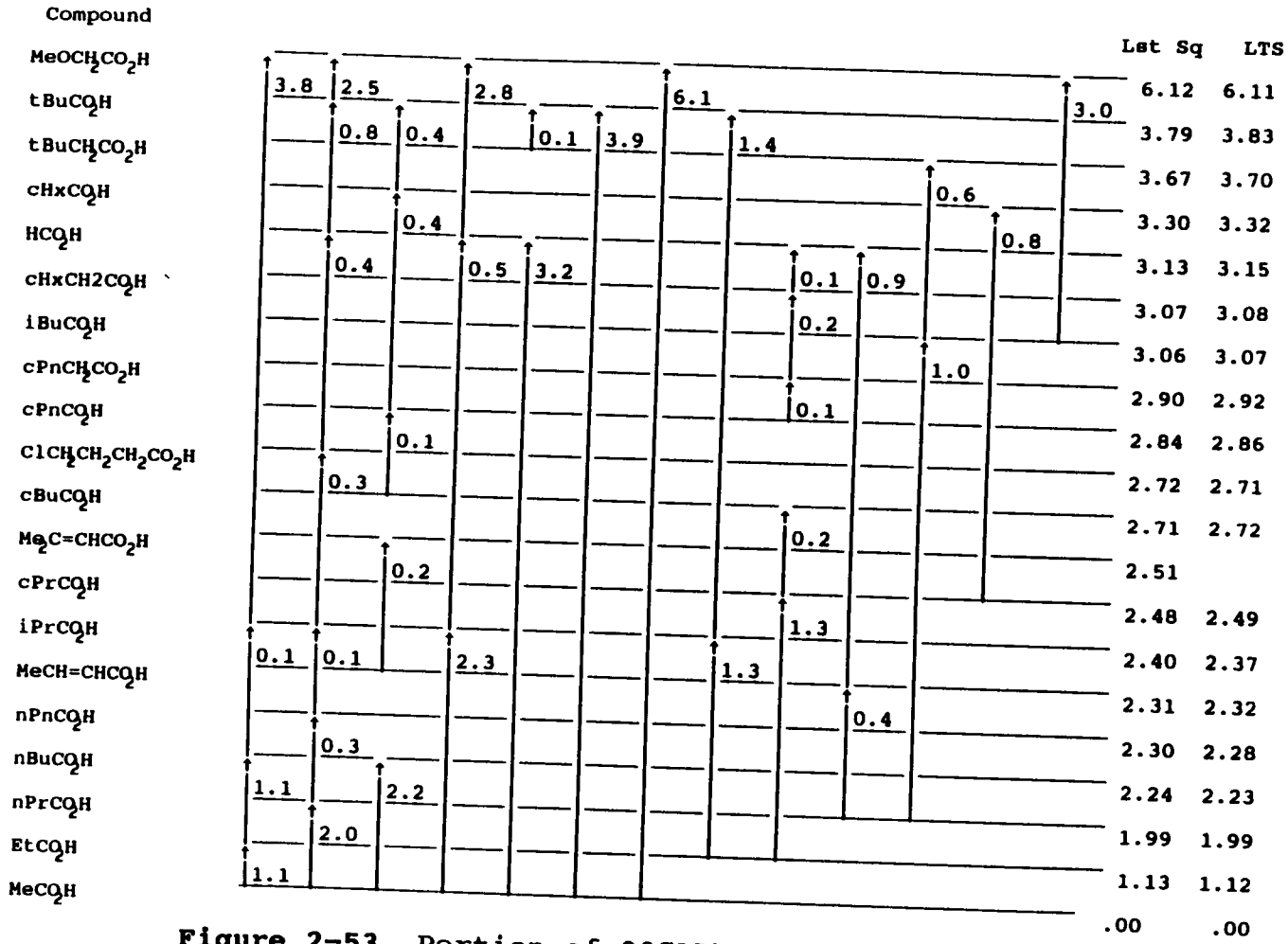


Figure 2-53. Portion of 89CAL\REN: Nineteen Compound Scale. Source: Caldwell, G.; Renneboog, R.; Kebarle, P.; *Can. J. Chem.* **1989** 67, 611-18.

Aliphatic carboxylic acid acidities, based on H^+ transfer.

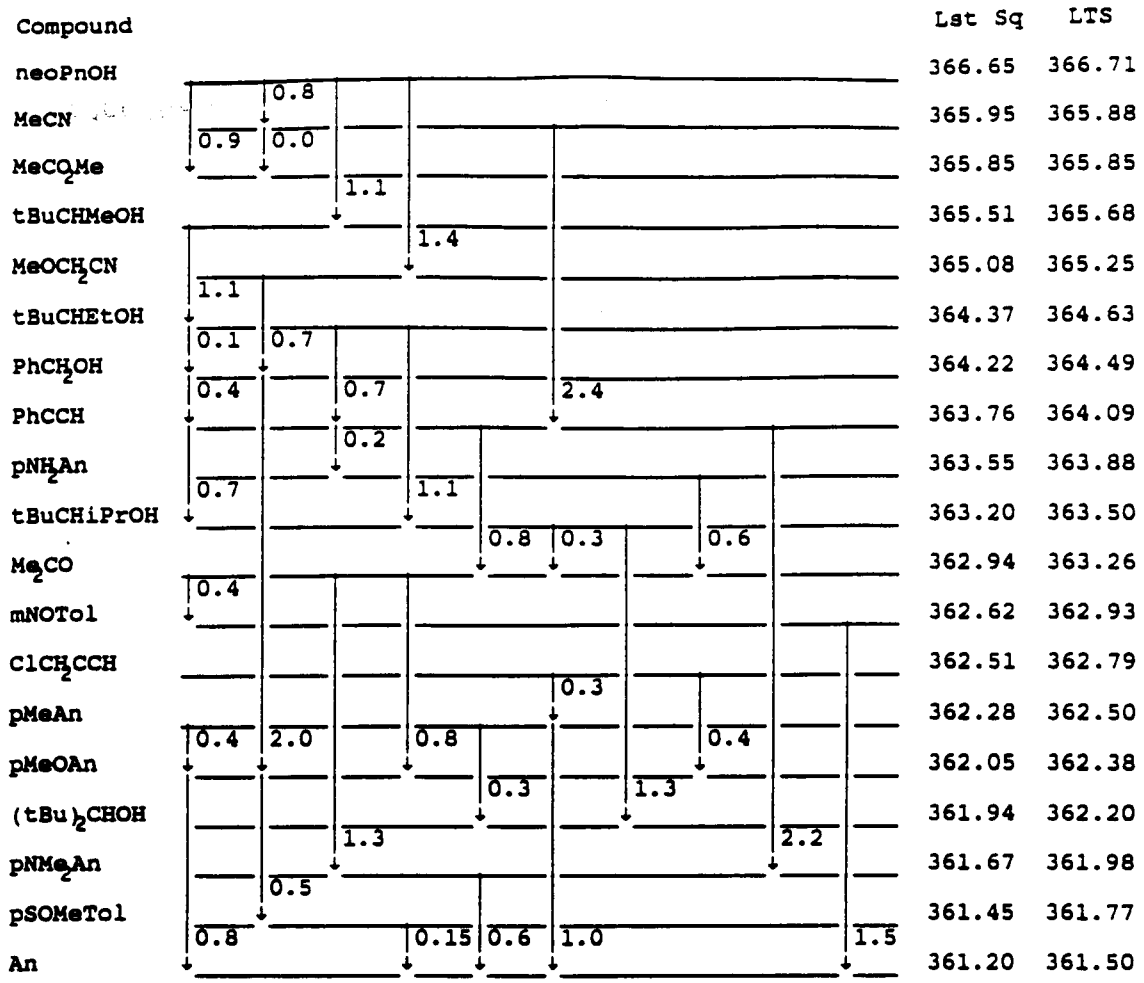


Figure 2-54. Portion of 74BAR\SCO: Nineteen Compound Scale. Source: Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979** 101, 6046-56.

Scale has a few of the more questionable equilibria data points removed. Also removed are a few measurements to break the scale into smaller fusion sets.

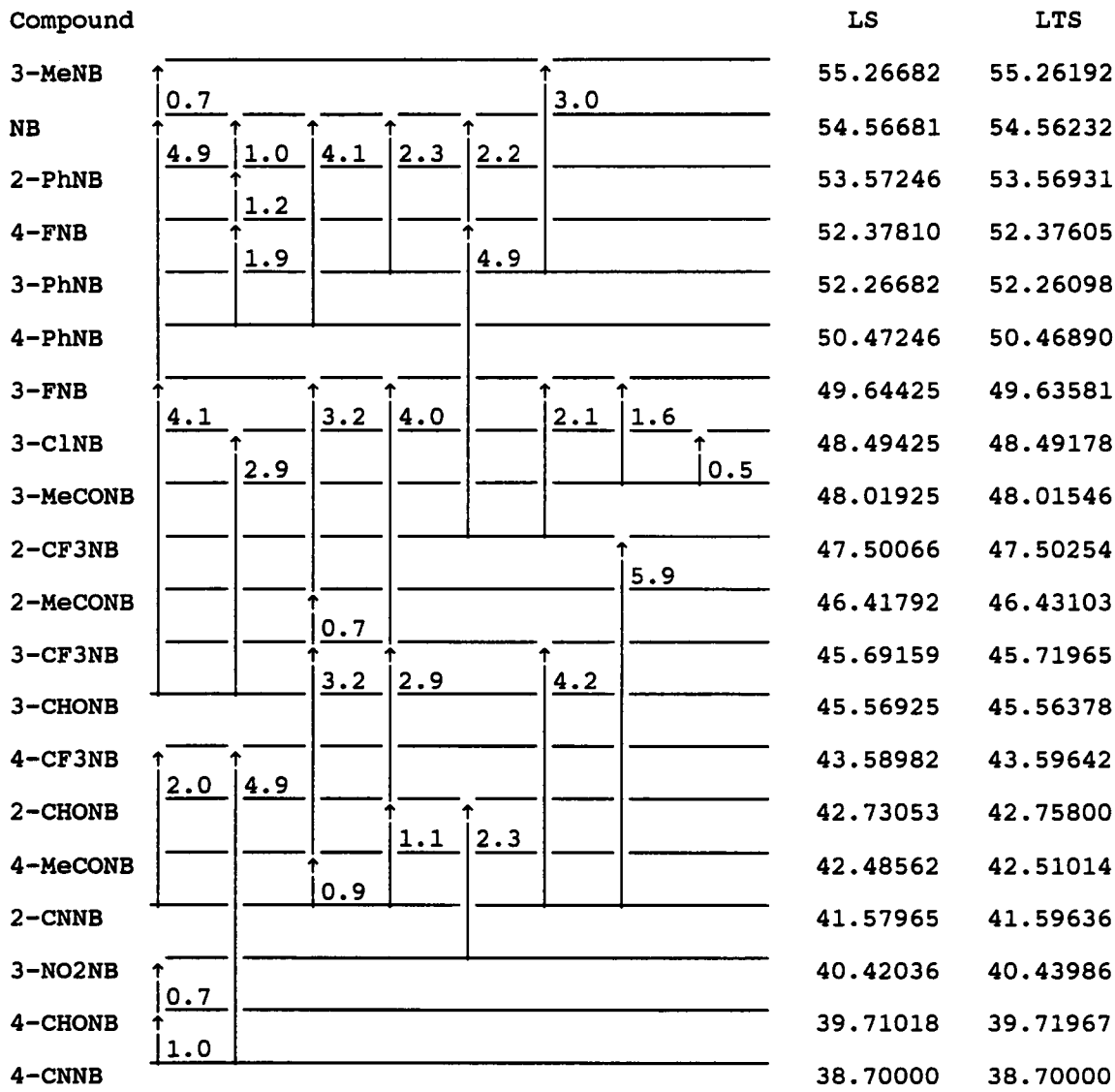


Figure 2-55: Portion of 89CHO/KIS: Twenty Compound Scale. Source: Chowdhury, S.; Kishi, H.; Dillow, G.W.; Kebarle, P. *Can. J. Chem.* **1989** *67*, 603.

two pathways leading from SF_6 to p-cyanonitrobenzene: the two disagree by 0.3 kcal/mol. Another influential compound where the differences in experimental affinities greatly affect the total residual is maleic anhydride. The values obtained by the iterative smoothing solution are slightly larger than those from the LS solution, but the iterative solution is decreasing with each successive pass. The LTS algorithm removes two equilibria as suspicious: maleic anhydride to p-$CNC_6H_4NO_2$ and SF_6 to m-$NO_2C_6H_4CF_3$. Both of the equilibria are questionable, although there are not enough measurements to make absolute evaluations. The result of removing these two measurements by the LTS algorithm increases the affinity of SF_6 and maleic anhydride by more than 0.1 kcal/mol.

The Li^+ affinity scale of Woodin and Beauchamp[19] is the second literature fusion set of ten compounds. Good agreement is shown among the equilibria using the boundary solution test. Furthermore, the large LS residual is only 0.07 kcal/mol. There is fair agreement between the iterative and the LS solutions. The LS values agree with those given by Woodin and Beauchamp.[19] By the 65^{th} pass of the iterative algorithm there is a noticeable shrinking (0.02 cal/cycle) of the range value from its previous value. The LTS algorithm temporarily removes the suspicious values of two equilibria: HCN-H_2O and Me_3N-NH_3. In the first case, either of the equilibria involving water may be a

truly suspicious measurement since there are only two measurements to water within the scale.

The solution for the ten-compound synthetic scale behaves exactly as the solutions for the lower case analogues.

The first eleven-membered case comes from the 77WOL/STA study.[6] The fusion set from H_2CO to CF_3CO_2Et shows good agreement among the fourteen measurements. The iterative smoothing solutions shrinks slightly compared to the LS solution. The LTS algorithm chooses to remove two measurements. Both are reasonable candidates for further study even though the absolute error is small.

The second eleven-member study is one of HPMS acidities for substituted benzoic acids.[1] The agreement among the seventeen measurements is excellent. The iterative solution again shrinks with the increasing number of passes necessary to find the solution. The LTS solution again chooses two measurements to isolate as suspect. The principal effect of this decision is to lower the affinity of o-fluorobenzoic acid by 0.1 kcal/mol.

The eleven-member synthetic scale poses a new problem not seen in the smaller synthetic scales. Here the LTS algorithm chooses to remove all references to the mock compound, a1, in determining the solution. Because the synthetic scale is well constructed in terms of overlaps, having three paths for each compound, it is unlikely that

the advice regarding scales of this size as given by the current LTS algorithm will be foolproof. The LTS algorithm also removes all three measurements involving a1 in its analysis of each of the synthetic scales of twelve to fifteen compounds.

One twelve compound fusion set is found in the literature, that being from Meot-ner's and Sieck's 1991 study.[17] The ΔH and ΔG scales from this work have two different fusion set structures because of the entropy changes. The twelve-compound ladder is the one of enthalpies. There is little agreement among the equilibria according to the boundary solution test: the difference per compound here is almost 0.27 kcal/mol, where a good agreement difference would be less than 0.05 kcal/mol. No additional information can be obtained from the iterative solution which was not present in the LS solution. The LTS solution picks out three different equilibria as possibly being suspicious: $MeNH_2$ - thiazole, 3F-pyridine - $EtNH_2$, and $EtNH_2$ - $tBuNH_2$. Each of these equilibria represent a possible suspicious ΔH_{eq} value, but there are other possible choices for re-investigation as well.

The only thirteen-membered set is part of Cumming and Kebarle's study on gas phase acidities.[2] The agreement between equilibria is marginal according to the boundary difference determination. Only eight of the twenty-two measurements between pyrrole and acetone show a LS residual

of less than 0.1 kcal/mol. The shrinking pattern seen in the iterative solution is somewhat changed here indicating that the iterative solution might be slightly more sensitive to a localized error. The LTS algorithm selects three measurements for isolation. Equilibria involving $PhSO_2n$-Pr were particularly distrusted.

Two fourteen-compound scales appear in the literature. The first is part of the gas phase affinities of acetophenones toward the TMS cation.[12] The twenty-two measurements between C_6H_5CHO and o-methoxyacetophenone show reasonable agreement between the equilibria: one LS residual is greater than 0.1 kcal/mol. The shrinking iterative LS solution is only slightly larger than the solution obtained from LS. There are differences in the shrinking pattern indicating some detailed, localized influence on the iterative matrix not observable by the global LS. Accordingly, the interpolants from the iterative matrix more closely resemble the calculated values projected by the iterative solution than they resemble those projected by the LS algorithm.

The second fourteen-compound scale is part of the fluoride affinity compilation.[30,36] The twenty measurements from tBuOH to F_2CO demonstrate poor agreement, both in terms of the number of LS residuals greater than 0.1 kcal/mol and by the boundary difference test. The shrinking iterative solution does not reach the exact LS solution before

terminating in the zero second-order derivative mode. As with many others of these scales, the interpolants more closely resemble the iterative projections than they do the LS projections. Only the Me_3SiF - PF_3 pair is considered suspect by LTS, although there are several measurements which still have large residuals after the removal of this pair. The effect of removing this pair on this scale is to increase the relative affinity of Me_3SiF by 0.2 kcal/mol and the compounds between Me_3SiF and F_2CO by 0.1 kcal/mol.

The only fifteen-membered scale found in the literature was the upper portion of the gas phase acidity scale of substituted phenols.[1] Fair agreement exists among the equilibria constructing the fusion set from p-chlorophenol to p-aminophenol, but there is still a large number of residuals larger than 0.1 kcal/mol. The LTS algorithm attempts to remove three measurements in its analysis of the fluoride compilation. Unfortunately, two of the three omitted values are the scale's entire ties to p-aminophenol. Thus the LTS solution should be completely and categorically rejected as having insufficient evidence to have a legitimate solution.

The Yamdagni and Kebarle work[7] on gas phase basicities between ethanol and ammonia has sixteen compounds and thirty-one equilibria. The equilibria do not have good agreement according to the boundary difference test. The largest of the LS residuals belong to the acetone - methyl

acetate pair. That pair and methyl acetate - ethyl acetate were chosen as the most suspicious according to the LTS algorithm. The principal effect of removing these equilibria is to increase the basicity of acetone by 0.1 kcal/mol.

The other sixteen-compound work contains twenty-two measurements. The fusion set between methyl chloride and sulfur dioxide is the larger of the two main fusion sets which make up the compilation of chloride affinities.[29,34-36] The agreement between the equilibria is marginally acceptable according to the boundary difference test. By the time the iterative solution converges, the solution's range has fallen to less than the LS solution's range. The LTS solution becomes useless since the algorithm removed all reference to $COCl_2$.

The first eighteen-membered scale is also part of the fluoride compilation.[29,36] There is a fair amount of agreement within the fusion set, but this fusion set has already had one bad equilibrium removed: 3-fluoropropylene oxide - water. That one measurement greatly disturbs the rest of the scale, and Larson and McMahon weighed this equilibrium very lightly when they were assigning their own affinities.[29] Thus it seemed logical to remove this value completely from consideration before analyzing the rest of the scale. The 136 iterative passes allow the scale to expand to a point where each value is slightly larger than

the corresponding LS value. This fusion set is one of the few where the scale expands rather than contracts with each additional pass. The LTS parameters declare four equilibria to be in error: each declaration is valid. The effect on the rest of the scale would be hard to determine. The most likely method for determining the effect of the missing equilibria for scales as large as this one would be to re-evaluate the LS algorithm over the data remaining after the first LTS solution. Then a comparison between the LS solution of the altered scale and the original LTS solution could be performed.

The second of the eighteen-membered fusion sets is the Caldwell, Renneboog, and Kebarle study of proton transfer reactions among aliphatic carboxylic acids.[4] The agreement among the acids within the scale is fair according to the boundary difference test. Only six of the thirty-five compounds between $MeOCH_2CO_2H$ and F_2CHCO_2H show a LS residual greater than 0.1 kcal/mol. The differences between the LS and iterative solutions due to the shrinking of the latter as it approaches convergence are small. The shrinking of the iterative solution tends to be localized about $HOCHMeCO_2H$. The LTS solution rejects three measurements which usually indicates a compound has been entirely removed from the scale. The resulting solution is still valid as no compound is removed from the scale.

The first of the nineteen membered scales is the first half of the Caldwell, Renneboog, and Kebarle study:[4] namely from $MeCO_2H$ to $MeOCH_2CO_2H$. The differences between the minimum and the maximum vectors for $MeOCH_2CO_2H$ is only 1.0 kcal/mol indicating a reasonable agreement between most of the compounds. The LTS solution trims three measurements from the ladder in its solution. While these affinities do have sizeable residuals, there are still more which are good candidates for reinvestigation. Overall, the two fusion sets construct a fairly good scale in terms of measurements: there are relatively few pendants, and each fusion set has almost twice as many measurements as there are compounds.

The second of the nineteen-membered scales is the portion of the gas phase acidity scale from neopentanol to aniline.[3] The boundary difference test on these compounds indicates that some disagreement within the thirty-three equilibria is to be expected. The iterative smoothing algorithm does not converge until after 129 cycles. There is ample room for rounding errors in addition to the scale shrinkage that occurs. The LTS solution rejects the MeCN - PhCCH and pMeAn - pMeOAn pairs. Part of the differences observed between the LS and LTS solutions is due to the inability of the simplex to settle on nineteen individual values given the minimum step size with which it works.

Only one twenty-compound scale was noted in the literature, that of 89CHO/KIS.[16] The agreement between these

electron affinity equilibria is remarkably good. The scale covers some 16 kcal/mol between 4-cyanonitrobenzene and 3-nitrotoluene. The iterative solution expands rather than contracts in this fusion set, just as it did for part of the fluoride affinity work. The LTS algorithm suggests four plausible equilibria for re-investigation, but the worst pair has less than 0.1 kcal/mol of residual error.

The twenty-compound synthetic scale behaves as the synthetic scales of eleven through fifteen compounds did: the LTS solution completely removed all trace of a1 from the scale thereby invalidating the entire solution.

J. Large Fusion Sets.

From the previous discussion, one finds less usable information as the size of the fusion set increases. Fortunately only four literature scales were found having fusion sets which exceeded twenty compounds. Each of these scales are quite complex, with serious disagreements among the affinities in each of these ladders. The results of the LS algorithm and boundary solution algorithms on these large fusion sets is given in Table 2-12. No graphs are presented here due to typesetting space constraints.

The first of these ladders is a study by Meot-ner and Sieck.[17] The ΔG and ΔH are represented in Table 2-12. The 28 compound ΔG scale runs from benzene to 2-fluoropyridine.

Table 2-12: Least Squares and Boundary Solution Summaries for Scales Larger than Twenty Compounds.

Scale	# Cpd	# Eq	Fusion Set	# Pert	Rms	Range:LS	Minimum	Maximum
79BAR/SCO[a]	28	50	An - mClAn	30	1.537076	6.4728	4.1500	7.7500
91MEO/SIE[b] ΔG SCALE	28	55	C_6H_6 - 2F-pyridine	37	1.877525	30.5750	27.4000	34.5000
91MEO/SIE ΔH SCALE	31	64	C_6H_6 - 2F-pyridine	50	3.016972	33.7890	33.7890	38.3000
88CUM/KEB[c]	34	65	CCl_2HCO_2H - pyrrole	61	13.81576	18.25686	28.2568	20.0904
79BAR/SCO[a]	51	100	mClAn - Phenol	26	1.004992	9.4538	8.0991	10.6000
79BAR/SCO	114	217	EtOH - Phenol	82	2.177751	25.4722	20.5500	30.1500
88BOR[d]	844	1735	T11 - Y1AP1	104	2.170859	27.8768	27.8768	31.4260

Notes: Affinity values are in kcal/mol except for 88BOR which is in pK units. # Pert. refers to the perturbation calls made by the least squares algorithm.
Sources: [a] Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979** 101, 6046.; Lias, S.G.; Bartmess, J.E.; Liebman, J.F.; Holmes, J.L.; Levin, R.D.; Mallard, W.G.; *J. Phys. Chem. Ref. Data*, **1988**, *17*, Suppl. 1. [b] Meotner, M.; Siek, L.W. *J. Am. Chem. Soc.* **1991** *113*, 4448. [c] Cumming, J.B.; Kebarle, P.; *Can. J. Chem.* **1978** *56*, 1. [d] Bordwell, F.G.; *Acc. Chem. Res.* **1988** 21, 456.; We thank Professor Bordwell for access to his computer database of over 1100 acids in a personal communication.

According to the boundary solution test, there are serious disagreements among the equilibria. The differences in the minimum and maximum range values averages to more than 0.25 kcal/mol disagreement per compound. There are many residuals after the LS solution which are greater than 0.1 kcal/mol. In fact, most of the equilibria disagree. The biggest problems involve two equilibria where the calculated and experimentally measured affinities differ by more than 0.5 kcal/mol. These equilibria are $Me_2C{=}CH_2$ to Et_2CO and $CF_3CH_2NH_2$ to cPrCOMe. The iterative solution is generally larger than the LS solution, but the iterative solution range was shrinking at the rate of 1.3 cal/mol per cycle when the iterative routine terminated. The LTS removes four measurements in its attempt to find a solution. Each of the four are reasonable candidates for further study, but the entire scale suffers from the uncertainties in ΔG_{eq} values.

The same set of extreme compounds carries even more compounds in the ΔH scale.[17] There is again poor agreement among the equilibria which comprise this scale. More than three-fourths of the equilibria bear a residual greater than 0.1 kcal/mol after the LS routine. Ten of these measurement disagree with their calculated values by more than 0.5 kcal/mol. The iterative solution values are slightly less than those obtained from the LS solution. The iterative solution values are generally increasing with each successive cycle. The LTS solution suggests four measurements as

highly suspicious and in need of investigation. Other measurements are also likely: as the size of the fusion set increases, the difficulty in isolating a truly suspicious measurement increases.

The next ladder is taken from the gas phase acidity work of J.B. Cumming and P. Kebarle.[2] The thirty-four acids ranging from CCl_2HCO_2H to pyrrole are not well ordered. Of the sixty-five equilibria in the ladder, thirty-three have LS residuals of more than 1.0 kcal/mol. Only four have residuals of less than 0.1 kcal/mol, which is the cutoff level for most ladders.

The minimum and the maximum boundary algorithms must default with an error status due to the nature of the scale. The iterative routine can now supply additional information: the solution does not merely repeat the LS solution. The smoothing algorithm cycles through ten different perturbations of the order of the compound affinities. Each time the number of errors in the row monotonicity is determined. The fewest errors in the monotonicity check occur after the fourth re-initialization, but the default settings constrain the program to do ten re-initializations. To obtain the iterative matrix with the fewest errors the parameter must be changed from the default settings of ten re-initializations to only four.

The LTS algorithm suggests re-investigation of six equilibria. In the light of the size of the error involved

in this scale, many more equilibria should probably be remeasured.

The third large ladder is the gas phase acidity scale from MeOH to Phenol, which is actually a compilation from several sources and laboratories.[3,77] The fusion set between EtOH and phenol contains 114 compounds. To derive useful information from this one fusion set, though important, would be both complicated and time consuming on a PC. Fortunately the ladder can be partitioned into slightly more manageable fusion sets. These fusion sets are (1) EtOH - tBuOH, (2) tBuOH - neoPnOH, (3) neoPnOH - aniline, (4) aniline - m-chloroaniline, and (5) m-chloroaniline - phenol. Analysis of these smaller fusion sets can be performed by selectively removing a few measurements which cross over one of the fusion points. Selective removal of one or more values should not be done if one needs exact quantities, but rather if the desired goal is the isolation of outlying equilibria. Even with the use of this technique, the last two fusion sets have more than twenty compounds, making the analysis difficult to interpret. In both cases, attempts to further split the fusion sets would require the removal of several equilibria rather than one of two measurements. Interpretations of the first three fusion sets occur in the previous sections on the smaller fusion sets.

The aniline to m-chloroaniline fusion set[3,77] contains 28 compounds and is not well ordered. The boundary

solutions terminate with an error status. The iterative smoothing routine again plays an important role by suggesting an affinity order different than the one obtain from the LS vector. The equilibria rejected by LTS are candidates for restudy (as is much of this section of the scale), but the LTS solution should not be considered as being correct.

The m-Chloroaniline - phenol fusion set[3,77] is even larger than the previous one: it contains fifty-one compounds. It is slightly better ordered, but the iterative re-initialization algorithm is invoked since, after the tenth iteration, the solution edge is found not to be monotonically increasing. After re-initialization, the new **F** matrix is one which increases rather than decreases with each iterative pass. The LTS algorithm suggests five equilibria for special attention, but again the solution obtained is probably incorrect for a scale of this size. The entire scale of 114 compounds is, of course, not well ordered based on the areas within the scale where large errors in agreement occur. The iterative technique runs through its ten re-initialization cycles, a comparison of the number of row monotonicity errors and the corresponding ranges show that the algorithm is merely cycling between three initial matrices. The LTS program suggests eighteen different suspect equilibria as shown in Table 2-13. If too many compounds are removed by the LTS algorithm, the

Table 2-13: Remaining equilbria after the LTS treatment of the mClAn to Phenol Fusion Set.

		Measured	Calculated	Residual
PhCCH	mMeAn	1.9	2.65008	-.75008
mMeAn	An	.5	-.11901	.61901
PhCOMe	CF_3CH_2OH	.5	-.07221	.57221
Ph_2NH	Phenol	2.1	1.55768	.54232
EtCHO	pFAn	1.7	1.16932	.53068
mCF_3An	oNO_2Tol	.1	.62976	-.52976
An	PhNHMe	1.7	1.18970	.51030
pFAn	CF_3CH_2OH	2.6	2.13257	.46743
MeCH=NOH	CF_3CH_2OH	2.1	2.55847	-.45847
An	Me_2C=NOH	.0	.42225	-.42225
Me_2CO	mMeAn	2.3	1.88793	.41207
An	PhNHEt	2.2	1.78971	.41029
PhNHEt	oFAn	1.7	1.35163	.34837
tBuCH=NOH	CF_3CH_2OH	.9	1.22872	-.32872
MeSH	oNO_2Tol	2.0	1.67686	.32314
Me_2C=NOH	EtCHO	.4	.07701	.32299
PhNHMe	pFAn	.8	.47888	.32112
Me_2C=NOH	MeCH=NOH	.5	.82043	-.32043

Notes: All values are in kcal/mol.
Source: Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979**, *101*, 6046. ; Lias, S.G.; Bartmess, J.E.; Liebman, J.F.; Holmes, J.L.; Levin, R.D.; Mallard, W.G.; *J. Phys. Chem. Ref. Data*, **1988**, *17*, Suppl. 1.

resulting scale is no longer interconnected as proved by taking the rest of the list and creating a new data file from the results.

The largest of all the fusion sets is the compilation of all the solution phase acidities measured by graduate students under the direction of F. G. Bordwell.[37,38] This analysis had to be done on a mainframe as it is comprised of about 1100 compounds and 2300 equilibria. Portions of Bordwell's data had to be manipulated prior to analysis by the LADDER program. First, some acids had both a regular code within the database as well as an indicator code because they were acid-base indicators. The data for the scale was adjusted such that each compound had only one code name. Second, references to ion pairing equilibria were removed. Finally, some of the more obvious disagreements within the scale were removed prior to further analysis. Because portions of the work have not been previously published, no detailed discussion will be included here in deference to the original researchers. The LS range for the entire scale through DMSO is in fair agreement with Bordwell's estimation, based on the anchoring acid pentachlorobenzenethiol. This acid is one of the few in the scale with a low enough pK_a as to be directly measured. The LS calculated value for DMSO is within half a pK unit of the value determined by Bordwell.[37] Professor Bartmess has forwarded a copy of the analysis to Professor Bordwell.

K. Conclusions on Fusion Sets

For most fusion sets the LS algorithm provides a reasonable solution to the experimental data. The solutions from the iterative smoothing algorithm and the LS algorithm tend to be similar. The two solutions tend to differ greatly only when the iterative smoothing algorithm has been re-initialized, making the affinity order indices slightly different from those set in the initial **F** matrix by the LS ordering of the affinities.

The minimum/maximum vector solutions can provide an estimate as to the agreement between the equilibria within the fusion set. Acceptable agreement is 0.05 kcal/mol difference per each compound within the fusion set.

The LTS algorithm can make valid suggestions regarding suspicious equilibria. The simplex algorithm used is efficient providing there are eight or fewer compounds within the fusion set. For fusion sets larger than eight compounds, the results obtained from the simplex algorithm may be inaccurate as an entire compound could be eliminated from the analysis.

V. WEIGHTING SCHEMES

Weighting the LS problem is another tool for evaluating the equilibria. Details of the analysis were discussed in

Chapter One. To evaluate the weighting schemes, the two fusion sets that make up 93DEC/GAL were investigated.[5] Weighting schemes included: (1) the uniform weight, (2) the reciprocal of standard deviation weight, (3) the reciprocal affinity weight, (4) the reciprocal square of the affinity weight, (5) one of the Gaussian schemes, and (6) the influence weight. The Gaussian weighting scheme chosen was the HPMS scheme even though the original work was performed in FTMS. The rationale for using the more relaxed of the two Gaussian weightings is that the original work is reported in kJ/mol, rather than in kcal/mol. This is the only literature scale studied which was tabulated in terms of kJ/mol. Because an equilibrium affinity reported as 4.2 kJ/mol is equal to 1.0 kcal/mol, the affinity data (and any residuals) reported in kJ/mol are much smaller than they at first appear to the experimentalist used to working in kilocalories.

The reason for the use of this scale is that this is only literature reference studied where a standard deviation is reported for each affinity. Most scales have omitted this information or have not performed multiple measurements on each equilibrium.

Three notes about the interpretation of the weighting schemes actually precede the results shown in Tables 2-14 and 2-15. First, looking at the output of the perturbation section of the LS algorithm can be confusing. When the

Table 2-14: Results of Weighting Schemes on the GeH_4 - EtSH fusion set.

Compound	Uniformly weighted[a]	Reciprocal Std dev[b]	Reciprocal ΔG^{-1}	Reciprocal ΔG^{-2}	Gauss	Influence
MeSH	0.0000	0.4874	0.4957	0.5000	0.4946	0.3015
GeH_4	0.0175	0.0000	0.0000	0.0000	0.0000	0.0000
pyrrole	0.7100	0.3045	0.2511	0.2500	0.2590	0.4355
EtSH	7.6825	7.5502	7.5446	7.4035	7.1182	7.7074

Notes: [a] Values are in kJ/mol. [b] Weights are given names of the w_i in the equation of finding the solution $\mathbf{x}$ such that $\{n^{-1}\Sigma\ w_i(y_i - \mathbf{A}_i\mathbf{x})^2\}^{\frac{1}{2}}$ is a minimum.

Source: Decouzon, M.; Gal, J.; Gayraud, J.; Maria, P.; Viaglo, G.; Volpe, P. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 54.

Table 2-15: Results of Weighting Schemes on the EtSH - H_2S fusion set.

Compound	Uniformly weighted[a]	Reciprocal Std dev[b]	Reciprocal ΔG^{-1}	Reciprocal ΔG^{-2}	Gaussian	Influence
EtSH	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
nPrSH	3.4890	3.7290	3.4641	3.4181	3.4182	3.5492
nBuSH	5.4903	5.3234	5.3973	5.4045	5.4941	5.4812
$nC_6H_{13}SH$	7.7607	7.7505	7.8121	7.8644	7.9429	7.8584
tBuSH	10.3680	10.5278	10.5406	10.6482	10.7211	10.3690
H_2S	16.5543	16.5219	16.7149	16.8505	16.9852	16.6164

Notes: [a] Values are in kJ/mol. [b] Weights are given names of the w_i in the equation of finding the solution $\mathbf{x}$ such that $\{n^{-1}\Sigma\ w_i(y_i - \mathbf{A}_i\mathbf{x})^2\}^{\frac{1}{2}}$ is a minimum.

Source: Decouzon, M.; Gal, J.; Gayraud, J.; Maria, P.; Viaglo, G.; Volpe, P. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 54.

weights are provided, the vector of weighted affinities replaces the vector of uniformly weighted affinities. While the perturbation algorithm still functions as it should, and one can still obtain information from the perturbation effects, the algorithmic output is more difficult to read as the weighted affinities and not the measured affinities are given in the output. The best reference here would be the uniformly weighted affinity case, which should be calculated first.

Second, when the weighted LS is performed on the DEC VAX system, the IMSL subroutine LSBRR will give a slightly different solution than the one calculated by the PC version of the program. The two iterative LS solutions do agree to 0.01 J/mol, which is far better than the experimental precision of the chemistry involved (0.4 kJ/mol). The LS residuals show less agreement, with the absolute residual sum differing by up to 29 J/mol for the HPMS Gaussian weighting scheme and by 1.66 kJ/mol for the method of reciprocal standard deviations. The difference is due to the way the residuals are calculated by the iterative LS solution. As generalized software, the subroutine LSBRR[91] treats the nonuniformly weighted problem exactly as it does the uniformly weighted case. According to Wampler,[61,62] the iterative LS algorithm must update the weighted residuals for each subsequent iteration. Since LSBRR updates the residuals but not the weighted residuals, the residual rms

reported by the encoded LS algorithm written into the LADDER program is the truer value. If the residuals were to be calculated based only on the final solution as in the LSBRR subroutine, then there is much better agreement between the two sets of residual values. As most researchers will be more interested in the chemistry rather than the mathematics, further discussion of the differences due mathematical detail is currently unnecessary. The program does a comparison of the two iterative procedures and takes as its final iterative solution the one having the lower residual rms. As previously stated, the difference between the two solutions is so small that either solution would be valid.

The third note is that there is some dependence of the other algorithms on LS. The minimum/maximum vectors initially begin by supposing that the LS solution is correct, and then attempt to put an upper and a lower boundary on each LS determined affinity. The iterative scheme is slightly dependent on what order the compounds are in, even though the algorithm has the capacity to rearrange the order. The simplex solution initially depends on the LS vector if there is more than seven compounds within the scale.

The examination of the first fusion set of 93DEC/GAL shows that the different weighting schemes produce not only different solutions to the problem, but they have differing

affinity orders as seen by Table 2-14. Here, the uniformly weighted LS has been skewed so it results in a different affinity order than the weighted schemes. Three of the weighting schemes favor smaller equilibrium affinities (in this case ΔG_{eq}) values: the Gaussian scheme and the two reciprocal ΔG_{eq} schemes. Therefore the MeSH - EtSH measurement at 8.2 kJ/mol is weighted less heavily than the smaller values. Because the larger affinities also were measured with less precision than the smaller ones, the weighting based on the inverse of the standard deviations follows the same trend. The influence weighting scheme is convoluted. Here all of the weights are the same except for the GeH4 - EtSH pair, which is given less weight because each of these compounds appear with the greatest frequency. The solution obtained from influence weighting scheme is not as skewed as the uniform LS solution, but the values obtained are different than those obtained from the other schemes.

For the second fusion set of 93DEC/GAL,[5] the values have better agreement between the weighting schemes as is shown by Table 2-15. Here the agreement between the uniform and most of the other schemes is fairly good. The reciprocal standard deviation weighted solution is fairly close to the uniformly weighted solution except for nPrSH and nBuSH. It should be noted that the standard deviation for the nBuSH - $nC_6H_{13}SH$ equilibrium is largest within the

set, explaining the difference for nBuSH. The inverse ΔG_{eq} weighting scheme and the Gaussian are all similar as expected: the larger the ΔG_{eq}, the lower the weight the equilibrium is given. The influence weight is unique, showing the largest difference from the uniformly weighted LS for the compound hexane thiol. Hexane thiol is the most frequently used compound in the scale, so that the influence of this compound is sizable. The relative acidity in the influence weighted scheme is affected most by the pathway EtSH → nBuSH → $nC_6H_{13}SH$ where the weights are smaller than the direct path of EtSH → $nC_6H_{13}SH$.

The effects of the different weighting schemes on the other algorithms should be discussed. The boundary vector solutions are somewhat dependent on the LS ordering, although the proper minimums and maximums are calculated for the uniformly weighted case. There is a slight disagreement between values as the LS value is occasionally used within these two vectors.

The iterative scheme is initially dependent on the LS order, so that scheme shows some dependence on the weighting method even though the measurement matrix **F** (as the algorithm now stands) is uniformly weighted. For both fusion sets, the iterative scheme after uniformly weighted LS calculation reproduces the uniformly weighted LS solution. The iterative schemes after the other weighting methods also tend to produce the same solution as the

uniformly weighted LS. Thus the iterative scheme may serve as a continual marker of the uniformly weighted LS solution, provided that, in the uniform case run, the iterative scheme and the LS solution agree.

When examining the role of weight on the LS solution, one should abandon any examination of the LTS solution. The nature of the simplex algorithm obscures any reasonable solution.

The best choice of weighting is dependent on availability and the type of scale being investigated. The uniformly weighted LS should always be run first. If the standard deviations are known, then this weighting is a good alternative to the uniformly weighted LS. If one or more of the affinities is significantly larger than the rest, one of the three descending weighting schemes can be used. The Gaussian scheme seems rational in that the errors in the measurement are mainly a function of the uncertainty in the pressures. The exponential scheme compensates for errors in the logarithmic affinities. The influence weight should be used only when there are significantly influential points in one particular region of the scale. As most of the scales have the same relative influence for individual equilibria, this choice of weightings may be never used.

VI. RANGE GROWTH/DIMINISHING IN ITERATIVE SMOOTHING

As has been indicated throughout the sections IV. and V., the iterative smoothing solution frequently mimics the LS solution. It was also noted that the values did not exactly match between the two algorithms. Sometimes the solution returned from the iterative algorithm was larger than the one returned by LS. Sometimes the iterative smoothing algorithm resulted in a slightly smaller affinity value for most of the compounds within a given fusion set.

In order to try to predict the behavior of the iterative solution toward gradual shrinking or expansion, the scales so far described were recalculated with no available termination routine for the iterative smoothing algorithm except for the number of passes. In each case 1000 cycles were performed on each fusion set which is well above the normal number of iterative passes usually necessary for the algorithm to terminate.

Four different patterns were observed. The scale could become truly static where each of the values remains unchanged after an initial number of cycles. Second, the scale could expand with each successive cycle. Third, the scale could contract with each successive cycle. Fourth, the range value could fluctuate about a value in an asymptotic sense.

The patterns proved to be fairly complicated. Part of the behavior is attributed to the number of known equilibria within the **F** matrix. If the matrix is very sparse, where only one overlapping Hess cycle occurs, the iterative solution is most likely to reach a static value. The three compound scales all reached true stasis, where each value remained unchanged with multiple iterations. Most of the four compound fusion sets also reach true stasis. In the three and four compound cases, a large portion of the iterative array is filled by measured values. Thus it appears that the expansion and contraction is affected by the number of overlaps within the scale.

Another part of the behavior must be related to the overall rounding considerations given in the binary arithmetic. Close analysis of fusion sets of five to eight compounds listed in Tables 2-16 through 2-19 showed that many of the eight compound scales reached a true stasis. Only a few of the eight compound scales either expanded or contracted with successive passes, whereas very few of the five, six or seven compound fusion sets reached a truly static point.

There also seems to be a tendency for the iterative solution to in size until it approximates the LS solution if the first pass of the iterative solution is small compared to the LS solution. If the first pass iterative solution is larger than the LS solution, then repetitive cycles diminish

Table 2-16: Range Termination in Iterative Smoothing for the Five Compound Case.

Scale	# Eq	Fusion Set	Termination Status
82JON/STA[a]	5	iBuBr - sBuBr	static
82UPP/STA[b]	6	EtOH - EtCN	increasing
87KEB/CHO[c]	7	C_6F_6 - NB	decreasing
79BAR/SCO[d]	7	MeOH - EtOH	alternating
77WOL/STA[e]	7	nPrCHO - HCO_2Et	alternating
85MCM/KEB[f]	7	CF_3CN - H_2S	increasing
94KOP/TAF[g]	7	$(C_4F_9SO_2)$TfNH - Tf_2NH	increasing
94KOP/TAF	7	(CF_3CO)TfNH - Tf_2CH_2	alternating
SYNTHETIC	9	a2 - a5	static

Sources:[a] Jones, R.L.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 2296. [b] Uppal, J.S.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 1235. [c] Kebarle, P., Chowdhury *Chem. Rev.* **1987** *87*, 513. [d] Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979** *101*, 6046. [e] Wolf, J.A.; Staley, R.H.; Koppel, I.; Taagepera, M.; McIver, R.T., Jr.; Beachamp, J.L.; Taft, R.W. *J. Am. Chem. Soc.* **1977** *99*, 5417. [f] McMahon, T.B.; Kebarle, P. *J. Am. Chem. Soc.* **1985** 107, 2612. [g] Koppel, I.A.; Taft, R.W.; Anvia, F.; Zhu, S.-Z.; Hu, L.-Q.; Sung, K.-S.; DesMarteau, D.D.; Yagupolskii, L.M.; Yagupolskii, Y.L.; Ignat'ev, N.V.; Kondratenko, N.; Volkonskii, A.Yu.; Vlasov, V.M.; Notario, R.; Maria, P.-C.; *J. Am. Chem. Soc.* **1994**, *116*, 3047.

Table 2-17: Range Termination in Iterative Smoothing for the Six Compound Case.

Scale	# Eq	Fusion Set	Termination Status
82KAP/STA[a]	7	Me_2O - EtCHO	increasing
76AUE/WEB[b]	7	Et_3N - nBu_3N	increasing
93DEC/GAL[c]	8	EtSH - H_2S	static
82JON/STA[d]	8	HCO_2nBu - Me_2CO	increasing
94KOP/TAF[e]	8	m-CF_3PhCH$(CN)_2$ - p-ClPhCH$(CN)_2$	increasing
94KOP/TAF	9	PhCH$(CN)_2$ - CF_3CO_2H	decreasing

Sources: [a] Kappes, M.M.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 1813. [b] Aue, D.H.; Webb, H.M.; Bowers, M.T. *J. Am. Chem. Soc.* **1976** *98*, 311. [c] Decouzon, M.; Gal, J.; Gayraud, J.; Maria, P.; Viaglo G.; Volpe P. *J. Am. Soc. Mass Spectrom.* **1993** *4*, 54. [d] Jones, R.L.; Staley, R.H. *J. Am. Chem. Soc.* **1982** *104*, 2296. [e] Koppel, I.A.; Taft, R.W.; Anvia, F.; Zhu, S.-Z.; Hu, L.-Q.; Sung, K.-S.; DesMarteau, D.D.; Yagupolskii, L.M.; Yagupolskii, Y.L.; Ignat'ev, N.V.; Kondratenko, N.; Volkonskii, A.Yu.; Vlasov, V.M.; Notario, R.; Maria, P.-C.; *J. Am. Chem. Soc.* **1994**, *116*, 3047.

Table 2-18: Range Termination in Iterative Smoothing for the Seven Compound Case.

Scale	# Eq	Fusion set	Termination Status
FDATA[a]	9	F_2CO - AsF_3	decreasing
79BAR/SCO[b]	11	tBuOH -neoPnOH	increasing
77WOL/STA[c]	11	MeOAc -$CF_3CH_2NH_2$	increasing
77WOL/STA	13	MeCHO - nPrCHO	increasing
88DIL/KEB[d]	14	C_6F_6 - $C_6F_5NO_2$	decreasing
SYNTHETIC	21	A4-A7	increasing

Sources: [a] Larson, J.W.; McMahon, T.B. *J. Am. Chem. Soc.* **1983** *105*, 2944.; *J. Phys. Chem.* **1984** *88*, 1083.; *J. Am. Chem. Soc.* **1985** *107*, 766. [b] Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979** *101*, 6046. [c] Wolf, J.A.; Staley, R.H.; Koppel, I.; Taagepera, M.; McIver, R.T., Jr.; Beachamp, J.L.; Taft, R.W. *J. Am. Chem. Soc.* **1977** *99*, 5417. [d] Dillow, G.W.; Kebarle, P. *J. Am. Chem. Soc.* **1988** *110*, 4877.

Table 2-19: Range Termination in Iterative Smoothing for the Eight Compound Case.

Scale	# Eq	Fusion Set	Termination Status
Taft[a]	10	m-CNPhCH(CN)$_2$ - m-CF$_3$PhCH(CN)$_2$	static
94KOP/TAF[b]	11	m-CNPhCH(CN)$_2$ - m-CF$_3$PhCH(CN)$_2$	decreasing
88OPE/TEW[c]	11	nPrOH - Me$_2$CO	static
77WOL/STA[d]	12	HCO$_2$Et - MeOAc	static
77MCM/KEB[e]	12	CF$_2$HCO$_2$H - CH$_2$FCO$_2$H	decreasing
79BAR/SCO[f]	13	EtOH - tBuOH	static
Cl data[g]	13	SO2 - HCO$_2$H	increasing
94KOP/TAF[b]	13	CF$_3$COSH - CF$_3$CO$_2$H	increasing
76YAM/KEB[h]	14	mC$_6$H$_4$F$_2$ - MeCHO	decreasing
85MCM/KEB[i]	14	SO$_2$F$_2$ - CF$_3$CN	static
91MEO/SIE[j]	14	3F pyr - Me$_3$N	static

Sources: [a] Taft, R.W., personal communication. [b] Koppel, I.A.; Taft, R.W.; Anvia, F.; Zhu, S.-Z.; Hu, L.-Q.; Sung, K.-S.; DesMarteau, D.D.; Yagupolskii, L.M.; Yagupolskii, Y.L.; Ignat'ev, N.V.; Kondratenko, N.; Volkonskii, A.Yu.; Vlasov, V.M.; Notario, R.; Maria, P.-C.; *J. Am. Chem. Soc.* **1994**, *116*, 3047. [c] Operti, L.; Tews, E.C.; Freiser, B.S. *J. Am. Chem. Soc.* **1988** *110*, 3847. [d] Wolf, J.A.; Staley, R.H.; Koppel, I.; Taagepera, M.; McIver, R.T., Jr.; Beachamp, J.L.; Taft, R.W. *J. Am. Chem. Soc.* **1977** *99*, 5417. [e] McMahon, T.B.; Kebarle, P.; *J. Am. Chem. Soc.* **1977** *99*, 2222. Lower portion of phenol scale. [f] Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979** *101*, 6046. [g] Larson, J.W.; McMahon, T.B. J. Am. Chem. Soc. **1985** *107*, 766.; *J. Am. Chem. Soc.* **1984** *106*, 517.; *Can. J. Chem.* **1984** *62*, 675. *J. Phys. Chem.* **1984** *88*, 1083. [h] Yamdagi, R.; Kebarle, P. *J. Am. Chem. Soc.* **1976** *98*, 1320. [i] McMahon, T.B.; Kebarle, P. *J. Am. Chem. Soc.* **1985** *107*, 2612. [j] Meotner, M.; Siek, L.W. *J. Am. Chem. Soc.* **1991** *113*, 4448. Affinities here are ΔG's.

the solution values toward the LS values. However, this does not explain the cases where the range fluctuates about a mean value. There are clearly unaccounted factors present in determining the behavior of the iterative algorithm.

VII. RESULTS ON ENTIRE SCALES

Demonstration of the program would not be complete without comparing the relative affinity values obtained from the program with those already found in the literature. Discussion will now proceed on a scale by scale basis rather than focusing on individual fusion sets. To simplify the discussion, three areas will be addressed: (1) Does the relative affinity range obtained from the LS algorithm reasonably match the range found in the literature? (2) Is there disagreement between the relative affinities on an individual compound basis? (3) What can be done to improve the scale?

Agreement between the range of relative affinities found in the literature should agree with the LS solution of the data for most scales to within a few tenths of a kcal/mol. A comparison between the reported relative affinity range and the affinity range according to the LS analysis is given in Table 2-20 for several scales.

An individual relative affinity is considered to be significantly different if the value obtained from the LS

Table 2-20: Comparison of the Relative Affinities Calculated by Least Squares with Those Reported in the Literature.

Scale	Compounds	LS Rel Affinity	Literature Rel Affinity
77MCM/KEB[a] phenol	CF_2HCO_2H - oOHPhOH	8.9	9.0
77MCM/KEB[a] phenol	pClPhOH - pNH2PhOH	11.1	10.8
77MCM/KEB[a] benzoic acid	HCO2H - $mClPhCO_2H$	10.0	10.2
79BAR/SCO[b]	MeOH - PhOH	28.6	28.1
79BAR/SCO[b]	MeOH - tBuOH	5.9	5.9
89CAL/REN[c]	$MeCO_2H$ - $CF_3CF_2CF_2CO_2H$	26.2	26.6
77WOL/STA[d]	H2O - NH3	30.8	31.4
76YAM/KEB[e]	H2O - NH3	32.0	31.8
85MCM/KEB[f]	CH4 - H2S	33.8	33.0
92MIS/KAN[g]	pCF3 - pOMe	12.5	12.4
87KEB/CHO[h]	C6F6 - 2,3Cl2NpQ	38.1	38.7
89CHO/KEB[i]	4-CNNB - 3-MeNB	16.6	16.6
78WOO/BEA[j]	Me2NH - H2O	6.7	6.8
82UPP/STA2[k]	MeCOEt - MesH	10.9	10.9
88OPE/TEW[l]	MeCOEt - MeOH	7.7	7.7
82UPP/STA[m]	nPrCO2Et - MeCHO	11.0	11.2
88DIL/KEB[n]	C6F6 - C6F6NO2	12.6	12.6

Notes: All Affinities are in kcal/mol.
Sources: [a] McMahon, T.B.; Kebarle, P.; *J. Am. Chem. Soc.* **1977**, *99*, 2222. [b] Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979**, *101*, 6046.; Lias, S.G.; Bartmess, J.E.; Liebman, J.F.; Holmes, J.L.; Levin, R.D.; Mallard, W.G.; *J. Phys. Chem. Ref. Data*, **1988**, *17*, Suppl. 1. [c] Caldwell, G.; Renneboog, R.; Kebarle, P.; *Can. J. Chem.* **1989**, *67*, 611. [d] Wolf, J.A.; Staley, R.H.; Koppel, I.; Taagepera, M.; McIver, R.T., Jr.; Beachamp, J.L.; Taft, R.W. *J. Am. Chem. Soc.* **1977**, *99*, 5417.

Table 2-20: (continued)

Sources: [e] Yamdagi, R.; Kebarle, P. *J. Am. Chem. Soc.* **1976**, *98*, 1320. [f] McMahon, T.B.; Kebarle, P. *J. Am. Chem. Soc.* **1985**, *107*, 2612. [g] Mishima, M.; Kang, C.H.; Fujio, M.; Tsuno, Y. *Chem. Lett.* **1992**, 493. [h] Kebarle, P.; Chowdhury, S. *Chem. Rev.* **1987**, *87*, 513. [i] Chowdhury, S.; Kishi, H.; Dillow, G.W.; Kebarle, P. *Can. J. Chem.* **1989**, *67*, 603. [j] Woodin, R.L.; Beachamp, J.L. *J. Am. Chem. Soc.* **1977**, *100*, 501. [k] Uppal, J.S.; Staley, R.H. *J. Am. Chem. Soc.* **1982**, *104*, 1238. [l] Operti, L.; Tews, E.C.; Freiser, B.S. *J. Am. Chem. Soc.* **1988**, *110*, 3847. [m] Uppal, J.S.; Staley, R.H. *J. Am. Chem. Soc.* **1982**, *104*, 1235. [n] Dillow, G.W.; Kebarle, P. *J. Am. Chem. Soc.* **1988**, *110*, 4877.

analysis disagrees from the literature's relative affinity by more than ±0.2 kcal/mol. Only selected scales will be discussed here.

A. Phenol Acidities[1]

The major problem of the phenol acidity scale is that it is actually three isolated scales. The LS solution for the most acidic portion between CF_2HCO_2H and catechol agrees fairly well with the literature value. With the scale anchored on CF_2HCO_2H, all disagreements between the reported and calculated relative affinities were below the ±0.2 kcal/mol limit.

The weakest portion of the reported scale from p-ClPhOH to p-NH_2PhOH has an 11.06 kcal/mol range according to the LS algorithm while the reported relative affinity range is only 10.8 kcal/mol. Only the very weakest acid p-NH2PhOH (at 37.8 kcal/mol) disagrees with the literature value (37.5 kcal/mol). The error still is reasonably small. Each ladder has already been statistically treated since there is only one fusion set in each portion of the scale.

B. Benzoic Acid Acidities[1]

McMahon and Kebarle also reported a scale based on substituted benzoic acids. Like the phenol work, the scale

is made up of one solitary fusion set attached to many pendant measurements. The number of pendant measurements here cause a problem in the program. When the program tries to reconnect all of the pendant measurements to the scale, it cannot find attachment points for several linked pendant compounds.

C. 79BAR/SCO[3,77]

Since the data file including 79BAR/SCO contains both additional and corrected equilibria, judgment on the accuracy of these researchers in their assessment of the relative affinities is not strictly feasible without going back to the original data. The original scale has been replicated from MeOH to tBuOH in the course of breaking the larger gas phase acidity scale into manageable fusion sets. The agreement between the LS solution and the relative affinities reported is excellent over this small range of compounds.

D. 89CAL/REN[4]

This gas phase acidity work for substituted carboxylic acidities shows errors in the assessment of ΔG relative to acetic acid. Not only is the calculated LS range 0.4 kcal/mol smaller than the literature value, but several

compounds show differences in the relative affinities between the reported value and the LS value. Some of this error is due to biasing from previous work. For example, there are four possible pathways connecting $CH_3OCH_2CO_2H$ and $ClCH_2CH_2CO_2H$. The values obtaining from summing along each path are 1.3, 1.2, 1.4, and 2.3 kcal/mol respectively. In light of the other three pathways, the pathway $CH_3OCH_2CO_2H \rightarrow PhCO_2H \leftarrow ClCH_2CH_2CO_2H$ leading to the value 2.3 kcal/mol is highly questionable. The mean value from the affinity summation from these four paths is 1.55 kcal/mol, which is skewed in favor of the large value. Since the skewed relative affinity for $ClCH_2CH_2CO_2H$ agreed with Kebarle's earlier paper,[2] the skewed relative affinity of 7.7 kcal/mol was reported as correct. The LS calculated affinity is 7.4 kcal/mol.

Other compounds where disagreement exceeds more than 0.2 kcal/mol include four which may have had their results partially skewed to favor the same earlier work.[2] These acids are $ClCH_2CO_2H$, $BrCH_2CO_2H$, F_2CHCO_2H, Cl_2CHCO_2H, and CF_3CO_2H. With the exception of $Cl_2CH_2CO_2H$, the relative affinity reported in the more recent work[4] is smaller than in the older work.[2] The lowering of the four compound affinities from the first to the second works was not enough to agree with the calculated LS affinities. There are four other compounds ($MeCHBrCO_2H$, $EtCHBrCO_2H$, $CF_3CH_2CO_2H$, and $CF_3CF_2CF_2CO_2H$) where the differences in the LS relative

affinity and the reported affinity exceed 0.2 kcal/mol. $EtCHBrCO_2H$ and its pendant companion $MeCHBrCO_2H$ are influenced by $BrCH_2CO_2H$. $CF_3CH_2CO_2H$ is influenced by F_2CHCO_2H, and $CF_3CF_2CF_2CO_2H$ is a pendant acid connected to CF_3CO_2H. Therefore almost every disagreement between the LS solution and the reported affinities had to do with the experimenter bias based on earlier work.

E. 77WOL/STA[6]

There are also clear problems between the LS values and the literature relative affinities in 77WOL/STA. The problem here does not seem to be one of serious errors but rather the experimenters seemed biased to larger relative affinities than can be justified by LS. That is, the researchers tended to assign affinities based on the largest possible value without any explanation for choosing the larger values. The biased approach was not consistently followed, but occurs frequently enough within the scale so that almost all of the LS relative affinities between H_2O and HCO_2nBu are smaller than the reported affinities by over 0.2 kcal/mol.

F. 76YAM/KEB[7]

In contrast to the work in section E., the problem of agreement between the LS and literature affinities for Yamdagni and Kebarle's basicity scale between H_2O and NH_3 appears in only one fusion set. The fusion sets between NH_3 and propylene all show agreement between the reported and LS relative affinities to within 0.2 kcal/mol. However, the three compound fusion set: m-difluorobenzene - 1,3,5-trifluorobenzene - formic acid has a serious error in it. The measurement between m-difluorobenzene and formic acid is 3.5 kcal/mol, while sum of the ΔG_{eq} values for the other pathway is 3.9 kcal/mol. Yamdagni and Kebarle list the affinity difference between m-difluorobenzene and formic acid as 3.4 kcal/mol, which is smaller than either direct measurement or summed measurements. The relative affinities for H_2S, CF_3CO_2H, and H_2O toward formic acid are in agreement with the calculated value. Therefore there is really only one error in agreement between the LS and reported values: the relative affinity of formic acid should be increased to 22.9 kcal/mol.

G. 85MCM/KEB[10]

There are three fusion sets within this basicity scale. Only the H_2S to CF_3CN fusion set has correctly assigned

relative affinities according to the LS routine. However, the reported relative affinity for CF_3CN based on the second fusion set between CF_3CN and SO_2F_2 is incorrect by 0.9 kcal/mol according to the LS criterion.

Both the CF_3CN to SO_2F_2 and SO_2F_2 to CH_4 fusion sets have serious problems in agreement between the individual equilibria. The reported relative affinities are therefore not expected to be in good agreement with the LS solution affinities.

H. 92MIS/KAN[12]

The agreement between the LS solution and the reported affinities in the basicities of substituted acetophenones toward TMS is excellent. All calculated affinities agree to within 0.2 kcal/mol of their reported values.[12]

I. Electron Affinity Papers[15,16]

Of the four fusion sets which construct an electron affinity scale between C_6F_6 and 2,3-dichloro-1,4-napthaquinone,[15] only the fusion set between NB and 4-CNNB shows serious disagreement between the LS solution and the reported relative affinities. The fusion set range reported by Kebarle and Chowdhury[15] is greater than that obtained from the LS solution by 0.6 kcal/mol. In fact, the

reported affinity difference of 16.5 kcal/mol between these two compounds is close to the difference calculated by the maximum vector algorithm.

The agreement between the calculated and reported affinities is excellent in a subsequent electron affinity paper by Chowdhury *et. al.*[16]

J. Metal Ion Affinities

The reported affinities listed in 78WOO/BEA,[19] 82UPP/STA,[22] 82UPP/STA2,[23] and 88OPE/TEW[27] are in excellent agreement with the values calculated from the LS algorithm. Critical evaluation of the binding energies presented in 82JON/STA,[21] 82KAP/STA,[24] and 82KAP/STA2[25] was not done since the relative ΔH values were reported rather than the relative ΔG values. The data in the last three scales are ΔG_{eq}'s, and the ΔH_{eq} corrections were not calculated.

K. Anion Affinities

The three papers which make up Larson's and McMahon's compilation of fluoride affinities[29,30,36] were not treated on an individualized basis. Similarly, the chloride compilation was not separated as to the individual papers.[29,34-36] Therefore assessment of the reported affinities presented in these papers was not possible.

The evaluation of the cyanide affinity scale of Larson and McMahon[32] led to its correction from the original reference data.[97] The initial scale as published contained several errors. Pyrrole is listed in the graph even though no equilibrium exists which connects pyrrole with the other compounds: the connecting measurement had not been published. There were also considerable errors in the assignment of relative ΔG values that pertain to pendant measurements. For example, $CHCl_3$ is only connected to CF_2HCH_2F by a pendant measurement of 0.1 kcal/mol. If the relative affinity of CF_2HCH_2F were 11.0 kcal/mol, then the assigned ΔG for $CHCl_3$ is 10.9 kcal/mol and not 10.8 kcal/mol as reported. The errors in the assignment of pendant values become important since the anchoring compound is tied to the fusion set through a chain of pendant measurements. The relative range for H_2O to FCH_2CH_2OH is 5.0 kcal/mol from the assigned relative ΔG values, but the range should be 5.6 kcal/mol based on the equilibrium ΔG_{eq}'s.
The relative fluoride affinities of Dillow and Kebarle[31] are in excellent agreement with their LS values.

VIII. OVERALL SCALE CONSTRUCTION

Based on the conclusions of the previous sections, the scales have been separated into four categories. The first category is the pendant scale which is unacceptable for

publication. Further work is necessary to tie such pendant measurements together where possible.

The second category is the preliminary scale. Preliminary scales are scales in which there may be many pendant measurements, but there are also a few fusion sets. Frequently these scales are constructed of fusion sets from three to six compounds. While acceptable for publication, it is desirable to do further work to tie any pendant compounds to other compounds within the scale and thus remove as many pendants as possible.

The third category is the small fusion set scale. Small fusion set scales contain very few pendant measurements and have multiple small fusion sets, where each fusion set contains four to eight compounds. When the construction of a new scale is a primary goal of the paper, the new scale should attempt to be a small fusion set scale. It is best if at least three paths exist between compounds within each ladder. Fusion compounds should be compounds which have accurately known absolute affinities whenever possible.

The fourth category is the large scale. Large scales come from extended work in an area resulting in a continuation of small set scale research. Here the fusion points from the small scale research have been crossed by multiple equilibria forming fusion sets of twenty compounds or more. Finding errors in large scales is not a simple

investigation for program LADDER. Very little information can be found regarding suspicious equilibria due to the nature of the simplex algorithm. Further work in the source code should eliminate using the simplex algorithm in favor of directly performing the LTS analysis using the current LS routines. The only way currently to use the LS routine to solve the LTS problem is to remove the isolated equilibria and create a second data file containing the remaining equilibria. This is a time consuming process in comparison to the LS algorithm's solution time. Further research is required to obtain a useful algorithm for isolating measurement errors within large fusion sets.

CHAPTER 3

THE USE OF DISSOCIATIVE ATTACHMENT TO PRODUCE PRIMARY ANIONS IN ION CYCLOTRON RESONANCE MASS SPECTROMETRY

I. INTRODUCTION

Ions can be created in the gas phase by one of five processes: (1) electron ionization,[98,99] (2) photoionization,[100] (3) electron attachment,[98] (4) dissociative attachment,[101] or (5) collision with a previously charged species.[102,103]

Cations are readily formed by electron ionization and photoionization of a neutral gas. In each case the ionizing energy is absorbed by the neutral molecule which then stabilizes itself through the emission of an electron. Electron ionization (EI) is relatively facile for cation production. In the typical EI experiment, the electrons are beamed into the gaseous sample at energies between 30 and 100 eV.[99] If the electron passes near enough to the electron cloud surrounding a neutral molecule, an easily excited electron from that cloud can then be ejected. In the photoionization process,[100] the molecule is stimulated by high flux incident radiation (often a laser) at a known photon energy. The excited molecule then releases energy by

ejecting an electron. In mass spectrometry, the use of a charged bath gas to ionize other neutral species is termed chemical ionization, (CI).

Anions can only be created by one of three methods: (1) electron capture (EC), (2) dissociative attachment (DA), or (3) the reaction of the neutral species with a previously formed ion, NCI.[103] EC is a useful tool for analyzing compounds of environmental concern,[104] and NCI also plays an important role in organic mass spectrometry.[105] However, the present study limits itself to the dissociative attachment process. The study utilizes the chemistry of relatively weak bonds in order to generate primary ions by dissociative attachment.

For a molecule to undergo dissociative attachment with a large enough probability to be useful in conventional intermediate passage mode ICR spectrometry, there are four criteria which are likely to be helpful. First, the molecule should have a low electron affinity which can be estimated by a bound Lowest Unoccupied Molecular Orbital (LUMO).[106,107] Second, the excited molecular ion must live long enough without ejecting the electron for the molecule to dissociate at a weak bond. Third, the molecule should have a structure such that the ion formed does not react with its own neutral precursor to form unwanted secondary ions.[106] Finally, the separation techniques following the synthesis of the precursor molecule should eliminate any

interfering ions. Each of these criteria will now be examined in detail.

The first criterion of both EC and DA is whether the neutral molecule can capture the incident low energy electron. The capture process within the molecule takes place between two electronic states: the ground state neutral molecule and the excited state molecular anion.[101] The vertical transition from the ground state neutral molecule to an excited state molecular anion is favored if the molecule has low energy, unoccupied molecular orbitals where the electron may attach. Semi-empirical molecular orbital calculations help determine whether a compound meets this criterion. An earlier study by Crowder and Bartmess[106,107] found compounds undergoing dissociative attachment often have a bound LUMO. The electron capture cross section of a neutral species is much smaller than a typical positive ionization cross section.[99] Therefore electron capture and any subsequent dissociative attachment (DA) by the radical anion are often much less efficient than electron ionization in terms of cross section. Unless the compound has a sizeable EC cross section, very little signal will be obtained from attempts to generate the primary anion. Only a few moieties have been found to show consistently good electron capture at sub-microtorr pressures. The primary ion should be the base peak of the spectra or at least a fairly intense peak. Most of the molecules which

undergo efficient electron capture have an extended pi cloud to capture the electron. Small molecules such as H_2O and CCl_4 obviously do not have extended pi clouds, but these molecules do undergo EC and DA.[101] These molecules also have highly electronegative atoms within the molecule which could attract the low energy electron. The size of the dissociative attachment cross section for H_2O is within an order of magnitude of the DA cross section sizes for the monohalobenzenes.[101] For CCl_4, it has been postulated that the neutral ground state potential energy surface is overlapped by the excited CCl_4^{-*} state in an area where the Franck-Condon region is repulsive.[101] Carbon tetrachloride has a DA cross section almost three orders of magnitude larger than that of chlorobenzene.

The second criterion separates dissociative attachment from electron attachment, (EA). For DA to occur, the molecular anion must live long enough to dissociate. As shown by the scheme (3.1), there are three pathways by which the excited molecular anion from electron capture (EC) might relax: EA, DA or autodetachment, (AD).

$$\text{(3.1)} \qquad AB + e^- \underset{AD}{\overset{EC}{\rightleftharpoons}} [AB^-]^* \xrightarrow{DA} A + B^-$$

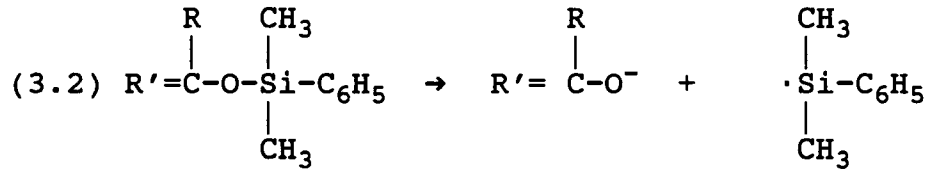
$$[AB^-]^* \xrightarrow{EA} [AB^-] + h\nu$$

The first pathway through which the excited species can release energy is by the ejection of the electron, or autodetachment (AD). (Autodetachment is the opposite of the initial capture process). The second pathway is radiative emission. Radiative emission may occur if the molecule has extensive rovibrational modes which can emit light in the infrared region. This pathway is relatively rare, and occurs on a time scale (milliseconds) slower than most DA reactions.[108,109] The third pathway, dissociation, is the one being studied in detail. The excited state radical anion has absorbed or transferred energy to a relatively weak bond. The resultant bond dissociation releases the excitation energy from the anion by imparting it to a neutral fragment. For the dissociative attachment pathway to be efficient, the rate of dissociation must be larger than either of the other relaxation pathways.

The criterion not mentioned in the previous work[106] is that the parent molecule providing the primary ion must be free of structural groups which can react bimolecularly with the product anion. For anions, the neutral parent often must be methylated at any undesirable acidic site. For example, an attempt to form $(CH_3)_2C{=}NO_2^-$ should be done from $C_6H_5N{=}NC(CH_3)_2NO_2$. Using the nonmethylated azo compound yields $CH_2NO_2^-$, which could deprotonate $C_6H_5N{=}NCH_2NO_2$ to yield $C_6H_5N{=}NCH{=}NO_2^-$.

In addition, separation techniques following the neutral molecule synthesis should be able to remove contaminating ions which might interfere with the production of the desired primary ion. For example, iodo compounds are often used in commercial synthesis and the trace amounts of iodide will be found in the anionic ICR spectra. Fortunately HI is a strong gas phase acid[77] as well as a strong aqueous acid. Any compound more acidic than HI would protonate iodide through proton transfer. Therefore any compound with a gas phase acidity stronger than HI should not have any trace amounts of iodide within its synthesis or else the iodide could interfere with the production of the conjugate base of that compound. Thus the material must be synthetically pure such that trace impurities show no interferences with the dissociative attachment process. This criterion can normally be met through preparative GC separation techniques. Double resonance ejection of the interfering ion is also possible using the ICR spectrometer.

Dissociative attachment of aromatic diazo compounds has previously been studied.[106] The dissociative attachment process to create specific anions is expanded here to create enolates and sulfonates. For the enolates, the actual molecules are formed by the reaction type shown in reaction (3.2).



(3.2)

where R and R′ may be hydrogen or alkyl groups. The pi cloud of the phenyl group in the molecules to be studied should provide the low lying unoccupied molecular orbitals necessary for efficient electron capture, followed by dissociation of the reactive Si-O bond. It is hoped that the molecule will dissociate into the desired stable enolate anion and the dimethylphenylsilyl free radical. For the sulfonates listed in Figure 3-1, the polarizability of sulfur and the pi cloud of the sulfur-oxygen double bond serve as a good electron trap. The catenated sulphur (S-S) bond in methyl methanethiosulfonate is weak, allowing the molecule to react by the elimination of the methylthio radical. Similarly, methyl is a sufficiently good leaving group to stabilize charged, excited state dimethylsulfonate into $MeOSO_2^-$.

(a) (b)

Figure 3-1: Sulfur Compounds Studied. (a) Dimethyl sulfate (b) Methyl Methanethiosulfonate

II. EXPERIMENTAL

Two methods were used to synthesize the silyl enol ether, both adapted from the literature.[109,110] Cazeau's technique was used for the 2-methyl-1-propenyl ether,[110] but the more rigorous lithium isopropyl amide (LDA) technique was necessary for the production of the cyclohexenyl ether. The products were confirmed by NMR spectroscopy.

A. 2-methyl-1-propenyl dimethylphenylsilyl ether, 2MPDMPS:

In a three-necked round bottom flask, 2 mL of chlorodimethylphenylsilane (12.5 mmol) were added to 10 mL of pentane under N_2. A mixture of 1 mL isobutyraldehyde (11.0 mmol), 1.70 g NaI (11.3 mmol), and 1.35 g of 4-dimethylamino-pyridine (11.1 mmol) in 10 mL of acetonitrile was added dropwise with stirring over the course of twenty minutes. The solution was stirred under N_2 for 24 hours. To the 2MPDMPS solution 10 mL of cold pentane was added, followed by the immediate addition of ice-cold water. The aqueous and organic layers were then separated. The aqueous layer was subsequently washed twice with 10 mL of cold pentane, and the pentane washes were added to the organic phase. The combined organic phases were washed three times with ice-cold water, with the aqueous phase

separated each time. Drying of the organic layers was overlooked after the separation from the aqueous layer. The pentane of the organic layer and any residual water was removed under vacuum at 70 °C for 20 minutes. With all the volatile organics removed, the resulting solution was further purified by preparative GC. Nuclear magnetic resonance spectroscopy showed that the first peak that eluted from the QF-1 column was the bis-dimethyl-phenylsilyl ether, $(PhMe_2Si)_2O$. The second peak from the QF-1 column was the desired silyl enol ether. Some decomposition on the QF-1 column was observed. The silyl enol ether peak was collected for ICR determination. The NMR spectral data taken on a Varian Instrument T-60 in deuterated chloroform showed the absorptions listed in Table 3-1. These shifts were in fair agreement with the trimethylsilyl ether work of Cazeau *et. al.*[110] The phenyldimethyl ether was not entirely pure according to the integration of the NMR spectrum. Resolution on this untuned T-60 was poor enough that multiplicity assignment and the evaluation of J constants was not possible.

B. Cyclohexenyl dimethylphenylsilyl ether:

In the three-necked 50 mL round-bottom flask 10 mL of dry tetrahydrofuran was added to a few crystals of triphenylmethane. To the system under N_2, 4.6 mL of a 1.6 M

Table 3-1: NMR Spectral Constants for 2-Methyl-1-propenylphenyldimethylsilyl ether and 2-Methyl-1-propenyltrimethylsilyl ether.[c]

Assignment	$Ph(Me_2)SiOCH=CME_2$ [a] δ : Mult.[b]	$Me_3SiOCH=CME_2$ [d] δ :Mult.[c]
$SiMe_3$		0.17 : s
$SiMe_2$	0.33 : s	
CMe_2	1.53 : m	1.55 : s
vinyl	4.40 : m	5.93 : m
phenyl	7.33 : m	

Notes: [a] Qualitative, slightly impure spectrum on an untuned Varian T-60. [b] δ (ppm), Multiplicity guessed, J values not possible due to lack of resolution, Integration showed impurities still present in the sample. [c] δ (ppm), Multiplicity, J splitting and integration not reported. Source: [d] Cazeau, P.; Duboudin, F.; Moulines, F.; Babot, O. Dunogues, J. *Tetrahedron* **1987**, *43*, 2075-88.

solution of butyllithium in hexanes (7.38 mmol butyllithium) were added via stainless steel cannula. The resulting reddish solution was cooled with an ice bath, then 1 mL of diisopropylamine (7.14 mmol) was added via syringe. The resulting lithium diisopropylamide solution was further cooled in a dry ice/acetone bath and stirred for one hour. With the addition of approximately 0.45 mL cyclohexanone (4.3 mmol), the red color of triphenylmethide was no longer seen. The solution was warmed to 0 °C, then 2 mL of chloro-dimethylphenylsilane (12.5 mmol) was added via syringe. The solution was allowed to stir for four hours to enable the reaction to go to completion. Then 15 mL of cold pentane

was added with stirring, followed by 15 mL of 1 M sodium bicarbonate. The organic layer was then removed, and without drying (overlooked by accident), the more volatile organics and any residual water were removed under vacuum at 70 °C for twenty minutes. The NMR spectra revealed the presence of the vinyllic hydrogen at 4.7 ppm. Comparison of the integration of vinyllic hydrogen to phenyl hydrogen revealed that the solution contained approximately 15% of the desired silyl enol ether. Short path distillation improved this ratio to 1:3, with the silyl enol ether coming off under vacuum between 170 - 185 °C. None of the desired silyl enol ether was present in the fraction which boiled above 200 °C at mechanical vacuum pump pressure. The silyl enol ether fraction, though still having impurities such as $(PhMe_2Si)_2O$, was pure enough for qualitative work on the ICR. The NMR shifts for the cyclohexenyl dimethylphenylsilyl ether and its trimethylsilyl analogue are reported in Table 3-2. The NMR shifts are again in fair agreement with Cazeau's spectra on the analogous trimethylsilyl enol ether.[110] Evaluation of multiplicity and J constant were not feasible due to the poor resolution of the Varian T-60.

Table 3-2: NMR Spectral Constants for Cyclohexenyl-phenyldimethylsilyl ether and Cyclohexenyl trimethylsilyl ether.

Assignment	$Ph(Me_2)SiOCH=C_6H_9$ [a]	$Me_3SiOCH=C_6H_9$ [d]
	δ : Mult.[b]	δ :Mult.[c]
$SiMe_3$		0.15 : s
$SiMe_2$	0.33 : s	
ring $-CH_2-$	1.5-1.9 : m	1.3-2.2 : s
vinyl	4.75 : m	4.75 : m
phenyl	7.15 : m	

Notes: [a] Qualitative, slightly impure spectrum on an untuned Varian T-60. [b] δ (ppm), Multiplicity guessed, J values not possible due to lack of resolution, Integration showed impurities still present in the sample. [c] δ (ppm), Multiplicity, J splitting and integration not reported. Source: [d] Cazeau, P.; Duboudin, F.; Moulines, F.; Babot, O. Dunogues, J. *Tetrahedron* **1987**, *43*, 2075-88.

C. Sulfates:

The samples of dimethyl sulfate and methyl methanethiosulfonate were purchased from Aldrich Chemical and were used without further purification. An attempt was made to study diphenyl sulfone but the compound had insufficient room temperature vapor pressure to study in the ICR.

D. Instrumentation:

The ICR spectrometer used in this study is a home-built instrument which has been previously described.[112] A 1.0

Tesla water-cooled electromagnet provides the ion trapping field. The McIver type reaction cell[113] is connected to a capacitance bridge detector. The pressure of neutral compounds within the cell is 1 to 4 microtorr with a background pressure of 10-20 nanotorr. Operation of the instrument is computer controlled: for rf frequency sweep generation by a Rockland 2MHz synthesizer, for data acquisition, and for data averaging. For the experiments in this chapter intermediate passage detection is used.[112] At the time of these experiments, double resonance ejection to remove ionic interferences was not feasible. Samples underwent several freeze-pump-thaw cycles to remove trace gaseous impurities.

III. Results

A. ICR Spectra of 2-methyl-1-propenyl dimethylphenylsilyl ether:

With a background pressure of 0.02 μtorr, the silyl enol ether sample was introduced into the cell to a total ion gauge pressure of 2.1 μtorr. The estimated true pressure based on the rough Baratron Capacitance Manometer to Ion Guage ratio[114] was 0.2 μtorr. The filament on the ion gauge was not turned off, so some pyrolysis was possible. The filament current was adjusted to 3.0 amps to maximize the

signal without causing distortion. However, no record of the actual emission current was obtained. The unquenched cation spectrum was observed first. In an unquenched spectrum no ejection of ions takes place after the initial grid pulse which allows electrons to enter the cell. Therefore the unquenched spectrum is a product of both primary and secondary ion processes. Table 3-3 shows the major cationic peaks, with the most abundant being $PhMe_2Si$ at 72 mV rf.

The assignment of the peak at a m/z ratio of 126 is in question. HI would normally be expected at 128 amu, while the (M+1) ion of DMAP would only be 123. Instrument resolution was less than 1 mass unit. Because of the presence of I^- in anionic spectra, the tentative assignment is for the ion at 126 m/z to be HI^+.

Table 3-3: Major Cationic Peaks from 2-methyl-1-propenyl dimethylphenylsilyl ether.

AMU	% Intensity	Identity
197	22	Me_2ClCH_2OH
151	46	$PhMe_2SiOH$
136	100	$PhMe_2Si$
126	42	HI ? DMAP (M+1) ?
109	18	DMAP - 15 ?
78	15	benzene (M+)
72	8	$Me_2C=CHOH$
43	17	iPr

After the initial cationic spectrum was taken, the peak intensities of the anions were studied in quenched mode at a 40 msec delay as a function of the electron energy. DA typically occurs for low energy electrons, usually less than 15 V acceleration. The electron energies given in Figure 3-2 are the difference between the potential applied to the filament and the side plate potential. It is not the classical definition of electron energy. The studied range was from -2 V to +7 V of bias potential. Electron energy readings less than approximately 0 V indicates the flooding of thermal electrons into the cell. The electron energy as an average value does not account for the potential drop across the filament. The graph of the observed intensities as a function of electron energy is depicted in Figure 3-2.

B. ICR Spectra of Cyclohexenyl dimethylphenylsilyl ether:

The ether to be studied has a low vapor pressure at room temperature. Therefore, the usual method of entering the gaseous neutral via a leak valve had to be abandoned in favor of a method which allows the neutral reagent to be within the confines of the cell. A small amount of sample was introduced into a short section of capillary tubing and placed onto the solid probe. The probe is then inserted into the instrument close to the cell. The proximity of the

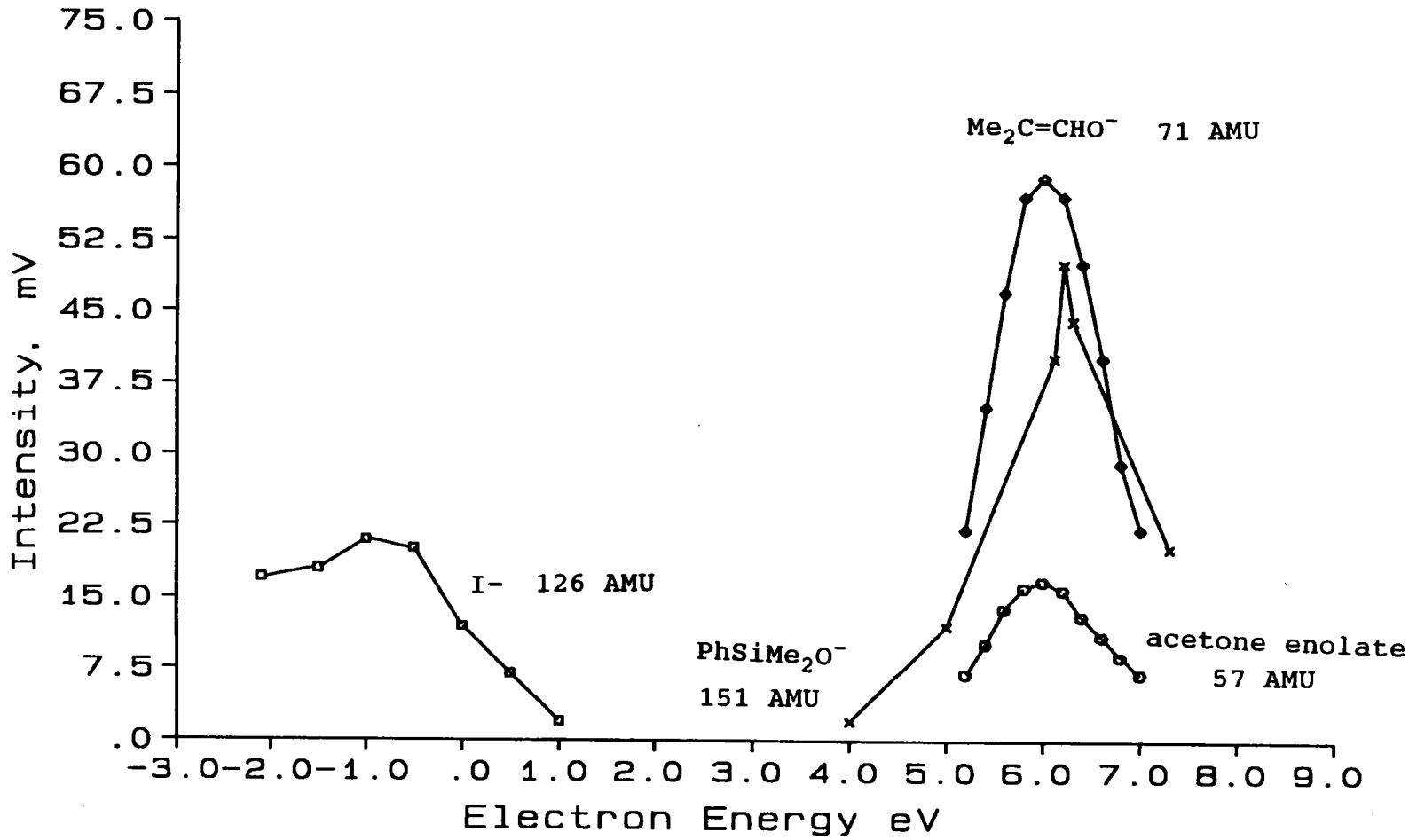


Figure 3-2: Anions obtained from 2-methyl-1-propenyl phenyldimethylsilyl enol ether.

probe to the cell increases signal detection of the relatively nonvolatile species.

The sample was placed on a solid probe and introduced into the cell at an ion gauge pressure of 20 μtorr. During the course of the experiments, the pressure would occasionally drop as low as 4 μtorr, though for these qualitative results, quantitative knowledge of the reactant pressures was not critical. The estimated true pressure based on the 4 μtorr ion gauge reading is 0.3 μtorr.

As with the 2-methyl-1-propenyl ether, the cation spectrum was first obtained in the unquenched mode. The filament current was adjusted to maximize signal without distortion. For both the cationic and the anionic studies, the mass range scanned was from 80 AMU to 300 AMU, and the ion gauge filament was turned off to prevent pyrolysis. For cations, a maximum in signal intensity was reached at an electron energy of 16 V for the silyl enol ether (M+1) peak. The observed mass/charge ratios correlate well with the expected fragment structures of the ether as shown by Table 3-4. The base peak in the spectra was the (M+1) cation of the silyl enol ether at 245 mV rf.

Possible impurities included bis-dimethylphenylsilyl ether $(PhMe_2Si)_2O$ and the tertiary amine N'N'N'-diisopropyl-dimethylphenylsilyl amine (DIDMPSA). The anions were studied as a function of the electron energy.

Table 3-4: Major Cationic Peaks from Cyclohexenyl dimethylphenylsilyl ether.

AMU	%Intensity	Identity
287	2	BDMPSE (M+1)
154	8	$PhMe_2SiOH$ (M-1)
234	100	$PhMe_2SiO$-cHexene (M+1)
102	10	$(iPr)_2NH$ (M+1)
156	14	$PhMe_2SiOH$ (M+1)
100	2	$(iPr)_2NH$ (M-1)

Figure 3-3 charts the anionic intensity as a function of the electron energy. To clarify the observed species in Figure 3-3, Figure 3-4 shows an expanded version of a small portion of Figure 3-3.

C. ICR Study of Dimethylsulfate:

In unquenched mode, dimethyl sulfate was leaked into the cell to a pressure of 3.6 μtorr. The estimated true pressure based on the Baratron factors is 0.91 μtorr. Four peaks were observed in the anionic unquenched mode as shown in Figure 3-5: (a) 110 amu, the (M-15) anion, (b) 95 amu, corresponding to $MeOSO_2^-$, (c) 80 amu, corresponding to SO_3^-, and (d) 46 amu, which was attributed to an impurity of NO_2^-, presumably from a previous days experiment. Another possible assignment for the appearance of m/z 46 might be

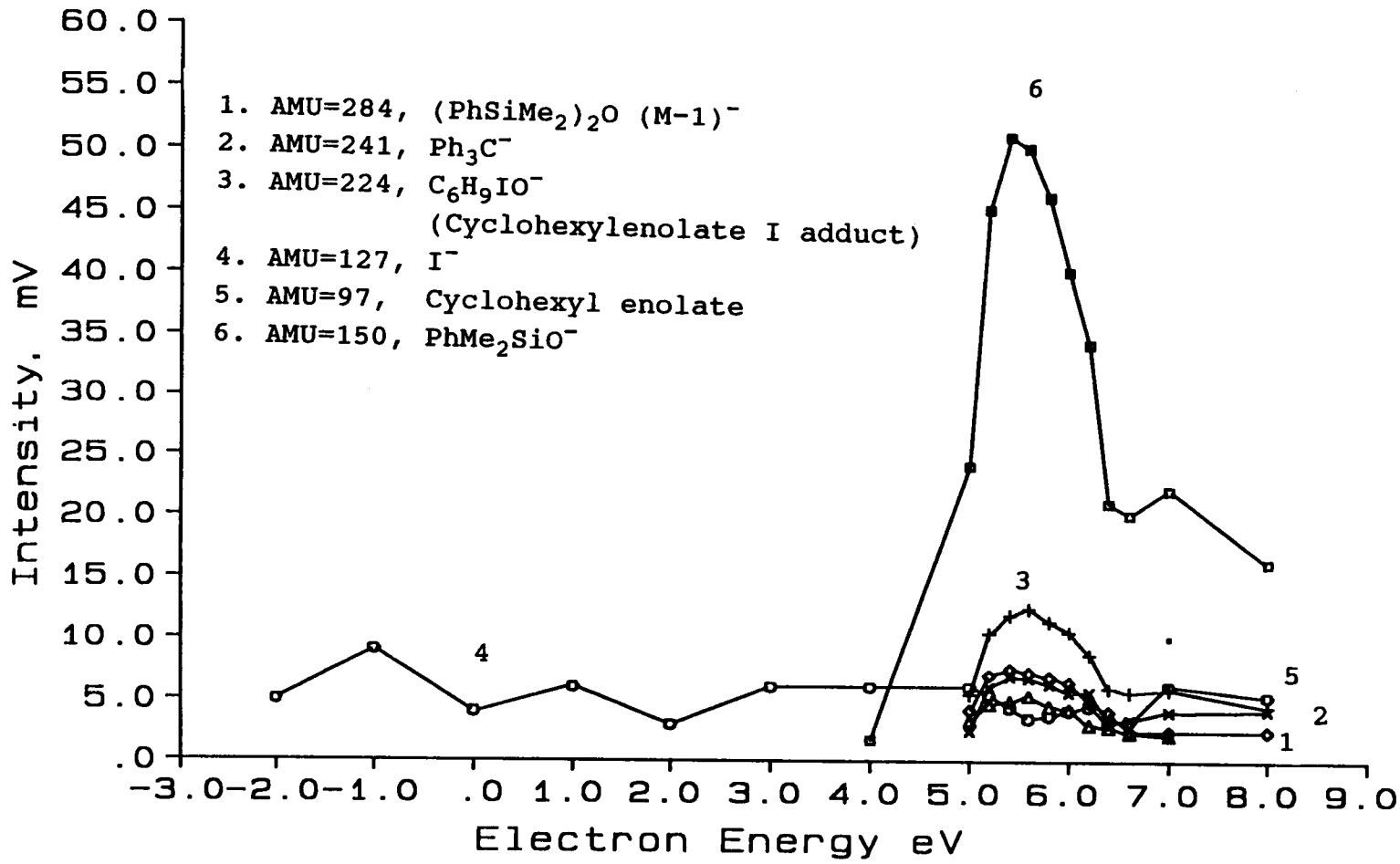


Figure 3-3: Anions obtained from cyclohexenyl phenyldimethylsilyl enol ether.

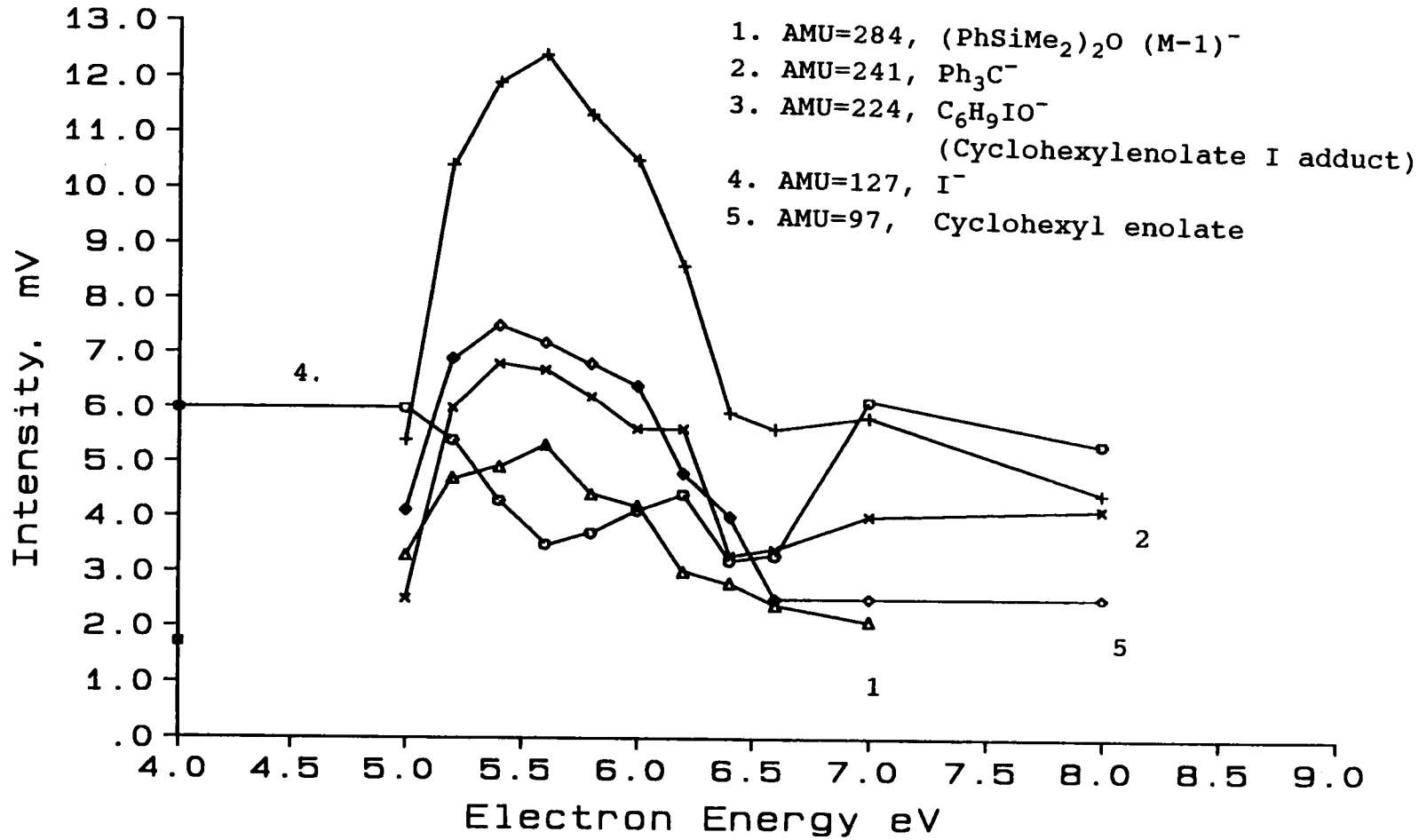


Figure 3-4: Expansion of the low intensity region greater than 4 eV from Figure 3-3.

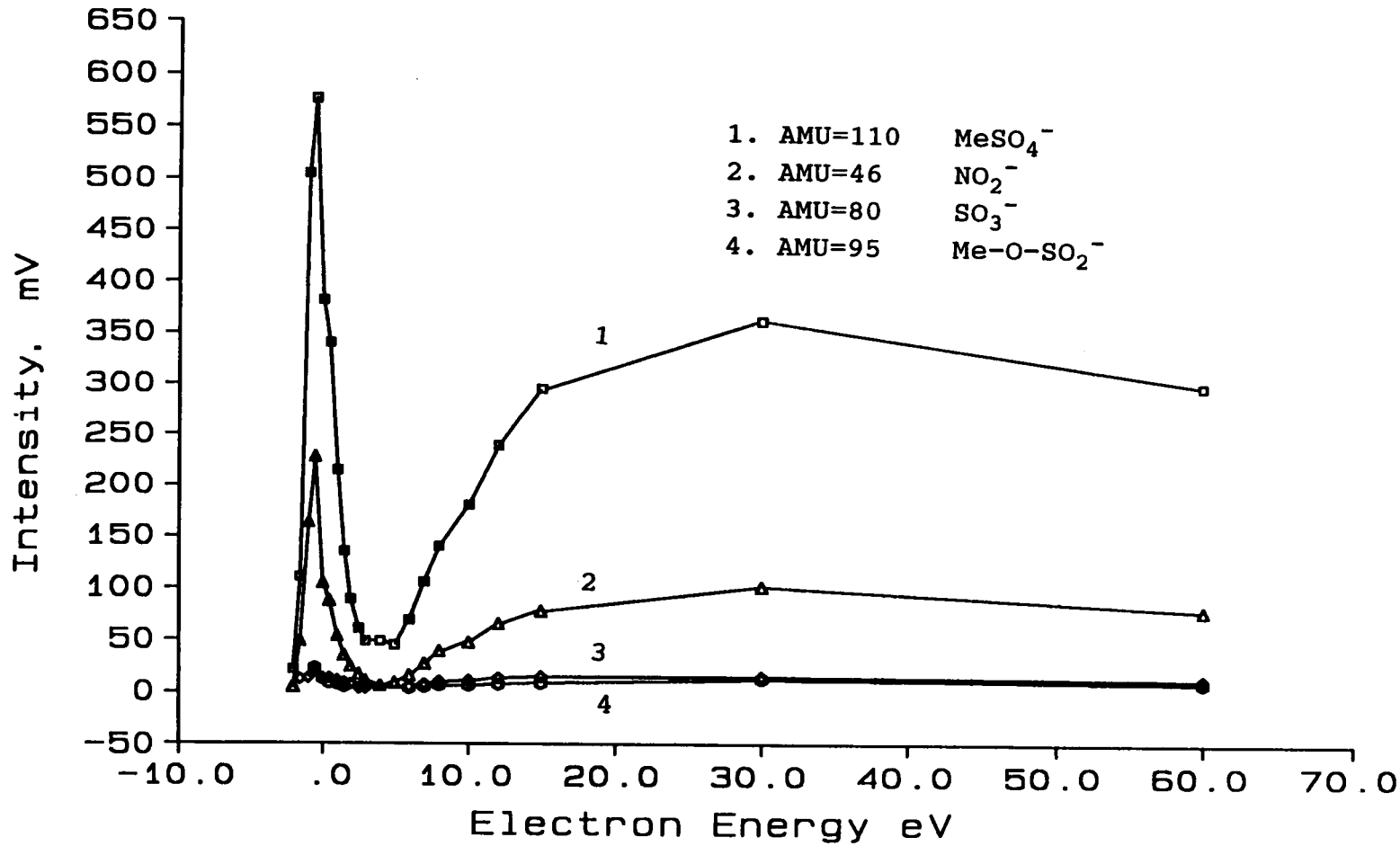


Figure 3-5: Anions obtained from dimethylsulfate

the presence of MeS^-, although this is a far-fetched possibility. Another possible ion for this m/z ratio might be ethoxide. It could be possible that traces of ethanol were present in the commercial dimethyl sulfate and the material was used as purchased. There is some concern that the presence of the ion at 46 AMU might have invalidated the purely DA process through a bimolecular reaction. Unfortunately double resonance ejection to eliminate this m/z ratio was not feasible on the ICR at the time of the study.[115] In the 40 msec delay quenched mode, the anionic spectra was measured as a function of the electron energy. The results of this study are given in Figure 3-5.

D. ICR Study of Methyl Methanethiosulfonate:

In the unquenched cationic mode, methyl methanethiosulfonate was leaked in to a total pressure of 1.8 μtorr. The estimated pressure based on the Baratron factors was 0.37 μtorr. The spectrum was obtained at 30 eV of ionization energy: the cationic peaks included are those found in Table 3-5. The base peak in the cationic spectrum was the (M+15) peak at 40 mV rf.

The anionic spectra were observed at electron energies varying from +1.8 to -70 eV. The base peak corresponded to $MeSO_2^-$, while traces of $MeSO_2S^-$ and SO_2^- were also observed

Table 3-5: Major Cationic Peaks from Methyl methanethiosulfonate.

Amu	% Base	Cation
141	100	M+15
127	45	M+1
108	8	M-18
95	53	M-31
94	55	M-32
81	13	M-45

in Figure 3-6. The gas phase acidity of the conjugate acid of the major peak, $MeSO_2^-$, was bracketed between dichloroacetic acid at 321.9 kcal/mol and trifluoroacetic acid at 317.4 kcal/mol.[77]

IV. Discussion

A. Silyl Enol Ethers

The first problem encountered with the synthesis of the silyl enol ethers was due to the steric hindrance of the phenyl group, in contrast with the methyl group of the chlorotrimethylsilane. To study the effects of steric hindrance, Cazeau's synthesis of 2-methyl-1-propenyl trimethylsilyl ether was repeated. The reaction proceeded more rapidly (and cleanly) with the trimethylsilyl ether than for the phenylated analogue. Thus the phenyl group

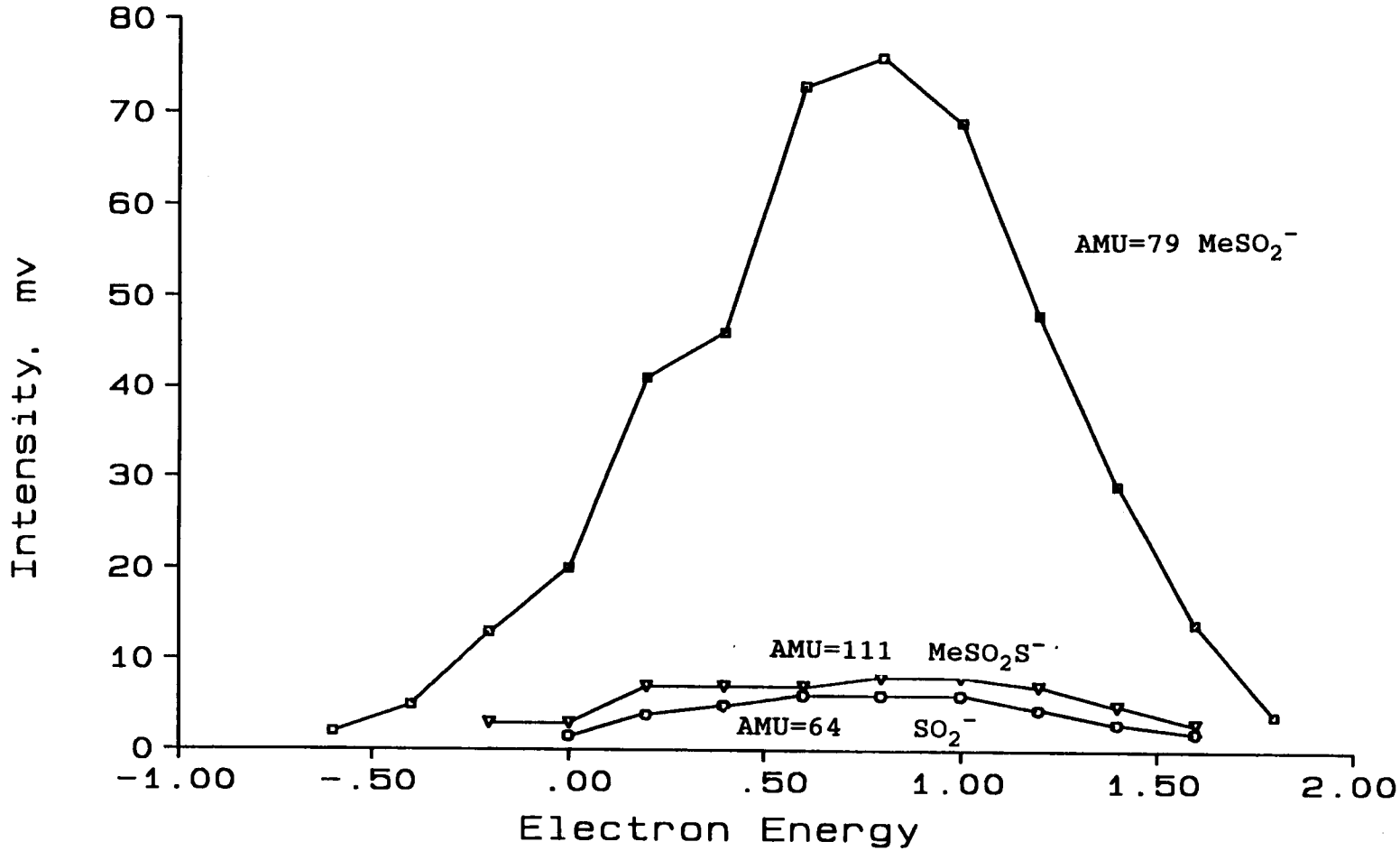


Figure 3-6: Anions obtained from methylmethanethiosulfonate

sterically hinders the reaction to make 2MPDMPS. Allowing Cazeau's method for trimethylsilyl enol ethers to react for a much longer time than recorded in the literature[110] proved to be without advantage. Thus the use of triethyl amine as the abstracting base from the Cazeau synthesis[110] was abandoned in favor of the much stronger base dimethyl-aminopyridine (DMAP). Use of DMAP seemed to work effectively for the 2-methyl-1-propenyl ether at a 24 hour reaction time. The use of DMAP did not work well with the cyclohexenyl ether, thus the more rigorous LDA synthesis was necessary.

Another problem occurred in the isolation of the desired ether from the by-products, mainly $(PhMe_2Si)_2O$. The bis-silyl ether is readily synthesized by the reaction of water with chlorodimethylphenylsilane in the presence of base. Two possible corrections are needed in the previous syntheses of the silyl enol ethers. First, the $(PhMe_2Si)_2O$ may be eliminated as a major by-product if the silyl chloride is used as the limiting reagent in the syntheses. Further, drying the organic layer before distillation will eliminate further decomposition of the reactive enol ether.

Fractional distillation under mechanical pump vacuum failed to satisfactorily separate the mixture. Both enol ethers start to decompose on the acid washed silica column of the gas chromatograph, though the 2-methyl-1-propenyl ether was collectible. Freezing out the cyclohexenyl ether

fraction from pentane at dry ice temperatures did not work; the fraction remained soluble in pentane at this temperature.

Upon electron bombardment, the creation of the enolate ion is observed for both systems, though it is yet to be determined whether the enolate is a primary or secondary ion. Both systems contained several impurities in the sample: notably some iodide containing ones, and the bis-silyl ether. It is unlikely that the iodide interferes with the production of the desired enolates, which tend to be more basic than iodide.[77] In both systems the dimethyl-phenyl siloxide ion is formed and may interfere with the production of the enolate. The intensity of I^- at 6 V is minimal for the 2-methyl-1-propenyl ether but considerable for the cyclohexenyl ether. The iodide contamination must be present in the reagents used in the synthesis of the original dimethylphenylsilyl chloride, since the LDA route does not use iodine to improve the kinetics of the reaction. Further work is necessary to determine if aromatic silyl enol ethers are a route for synthesizing low energy enolates within an ICR spectrometer.

In keeping with the dissociative attachment criterion regarding the molecules having low energy or bound LUMO's, a semi-empirical molecular orbital study was performed on the enol ethers. The enol ethers are compared to a number of

other compounds which have been previously studied by the same laboratory. The listing is in Table 3-6.

The LUMO is unbound in both the 2-methyl-1-propenyl ether and the cyclohexenyl ether. There has been some disagreement in the viability of using the LUMO of the neutral for predicting the occurrence of EC. Crowder[106] reported that there are cases where the anion is formed from an unbound MNDO LUMO. Acetic anhydride, for example, is unbound by 0.3 eV, but acetate is produced. The adiabatic electron affinity can be estimated through semi-empirical calculation using the difference in the heats of formation for the radical anion and the neutral precursor. For the

Table 3-6: LUMOS for Selective Precursors from MNDO Calculations.

Molecule	LUMO eV[1]	Anion Produced?
PhN=NH	-0.530	yes
$PHN=NCH_2COCH_3$	-0.647	yes
$PHN=NCH_2NO_2$	-1.17	yes
$(CH_3CO)_2O$	0.392	yes
$CH_3CONHPh$	0.128	no
$Ph(Me)_2Si-O-CH=CMe_2$	0.099	probable
$Ph(Me)_2Si-O$-cyclohexene	0.132	probable

[1] Calculations of all but the silyl compounds in C.A. Crowder, MS Thesis, Univ. of Tennessee-Knoxville, May 1989, pp. 29, 35.

2-methyl-1-propenyl ether, the electron affinity is 0.13 eV which is fairly close to the LUMO energy of the neutral species. Two things must be remembered when evaluating either calculation. First, the calculation is not likely to be precise. The error in the calculated ionization energies from MNDO calculations using Koopman's theorem is substantial for some compounds.[116] If the HOMO energies are incorrect, then the LUMO energies are probably also incorrect. Therefore a compound with a slightly unbound LUMO by calculation may still undergo EC. It should also be remembered that the calculation is merely a tool. Synthesizing and purifying a new compound for a DA study is time consuming. If the LUMO is bound or just slightly positive, the molecule may be worth synthesizing for the DA study. If the LUMO is sufficiently unbound according to the calculation, a time consuming synthesis can be avoided.

B. Sulfates:

Semi-empirical calculations were performed on the hydrogen and the methylated compounds of a variety of sulfur oxides. The study used three different Mopac Hamiltonians: MNDO,[116] AM1,[117] and PM3.[118,119] The purpose for the study was actually twofold. The first part of the study was to verify the role of increasing sulfur valency toward having the

bound LUMO. The role of sulfur catenation to stabilize the molecule toward DA was also of interest.

After the introduction of the PM3 Hamiltonian,[118] an exhaustive study was performed using that Hamiltonian. The resulting HOMO and LUMO energies for the methyl ethers, thioethers, and sulfur oxides are given by Table 3-7. The PM3 Hamiltonian was used to calculate the ionization energies for catenated sulfur compounds since it is now known that methyl methanethiosulfonate readily undergoes DA.

The rationale for favoring PM3 over MNDO or AM1 is that the calculated heats of formation using PM3 are reportedly better for sulfur (IV) and sulfur (VI) than for using MNDO,[118] even with the revised MNDO sulfur parameters of 1986.[120] When first developed, both the original MNDO and AM1 Hamiltonians did not include the sulfur parameters capable of handling the inclusion of sulfur's 3d AO's in hyperconjugation. Because both molecules studied here have hypervalent sulfur, accurate studies capable of handling hypervalent sulfur were required. MNDO (QCPE 353),[124] or Modified Neglect of Differential Overlap, was first developed by M. J. S. Dewar from Pople's NDDO program. The version used in this study was the PC version created by Janetta D. Bowden and G. Scott Owen.[125] The parameters of the PC program for carbon, hydrogen and oxygen were part of Dewar's original work[116] while the sulfur parameters were those revised in 1986.[120] The AM1 calculations required a

Table 3-7: PM3 calculations on Methyl Ethers, Thioethers and Sulfur-Oxygen Compounds.

Hill	Isomer	ΔH_f kcal/mol	HOMO, eV	LUMO, eV
C_2H_6O	Me-O-Me	-46.9	-10.67	3.03
C_2H_6S	Me-S-Me	-10.5	-8.87	0.42
C_2H_6S2	Me-S-S-Me	-3.9	-9.41	-2.03
C_2H_6OS	Me-S-O-Me	-45.4	-9.12	0.09
C_2H_6OS	Me-SO-Me	-38.3	-9.33	0.17
$C_2H_6OS_2$	Me-SO-S-Me	-28.3	-9.42	-1.98
$C_2H_6O_2S$	Me-O-S-O-Me	-83.9	-9.37	0.06
$C_2H_6O_2S$	Me-SO-O-Me	-81.7	-9.75	0.05
$C_2H_6O_2S$	Me-SO_2-Me	-75.5	-11.18	-0.06
$C_2H_6O_2S_2$	Me-S-SO-O-Me	-65.6	-9.47	-2.02
$C_2H_6O_2S_2$	Me-SO_2-S-Me	-53.0	-9.86	-2.07
$C_2H_6O_3S$	Me-O-SO-O-Me	-125.6	-10.06	0.12
$C_2H_6O_3S$	Me-SO_2-O-Me	-123.6	-11.50	-0.09
$C_2H_6O_3S_2$	Me-S-SO_2-O-Me	-96.6	-10.07	-2.29
$C_2H_6O_4S$	Me-O-SO_2-O-Me	-170.9	-11.69	-0.11
CH_3O	Me-O^-	-37.9	-1.79	10.62
CH_3S	Me-S^-	-22.1	-2.29	7.46

Table 3-7: (continued)

Hill	Isomer	ΔH_f, kcal/mol	HOMO, eV	LUMO, eV
CH_3S_2	Me-S-S^-	-32.9	-2.48	4.17
CH_3OS	Me-O-S^-	-59.6	-2.36	6.57
CH_3OS	Me-S-O^-	-58.8	-1.77	6.84
CH_3OS_2	Me-SO-S^-	-64.2	-3.48	3.72
CH_3OS_2	Me-S-SO^-	-71.7	-2.99	4.03
CH_3O_2S	Me-O-SO^-	-104.8	-2.08	6.22
CH_3O_2S	Me-SO_2^-	-107.6	-3.10	6.46
$CH_3O_2S_2$	Me-O-SO-S^-	-109.5	-3.81	3.49
$CH_3O_2S_2$	Me-S-SO_2^-	-116.1	-3.96	3.96
$CH_3O_2S_2$	Me-SO_2-S^-	-101.3	-4.62	3.21
CH_3O_3S	Me-O-SO_2^-	-161.9	-3.62	6.05
CH_3O_3S	Me-SO_3^-	-166.5	-5.44	6.40
$CH_3O_3S_2$	Me-O-SO_2-S^-	-148.6	-4.87	2.98
$CH_3O_3S_2$	Me-S-SO_3^-	-159.7	-4.80	3.96
CH_3O_4S	Me-O-SO_3^-	-222.0	-6.08	6.24

mixing of the AM1 and MNDO Hamiltonians and thus were deemed not to be as accurate as the PM3 calculations.

A few trends are noted from the data in table 3-7. First, changing an oxygen to a sulfur within a compound always decreases the energy of the LUMO by about 2 eV. Second, sulfones have more tightly bound LUMO's than the corresponding sulfoxides. Third, the addition of a catenated sulfur (compare structures 2 and 3) always makes LUMO lower in energy. The average decrease in the LUMO energy is 2.17 eV over the compounds studied in the series. Therefore it could be expected that catenated sulfur compounds will undergo DA at very low energies.

In keeping with the discussion on the viability of using the LUMO energies of the neutral toward predicting successful EC, the adiabatic electron affinities were calculated for both dimethyl sulfate and methyl methane thiosulfonate. The adiabatic electron affinities using the PM3 Hamiltonian were 1.48eV for dimethyl sulfate and 2.84eV for methyl methane thiosulfonate.

The corresponding enthalpies of formation, HOMO and LUMO energies are listed in Table 3-8 for the hydrogen analogues of Table 3-7. The same trends noted for the dimethyl compounds are also observed for the analogous hydrogen compounds.

As a final note on the PM3 calculations, the heats of reaction for the dissociative process are given in Table 3-9

Table 3-8: PM3 Calculations on the Hydrogen Analogues of Table 3-7.

HILL	ISOMER	ΔH_f kcal/mol	HOMO, eV	LUMO, eV
H_2O	H-O-H	-53.4	-12.32	4.06
H_2S	H-S-H	-0.9	-9.63	0.55
H_2S_2	H-S-S-H	8.6	-9.86	-2.33
H_2OS	H-S-O-H	-42.6	-9.42	0.08
H_2OS	H-SO-H	-25.5	-9.47	0.27
H_2OS_2	H-SO-S-H	-15.2	-9.70	-2.27
H_2O_2S	H-O-S-O-H	-94.4	-9.31	0.27
H_2O_2S	H-SO-O-H	-79.8	-9.84	0.08
H_2O_2S	H-SO_2-H	-68.2	-11.22	0.31
$H_2O_2S_2$	H-O-SO-S-H	-68.9	-9.74	-2.16
$H_2O_2S_2$	H-SO_2-S-H	-44.8	-10.40	-2.13
H_2O_3S	H-O-SO-O-H	-140.4	-10.30	0.30
H_2O_3S	H-SO_2-O-H	-126.6	-11.81	0.02
$H_2O_3S_2$	H-O-SO_2-S-H	-100.3	-10.58	-2.37
H_2O_4S	H-O-SO_2-O-H	-188.1	-12.56	0.17
HO	H-O^-	-17.5	-1.01	15.38

Table 3-8: (continued)

HILL	ISOMER	ΔH_f kcal/mol	HOMO, eV	LUMO, eV
HS	HS^-	-15.9	-2.37	8.09
HS_2	$H-S-S^-$	15.7	-1.67	3.45
HOS	$H-O-S^-$	-62.8	-2.19	7.11
HOS	$H-SO^-$	-50.9	-1.75	7.24
HOS_2	$H-SO-S^-$	-57.6	-3.44	3.71
HOS_2	$H-S-SO^-$	-67.1	-2.90	4.09
HO_2S	$H-O-SO^-$	-109.8	-1.83	6.76
HO_2S	$H-SO_2^-$	-98.4	-2.93	6.69
HO_2S_2	$H-O-SO-S^-$	-116.4	-3.71	3.62
HO_2S_2	$H-SO_2-S^-$	-96.3	-4.58	3.31
HO_2S_2	$H-S-SO_2^-$	-112.8	-4.03	4.01
HO_3S	$H-O-SO_2^-$	-168.4	-3.56	6.59
HO_3S	$H-SO_3^-$	-159.1	-5.14	6.54
HO_3S_2	$H-O-SO_2-S^-$	-157.6	-4.82	3.06
HO_3S_2	$H-S-SO_3^-$	-156.7	-5.14	3.99
$HO4_S$	$H-O-SO_3^-$	-229.8	-6.00	6.75

Table 3-9: Decomposition Products from Dimethyl sulfate and Methyl methane thiosulfonate.

Reaction	ΔH_{rxn} [a], kcal
$Me\text{-}O\text{-}SO_2\text{-}O\text{-}Me + e^- \rightarrow MeSO_4^- + Me\cdot$	-16.85
$Me\text{-}O\text{-}SO_2\text{-}O\text{-}Me + e^- \rightarrow MeSO_3^- + OMe\cdot$	-47.65
$Me\text{-}SO_2\text{-}S\text{-}Me + e^- \rightarrow MeSO_2S^- + Me\cdot$	-13.95
$Me\text{-}SO_2\text{-}S\text{-}Me + e^- \rightarrow MeSO_2^- + SMe\cdot$	-22.06
$Me\text{-}SO_2\text{-}S\text{-}Me + e^- \rightarrow Me\cdot + {}^-SO_2\text{-}S\text{-}Me$	-28.82

[a] The values used to calculate the reaction were either the PM3 Heats of formation or, in the case of radicals, Benson, S. *Thermochemical Kinetics*, 2nd. ed. New York: Wiley And Sons, 1976, p 299.

for a few DA products. Not all of the ions listed in Figures 3-5 and 3-6 are listed here. As can be seen from the table, each of the simple reactions are exothermic. For methyl methanethiosulfonate, although the loss of ·SMe is less exothermic than the loss of ·Me to form $MeSSO_2^-$, the fairly weak sulfur-sulfur bond seems to provide the best channel for dissociation in terms of production of the ion. $MeSO_2^-$ gives a more intense signal than $MeSSO_2^-$.

The production of $MeSO_3^-$ is not without precedent. Lum and Grabowski studied the molecule Me_2SO_4 extensively under the conditions of flowing afterglow mass spectrometry.[122] Lum and Grabowski report that nucleophiles stronger than and including HS^- will form the $MeOSO_3^-$ ion.[122] Nucleophiles more basic than and including OH^- will also form $MeSO_2^-$.

The relative instability of the S-S bond in methyl methanethiosulfonate is also known.[123] In fact, the ease of SCH_3 abstraction is used in biochemistry to protect sulfhydryl groups.[123] The gas phase acidity of the conjugate acid of $MeSO_2^-$ from the DA of methyl methanethiosulfonate was bracketed in this study between 317.4 and 321.9 kcal/mol. The final portion of this chapter focuses its attention to the structure of the conjugate acid. There are two protonation sites as shown in Figure 3-7: the sulfur and one of the oxygens. To ascertain which site in $MeSO_2^-$ is protonated by trifluoracetic acid, calculations were performed using both semi-empirical and *ab initio* calculations. To reduce costly computer time for the *ab initio* work, the methyl group in $MeSO_2^-$ was replaced by a hydrogen. Similarly, a hydrogen replaced the methyl group in each of the compounds listed in Figure 3-7. The analogous hydrogen structure to Figure 3-7 (a) will hereafter be referred to as H_2SO_2, while the hydrogen counterpart to Figure 3-7 (b) will be referred to as HSOOH.

Figure 3-7: Possible Protonation Sites in $MeSO_2^-$: (a) Sulfur (b) Oxygen

The geometries for HSO_2H and HSO_2^- were constrained to have C_{2v} symmetry.

Geometries for H_2SO_2 from the MNDO semi-empirical calculations are compared to those found from the STO-3G and STO-3G* basis sets in Table 3-10. In terms of geometry, the MNDO calculations are not in good agreement with either of the STO-3G *ab initio* studies. The MNDO calculations have been calibrated to approximate experimental quantities, but the STO-3G calculations will not in many cases give such results. The initial calculations on the STO-3G level did not include the d AO's for the hypervalent sulfur. Although Dewar did not include sulfones and sulfoxides in his optimized parameters,[116] the calculations with his revised parameters would be better than the omission of the 3d AO's of sulfur in the *ab initio* calculation. One reference does include the 3d AO's,[126] but this can be regarded as only a patch-up of the STO-3G basis set. The agreement between the two *ab initio* calculations for H_2SO_2 is somewhat close

Table 3-10: Geometric Parameters for H_2SO_2.

Parameter	MNDO	STO-3G	STO-3G*[a]
r H-S, Å	1.543	1.342	1.363
r S-O, Å	1.585	1.912	1.454
∠ H-S-O	100.7	105.3	107.8
∠ O-S-O	136.6	134.4	126.1

Source: [a]R.J. Boyd and J.P. Szabo *Can. J. Chem.*, **60**, (1982) 730-4.

except for the S-O bond. Inclusion of the 3d AO's gave S-O a more double character.

Higher level basis sets for *ab initio* calculations on H_2SO_2 and HSOOH were not found in the literature. The molecules HSOOH and H_2SO_2 were also calculated using a PC version of MOPAC[125] with the AM1 Hamiltonian. PM3 calculations were also performed on the DEC VAX cluster system using the latest update of MOPAC, MOPAC5.[127] Again, the PM3 Hamiltonian becomes the most reasonable of the three semi-empirical methods in terms of the heats of formation for sulfur (VI).[118]

The *ab initio* calculations were performed using Gaussian 86.[128] Each calculation began with the MNDO starting geometry. No zero point energy calculations were performed, the structures were simply optimized at the corresponding levels. A comparison of STO-3G, 6-31G** and 6-31G** using Møller-Plesset second order perturbation theory is found in Table 3-11.

The results from the 6-31G** basis set seems to be in fair agreement with the STO-3G set for H_2SO_2 with one important exception. The higher level basis sets calculate the LUMO as being bound, in accordance with the observed dissociative attachment spectra of the analogous dimethyl sulfate. The Møller-Plesset second order perturbation interactions on the 6-31G** set gives the sulfur-oxygen bond

Table 3-11: *Ab initio* calculations on H_2SO_2, HSOOH and HSO_2^-.

Compound	Description	STO-3G	6-31G**	MP2/6-31G**
H_2SO_2	HOMO eV	-0.29778	-0.28409	-0.4783
	LUMO eV	0.35738	-0.16132	-0.1755
	E_{tot} Hartrees	-541.697	-547.842	-548.816
HSOOH	HOMO eV		-0.40391	-0.4106
	LUMO eV		0.15269	0.1371
	E_{tot} Hartrees		-548.317	-548.841
HSO_2^-	E_{tot} Hartrees		-547.747	-548.270

more double bond character and further decreases the energy of the HOMO for the H_2SO_2 molecule.

A comparison of the isomer HSOOH to H_2SO_2 shows H_2SO_2 is 0.025 Hartrees (16 kcal/mol) less stable than HSOOH at the MP2/6-31G** level. The electronic stability is in rough agreement with PM3 calculations on these isomers (see Table 3-8) with the ΔH_f of HSOOH being more stable by 11.6 kcal/mol. The total electronic energies at the HF/6-31G** level should not be trusted. Here the H_2SO_2 is less stable than HSOOH by more than 329 kcal/mol.

From the difference between the total electronic energies of the neutral relative to the anion, the ΔE_{acid} can be calculated. For the MP2/6-31G** calculations which were deemed the more reasonable, the ΔE_{acid} is 342.6

kcal/mol for H_2SO_2. For the more stable HSOOH, the ΔE_{acid} is 358.3 kcal/mol. These are in reasonable agreement with the experimentally observed value of 317-321 kcal/mol. The calculation is expected to come closer to the experimental value if the even higher MP2/6-31$^+$G** level were to be used. The PM3 calculation gives a good estimate of the experimental value. From the calculated heats of formation, the ΔH_{acid} is 335 kcal/mol.

Further comparison of the geometric data for H_2SO_2 and HSO_2^- appear as Table 3-12. Major geometric changes occur when the H_2SO_2 is generated from HSO_2^-. Large scale geometric changes also occur when HSOOH is formed from the anion.

Table 3-12: *Ab initio* MP2/6-31G** geometries of H_2SO_2 and HSO_2^-.

Description	H-S length in Å	S-O length in Å	H-S-O bond angle, in degrees	O-S-O bond angle, in degrees
H_2SO_2	1.351	1.457	107.65	124.39
HSO_2-	1.419	1.520	100.94	115.78
Difference	0.07	0.06	-6.71	-8.61

CHAPTER 4

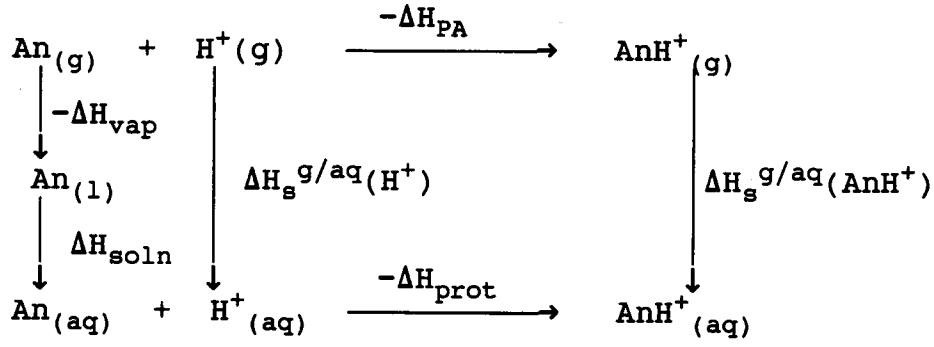
THE ENTHALPIES OF SOLUTION FOR SELECTED SUBSTITUTED ANILINES

I. INTRODUCTION.

Prior to the onset of modern gas phase research in ion thermodynamics, the characteristic properties of a cation or an anion were measured in a variety of solvents. Many reactions were found to be solvent dependent: for example, protic solvents slow down anionic S_N2 reactions relative to polar aprotic solvents.[129] Studies of amines were of particular interest. The basicity of amines is not a direct, linear function of the degree of alkylation: secondary and primary amines are more basic than ammonia, which is more basic than tertiary amines.[130] The thermodynamics of reactions for substituted anilines has also been investigated.[131]

The advent of accurate gas phase ion-molecule thermodynamic measurements allowed the intrinsic properties of compounds to be measured without any solvent. Reaction thermochemistry could now be separated into the intrinsic properties of the molecule and those associated with the solvated molecule. The intrinsic proton affinity, $-\Delta H_{PA}$, is

given for the prototypical species, An, in the top line of Scheme 1.

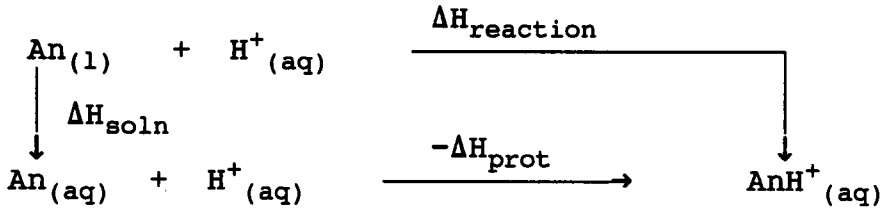


Scheme 1

Many of the quantities needed to complete scheme 1 are available. The experimental intrinsic basicities, expressed in ΔG's, for many anilines are known.[77,131] The intrinsic proton affinities are then determined from van't Hoff plots of the gas phase basicity at several temperatures[15] or by extra-thermodynamic assumptions.[7] The aqueous enthalpies of basicity, $-\Delta H_{prot}$, are known from the van't Hoff temperature dependence of the pK_a of the corresponding conjugate anilinium acid,[132] or from direct calorimetry.[133,134] The heats of vaporization, ΔH_{vap}, of substituted anilines are either known or may be derived from the Clausius-Clapeyron equation based on changes in vapor pressure with respect to temperature.[135-137] The enthalpy of taking H^+ from the gas phase into water, $\Delta H_s^{g/aq}(H^+)$, has been estimated from the electrochemical works of Randle,[138] Trassatti,[139] and

Klots[140] to be -271.6 ± 0.4 kcal/mol. This leaves only two portions of the entire scheme without either direct measurement or a usable theoretical basis: the solvation of the neutral liquid anilines into water, ΔH_{soln}, and the solvation of gaseous aniliniums into water, $\Delta H_s^{g/aq}(AnH^+)$. If either can be determined, then the other can be determined via Hess's law.

The problem with calorimetrically obtaining the enthalpy of solution, ΔH_{soln}, for the neutral liquid anilines into water is primarily one of solubility. The hydrophobic phenyl moiety is very insoluble in water. To avoid the problem of solubilities, this work has measured $\Delta H_{reaction}$ for the reaction of neutral anilines into aqueous acid as shown by scheme 2.



Scheme 2

Knowing $\Delta H_{reaction}$ and ΔH_{prot} from the literature allows the calculation of the third quantity, ΔH_{soln} (for the process $An_{(l)} \rightarrow An_{(aq)}$), as shown by equation (4.1).

(4.1) $\Delta H_{soln} = \Delta H_{reaction} - \Delta H_{prot}$

Once this quantity is known, the ΔH for $An_g \rightarrow An_{(aq)}$ may be calculated using equation (4.2).

(4.2) $\Delta H_s^{g/aq}(An) = -\Delta H_{vap} + \Delta H_{soln}$

The final enthalpy of interest, the ΔH for $AnH^+_{(g)} \rightarrow AnH^+_{(aq)}$ can now be calculated from (4.3).

(4.3) $\Delta H_s^{g/aq}(AnH^+) = \Delta H_{prot} + \Delta H_s^{g/aq}(An) + \Delta H_s^{g/aq}(H^+) - \Delta H_{PA}$

II. Experimental.

A. Calorimeter:

The calorimeter used in these experiments has been previously described.[141] A 300 ml Dewar served as the reaction flask and was covered by a 3 cm cap of Teflon through which four holes were drilled. Through the center hole a glass rod stirred the solution via a Transicoil PN121-115 electric motor which was regulated to 1000 rpm. On the submerged end of the glass rod, two 1 cm Teflon disks which had been partially sliced were mounted to agitate the solution. The stirring was conducted at speeds allowing little observable vortex action.

The second hole of the 3 cm Teflon cap was reserved as a solute injection port. All samples were injected by the use of a 5 ml disposable syringe, modified such that a 0.7 cm thick Teflon disk sealed the syringe. A small well was drilled into the face of one side of the disk. Solid and liquid samples were placed into the well of this cup. Then a thin Teflon disk was placed over the cup to prevent any volatilization. Sample sizes typically ranged from 30 - 70 mg, resulting in solution concentrations of at most 5 mM after the injection of a second sample into the solvent. Once the syringe assembly reached thermal equilibrium with the stirred solvent, the sample was expelled through the use of a glass plunger. This sample delivery system worked with better precision than the previous method: injection via a Hamilton syringe. The Hamilton syringe tended to leak solute into the solution, resulting in reaction enthalpies which were too small. Tests of this new injection method resulted in almost no observable heat gain or loss (about 0.01 cal) with a blank injection. However, two separate problems occurred with the use of this method. First, a small air bubble could form at the top of the well in the cup preventing efficient solute transfer. Poor solute transfer was also observed whenever the cup landed upside down in the bottom of the Dewar. Both problems were magnified by the limited solubility of the solute.

The third hole in the cap served as a mount for the calorimetric reference. A 56 ohm resistor was placed in a 5 mm diameter glass tube. A small amount of silicone oil was placed into the tube to provide for efficient heat transfer. The tube was sealed then with a small amount of cork at the open end. The reference heat was provided by a 12.0 V pulse for 0.188 s, as regulated by the LM555 one-shot electronic timer. Reference heat injections were conducted before and after solute injection. Any trials having disparate reference injections were not considered as useful data.

The fourth hole in the cap contained a Fenwall GA51M2 glass encased thermistor which was placed inside a small amount of silicone oil within the shaft of a 5 mm NMR tube. The thermistor is one arm of a Wheatstone bridge. The output of the Wheatstone bridge was fed to a Siliconix 7600 operational amplifier, set up as an invertor with a gain of 10. The signal, after further tenfold gain via a LM356 operational amplifier, was then fed to a strip chart recorder, operating at 50 mV full scale. The output signal was estimated at 5 mV/mdeg temperature change within the Dewar. The heat of stirring could not be completely avoided within the system. A second LM356 operational amplifier acting as an integrator was used to generate a ramp voltage. This ramp voltage was subtracted from the bridge output voltage to compensate for the heat of stirring, thereby allowing a fairly flat baseline on a strip chart recorder.

The accuracy of the calorimeter to ± 0.1 kcal/mol had been recently checked by Wilson with repeated measurements of the enthalpy of solution of KCl, at 4.2 ± 0.02 kcal/mol.[141] Similar measurements on KCl verified the experiments performed by Wilson.

B. Compounds:

Liquid anilines were distilled in an inert atmosphere to yield a clear liquid and used within 24 hours. Beyond this time, a slight yellowing of the liquid was observed due to a reaction with air, making it necessary to redistill the material. An impure sample usually resulted in a numerically smaller reaction enthalpy than the pure compound. Freezing the sample to avoid air contamination proved futile. Solid anilines were sublimed under mechanical pump vacuum near the melting point of the aniline. Several hours of sublimation were required to obtain sufficient crystals for study. Enthalpies for the solid anilines were measured within 48 hours after sublimation. No discoloration was observed over the course of a few days for the solid samples when stored in a desiccator. The delay in performing calorimetry on the sublimed compounds was due to laboratory efficiency. It was often more efficient to let the aniline sublime over the

course of several hours, and during this time calorimetric measurements on other compounds were taken.

III. Results.

A. Neutral Solution:

The results of the calorimetry experiments are given by Tables 4-1 and 4-2. The term ΔH_{prot} refers to the enthalpy change when the aqueous aniline gains a proton to form the corresponding anilinium ion, as depicted by the bottom line of Scheme 2. The protonation enthalpies for aniline bases are often written in the reverse direction as $\Delta H_{ionization}$ for the conjugate anilinium acids.[130]

Table 4-1: Experimental Heats of Solution, N-Methylated Cases as in Equation (4.1).

Compound	ΔH_{prot} kcal/mol	ΔH reaction:this work, kcal/mol	ΔH_{soln} kcal/mol
Aniline	-7.17 ± .10 [a]	-7.05 ± 0.06 (3)[c]	0.12 ± 0.11
N-Methyl Aniline	-5.98 ± 0.10[b]	-6.19 ± 0.10 (5)	-0.21 ± 0.14
N,N-Dimethyl Aniline	-4.92 ± 0.10[b]	-6.00 ± 0.22 (6)	-1.08 ± 0.24

Sources: [a] Mean value from: W.F. O'Hara. *Can. J. Chem.*, **1968**, *46* 1965.; Everet, D.H.; Wynne-Jones, W.F.K. *Trans. Faraday Soc.* **1939** *35*, 1380.; Zawidzki, T.W.; Papee, H.M.; Canady, W.J.; Laidler, K.J. *Trans. Faraday Soc.* **1959** *55*, 1738.
[b] P. deMaria, A. Fini, and F.M. Hall, *Gazz Chim. Ital.*, **1983**, *113*, 69.
[c] Number of runs in the study.

Table 4-2: Experimental Heats of Solution: Ring Substituted Cases as in Equation (4.1).

Compound	ΔH_{prot} kcal/mol	ΔH reaction:this work, kcal/mol	ΔH_{soln} kcal/mol
Aniline	-7.17 ± 0.10[a]	-7.05 ± 0.06 (3)[c]	0.12 ± 0.11
m-Me	-7.58 ± 0.02[b]	-7.12 ± 0.10 (4)	0.46 ± 0.10
m-OMe	-7.11 ± 0.02[b]	-5.32 ± 0.16 (4)	1.79 ± 0.16
m-F	-6.56 ± 0.04[b]	-5.90 ± 0.15 (6)	0.66 ± 0.16
m-Cl	-6.45 ± 0.03[b]	-5.46 ± 0.13 (6)	0.99 ± 0.13
m-CN	-5.74 ± 0.06[b]	0.39 ± 0.10 (6)	6.13 ± 0.12
p-OMe	-8.51 ± 0.01[b]	-3.38 ± 0.05 (4)	5.13 ± 0.05
p-Me	-7.86 ± 0.11[b]	-3.38 ± 0.19 (7)	3.98 ± 0.21
p-F	-7.76 ± 0.04[b]	-7.02 ± 0.10 (3)	0.74 ± 0.11
p-Cl	-6.67 ± 0.09[b]	-1.53 ± 0.18 (6)	5.14 ± 0.20
p-NO_2	-4.12 ± 0.01[b]	0.00 ± 0.05 (3)	4.12 ± 0.05

Sources: [a] Mean value from: W.F. O'Hara. *Can. J. Chem.*,**1968**, *46* 1965.; Everet, D.H.; Wynne-Jones, W.F.K. *Trans. Faraday Soc.* **1939** *35*, 1380.; Zawidzki, T.W.; Papee, H.M.; Canady, W.J.; Laidler, K.J. *Trans. Faraday Soc.* **1959** *55*, 1738.
[b] C.L. Liotta, E.M. Perdue, and H.P. Hopkins, Jr. *J. Amer. Chem. Soc.*,**1973**, *95*, 2439-45.
[c] Number of runs in the study.

One problem to be addressed was the range of ΔH_{prot} values for previously reported in the literature at 298 K for unsubstituted aniline the values range from 5.78 kcal/mol to 7.43 kcal/mol.[146] The procedure followed here to estimate the best value for aniline ΔH_{prot} was that of Larson and Hepler:[147] the best value is the mean of three careful works.[133,143,144]

The enthalpy of solution found for each of the ring-substituted anilines in this study was endothermic as would be expected from the low solubility of each compound in water. Successive methyl substitutions on the nitrogen moiety resulted in ΔH_{soln} becoming more exothermic, even though the N-methyl substituted compounds are less soluble in water than aniline. The trend toward more exothermic ΔH_{soln} with increasing methyl substitution is also found in the corresponding methyl amines: the more substituted the amine, the more exothermic the solution enthalpy. The neutral heats of solution from the literature[153] for the comparable methyl amines are Me_3N -13.23; Me_2NH -12.69, and $MeNH_2$ -10.82 kcal/mol.

For the ring substituted compounds, there is a fair degree of correlation (correlation coefficient, r, = 0.962) of the ΔH_{prot} of the anilines with the Hammet acidity constants[148] shown in Figure 4-1. Most of the compounds listed in Table 4-2 are from the work of Liotta and coworkers. They show that the ΔG_{prot} and the ΔH_{prot} are also well correlated for the compounds studied.[134] The large residual value of p-nitroaniline is due to increased resonance in that species.[148] Study of Figure 4-1 reveals a y-intercept which corresponds to the ΔH_{prot} for aniline at -7.51 kcal/mol, which is in fair agreement with the value measured by Liotta and coworkers at -7.43 kcal/mol.[134]

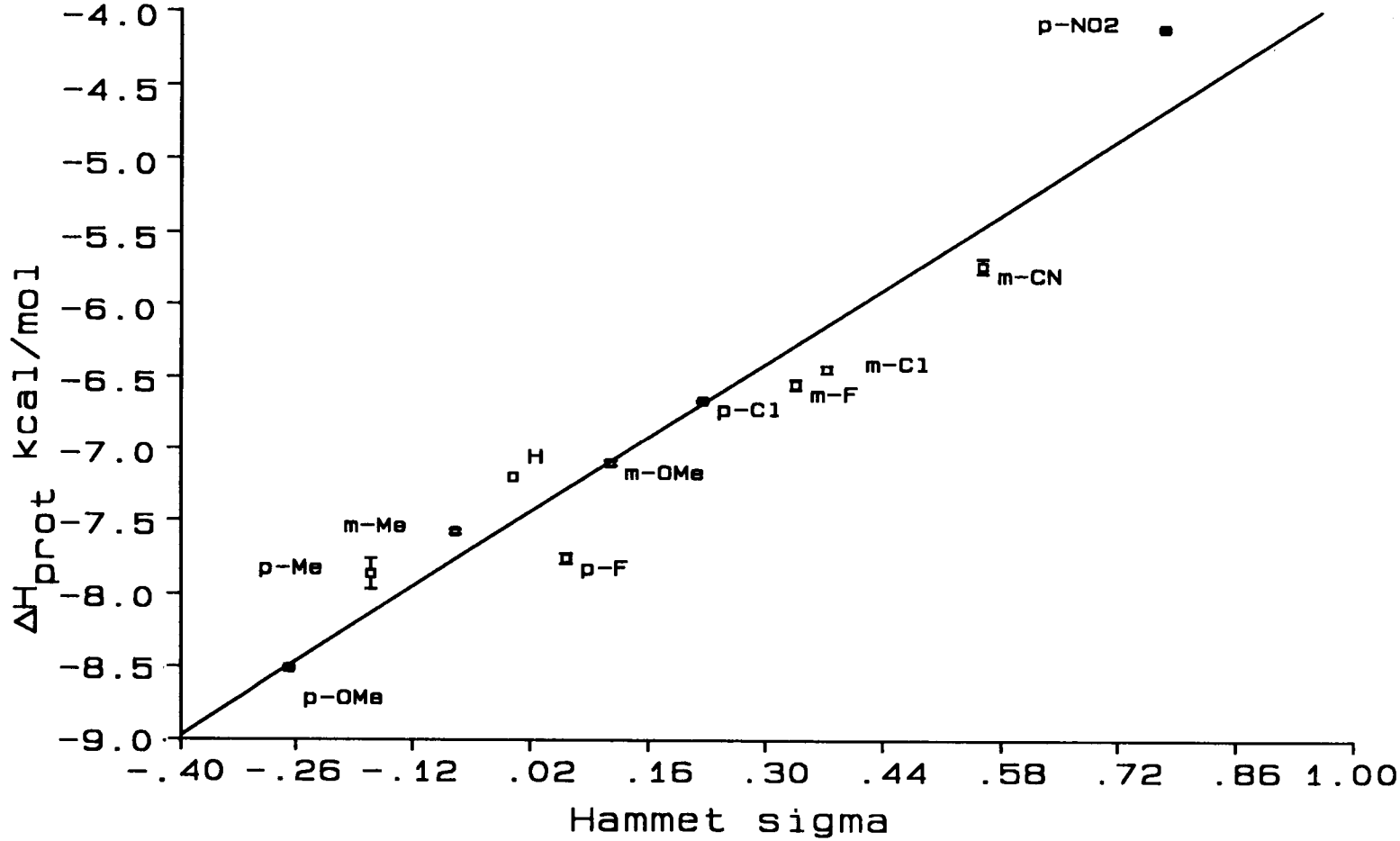


Figure 4-1: Plot of ΔH_{prot} vs. Hammet δ constants for Anilines.

Although the data in Figure 4-1 seems to be internally consistent, the measured ΔH_{prot} for aniline is lower than that proposed by other researchers.

B. Solvation of Substituted Aniliniums:

In order to complete the investigation of the solvation enthalpy of the anilinium ions, the heats of vaporization of the corresponding anilines are required. Heats of vaporization, ΔH_{vap}, can be estimated by using the Clausius-Clapeyron equation, given the vapor pressures of the aniline at various temperatures.[137] Only five of the thirteen anilines present in this study could have the ΔH_{vap} calculated in this manner. Further work is necessary to obtain accurate heats of vaporization for the remaining compounds which are liquids. The enthalpies of sublimation for the solid anilines were not found in the literature. The calculated ΔH_{vap} and the gas phase proton affinities, ΔH_{PA}, found in the literature are listed in Table 4-3.

From the data given in Tables 4-1 through 4-3, the solvation enthalpies of the five anilines and the corresponding anilinium ions can now be calculated using equations (4.2) and (4.3). The results of these calculations are listed in Table 4-4. In Table 4-4 there are two different series of data possible based on

Table 4-3: Proton Affinities and Vaporization Enthalpies for Anilines.

Compound	ΔH_{PA} kcal/mol[#]	ΔH_{vap} kcal/mol
Aniline	209.5 ± 0.7 [0.0][a]	13.3 ± 0.1[b]
N-methylaniline	218.1 ± 2.0* [8.6][c]	14.0 ± 0.1[b]
N,N-dimethyl aniline	223.4 ± 2.0*[13.9][d]	12.6 ± 0.05[b]
m-methylaniline	213.4 ± 2.0*[3.9][e]	15.5 ± 0.1[b]
m-chloroaniline	207.2 ± 0.7 [-2.3][a]	14.8 ± 0.1[f]

Notes: [#] Values in brackets the are the ΔH_{prot} relative to aniline. * Unpublished errors assumed to be ± 2.0 kcal/mol. Sources: [a] Lau, Y.K.; Nishizawa, A.T.; Brown, R.S., Kebarle, P. *J. Am. Chem. Soc.* **1981**, *103*, 6291-95. [b] T. Boublik, V. Fried, and E. Hala. *The Vapor Pressures of Pure Substances*, 2nd. ed. Amsterdam: Elsevier, 1984. [c] Lau, Y.K.; Saluja, P.P.S.; Kebarle, P.; Alder, R.W. *J. Am. Chem. Soc.* **1978**, *100*, 7328. [d] Taft, R.W. *Prog. Phys. Org. Chem.* **1983**, *14*, 328. [e] Summerhays,K.D.; Pollack, S.K.; Taft,R.W. Hehre, J. *J. Am. Chem. Soc.* **1977**, *99*, 4585 [f] J. Walton. "Aromatic Nitrogen Compounds. Critical Parameters," in *Engineering Sciences Data* London:Engineering Sciences Data Unit, 1978.

Table 4-4: Solvation Enthalpies for Anilines and Aniliniums.

Compound	$\Delta H_s^{g/aq}(An)$ kcal/mol {by Eq. (4.2)}	$\Delta H_s^{g/aq}(AnH^+)$ kcal/mol {by Eq. (4.3)}
Aniline	-13.2 ± 0.1	-82.5 ± 0.5
N-methylaniline	-14.2 ± 0.2	-73.3 ± 1.1
N,N-dimethylaniline	-13.7 ± 0.2	-66.8 ± 1.2
m-methylaniline	-15.0 ± 0.2	-80.8 ± 1.2
m-chloroaniline	-13.8 ± 0.2	-84.6 ± 0.5

substitution position: N-methylation and ring substitution. Methylating the nitrogen moiety decreases the solvation enthalpy of the anilinium ions by a loss of

approximately eight kcal/mol for every additional methyl group. With ammonia as the reference zero for $\Delta H_s^{g/aq}(NH_3)$, increasing methylation makes the solvation enthalpy more endothermic.[151]

Solvation enthalpies for aniline and the two N-alkyl anilines are also highly correlated (correlation coefficient 0.9993) with the gas phase proton affinities, as shown in Figure 4-2. The more alkyl groups placed on the aniline nitrogen, the more poorly solvated the ion is. Traditional inductive effects should play a role in stabilizing the aniline or amine nitrogen.[13] However, with each addition methyl group there is one less site on the anilinium/amine for hydrogen to bond exchange with the solvent.[152]

In contrast to the highly positive correlation between solvation enthalpies and the proton affinity, there is poor correlation between the solvation enthalpies of the neutral anilines and their anilinium ions. Aue and coworkers argue that the solvation enthalpy of an ion can be separated into two terms.[153] The first term is a bulk solvation term related to van der Waals interactions with the solvent cavity. The bulk solvation term for the aniliniums is expected to closely resemble the solvation of the neutral aniline. The second term is a purely electrostatic term stemming from the charged anilinium. Therefore the solvation enthalpy of an ion can be separated into an electrostatic term and a general solvation term.[13] The

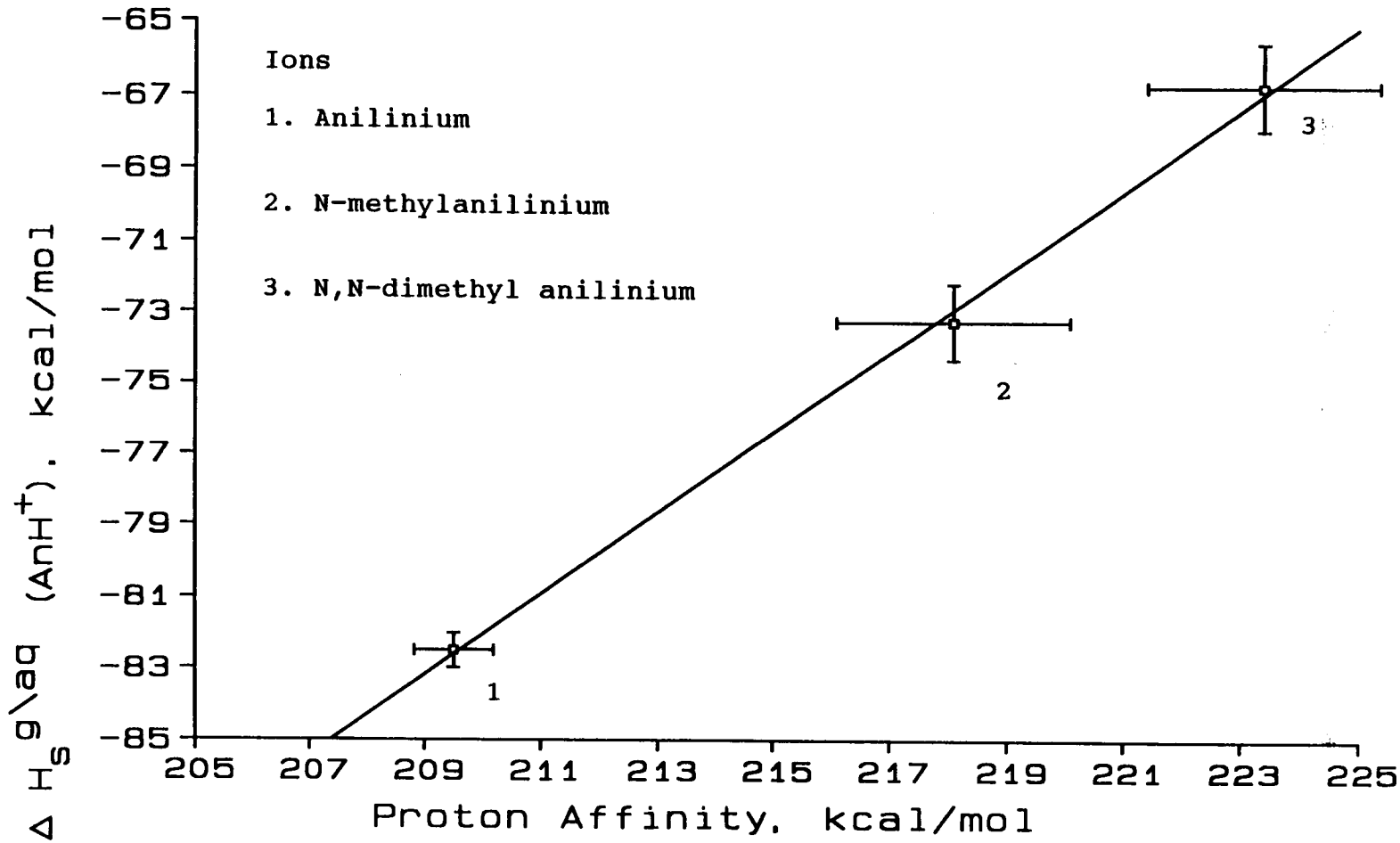


Figure 4-2: Plot of the bulk $\Delta H_s^{g\backslash aq}$ (AnH^+) vs. Proton Affinities for N-Methylated Aniliniums.

general solvation term is proportional to the solvation enthalpy of the neutral aniline. Thus the electrostatic term can be calculated to a first approximation by subtracting the solvation enthalpy of the neutral from the solvation enthalpy of the ion as shown be equation (4.4).

$$(4.4)\quad \Delta H_s^{g/aq}(AnH^+) \approx \Delta H_s^{g/aq}(An) + \Delta H_s^{g/aq}(AnH^+)^e$$

Wilson and Bartmess[141] found the electrostatic term correlated well with the proton affinities for aliphatic carboxylate anions. The electrostatic solvation enthalpy terms are given in Table 4-5 as $\Delta H_s^{g/aq}(AnH^+)^e$. For the anilines, correlation between the electrostatic terms of the solvation enthalpy and proton affinity is slightly greater (correlation coefficient 0.9999 in Figure 4-3) than when the bulk solvation enthalpy is considered against the proton affinity. Therefore more compounds are needed in order to validate the argument of Aue and coworkers in the N-alkylation studies. The principle barrier toward including other nitrogen-substituted alkyl anilines is the general lack of solubility of these compounds.

Table 4-5: Bulk and Electrostatic Solvation terms For N-methylated Anilinium ions.

Compound	$\Delta H_s^{g/aq}(An)$ kcal/mol	$\Delta H_s^{g/aq}(AnH^+)$ kcal/mol	$\Delta H_s^{g/aq}(AnH^+)^e$ kcal/mol
Aniline	-13.2 ± 0.1	-82.5 ± 0.5	-69.3 ± 0.5
N-methylaniline	-14.2 ± 0.2	-73.3 ± 1.1	-59.1 ± 1.2
N,N-dimethylaniline	-13.7 ± 0.2	-66.8 ± 1.2	-53.1 ± 1.2

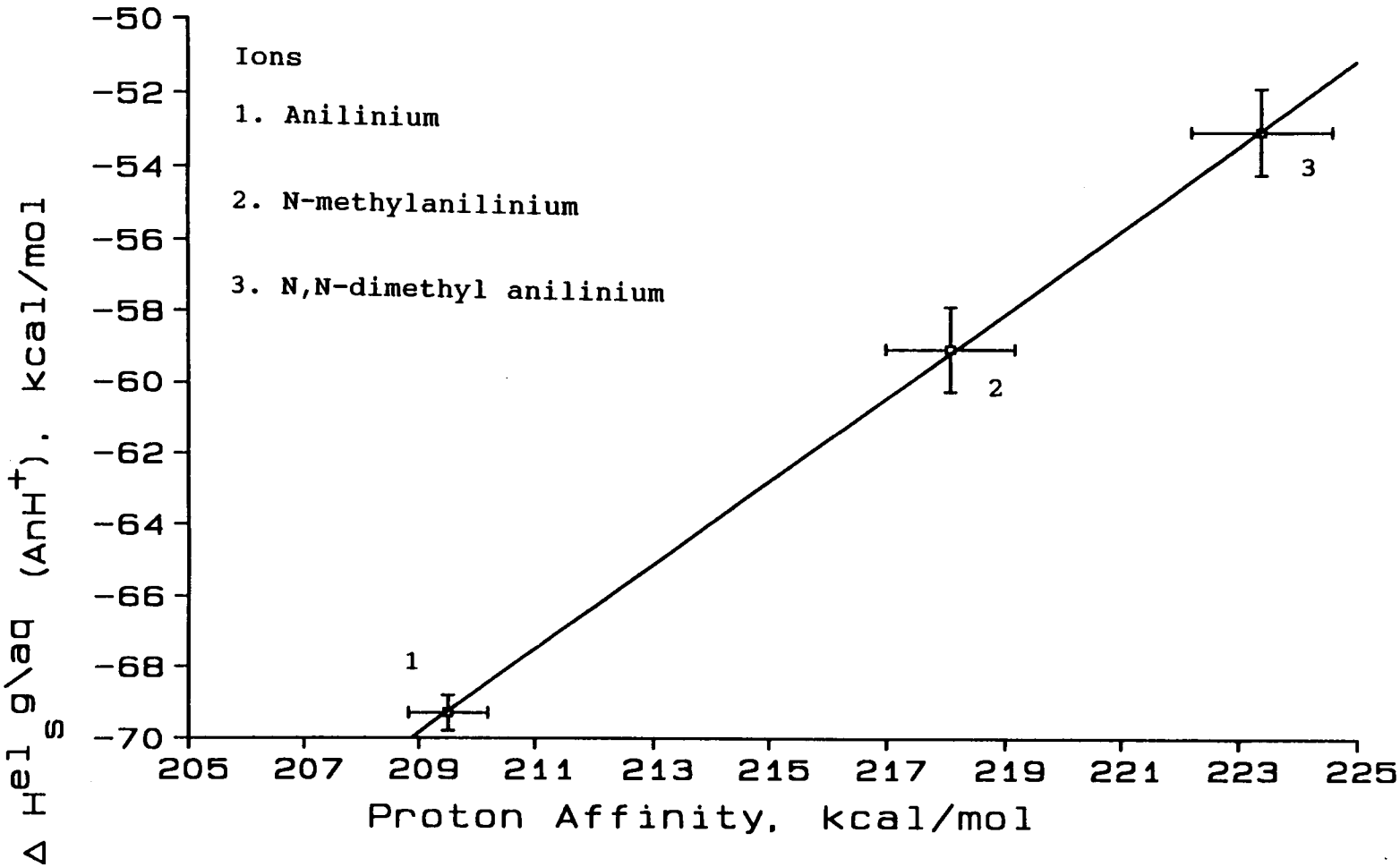


Figure 4-3: Plot of the electrostatic $\Delta H_s^{g\backslash aq}$ $(AnH^+)^e$ vs. Proton Affinity for N-Methylated Aniliniums.

The conclusions presented for the solvation enthalpies of the ring-substituted compounds is also limited by the lack of data. The lack of data, primarily ΔH_{vap} information, is unfortunate since Kebarle and coworkers[131] report that in some anilinium ions, protonation occurs on the ring while in others it seeks to protonate the nitrogen moiety. The determination of the proton position in solution is limited by the lack of measurements, particularly for vaporization and sublimation enthalpies for most of the compounds represented in Table 4-2.

For the three compounds where full data are available, there is a definite trend between the ring substituted anilinium solvation enthalpies given in Table 4-6 and the gas phase proton affinities, shown by Figure 4-4 ($r = 0.9783$). Each of these three compound is protonated on the nitrogen moiety according to the work of Kebarle's group.[131] Using the arguments of Aue and coworkers to calculate the effect of the electrostatic term results in an improved correlation (coefficient 0.9969), as shown by Figure 4-5.

For the ring substituted compounds, the electrostatic solvation term may be calculated. Comparison of equations (4.3) and (4.4) reveal a complete lack of dependence of $\Delta H_s^{g/aq}(AnH^+)$ on $\Delta H_s^{g/aq}(An)$. Therefore the electrostatic solvation enthalpies can be calculated for most of the ring substituted anilines through equation (4.5).

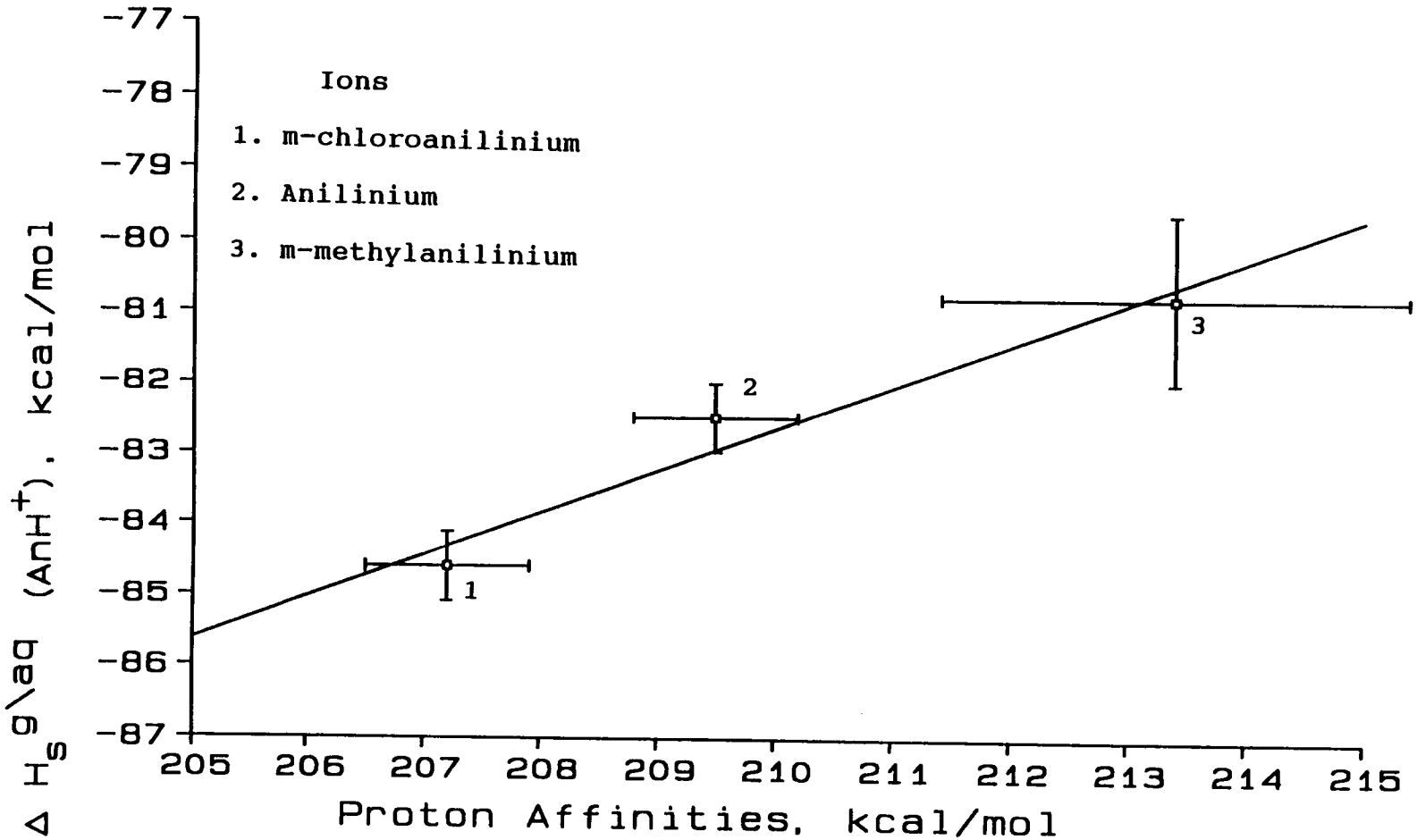


Figure 4-4: Plot of the bulk $\Delta H_s^{g\backslash aq}$ (AnH^+) vs. Proton Affinities for Ring Substituted Aniliniums.

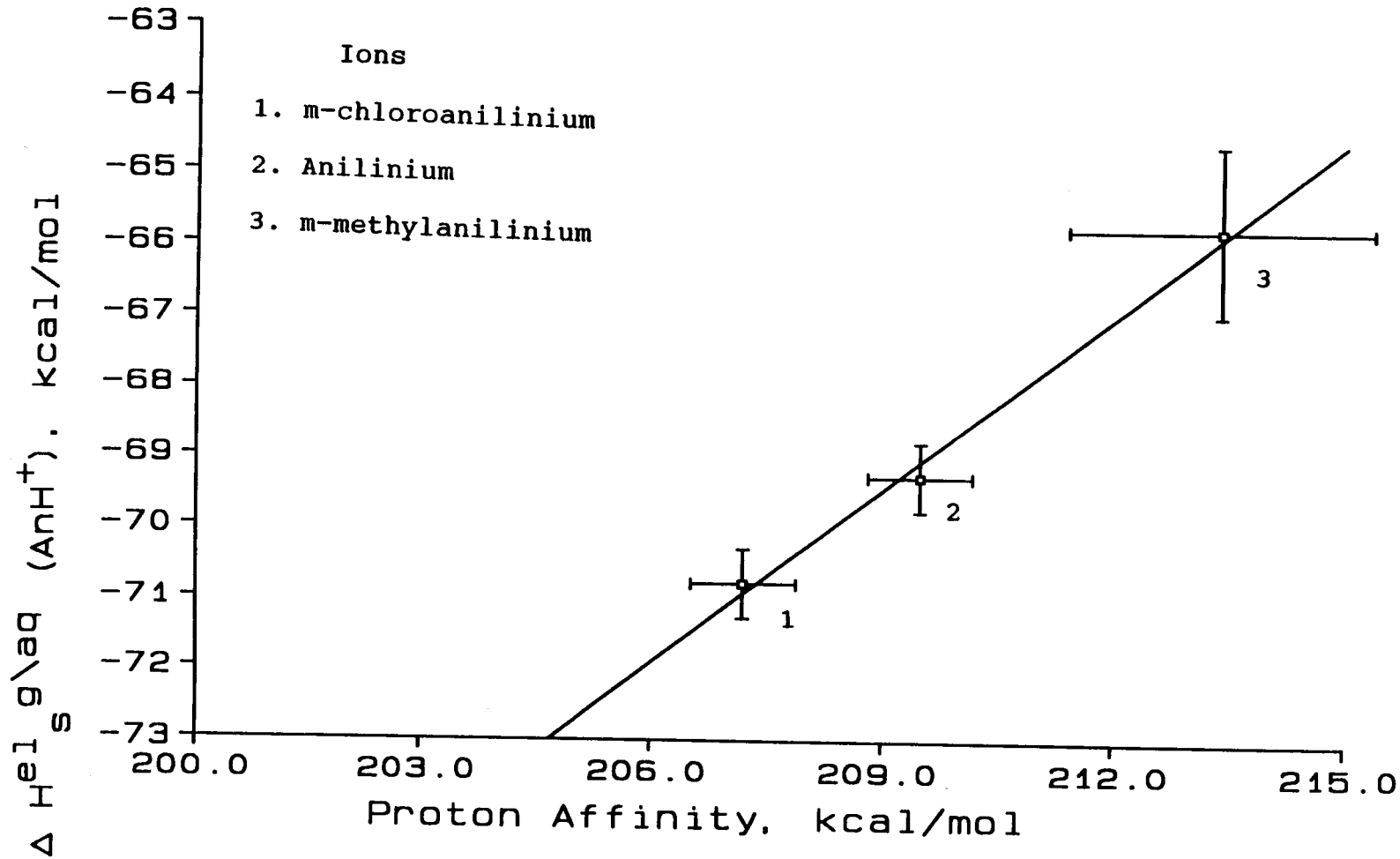


Figure 4-5: Plot of the electrostatic $\Delta H_s^{g\backslash aq}(AnH^+)^e$ vs. Proton Affinity for Ring Substituted Aniliniums.

(4.5) $\Delta H_s^{g/aq}(AnH^+) = \Delta H_{soln} + \Delta H_s^{g/aq}(H^+) - \Delta H_{PA}$

There is a high degree of correlation between the $\Delta H_s^{g/aq}(AnH^+)$ and the proton affinities as can be seen in figure 4-6.

Table 4-6: Bulk and Electrostatic Solvation terms For Ring Substituted Anilinium ions.

Compound	$\Delta H_s^{g/aq}(An)$ kcal/mol	$\Delta H_s^{g/aq}(AnH^+)$ kcal/mol	$\Delta H_s^{g/aq}(AnH^+)^e$ kcal/mol
Aniline	-13.2 ± 0.1	-82.5 ± 0.5	-69.3 ± 0.5
3-methylaniline	-15.0 ± 0.2	-80.8 ± 1.2	-65.8 ± 1.2
3-chloroaniline	-13.8 ± 0.2	-84.6 ± 0.5	-70.8 ± 0.5

Although the amount of data is too limited to make absolute predictions, it seems solvation affects the two series in a similar fashion. The inductive effects on the ring are electrostatic in nature and play a greater role than the stabilizing affects of the N-alkylation. The N-methylation data are twice as sensitive toward the proton affinity as the ring-substitution. The electrostatic solvation term is more sensitive toward the proton affinity than the bulk solvation term as would be expected from a charged-charged process. Thus the electrostatic solvation terms for ring substituted ions would play a greater role in the solvation enthalpies of these ions than the N-alkylated counterparts.

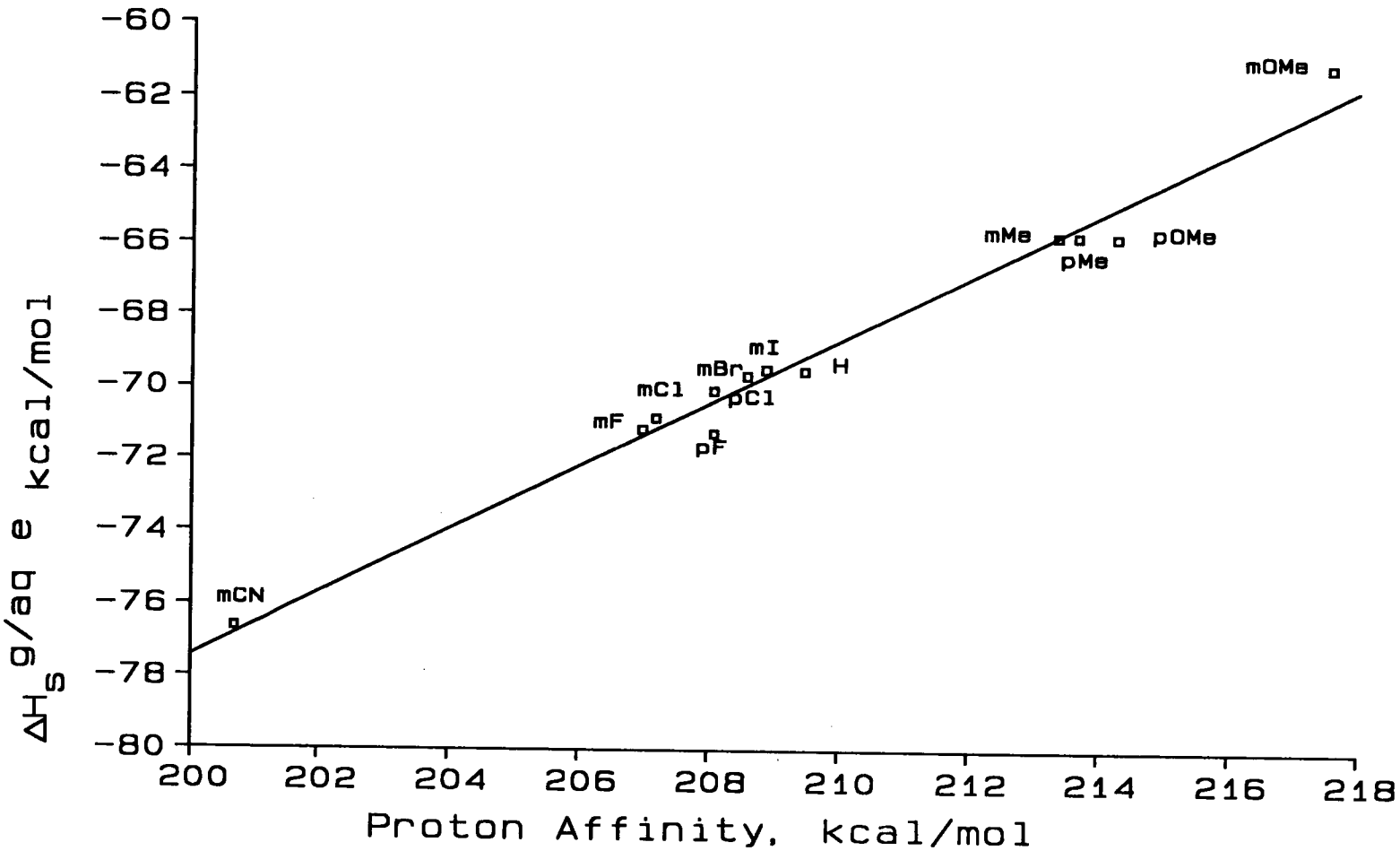


Figure 4-6: Plot of the electrostatic $\Delta H_s^{g\backslash aq}$ $(AnH^+)^e$ vs. Proton Affinity for Ring Substituted Aniliniums including those where heats of vaporization are not reported.

LIST OF REFERENCES

LIST OF REFERENCES

1. McMahon, T.B.; Kebarle, P.; *J. Am. Chem. Soc.* **1977**, *99*, 2222.

2. Cumming, J.B.; Kebarle, P.; *Can. J. Chem.* **1978**, *56*, 1.

3. Bartmess, J.E.; Scott, J.A.; McIver, R.T.; *J. Am. Chem. Soc.* **1979**, *101*, 6046.

4. Caldwell, G.; Renneboog, R.; Kebarle, P.; *Can. J. Chem.* **1989**, *67*, 611.

5. Decouzon, M.; Gal, J.; Gayraud, J.; Maria, P.; Viaglo G.; Volpe P. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 54.

6. Wolf, J.A.; Staley, R.H.; Koppel, I.; Taagepera, M.; McIver, R.T., Jr.; Beachamp, J.L.; Taft, R.W. *J. Am. Chem. Soc.* **1977**, *99*, 5417.

7. Yamdagi, R.; Kebarle, P. *J. Am. Chem. Soc.* **1976**, *98*, 1320.

8. Bohme, D.K.; Mackay, G.I.; Schiff, H., *J. Chem. Phys.* **1980**, *73*, 4976.

9. McMahon, T.B.; Kebarle, P. *J. Am. Chem. Soc.* **1986**, *108*, 6502.

10. McMahon, T.B.; Kebarle, P. *J. Am. Chem. Soc.* **1985**, *107*, 2612.

11. Fujio, M.; McIver, R.T., Jr.; Taft, R.W. *J. Am. Chem. Soc.* **1981**, *103*, 4017.

12. Mishima, M.; Kang, C.H.; Fujio, M.; Tsuno, Y. *Chem. Lett.* **1992**, 493.

13. Aue, D.H.; Webb, H.M.; Bowers, M.T. *J. Am. Chem. Soc.* **1976**, *98*, 311.

14. Rains, L.J.; Moore, H.J. McIver, R.T., Jr. *J. Chem. Phys.* **1978**, *68*, 3309.

15. Kebarle, P.; Chowdhury, S. *Chem. Rev.* **1987**, *87*, 513.

16. Chowdhury, S.; Kishi, H.; Dillow, G.W.; Kebarle, P. *Can. J. Chem.* **1989**, *67*, 603.

17. Meot-ner, M.; Sieck, L.W. *J. Am. Chem. Soc.* **1991**, *113* 4448.

18. Meot-ner, M.; Nelson, S.F.; Willi, M.R.; Frigo, T.B. *J. Am. Chem. Soc.* **1984** 106, 7384.

19. Woodin, R.L.; Beachamp, J.L. *J. Am. Chem. Soc.* **1977**, *100*, 501.

20. Corderman, R.R.; Beachamp, J.L. *Inorg. Chem.* **1977**, *16*, 3135.

21. Jones, R.L.; Staley, R.H. *J. Am. Chem. Soc.* **1982**, *104*, 2296.

22. Uppal, J.S.; Staley, R.H. *J. Am. Chem. Soc.* **1982**, *104*, 1235.

23. Uppal, J.S.; Staley, R.H. *J. Am. Chem. Soc.* **1982**, *104*, 1238.

24. Kappes, M.M.; Staley, R.H. *J. Am. Chem. Soc.* **1982**, *104*, 1813.

25. Kappes, M.M.; Staley, R.H. *J. Am. Chem. Soc.* **1982**, *104*, 1819.

26. Kappes, M.M.; Uppal, J.S.; Staley, R.H. *Organometallics* **1982**, *1*, 1303.

27. Operti, L.; Tews, E.C.; Freiser, B.S. *J. Am. Chem. Soc.* **1988**, *110*, 3847.

28. Solomon, J.J.; Field, F.H. *J. Am. Chem. Soc.* **1976**, *98*, 1025,1567.

29. Larson, J.W.; McMahon, T.B. *J. Am. Chem. Soc.* **1985**, *107*, 766.

30. Larson, J.W.; McMahon, T.B. *J. Am. Chem. Soc.* **1983**, *105*, 2944.

31. Dillow, G.W.; Kebarle, P. *J. Am. Chem. Soc.* **1988**, *110*, 4877.

32. Larson, J.W.; McMahon, T.B. *J. Am. Chem. Soc.* **1987**, *109*, 6230.

33. Larson, J.W.; Szulejko, J.E.; McMahon, T.B. *J. Am. Chem. Soc.* **1988**, *110*, 7604.

34. Larson, J.W.; McMahon, T.B. *J. Am. Chem. Soc.* **1984**, *106*, 517.

35. Larson, J.W.; McMahon, T.B. *Can. J. Chem.* **1984**, *62*, 675.

36. Larson, J.W.; McMahon, T.B. *J. Phys. Chem.* **1984**, *88*, 1083.

37. Bordwell, F.G.; *Acc. Chem. Res.* **1988**, *21*, 456.

38. Professor Bordwell generously provided access to his computer database of over 1100 acids in a personal communication, March 1987.

39. Skoog, D.A.; West, D.M.; Holler, F.J. *Fundamentals of Analytical Chemistry*, 6th ed. Fort Worth: Saunders Publishing, 1981, pp 212-35.

40. Andersen, T; Lykke, K.R.; Lineberger, W.C. *J. Chem. Phys.* **1987**, *86*, 1858.

41. Cooper, C.D.; Naff, W.T.; Compton, R.N. *J. Chem. Phys.* **1985**, *63*, 1858.

42. Dongarra, J.J.; Moler, C.B.; Bunch, J.R.; Stewart, G.W. *Linpack User's Guide;* Philadelphia:SIAM (Society for Industrial and Applied Mathematics), 1979.

43. Chandler, J.P. *Quantum Chem. Prog. Exchange* **1976**, *11*, program 307.

44. Press, W.H; Flannery, B.P.; Teulosky, S.A.; Vetterling, W.T. *Numerical Recipes: the Art of Scientific Computing* Cambridge: Cambridge Univ. Press, 1986, pp 229-34.

45. Lawson, C.L.; Hanson, R.J.; Kincaid, D.R.; Krogh, T.F. *ACM Trans. Math. Softw.* **1979**, *5*, 308.

46. Hanson, R.J.; Krogh, F.T. *ACM Trans. Math. Softw.* **1987**, *13*, 311.

47. Sorensen, D.N.; Bartmess, J.E. "The Proper Construction of 'Ladder' Affinity Scales," *Proceedings of the 37th ASMS Conference on Mass Spectrometry and Allied Topics*, Miami Beach, FL, May 17-22, 1989.

48. Sorensen, D.N.; Bartmess, J.E. "'Ladder' Algorithms for Constructing Equilibrium Affinity Scales," *Proceedings of the 39th ASMS Conference on Mass Spectrometry and Alied Topics*, Nashville, TN, May 19-24, 1991.

49. Kahaner, D.; Moler, C.; Nash, S. *Numerical Methods and Software*. Englewood Clifts, N.J.: Prentice-Hall, 1989, pp 190-234.

50. Kahaner, D.; Moler, C.; Nash, S. *Numerical Methods and Software*. Englewood Clifts, N.J.: Prentice-Hall, 1989, pp 41-75.

51. Golub, G.H,; Van Loan, C.F. *Matrix Computations*, Baltimore: Johns Hopkins Univ. Press, 1983, pp 52-79.

52. Venit, S.; Bishop, W. *Elementary Matrix Algebra*, 2nd ed. Boston: Prindle, Weber & Schmidt, 1985, pp. 52-62.

53. Shoemaker, D.P.; Garland, C.W., Nibler, J. *Experiments in Physical Chemistry*, 5th ed. New York: McGraw-Hill, 1989, pp. 801-27.

54. Hohn, F.E. *Elementary Matrix Algebra*. New York: Macmillan, 1958, pp 88-104.

55. Kahaner, D.; Moler, C.; Nash, S. *Numerical Methods and Software*. Englewood Clifts, N.J.: Prentice-Hall, 1989, pp 17-35.

56. Golub, G.H,; Van Loan, C.F. *Matrix Computations*, Baltimore: Johns Hopkins Univ. Press, 1983, pp 32-38.

57. Golub, G.H,; Van Loan, C.F. *Matrix Computations*, 2nd. ed. Baltimore: Johns Hopkins Univ. Press, 1989, pp 193-259.

58. Householder, A.S. *J. Assoc. Comput. Mach.* **1958**, *5*, 339.

59. Björk, Å. *BIT*, **1967**, 7, 257.

60. Björk, Å. *BIT*, **1967**, 7, 322.

61. Wampler, R.H. *ACM Trans. Math. Software*, **1979**, *5*, 457.

62. Wampler, R.H. *ACM Trans. Math. Software*, **1979**, *5*, 494.

63. Golub. G. *Numerische Mathematik*, **1965**, 7, 206.

64. Businger, P.; Golub. G. *Numerische Mathematik*, **1965**, 7, 269.

65. Golub. G.; Wilkinson, J.H. *Numerische Mathematik*, **1966**, *9*, 139.

66. Lawson, C.L. Hanson, R.J. *Solving Least Squares Problems*, Englewood Clifts, N.J.: Prentice-Hall, 1974, Ch 1.

67. Press, W.H; Flannery, B.P.; Teulosky, S.A.; Vetterling, W.T. *Numerical Recipes: the Art of Scientific Computing* Cambridge: Cambridge Univ. Press, 1986, pp 498-528.

68. Montgomery, D.C.; Peck, E.A. *Intoduction to Linear Regression Analysis*. New York: John Wiley and Sons, 1982, pp 99-104.

69. Montgomery, D.C.; Peck, E.A. Intoduction to Linear Regression Analysis. New York: John Wiley and Sons, 1982, pp 364-383.

70. Rousseeuw, P.J.; Leroy, A.M. *Robust Regression and Outlier Detection*, New York: John Wiley and Sons, 1987, pp 9-19.

71. Skoog, D.A.; West, D.M.; Holler, F.J. *Fundamentals of Analytical Chemistry*, 6th ed. Fort Worth: Saunders Publishing, 1981, pp 6-27.

72. Shoemaker, D.P.; Garland, C.W., Nibler, J. *Experiments in Physical Chemistry*, 5th ed. New York: McGraw-Hill, 1989, pp 29-82.

73. Chatterjee, S.; Hadi, A.S. *Sensitivity Analysis in Linear Regression*. New York:John Wiley and Sons, 1988, Ch 1.

74. Chatterjee, S.; Hadi, A.S. *Sensitivity Analysis in Linear Regression*. New York:John Wiley and Sons, 1988. Ch 3.

75. Hwang, J.D.; Winefordner, J.D. *Prog. Analyt. Spectrosc.* **1988**, *11*, 209.

76. Chatterjee, S,; Hadi, A.S. *Sensitivity Analysis in Linear Regression*. New York:John Wiley and Sons, 1988. Ch 4.

77. Lias, S.G.; Bartmess, J.E.; Liebman, J.F.; Holmes, J.L.; Levin, R.D.; Mallard, W.G.; *J. Phys. Chem. Ref. Data*, **1988**, *17*, Suppl. 1.

78. Lancaster, P.; Salkauskas, K. *Curve and Surface Fitting: An Introduction*. London:Academic Press, 1986.

79. Kahaner, D.; Moler, C.; Nash, S. *Numerical Methods and Software*. Englewood Clifts, N.J.: Prentice-Hall, 1989, pp 81-125.

80. EZplot, PKware, Inc, 1990,1991

81. Nyhoff, L.; Leestma, S. *Fortran 77 for Engineers and Scientists*. New York: MacMillan, 1985, p 271.

82. Nelder, J.A. Mead, R. *The Computer Journal*, **1965**, 7, 308.

83. Bartels, R.H.; Conn, A.R.; Sinclair, J.W. *SIAM J. Numer. Anal.* **1978**, *15*, 224.

84. Barrodale, I., Roberts, F.D.K. *SIAM J. Numer. Anal.* **1973**, *10*, 839.

85. Abdelmalek, N.N. *Math. of Computation* **1975**, *29*, 844.

86. Abdelmalek, N.N. *ACM Trans. Math. Software*, **1980**, *6*, 220.

87. Abdelmalek, N.N. *ACM Trans. Math. Software*, **1980**, *6*, 228.

88. Garvin, D.; Parker, V.B.; Wagman, D.D.; Evans, W.H. " A Combined Least Sums and Least Squares Approach to the Evaluation of Thermodynamic Data Networks," *NBSIR 76-1147*, Washington, D.C.: National Bureau of Standards, 1976.

89. Bartels, R.H.; Conn, A.R.; Charalambous, C. *SIAM J. Numer. Anal.* **1978**, *15*, 255.

90. Wilkinson, J.H. *The Algebraic Eigenvalue Problem*, Oxford: Clarendon Press, 1965, pp 94.

91. IMSL Inc. Sixth Floor, NBC Building, 7500 Bellaire Blvd., Houston, TX 77036.

92. Rousseeuw, P.J.; Leroy, A.M. *Robust Regression and Outlier Detection*, New York: John Wiley and Sons, 1987, pp 21-71.

93. Microsoft-Fortran version 3.2, Microsoft, Inc, Bellview, WA 98007.

94. VAX Fortran version 5.0, Digital Equipment Corporation, Maynard, Ma.

95. Taft, R.W. Personal Communication. Jul 1992.

96. Koppel, I.A.; Taft, R.W.; Anvia, F.; Zhu, S.-Z.; Hu, L.-Q.; Sung, K.-S.; DesMarteau, D.D.; Yagupolskii, L.M.; Yagupolskii, Y.L.; Ignat'ev, N.V.; Kondratenko, N.; Volkonskii, A.Yu.; Vlasov, V.M.; Notario, R.; Maria, P.-C.; *J. Am. Chem. Soc.* **1994**, *116*, 3047..

97. McMahon, T.B., Personal Communication correcting the Scale in Larson, J.W.; McMahon, T.B. *J. Am. Chem. Soc.* **1987**, *109*, 6230.

98. Christophorou, L.G.; Compton, R.N. *Health Phys.* **1967** *13*, 1277.

99. Melton, C.E. *Principles of Mass Spectrometry and Negative Ions*. New York: Marcell Dekker, Inc., 1970, Ch. 6 & 7.

100. Marr, G.V. *Photoionization Processes in Gases*, New York: Academic Press, 1967.

101. Christophorou, L.G. *Atomic and Molecular Radiation Physics*. London: John Wiley and Sons, 1971, 409-520.

102. McIver, R.T., Jr.; Lebford, E.B., Jr.; Miller, J.S. *Anal. Chem.* **1975** *47*, 692.

103. Hunter, R.L.; McIver, R.T., Jr. *Anal. Chem.* **1979** *51*, 699.

104. Budzikiewicz, H. *Mass Spectrom. Rev.* **1986** *5*, 345.

105. Zabik, S.E.; Watson, J.T.; Zabik, M.J. *Biomed. and Environ. Mass Spectrom.* **1990** *19*, 101.

106 Crowder, C.A. M.S. Thesis, University of Tennessee, 1988.

107. Bartmess, J.E.; Crowder, C. "New Sources of Electron Ionization Produced Anions," *Proceedings of the 34th ASMS Conference on Mass Spectrometry and Allied Topics*, Cincinnatti, OH June 8-13, 1986.

108. Caldwell, G.; Bartmess, J.E. *J. Phys Chem.* **1981**, *85*, 3571.

109. Dunbar, R.C. *Spectrochim. Acta* **1975**, *31A*, 797.

110. Cazeau, P.; Duboudin, F.; Moulines, F.; Babot, O. Dunogues, J. *Tetrahedron* **1987**, *43*, 2075.

111. House, H.; Czuba, L.J.; Gall, M.; Olmstead, H.D. *J. Org Chem.* **1969**, *34*, 2324.

112. Kiplinger, J.P. Ph.D. Dissertation, Indiana University, Bloomington IA, 1984.

113. McIver, R.T. *Rev. Sci. Instrum.* **1970**, *41*, 555.

114. Bartmess, J.E.; Georgiadis, R. *Vacuum* **1983**, *33*, 149.

115. McIver, R.T., Jr.; Dunbar, R.C.; *Int. J. Mass Spectrom. Ion Phys.* **1971** 7, 471.

116. Dewar, M.J.S.; Thiel, W.H. *J. Am Chem. Soc.* **1977**, *99*, 4899, 4907.

117. Dewar, M.J.S.; Zoebisch, E.G.; Healy, E.F.; Stewart, J.J.P *J. Am. Chem. Soc.* **1985**, *107*, 3902.

118. Stewart, J.J.P. *J. Comput. Chem.* **1989**, *10*, 209, 221.

119. Stewart, J.J.P. *J. Comput. Chem.* **1990**, *11*, 543.

120. Dewar, M.J.S.; Reynolds, C.H. *J. Comp. Chem.* **1986**, *2*, 140.

121. Benson, S. *Thermochemical Kinetics*, 2nd. ed. New York: Wiley And Sons, 1976, p 299.

122. Lum, R.C.; Grabowski, J.J. *J. Org. Chem.*, **1993**, *58*, 2029.

123. Welches, W.R.; Baldwin, T.O. *Biochem* **1980**, *20*, 512.

124. Dewar, M.J.S. *Quantum Chem. Prog. Exchange*, Indiana University, Bloomington, IN 47405, **1977**, *12*, program 353.

125. Program No. QCMP 004, Quantum Chemistry Program Exchange, Indiana University, Bloomington, IN 47405, wih extensive modifications by Bartmess, J.E for more efficient PC operation.

126. Boyd, R.J.; Szabo, J.P. *Can. J. Chem.* **1982**, *60*, 730.

127. Stewart, J.J.P MOPAC: A Semiempirical Molecular Orbital Program, QCPE 455 (1983) Version 5.0.

128. Frisch, M.J.; Binkley, J.S.; Schlegel, H.B.; Raghavachari, K.; Melius, C.F.; Martin, R.L. Stewart, J.J.P.; Bobrowicz, F.W.; Rohfling, C.M. Kahn, L.R.; Defrees, D.J.; Seeger, R.; Whiteside, R.A.; Fox, D.J.; Fleuder, E.M.; Pople, J.A. Carnegie-Mellon Quantem chemistry Pubishing Unit, Pittsburg PA., 1984.

129. March, J. *Advanced Organic Chemistry: Reactions, Mechanisms, and Structure*, 3rd ed. New York, Jown Wiley and Sons, 1985, pp 316-20.

130. Jones, F.M., III; Arnett, E.M. "Thermodynamics of Ionization and Solution of Aliphatic Amines in Water," in *Prog in Phys. Org. Chem.* Taft, R. W., ed. **1974** *11*, 263.

131. Lau, Y.K.; Nishizawa, A.T.; Brown, R.S., Kebarle, P. *J. Am. Chem. Soc.* **1981**, *103*, 6291.

122. Gandhi, A.H.; Patel, S.R. *Bull. Chem. Soc. Jap.* **1971**, *44*, 455.

133. W.F. O'Hara. *Can. J. Chem.*, **1968**, *46* 1965.

134. Liotta, C.L.; Perdue, E.M., Hopkins, H.P., Jr. *J. Amer. Chem. Soc.*,**1973**, *95*, 2439.

135. Hine, J. *Physical Organic Chemistry*. New York: McGraw-Hill, Ch 4, 1962.

136. Boubliek, T; Fried, V.; Hala, E. *The Vapor Pressures of Pure Substances*, 2nd ed. Amsterdam: Elsevier, 1984

137. Walton, J. "Aromatic Nitrogen Compounds, Critical Parameters" in *Engineering Sciences Data*. London:Engineering Sciences Data Unit, 1978.

138. Shoemaker, D.P.; Garland, C.W., Nibler, J. *Experiments in Physical Chemistry*, 5th ed. New York: McGraw-Hill, 1989, pp 219-29.

139. Randles, J.E.B. *Trans Faraday Soc* **1956**, *52*, 1573.

140. Trasatti, S. in *Modern Aspects of Electrochemistry*; Conway, E.; Bockris, J. Eds. New York, Plenum Press, 1979; Vol 13B.

141. Klots, C.E. *J. Phys. Chem.* **1981**, *85*, 3585.

142. Wilson, B.; Georgiadis, R.; Bartmess, J.E. *J. Am. Chem. Soc.* **1991**, *113*, 1762.

143. Everet, D.H.; Wynne-Jones, W.F.K. *Trans. Faraday Soc.* **1939** *35*, 1380.

144. Zawidzki, T.W.; Papee, H.M.; Canady, W.J.; Laidler, K.J. *Trans. Faraday Soc.* **1959** *55*, 1738.

145. de Maria, P.; Fini, A.; Hall, F.M. *Gazz Chim. Ital.*, **1983**, *113*, 69.

146. Christensen, J.J.; Hansen, L.D.; Izatt, R.M. *Handbook of Proton Ionization Enthalpies and Related Thermodynamic Quantities*. New York: John Wiley and Sons, 1976.

147. Larson, J.W.; Hepler, L.G. "Heats and Entropies of Ionization," in *Solute-Solvent Interactions*, Coetzee, J.F.; Ritchie, C.D., ed. New York: Marcel Dekker, 1969, pp 1-41.

148. Lau, Y.K.; Saluja, P.P.S.; Kebarle, P.; Alder, R.W. *J. Am. Chem. Soc.* **1978**, *100*, 7328.

149. Taft, R.W. "Protonic Acidities and Basicities in the Gas Phase and in Solution: Substituent and Solvent Effects," in *Prog. Phys. Org. Chem.*, Taft, R.W., ed. **1983**, *14*, 328.

150. Summerhays,K.D.; Pollack, S.K.; Taft, R.W.; Hehre, J. *J. Am. Chem. Soc.* **1977**, *99*, 4585.

151. Arnett, E.M.; Jones, F.M.,III; Taagpera, M.; Henderson, W.G.; Beauchamp, J.L.; Holtz, R.; Taft, R.W. *J. Am. Chem. Soc.* **1972**, *94*, 4724.

152. Trotman-Dickenson, A.F. *J. Chem. Soc.* **1949**, 1293.

153. Aue, D.H.; Webb, H.M.; Bowers, M.T. *J. Am. Chem. Soc.* **1976**, *98*, 318.

VITA

Daniel Nathan Sorensen was born in Canton, Ohio on September 16, 1964. He attended schools in the Canton Local school district and graduated Salutatorian from the 1982 graduating class of Canton South High School. After working a year to pay for college costs, he enrolled at Bob Jones University as a chemistry major. Between his Junior and Senior years at Bob Jones, he was introduced to the Chemistry Department at the University of Tennessee-Knoxville as a Science Alliance Summer Research Fellow. After graduating *cum laude* from Bob Jones University, Daniel came to UTK as a graduate student concentrating in physical chemistry. Finding the thermochemistry of ionic processes to be interesting, he joined the research group directed by Dr. John E. Bartmess. Also, that year, he joined both the East Tennessee Mass Spectrometry Discussion Group and the American Society for Mass Spectrometry. While at UTK he met and married Charlene Carpenter. They have one son, Andrew, born January 23, 1993.

Daniel is currently serving as an Assistant Professor of Chemistry on the faculty of Pikeville College, in Pikeville, KY.