TARGET FACTOR ANALYSIS OF GAS-CHROMATOGRAPHIC RETENTION INDICES

by

JOSEPH M. SOROKA

A dissertation submitted to the Graduate Faculty in Chemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1979

This manuscript has been read and accepted for the Graduate Faculty in Chemistry in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

5/21/79
date

Daryl J. Harris
Chairman of Examining Committee

21 May 1979
date

David C. Locke
Executive Officer

David C. Locke

Paul Saberfield

Albert Padovano
Supervisory Committee

The City University of New York

Acknowledgement

My feelings of fulfillment and satisfaction upon completing this dissertation can be matched only by the gratitude I feel towards the people who helped bring it to fruition.

To Professor Darryl Howery, my mentor, for generously lending his time, scientific expertise and editorial know-how to this thesis.

To Sarah, my wife and silent collaborator, whose patience and encouragement (not to mention, typing) sustained me throughout the most discouraging phases of this project.

And, finally, special thanks to my cheering section: my dedicated parents, Rabbi and Mrs. Samuel Soroka, our daughter, the ever-lovable Brendy J., Rabbi and Mrs. Simon Schwartz, my brothers and sister, for rooting me to the finish.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION

Most chemical problems are multifactor in scope. Many factors, some known and some most probably unknown, all contribute to a particular chemical measurement or reaction. As chemists we may succeed in elucidating how a system will react to changes in various parameters. However, a clarification of the importance and effect of all the parameters is seldom achieved. We hesitate to vary more than one or at most two parameters in a chemical experiment to determine chemical effects, lest the results leave us hopelessly confused. Ascertaining the true dependence of a measured quantity on each of the various experimental parameters is quite difficult.

Consequently, without a complete knowledge of the chemical interactions in a system, the chemist can not with confidence predict the effect of changing various parameters. Most times, his efforts become nothing more than "Let's try this." For example, chromatographers have for years been searching for an adequate explanation of solvent polarity and for a means of predicting which stationary phase would perform a needed separation.

To date, they have not been successful.

Even if the chemist is successful in amassing a large amount of data, he seldom knows how to interpret it fully. Recently, however, many of the techniques developed for data matrix manipulation in other fields, such as in the social sciences and psychology, have been adapted and applied to chemical problems. During this past decade, a new field called "chemometrics" has reached maturity.

One of the most powerful techniques that has been applied to interpretation of a large matrix of data is factor analysis. Factor analysis was originally developed for the social sciences by Thurstone and Spearman.[1-3] Malinowski and co-workers[4-6] succeeded in modifying factor analysis and applying it to chemical problems. Factor analysis enables the chemist to study many parameters simultaneously. Consequently, factor analysis allows the chemist to evaluate a problem which may have seemed hopelessly complicated.

One of Malinowski's major contributions was in developing a target testing procedure.[6] This procedure enables the chemist to test a parameter for its relevance in a problem. Thus, if a chemist suspects that a data vector may relate to one of the "factors" or parameters represented in a data matrix, the target testing procedure would allow him to determine if this is so.

There are various forms of factor analysis.[2]

For physical scientists, the power of the method resides in the target testing capability as developed by Malinowski. This method is known as target factor analysis (designated TFA in this thesis) and is to be distinguished from the earlier forms of factor analysis called principal component analysis (PCA) or abstract factor analysis (AFA).

In this thesis we will apply target factor analysis to various chemical problems. In Chapter II we discuss the theory of target factor analysis, including the important step of factor determination. In Chapter III the procedures of target factor analysis will be demonstrated by the use of a model problem. In Chapters IV through IX the application of target factor analysis to three chromatographic systems will be demonstrated. Some aspects of chromatography and solute-solvent theory will be discussed in Chapter IV. In Chapters V-VIII we will demonstrate how TFA has provided the first real insight into the solvent-related interactions for gas-liquid chromatography as well as extended insight into the solute-related interactions. In Chapter IX, TFA is used to explain a gas-solid chromatographic problem. Appendix A will demonstrate the application of TFA to an ion exchange problem.

The usefulness of TFA to the chemist has been immensely enhanced by the availability of a computer program originally written by Malinowski and coworkers[7] and since considerably expanded by us.[8] This program, called FACTANAL, is available from the Quantum Chemistry Program Exchange.

A brief discussion of this program's capabilities is included in Chapter III.

Three excellent reviews by Howery[9] and Weiner[10] have recently appeared in the literature documenting the myriad chemical problems successfully analyzed using factor analsis. Malinowski and Howery[11] are writing a monograph on the subject which should be available shortly.

The types of chemical problems approached by TFA include: linear free energy (Hammett-type) relationships,[4] activity coefficients,[5] gas phase acidities,[12] high performance liquid chromatography,[13] nuclear magnetic resonance,[6,14,15] drug activity,[16] polarography half-wave potential,[17] mass spectroscopy,[18] and infra-red solvent shifts.[19] Particularly well explored have been solute interactions in gas-liquid-chromatography.[20-27] In the following chapters we will delve into the factors influencing interactions in gas-liquid chromatography and in gas-solid chromatography.

CHAPTER II

THEORY OF ABSTRACT FACTOR ANALYSIS AND TARGET FACTOR ANALYSIS

A. Abstract Factor Analysis

In this chapter, we will discuss the theoretical basis of factor analysis. In the first part of this chapter we discuss the very important question of how to determine the proper factor space.

Target factor analysis requires that the data in a data matrix be explained using a particular type of equation.[2,6] As summarized in Eq. 1, each data point, d_{ik}, must be a linear sum of product terms, $r_{im}c_{mk}$:

$$d_{ik} = r_{i1}c_{1k} + r_{i2}c_{2k} + \ldots + r_{im}r_{mk} = \sum_{m=1}^{n} r_{im}c_{mk} \quad (1)$$

Here, r_{im} is the m-th cofactor term associated with i-th row designee of the data matrix, c_{mk} is the m-th cofactor associated with the k-th column designee of the matrix and n is the number of factors. Thus, each supposed contribution to the data point d_{ik} must be separable into two components: one depending only on physical and chemical characteristics of the row designee and the other reflecting the properties of the column designee.

In matrix notation, for a data matrix of r rows

and c columns,

$$[D] = [R] \cdot [C] \qquad (2)$$

where [D] is the observed r x c data matrix, [R] is the r x n row-designee cofactor matrix and [C] is the n x c column-designee cofactor matrix.

If one wishes to study chemical interactions, both the rows and columns of data are associated with chemical entities. For example, one can consider the r and c cofactors to be solute and solvent contributions respectively, and therefore, d_{ik} is the sum of various solute-solvent interactions terms. In gas-liquid chromatography, for example, the retention index of a particular solute on a particular solvent would be the sum of the contributions of the solute-solvent interactions for that pair.[28,35] Other examples include solute-solvent pairs in spectroscopic data,[19,38] donor-acceptor pairs in dissociation constants,[12,41] ion-ion pairs in ion exchange equilibrium,[31] radical-radical pairs in bond energy data and cation-ligand pairs in chelate formation constant.[32]

We will be mainly concerned with such entity-entity types of matrices. However, other types of data matrices are equally applicable to factor analysis. The r and c terms may represent the mass spectrum of a particular pure solute (r) and the partial pressure of that solute (c) in a mixture.[18,28-34] In this case, d_{ik} is the observed mass spectrum for the mixture. Data matrices of this type are known as entity-property matrices. Other entity-

property matrices in chemistry include such chemical problems as viscosities for solute-temperature pairs, conductivities for electrolyte-concentration pairs, spectral intensities for molecule-frequency pairs[36,37] and dosages for drug-disease pairs. Two other types of data matrices, entity-time and property-time matrices, are less common in chemistry.

Chemists tend to think of interactions in the sum-of-products formulation described in Equation 1. However, there are cases where Equation 1 is not valid. Often the chemist has no theoretical basis for using Equation 1. In such cases a factor analysis may be attempted empirically.

Since we will be concerned mainly with solute-solvent interactions, the following discussion will assume solutes for rows designees and solvents for columns designees. We realize that the formulation is equally valid for other types of data matrices.

Pretreatment of the data is often necessary. Either a correlation or a covariance matrix will be generated. Normalization of the data in the pretreatment step results in a correlation matrix. Unnormalized data will result in a covariance matrix. We may be interested in data centered about the raw data mean or centered about the data origin. Rozzett and Peterson[30] give an excellent discussion of data pretreatment and the proper use of each type of transformation.

Which pretreatment is used depends on the data. If, as is common, all of the variables are measured in the same units (for example, gas chromatographic retention index data) and all the errors are additive it is not necessary to pretreat the data. Thus, covariance about the origin would result.

Other data pretreatment methods include algebraic transformation of each data point, such as logarithm, exponential, square, square root and reciprocal transformations. These transformations should preferably be based upon sound scientific principles. For example, a logarithmic transformation of equilibrium constants should help to transform the data into the sum of product formulation necessary for TFA.

Multiplying data matrix $[D]$ by its transpose, $[D]^t$, gives $[C]$, the correlation (or covariance) matrix of dimension c x c:

$$[C] = [D]^t \cdot [D] \qquad (3)$$

The elements of $[C]$ in (3) are each summed over all solvent (column) designees and the diagonal matrix elements are each identified with a particular column designee. Each off-diagonal element is a sum of pairwise products of solvents effects for two solvents for all solutes (rows). An alternative correlation matrix:

$$[C] = [D] \cdot [D]^t \qquad (4)$$

could be formed in which case the diagonal elements are each identifiable with a particular solute (row) designee.

Correlation matrix $[\underline{C}]$ may be diagonalized by a diagonalization matrix $[B]$:

$$[B]^{-1}\,[\underline{C}][B] = [\lambda_j \quad \delta_{jk}] \tag{5}$$

where δ_{jk} is a Kronecker delta, and λ_j are the eigenvalues of the linear matrix equation

$$[\underline{C}]\ B_j = \lambda_j\ B_j \tag{6}$$

B_j is the jth eigenvector or the jth column of $[B]$ and λ_j is the eigenvalue associated with eigenvector B_j. The elements of the diagonalized matrix of $[\underline{C}]$ are abstract eigenvalues. Associated with each eigenvalue is an abstract eigenvector or abstract factor. A factor is customarily defined as a normalized eigenvector multiplied by the square root of the corresponding eigenvalue. The eigenvectors which result from this diagonalization may be thought of as a set of orthonormal "solvent basis vectors" (column designee vectors) which span the solvent-effect space. Real factors result from a transformation of real solvent basis functions and are the vectors with which, by means of a unitary transformation, we shall attempt to identify with particular solute (row designee) cofactors.[6,38]

We would like to define row and column matrices which would represent the influences of the solute and solvent parameters, respectively. We can rewrite Equations 5 and 6:

$$[B]^{-1}[\underline{c}][B] = [B]^{-1}[D]^{t}[D][B] = [B]^{t}[D]^{t}[D][B] \quad (7)$$

$$= [U]^{t}[U] = \lambda_{j}\delta_{jk} \quad (8)$$

where

$$[U] = [D][B] \quad (9)$$

Thus

$$[D] = [U][B]^{t} \quad (10)$$

By comparing this with Equation 2, we see that

$$[B]^{t} = [C] \quad (11)$$

and

$$[U] = [R] \quad (12)$$

Thus the transform of the diagonalization matrix $[B]$ is equivalent to the solvent (or column) cofactor matrix and the solute (row) cofactor matrix is given by Equation 12.

We have c eigenvalues for c eigenvectors. Factor analysis attempts to reproduce the original data with n factors (or eigenvectors) where n is less than c. This is known as factor compression. Thus we would now like to find the minimum number of linearly independent eigen-

vectors, B_j, necessary to reproduce $[D]$ within experimental error. We can start with the eigenvector B_1 associated with the largest eigenvalue λ_1 and attempt to reproduce the data:

$$[D]_1 = R_1 \; B_1 \tag{13}$$

where $R_1 = R_{i1}$ is a column vector, $B_1 = B_{1k}$ is a row vector and $[D]_1$ is the reproduced matrix using only the largest eigenvalue.

We continue by adding the next largest eigenvalue:

$$[D]_2 = |R_1 \; R_2| \cdot \begin{vmatrix} B_1 \\ B_2 \end{vmatrix} \tag{14}$$

and so forth until we have sufficient eigenvectors to represent the data satisfactorily. If we can reproduce the data satisfactorily using n eigenvectors,

$$[D]_n = |R_1 \; R_2 \; R_3 \ldots \; R_n| \cdot \begin{vmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{vmatrix} \approx [D] \tag{15}$$

we have determined the number of factors to satisfactorily span our data space. If $n < c$, then we have indeed reduced the number of terms in our factor equation. If, however, $n = c$, then we have as many factors as columns. This means that either our data does not have a factor analytical solution or that to span the factor space we do not have enough data.

B. Determination of the Correct Factor Space

Let us summarize briefly some of the more useful methods for determining the proper number of factors.

A realistic chemical criteria is to compare the reproduced and original data matrices point by point and accept the factor size necessary to reproduce the data within average experimental error. If x_{ik} is a data point in the original data matrix and x'_{ik} is the corresponding point in the reproduced matrix using m factors then the average rms error, $\bar{e}$, for that factor space is:

$$\bar{e} = \left[\sum_i (x_i - x'_i)^2\right]^{1/2} N^{-1} \qquad (16)$$

where N is the number of data points[42].

Another approach would be to compare the percentage of points with absolute errors greater than experimental. For example, we might accept a factor space where the percentage of data points with errors greater than experimental is only 2% of the total number of points, and reject a factor space where 5% of the errors are greater than experimental.

Most recently, Malinowski,[44,45] in a complete error analysis of abstract factor analysis described the error with three functions. A data point, d_{ik}, can be expressed as the sum of pure data, d^*_{ik}, and an error associated with that data point, e_{ik}. Thus:

$$d_{ik} = d^*_{ik} + e_{ik} \qquad (17)$$

If a data matrix contains no error the covariance matrix associated with the matrix can be decomposed by factor analysis into a sum of n factors, where n is less than r or c. Due to experimental error, FA yields an excessive number of factors equal either to the number of rows, r, or columns, c, of the data matrix, whichever is smaller. These excessive factors are associated with the small eigenvalues and their associated eigenvectors.

The error can be resolved into the imbedded error, $e^{\#}$, which mixes with the pure data in FA and can not therefore be removed, and an "extractable" error, e^{o}, which is removed by the factor analysis reproduction scheme. Thus:

$$d_{ik} = d^{*}_{ik} + e_{ik}{}^{\#} + e^{o}{}_{ik} \qquad (18)$$

The factors associated with the errors in the data consist of two types. The factors associated with the imbedded errors contain both pure data and error information; those associated with the extracted error are pure error factors.

For covariance about the origin, the experimental error is shown by Malinowski to be represented by the residual standard deviation (RSD) or Real Error (RE):

$$RE = RSD = \frac{\sum_{j=m+1}^{c} \lambda_j^{o}}{r(c-n)}^{1/2} \qquad (19)$$

Summing over all residual eigenvalues, i.e. over all eigenvalues less than the eigenvalue associated with the factor space under test, gives an estimate of the experimental error. The factor space where the Real Error function approximates experimental error, assuming it is known, should be the true factor space for the data.

The extracted error (XE) represents the extractable error, e^{o}_{ik}, and is associated with the pure error eigenvalue:

$$XE = \left[\frac{(C - N)}{C}\right]^{1/2} RSD \tag{20}$$

The imbedded error (IE) is associated with the error which mixes with the data and is therefore not removable from the data by factor analysis:

$$IE = \left[\left(\frac{n}{c}\right)\right]^{1/2} RSD \tag{21}$$

According to Malinowski's error theory,[44] the imbedded error function should reach a minimum at the true factor space. Because the extracted error is removed by the factor analysis scheme the reproduced data contains only the imbedded error. Factor analysis reproduction always leads to data improvement.

For correlation about the origin, the RSD takes the following slightly modified form:

$$RSD = \frac{\sum_{i}^{r} \sum_{k}^{c} d_{ik}^{2} \sum_{j=n+1} \lambda_{j}^{o}}{rc\ (c-n)}^{1/2} \tag{22}$$

Malinowski also discovered an empirical function called the Indicator (IND) function which reaches a minimum at the true factor space if all the errors are random and uniform in the data:

$$IND = \frac{RE}{(c-n)^2} \tag{23}$$

The IND function works quite well as we shall see on our real data. The imbedded error function is quite sensitive to the error distribution within the data and consequently often will fail to reach a minimum. In such cases the IND function may be sufficient to provide a determination of the factor space size.

Thus, even without any knowledge of the experimental error, one may be able to determine the true factor space using the IE and IND functions. In fact, once this is done a reliable estimate of the experimental error is obtained from the Real Error for that factor space.

We have found the average error, $\bar{e}$, and its complement, percentage of errors greater than experimental, together with the Malinowski error functions to be the best and most reliable criteria in determining the factor size. Though they have only recently been applied to real problems, the Malinowski functions present the greatest promise of providing a theoretically valid determination of the factor space size. The following chapters will provide many examples of the application of these criteria to real problems.

Several other criteria havebeen proposed for the determination of the factor size. Many texts[1-3] suggest accepting those factors with eigenvalues, λ, greater than 1.0 (for a correlation analysis); all eigenvalues less than 1.0 are assumed to account for an insignificant amount of the total variance of the data and are rejected. A similar criteria accepts all eigenvalues greater than the average eigenvalue, $\bar{\lambda}$.[42] Another criteria, the significant variance,[42] accepts all eigenvalues which are necessary to account for the fraction of the total variance thought to come from the desired information in the data.

For all of the real problems studied by us, the largest factor is the most dominant as evidenced by the size of the associated eigenvalue. For correlations, the first eigenvalue is always large; the following eigenvalues are almost always less than 1.0. Thus the eigenvalue greater than one, average eigenvalue and significant variance tests would lead us to conclude that only one factor is important, a fact which would ignore the contributions of the smaller factors. Rejection of factors because their respective eigenvalues are below an arbitrary value of 1.0 or below an average eigenvalue will invariably serve to eliminate some of the significant factors in the data. Similarly an accurate estimation of the total significant variance is hard to obtain. Is it 99.9%? Or perhaps, only 80%? Do we accept a factor if it accounts

for the 94th percentile of the variance or reject it?

Two other criteria, the Chi-square test[41,42] and the Exner function[43] have also been used. The Exner function has also been proposed as a criteria to compare test vectors. Both of these functions have proved to be inadequate for our purposes.

C. Target Factor Analysis

At this point, we have succeeded in defining the abstract cofactors in our data. These abstract factors are not recognizable from a physical or chemical viewpoint. We would like to transform the mathematical axes (factors) into axes which do have physical significance.

It is at this point that target factor analysis (TFA) differs from abstract factor analysis (AFA) or its most common form, principal component analysis (PCA). AFA techniques do not allow us to identify the abstract mathematical factors with physically meaningful vectors. However, much can be learned from an AFA.[1-3,29-41]

A unitary transformation of $[R]$ using a transformation matrix $[T]$ would provide a redefined rotated matrix $[\bar{R}]$

$$[\bar{R}] = [R]\ [T] \tag{24}$$

The inverse of $[T]$ can be used to locate the column solvent factor matrix in the new coordinate system

$$[\bar{C}] = [T]^{-1}[C] \tag{25}$$

Thus

$$[D] = [R][T][T]^{-1}[C] = [\bar{R}][\bar{C}] \tag{26}$$

Thus it is possible, upon a proper transformation, to find a row matrix which can be interpreted in chemical terms. There are an infinite number of possible transformations. However, only some of them will correspond to recognizable

real parameters. We, therefore, need to calculate the transformation columns, T_k, from real vectors. The target transformation method, through a least squares formulation, can test whether various physically significant parameters represent the cofactors in a problem.

Malinowski[6] has described a technique for ascertaining whether a suspected test quantity can be transformed, or rotated, into, and therefore associated with, a column of the $[\bar{R}]$ matrix. If the suspected test factor is a true factor, a rotation matrix $[T]$ can be found which will rotate a column of the $[\bar{R}]$ matrix, $\bar{R}_k$, into the suspected test factor, so that all of the elements of the rotated vector agree with the values of the test vector within experimental error. Malinowski's method is based on a least squares treatment and, to a chemist, is probably the most significant and powerful application for the factor analysis technique.

As derived by Malinowski,[6] the least squares vector transformer T_k, a column of $[T]$, is calculated from

$$T_k = \frac{1}{\lambda} \delta_{jk} \cdot [R]^t \cdot \bar{\bar{R}}_k \qquad (27)$$

$\bar{\bar{R}}_k$ is a vector composed of the suspected parameters.

Transformation of the axes is accomplished by the following

$$[\bar{R}] = [R][T] \qquad (28)$$

where $[T]$ is the transformation matrix and $[\bar{R}]$ is the row matrix in the new coordinate system. If a suspected

vector $\bar{\bar{R}}_k$ is a true vector, i.e., it represents one of the cofactors of the problem, then each element of $\bar{\bar{R}}_k$ must equal the corresponding element of $\bar{R}_k$ near experimental error. If $\bar{\bar{R}}_k$ is not a true vector, then the difference between $\bar{\bar{R}}_k$ and $\bar{R}_k$ will be large, and $\bar{\bar{R}}_k$ is not a cofactor. Thus, we can test whether any real test vector is a cofactor.

This transformation method is completely general and is applicable even if some $\bar{\bar{R}}_k$ values are omitted either purposely or because they are unknown. If $\bar{R}_k$ is indeed transformed the unknown values are predicted by the method. This method is known as free-floating.

Finally, in what is known as the combination step, we can attempt to reconstruct the original data matrix $[D]$ from the matrix of transformed real vectors $[\bar{R}]$:

$$[R][T][T]^{-1}[C] \stackrel{?}{=} [D] \qquad (29)$$

If the matrix resulting from the computation does not compare favorably with the original data matrix $[D]$ then the real matrix does not adequately span the factor space. This may be true because some of the test vectors are redundant, i.e., they represent the same cofactor.

Alternatively, the columns of matrix $[\bar{R}]$ may be n columns of the original data matrix termed "typical" columns. Such a combination would ascertain whether the data in the n columns adequately span the factor space. Those typical columns which do span the factor space will give a

satisfactory combination and are known as the "key combination set."[25]

Perhaps more intriguingly, once a satisfactory $[\bar{R}]$ is found, we can predict the value for d_{xk}, where x is a new row designee and k is one of the original column designees. The value for d_{xk} can be determined using the following equation:

$$d_{xk} = \sum_{m=1}^{m} r_{xm} \; c_{mk} \qquad (30)$$

where c_{mk} is the mth coefficient for the kth column designee of the combination based $[C]$ matrix, r_{xm} the corresponding value of cofactor m for row designee x. The $[C]$ matrix is obtained from TFA using equation 21. Assuming that the interactions or cofactors involved with designee x are the same as those in matrix $[\bar{R}]$, highly satisfactory predictions are possible.[21]

CHAPTER III

PROCEDURES OF TARGET FACTOR ANALYSIS

A. Steps in Target Factor Analysis

In this chapter we will describe the general procedures used in the target factor analysis of a typical chemical problem. We will make use of a computer program, FACTANAL, originally developed by Malinowski and coworkers and considerably modified by Howery and coworkers.[8]

In order to better illustrate the procedure, we will factor analyze a sample data matrix. As illustrated in Table 1, multiplying row matrix [R], a 7 x 3 (seven rows, three columns) matrix, by column matrix [C], a 3 x 6 (three rows, six columns) matrix, we generate data matrix [D], a 7 x 6 (seven rows, six columns) matrix, according to the following equation:

$$[R] \cdot [C] = [D] \qquad (31)$$

Matrix [D] is a pure three factor data matrix; each factor in matrix [D] is the product of a row cofactor of matrix [R] and a column cofactor of matrix [C].

The value of any data point, d_{ij}, in matrix [D] is:

$$d_{ij} = r_{i1}c_{1j} + r_{i2}c_{2j} + r_{i3}c_{3j} = \sum_{n}^{3} r_{in}c_{nj} \qquad (32)$$

Table I. Generation of sample data matrix from row matrix and column matrix

$$\begin{bmatrix} 40 & 15 & 70 \\ 30 & 35 & 45 \\ 80 & 85 & 100 \\ 70 & 20 & 10 \\ 20 & 95 & 0 \\ 10 & 60 & 65 \\ 8 & 90 & 30 \end{bmatrix} \times \begin{bmatrix} 0.9 & -0.1 & 0.0 & 0.6 & 0.4 & -0.3 \\ 0.1 & 0.2 & 0.5 & 0.0 & -.8 & 0.0 \\ 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \end{bmatrix} = \begin{bmatrix} 107.5 & 69.0 & 77.5 & 94.0 & 74.0 & 58.0 \\ 75.5 & 49.0 & 62.5 & 63.0 & 20.0 & 36.0 \\ 180.5 & 109.0 & 142.5 & 148.0 & 64.0 & 76.0 \\ 75.0 & 7.0 & 20.0 & 52.0 & 22.0 & -11.0 \\ 27.5 & 17.0 & 47.5 & 12.0 & -68.0 & -6.0 \\ 80.0 & 76.0 & 95.0 & 71.0 & 21.0 & 62.0 \\ 46.2 & 47.2 & 75.0 & 34.8 & -38.8 & 27.6 \end{bmatrix}$$

[R] [C] [D]

For example, the value of data point d_{23} for row 2, column 3 is calculated thusly:

$$d_{23} = r_{21}c_{13} + r_{22}c_{23} + r_{23}c_{23} = (30.0)\ (0.) + (35.5)\ (0.5) + (45.0)\ (1.0) = 62.5 \quad (33)$$

With actual problems, we usually have no idea of the individual row and column matrices. In a solute-solvent interaction problem, the matrices represent the representative solute and solvent interaction terms. It is these very matrices which we are interested in obtaining. The target transformation feature of TFA can provide the information necessary to generate these matrices.

Target factor analysis, as illustrated in Figure 1, involves the following steps:

1) Pretreatment - formation of a covariance or a correlation matrix [C] from a data matrix [D] of experimental data
2) Reproduction and factor determination -
   a) diagonalization of the correlation or covariance matrix and solution of the resulting eigenvalue problem to determine the eigenvectors associated with the factors which characterize the data and
   b) determination of the number of significant eigenvectors (abstract factors) needed to reproduce the data within experimental error

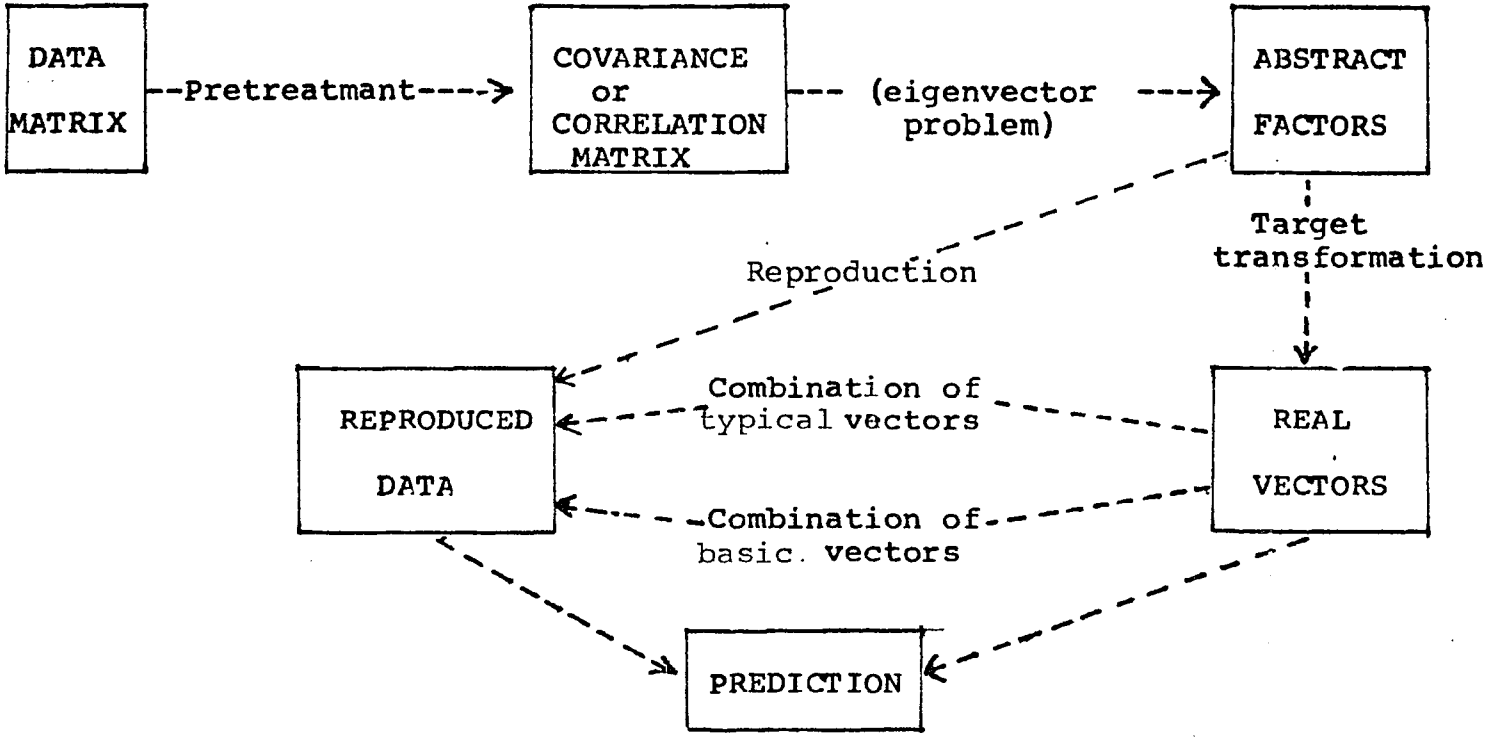
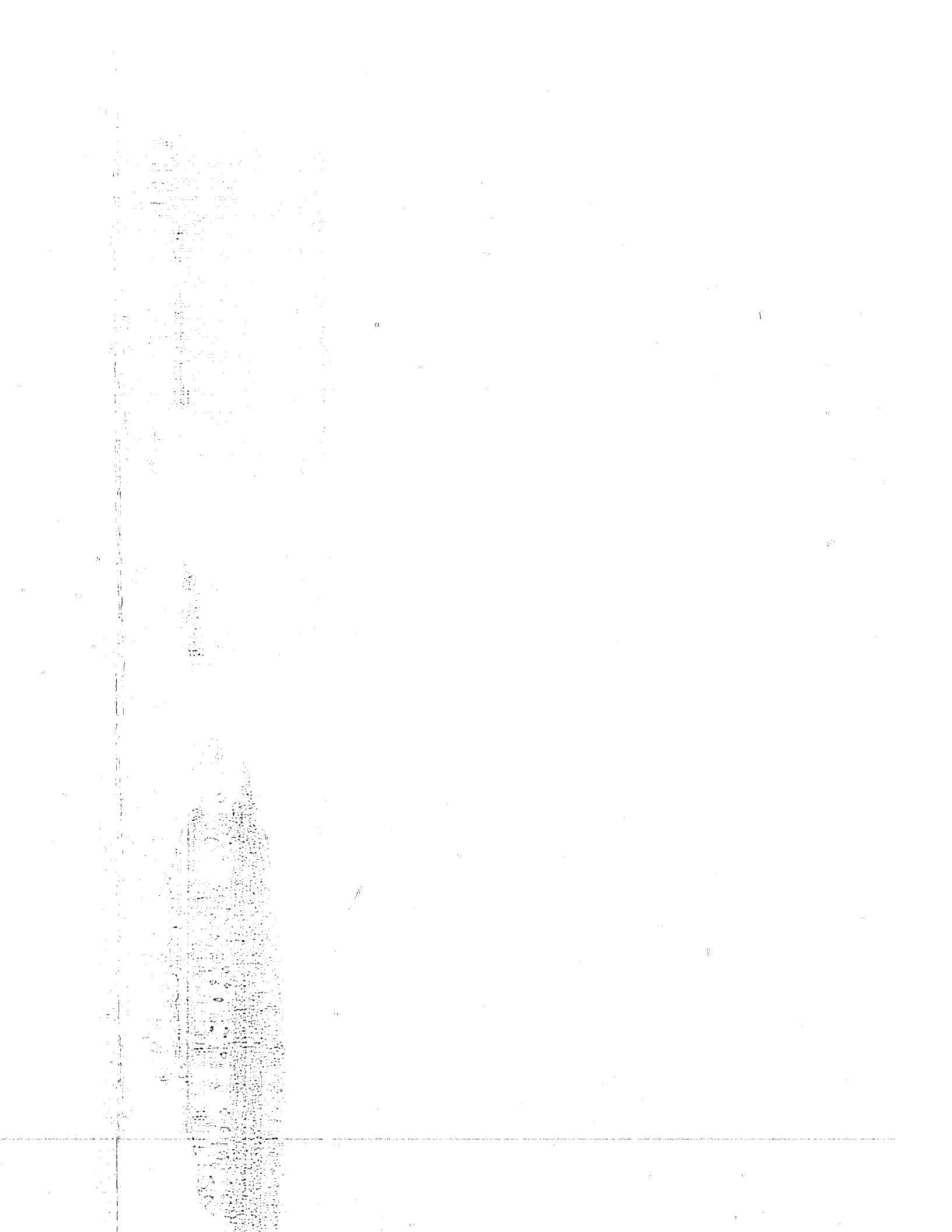


Figure 1 : The steps of target factor analysis



3) Target transformation - transformation of physically significant vectors to associate the abstract factors with real vectors

4) Combination - reproduction of the original matrix using sets of real vectors

5) Prediction of new data

We will now discuss each of these steps more fully using our model problem as an illustration.

1) Pretreatment:

Depending on the nature of the data, one may decide 1) to use the unnormalized data set and obtain a covariance matrix or, 2) to normalize the data set and consequently obtain a correlation matrix. Normalization is obtained by dividing each data point of the original matrix by the square root of the sum of squares for the column in which the data point appears. Normalization is necessary if the error in the data is proportional to the magnitude of the data or if the individual columns of the data matrix contain different units of measurement.[30] For our test data, covariance and correlation will provide the same information.

2) Reproduction:

We wish to determine the minimum number of abstract factors necessary to reproduce the original data. By eigenanalysis, the correlation or covariance matrix of the data matrix is diagonalized and a set of abstract eigenvalues related to the column designees is obtained.

Operation on the transpose of the original data matrix would produce a set of eigenvalues related to the row designees.

The complete set of c eigenvectors can be used to reproduce the original data exactly. However, in factor analysis, we wish to find a minimum number of factors or eigenvectors which would reproduce the original data to within some criteria. Most logically, in chemistry, the criteria would be experimental error. Thus, we attempt to reproduce the original data using successively smaller eigenvalues. First, only the largest eigenvector (which accounts for the most variance in the data) is used. Then, stepwise reproductions using the largest and second largest, then, the three largest, and so forth, are calculated till we succeed in reproducing the data satisfactorily. When this is done, we have determined the factor space of our data.

This step is illustrated in Table 2 for our test data. The table is a printout of the computer analysis for this step. In separate columns are listed for each factor 1)the eigenvalue associated with that factor, 2) the root mean square error (RMS) between the reproduced matrix using that number of factors and the original data matrix, 3)percentage of data points in the reproduced matrix with errors (when compared to the original data) within listed ranges and 4) the two largest error with their row and column locations. If the covariance option is used, values for

Table 2. Reproduction summary for sample problem

Factor	Eigenvalue	Rms error	Percentage of points within ranges <3.0	3.0-5.0	5.0-10.0	>10.0
1	1.99E 05	18.93	21.4	11.9	23.8	42.9
2	1.32E 04	8.77	38.1	16.7	21.4	23.8
3	4.02E 03	2.24E-04	100.0	0.0	0.0	0.0
4	2.06E-02	2.30E-04	100.0	0.0	0.0	0.0
5	1.32E-03	2.36E-04	100.0	0.0	0.0	0.0

Factor	Largest error	Row#	Column#	Second largest error	Row#	Column#
1	74.70	5	5	56.74	7	5
2	32.35	4	6	-28.72	4	1
3	.00	3	1	.00	4	2
4	.00	3	1	.00	3	2
5	.00	3	1	.00	3	2

Factor	Real error	Extracted error	Imbedded error	Indicator function
1	22.194	20.260	9.061	0.888
2	11.975	9.778	6.914	0.748
3	7.563E-03	5.348E-03	5.348E-03	8.404E-04
4	3.725E-02	2.151E-02	3.042E-02	9.314E-03
5	5.444E-02	2.223E-02	4.970E-02	5.444E-02

the Real, Extracted, and Imbedded errors and the Indicator function of Malinowski[43] are also tabulated.

For our test data (a true three factor space) we expect perfect agreement between the reproduction using three factors and our original data. As can be seen, three factors give a trivial rms error of .0002 with a largest error of .0005 probably due to computer round-off errors. The Malinowski Indicator function shows a minimum at three factors. The Real error at three factors is predicted to be .008, also insignificant. Indeed, we have confirmed that our data is a three factor space.

For most chemical problems, determination of the size of the factors space is not so clear. Consideration of the estimated experimental error, of the percentage of errors above experimental error and chemical intuition all play a role. The size of the largest errors may be important. If the largest errors are close enough to our idea of experimental error, then, though the percentage of errors above experimental may be large, the factor space may be pinned down. The Malinowski error functions may help in the decision. For most real problems the factor space is estimated to within 1 factor. Sometimes, the decision may be deferred until later in the analysis when other indications may point to a more definite determination.

Up to this point, we have accomplished, in general, a principal components analysis, the first step of an abstract factor analysis. The next steps involve target

transformation and real cofactor identification.

3) Target Transformation:

The heart of target factor analysis is the Least-squares target transformation facility, developed by Malinowski. Target testing allows one to relate physically and chemically significant parameters with the abstract vector cofactors of the space. In target transformation a suspected vector is tested. A least squares procedure minimizes, point by point, the difference between the test vector and the matrix of abstract eigenvectors obtained in step 2. The transformed best-fit vector is called the predicted vector. The overall point-by-point agreement between the test vector and the predicted vector determines whether the test vector represents one of the row cofactors of the space. How well they must agree before the transformation is considered successful is determined by one's knowledge of the test vector (i.e., its accuracy) as well as experience with target factor analysis.

Most vectors do not transform well; it is indeed a challenge finding a good test vector. However, even after one does obtain a test vector which transforms reasonably well, caution is necessary. A successful transformation means that the input vector is mathematically similar to some combination of the abstract mathematical cofactors. However, the vector may not have the chemical significance we have assigned to it. For example, is a molecular weight vector which transforms

well, testing the molecular weight or does it mathematically imitate a vector, the physical significance of which we have no idea?

Test vectors can include 1) physical vectors related to the overall physical properties of the row and column entities and 2) structural vectors based on detailed structural information of the entities. Examples of physical vectors include molecular weight, boiling point, refractive index, and dipole moment. Ingenuity plays an important role in developing structural vectors. Specific properties of entities based on even slight differences in structure can be tested. Examples include multiple bond uniqueness (where entities with a multiple bond are assigned a "1" and other entities are assigned a "0"), position of a particular functional group (i.e., hydroxyl groups) and number of carbon atoms.

Obviously, target transformation of an abstact factor, i.e., a column of the row matrix or a row of the column matrix (cofactors by definition) should test very well. This is illustrated for our sample problem in Table 3. As expected, agreement between the input vector, in this case, column 2 of [R], and the predicted vector is perfect. One data point of the input vector was purposely left blank (free-floated). Prediction of the free-floated value illustrates a novel application of the target transformation technique. If data points in a target vector are missing, the target transformation least squares scheme will predict appropriate values for the free-floated points.

Table 3. Target transformation of column 2 of row matrix R_1, one data point missing, (known value in parenthesis)

Row designee	Test vector	Predicted vector
1	15.0	15.0
2	35.0	35.0
3	(85.0)	85.0
4	20.0	20.0
5	95.0	95.0
6	60.0	60.0
7	90.0	90.0

A useful type of test vector is known as the uniqueness vector. By testing an input vector with a "1" for a particular row designee and "0"'s for all other row designees, one can determine if the particular row designee under test transforms to the abstract factors. If the uniqueness test for a particular row designee target transforms well, then that row designee uniquely represents the abstract factor. Most often the uniqueness test of a particular row designee does not test well. However, by studying the predicted vector much useful information can be obtained. For example, Table 4 illustrates the uniqueness test for row designee 6 from our sample problem. Whereas we have input a "1" for the data point corresponding to row designee 6, the predicted vector predicts only 0.37. Obviously, the uniqueness test for row designee 2 does not target transform well. Thus, row designee 6 does not adequately represent a cofactor of our sample problems. However, predicted points for some of the other row designees such as row designee 3 (=.20) are higher than expected. We interpret this to mean that designees with high predicted values correlate well with the tested designee. The uniqueness test tends to cluster similar designees. Thus, row designee 3 correlates well wtih row designee 6.

Table 5 summarizes these results for all seven row designees of the test problem. The uniqueness value for each row designee is listed with those designees which

Table 4. Uniqueness test for row designee 6, three factors

Original row designee	Test vector	Predicted vector
1	0.0	0.18
2	.0	.12
3	.0	.20
4	.0	-.31
5	.0	.09
6	1.0	.37
7	.0	.23

Table 5. Uniqueness test summary , three factors

ROW#	VALUE	CORRELATIONS ROW#(VALUE)[a]
1	0.335	3(0.311), 5(-.256), 6(0.175),
2	0.092	3(0.210),
3	0.507	1(0.311), 2(0.210), 4(0.243), 6(0.198),
4	0.707	3(0.243), 5(0.162), 6(-.313),
5	0.601	1(-.256), 4(0.162), 7(0.381),
6	0.369	1(0.175), 3(0.198), 4(-.313), 7(0.226),
7	0.388	5(0.381), 6(0.226),

[a] Correlations above |0.15| are listed.

correlate highly on that particular uniqueness test vector. Thus, row designee 4 is the most unique (value = .707) and correlates most with row designee 3 (.24), less with row designee 5 (.16) and negatively with row designee 6 (-.31).

Using these correlations obtained in the uniqueness test, we can construct new "correlation" vectors for testing via target transformation. These vectors test similarities that are apparent from the correlations. For example, a possible correlation vector based on the results of the uniqueness test for row designee 1 in Table 5 may have the form a value of "1" for row 1, 3 and 6 and "-1" for row 5 and "0"'s for all other rows. In essence, one is testing a vector based on "least squaring" onto the abstract cofactors matrix.

4) Combination:

Once we have found vectors which successfully target transform, we still must determine if we have found any sets of vectors which represent all of the abstract cofactors in our data space. It is also informative to know which vectors are redundant, i.e., represent the same cofactor and are therefore interchangeable. Do our sets of vectors adequately span the factor space? Will they, in combination, enable us to reproduce the original data satisfactorily?

The answer to these problems may be obtained in the combination step. Two types of combinations are possible.

We may take combinations of those target transformed vectors which we have found to satisfactorily transform unto the abstract matrix. Alternatively, we may take combinations of the "typical" vectors (the columns) of the original data matrix. Combinations of columns will help us determine those column entities which best represent the factor space and also those column entities which are most similar to each other.

Combinations of the real physical vectors that target transformed satisfactorily will enable us to determine the best set of parameters that represent the cofactors in the data space. Two sets of best sets are sought; one for the row entities, the other for the column entities. If we are able to adequately reproduce the original data using both sets, some information on the complete interactions involved may be obtained. Comparison with theory would then be possible.

Thus, a combination of the three columns, columns 1, 2 and 3, from the row matrix [R] reproduces the data matrix [D] with an rms error of only 0.07. By contrast, combinations of any two of these abstract cofactors (i.e., columns 1 and 2) with a random vector called POOR are unsatisfacory, resulting in an rms error of 45.57.

It is likely that more than one combination will give nearly equivalent results. Many vectors are redundant, representing the same abstract cofactors. Comparison of the best combinations (for example, the best 20 or 25

sets) obtained from a set of vectors would then show which vectors are substitutes for each other. This information, in the physical sense, can be highly informative. Additional information from further analysis of the best combinations obtained from a set of vectors may be obtained in what we call Pattern Tables. Such tables will be explained and illustrated in Chapters VI through VIII.

5) Prediction:

Once we have obtained good solutions to our problem, new data can be predicted according to equation 22 of Chapter II. One uses the key vectors involved in the best sets and the factor loading (column designee) matrix obtained in the combination step. Thus, new data often can be predicted with an accuracy equal to that furnished by the best solutions obtained in the combinations. Examples of successful predictions are illustrated in Chapters IV, V, VII, VIII, and IX.

The calculation of free-floated points in the target transformation step is another method of obtaining new data for unknown entities.

B. Introduction to FACTANAL, A Computer Program For Target Factor Analysis

During the course of these investigations, Malinowski's original TFA computer program[7] was expanded and modified considerably. The new program is now available from the Quantum Chemistry Program Exchange of Indiana University as Program #320.[8] Documentation was added to enable a user new to FACTANAL to understand and use the program. The control options were expanded to facilitate the routine analysis used in TFA such as combinations and target transformations. A simple test problem is included to illustrate the features available with the program.

The program, as available, is capable of handling 40 x 40 data matrices. However, expansion to matrices of larger sizes (eg., 60 x 60) is easily facilitated by modification of appropriate dimension statements.

FACTANAL is written in FORTRAN IV language, the computer language most familiar to chemists. It has been tested and used with both FORTRAN G and H compilers on both IBM 360 and IBM 370 systems. Options are available for transforming the original data matrix, ie., transpose and log of matrix as well as testing various functional formulas (square, reciprocal, square and exponential) of target vectors. It is possible to perform a TFA on both a data matrix (to test row designees) and its transpose (tests column designees) in one pass using an option

called INVPLU. Complete statistics including calculation of % errors, largest errors, average error, percent and number errors within ranges, Exner values and the Malinowski error functions, are calculated and tabulated automatically by FACTANAL for the reproduction step (c.f. Table 2). The uniqueness step is automatically run and tabulated in a convenient table with statistics (i.e., Exner values) (c.f. Table 5).

Other options automatically execute all possible combinations of typical vectors (i.e., columns of the data matrix) or alternatively all combinations of basic (real, physical) vectors with concise printout of results in convenient tables. Combinations of vectors are easily accomplished and retrieval of [R] matrices for prediction of new data is automatic. Borrowing a technique from pattern recognition, FACTANAL provides the option to run a K-nearest neighbors (K-NN) [66] analysis.

Contributions by this investigator to the program include:

1. Complete testing of program and optimization using FORTRAN G and H compilers.
2. Complete documentation with sample test problem.
3. Programming and testing of the following options:
   a. CONMAX - to execute TFA on selected rows and columns of a data matrix
   b. INV and INVPLU - to execute on data matrix and its transpose

c. UNIQUENESS TABLES

d. REPRODUCTION TABLES

e. Statistics including Exner function, Malinowski error functions, percent and number of errors within ranges, average (rms) error and largest errors

f. K-nearest neighbor analysis

These modifications have helped eliminate much of the preliminary drudge work in a typical TFA. With the above options, considerable information can be obtained after one or two executions of FACTANAL.

CHAPTER IV

SOME ASPECTS OF SOLUTE-SOLVENT INTERACTIONS APPLIED TO GAS CHROMATOGRAPHY

Introduction

The quantity directly measured in a gas liquid chromatography (GLC) experiment is the time it takes for a small sample of solute to pass through the separation column at a given flow rate of carrier gas and temperature and weight of stationary phase, i.e., the retention time. This depends on the partition coefficient for the solute-solvent pair in question. The partition coefficient, K, is defined[45-47] as the ratio of molar concentrations (c) of solute in the stationary (s) and mobile (m) phases at equilibrium:

$$K = \frac{c_s}{c_m} = \frac{n_s/v_s}{n_m/v_m} \qquad (34)$$

when n is the number of moles of solute and v is the volume of the phase in question. As long as the solute sample size is small enough, K can be assumed to be a constant.

The product of the retention time and the carrier

gas flow rate is the volume per unit time of gas swept through the column as the solute passes through. The specific retention volume, V_g, represents the volume (usually corrected to 0° C) of carrier gas required to elute one half the solute from a column containing one gram of solvent. It is related to K by

$$K = \frac{T\rho}{273.15} \quad Vg \tag{35}$$

where ρ is the density of the solvent or liquid phase at temperature T.

The net retention volume V_N is related to K by

$$V_N = K \; V_S \tag{36}$$

where V_S is the volume of the stationary phase. V_N is a function of stationary phase loading.

In gas solid chromatography, K may be defined in terms of the weight of the solid, W_s, or its surface area, A_s. Then V_s is replaced by W_s or A_s. The specific retention volume, V_g, is then:

$$V_g = \frac{V_n}{W_s} \frac{273}{T} \tag{37}$$

where V_s is corrected to 0° C.
At the column temperature, T:

$$V_g = \frac{K}{\rho_s} \frac{273}{T} \tag{38}$$

where the density of the solvent, ρ_s, is taken at T.

Under a given set of conditions Vn is a constant. It is characteristic of a given substance and can be used to identify it. Limitations arise from the inability

to reproduce a given set of constant conditions from one laboratory to another.

To overcome these difficulties, Kovats[48,49] introduced the retention idex, I_x. The retention index for a component x is:

$$I_x = 100\ n + 100 \cdot \frac{\log V_{n,x} - \log V_{N,n}}{\log V_{N,n+1} - \log V_{N,n}} \quad (39)$$

where $V_{N,x}$ is the retention volume of the sample component x, and $V_{N,n}$ and $V_{N,n+1}$ are the net retention volume for the two normal alkanes with carbon numbers n and n+1 which bracket the sample component. The normal straight chain hydrocarbons will have retention indices (RI) differing by multiples of 100 units. Two adjacent homologs in a series will have RI differing by 100 units.

The total change in free energy of mixing of a molecule from a gas phase to a liquid phase, ΔG_t^m, is often assumed to equal the sum of the individual free energy changes ΔG_i^m for each of the structural groups of the molecule:[50]

$$\Delta G_t^m = \sum \Delta G_i^m \quad (40)$$

It can also be shown,[45,50] rigorously for regular solutions and empirically for other types of solutions, that

$$\Delta G_t^m = \sum U_i\ V_i \quad (41)$$

where U_i and V_i are solute and solvent cofactors each characteristic of a specific type of interaction.

Since Log V_N is related to the free energy change,

$$\ln V_N = \ln K + \ln V_s$$

$$= \frac{\Delta G_t^m}{RT} + \ln V_s \qquad (42)$$

Then it is logical to assume that Kovats retention index, I_x, is also related to the free energy change. Therefore,

$$I \propto \sum_i U_i \cdot V_i \qquad (43)$$

An equation of this type was postulated originally by Rohrschneider.[51]

Thus we would expect GLC retention index data to be ideally suited to target factor analysis. In the following chapters we will describe the results of factor analysis on several of these real data matrices.

B. Role of the Solvent

Chromatographers have often tried to classify solvents on the basis of polarity. The solubility parameter, δ, derived from Hildebrand Solubility Theory, might be used as a measure of polarity in that its value generally increases with solutes that are called polar. The solubility parameter is defined[45,52] by

$$\delta_i = \left[\frac{\Delta E_i^v}{\bar{V}_i}\right]^{1/2} \qquad (44)$$

where δ_i measures the intermolecular energy of compound i per unit volume of pure liquid, ΔE_i^v is the net energy of vaporization, and $\bar{V}_i$ is the molar volume of the liquid.

The total molar energy of mixing of solute i in solvent j, ΔE^m_{ij}, for regular solutions can be shown[52] to be:

$$\Delta E^m_{ij} = \bar{V}_i (\delta_i - \delta_j)^2 \qquad (45)$$

Thus, deviations from ideality in mixing are related to the intermolecular forces in the pure liquids i and j as measured by the solubility parameters, δ_i and δ_j.

The solubility parameter theory is based on the assumption that only dispersion intermolecular forces are involved. Thus for mixtures of polar and non-polar molecules where specific interactions are important, the theory is generally less accurate.

Recently the solubility parameter has been expanded

to included individual parameters for each specific interaction. Keller, Karger and Snyder[52] summarize this by the equation:

$$\delta_{ij}^2 = \delta_{d_i}\ \delta_{d_j} + \delta_{p_i}\delta_{p_j} + \delta_{h_j}\delta_{h_i} \qquad (46)$$

where δ_{ij}^2, the total solubility parameter of solute i in solvent j is related to the sum of the product of the dispersive force contributions δ_d, polar contributions, δ_p, and hydrogen-bonding contributions, δ_h. δ^2, propertional to E, is thus shown to be equivalent to a sum of product terms; each term is associated with a specific interaction and each δ_i is associated with a particular entity. This is exactly the type of equation which lends itself to factor analysis.

The polar contribution δ_p, describes dipole-dipole and dipole-induced dipole interactions without proton sharing and may be further divided into an induction term, δ_{in} and an orientation term, δ_o. The acid-base term can also be divided into a proton acceptor term δ_a and a proton donor term δ_b. With these modifications, the above equation is shown by Keller[50,52] to be:

$$\delta^2 = \delta_d^2 + \delta_o^2 + 2\delta_{in}\delta_d + 2\delta_a\delta_b \qquad (47)$$

Care, however, is needed in applying this equation. The equation extends solubility parameter theory to solutions other than regular solutions where ΔS^m, the entropy of mixing, can not be ignored. In addition, important

parameters, such as, molecular size and shape, are ignored.

Another polarity scale, Rohrschneider's P* scale[53], measures the ability of the stationary phase to induce a dipole moment in butadiene. This scale is thus related to the dipole moment of the stationary phase. If a particular solute-solvent pair were to exhibit another type of interaction i.e., hydrogen bonding, the P* scale is not able to predict the correct polarity.

A more elaborate system proposed by Rohrschneider[51,54] is based on the splitting of the total polarity into various interaction forces. Five solutes are used as probes to characterize the liquid phase. Each solute is intended to measure a different type of solute-solvent interaction. The scale is developed using ΔI, the retention index differences between a polar liquid phase and squalane, the standard non-polar phase. ΔI represents the specific interactions forces (dipole-dipole, induction and other forces) that occur on the polar phase.

Rohrschneider characterized each interaction as the product of two factors: one dependent only on the solute, U_i, and the other on the solvent, V_i. For n interactions:

$$\Delta I = U_1V_1 + U_2V_2 + \ldots + U_nV_n = \sum_i^n U_iV_i \qquad (48)$$

For example, the ΔI for solute probe benzene measures dispersion forces and π-bonding; that for ethanol measures hydrogen bonding.

McReynolds[55] proposed a system similar to Rohrschneider's. However, he expanded the solute probe list to ten compounds and he substituted butanol for ethanol. McReynold's felt that Rohrschneider's five standard solutes did not span the entire range of interactions involved in chromatography.

Has McReynolds addition of five more solutes and substitution of one of them helped in spanning the full range of interactions? Kaplan et al.[56] feel that Rohrschneider's five solutes are enough to characterize all the interaction relations formed and indeed enough to calculate the retention index on new solutes, while Hartkopf et al.[60] identified only four functional probes as necessary. At least one manufacturer recommends that the first seven McReynold's probes should be sufficient in practice.[57]

Comparing the McReynolds or alternatively the Rohrschneider constants for the various solvents available in GLC one can notice that many of them exhibit the same apparent polarity, i.e., the solvents are redundant. Several authors have pointed out the necessity for standardization of stationary phases.[58,59] A small number of standard phases selected from the more than 700 used in the literature should be able to accommodate most separation problems. However, before such work can be completed a better understanding of the contribution of the stationary phases to retention is needed.

Many mathematical-statistical methods have been used to find redundant phases: target factor analysis, principal component analysis (a form of abstract factor analysis),[61-63] information theory,[64] pattern recognition,[65] K-nearest neighbor analysis (a form of pattern recognition)[66] and numerical taxonomy.[59,68] Most of these methods have been applied to find the sets of stationary phases that will provide the necessary separation of various compounds.

Other researchers have proposed thermodynamically-based polarity scales. Novak, et al.,[68] have proposed the Gibbs partial molar free energy of a methelene group ($\Delta G^{E}_{-CH_2}$) as a good criterion for the polarity of the stationary phase. Risby and coworkers[69] propose using the molar enthalpy of evaporation (ΔH^{S}_{V}) of the solute from solution and Figgins et al.[70] propose using the Gibb's free energy of mixing.

The polarity scales proposed by workers such as Novak, Risby, Figgins, and Rohrschneider (P* scale) are quite different from the polarity probes of McReynolds and Rohrschneider. McReynolds and Rohrschneider attempted to provide more than a single scale for the elusive "polarity" chromatographers use as a separation scale. The convenience of a single scale is obvious and indeed some researchers[71] condense the ten McReynolds constants to a single function: the sum of the ten constants.

Lowry et al.[71] compare many of these polarity measurements. They conclude that the sum of McReynolds constants

for the first five solute probes is perhaps the best measure of the "polarity." However, they point out that if other properties of the solute-solvent interactions are important then that polarity criterion is insufficient for stationary phase selection.

C. Role of the Solute

Much work is reported in the literature to relate the structure of the solute with its retention on a stationary phase. We report only some of the more general approaches. Takacs and coworkers[56,72] in a series of papers have attempted to link the retention index of a solute with molecular terms, I_m, and interaction terms, I_i. Defining the molecular term as a sum of atomic terms, I_a, and bond terms, I_b, the retention of solute x on a solvent, I_x, is:

$$I_x = I_m + I_i = I_a + I_b + I_i \qquad (49)$$

The interaction term I_i has been related to Rohrschneider's constants.[56]

Other workers have also correlated molecular structure and retention index. For example, Castello and D'Amato[73] have correlated the retention of alkyl iodides with their boiling point and molecular volume. They found that molecular volume effect offsets the boiling point effect. Castello and coworkers[74] found that the retention index of branched chain parafins depends on many physical properties of the solutes, especially the molecular volume. Streuller and Orloff[75] using multiple regression analysis, related log V_g with molecular weight, structural considerations (steric factors) and π-electron bonding of the solute. Most recently, Chretian and Dubois[76], in an extension of Takacs work, have applied

topological analysis to solute analysis.

Target factor analysis has been successfully applied by Howery and coworkers to test many molecular structure and physical properties of solutes. Data matrices involving hydrocarbons,[24] esters,[26] alcohols,[21] ethers[25] and other functional groups[20,23,27] have been studied.

Parameters which target transformed well and are associated with the solute cofactors in several of these problems included molecular weight[21, 23-25], molar refraction[21,23-25], molar heat of vaporization[20,21-23,24], boiling point[21,25,27], and carbon number.[21,25] Other vectors found important included unsaturation uniqueness for alcohols[21] and ethers[25], logarithm of vapor pressure and hydroxyl group position for alcohols,[21] and triple bond uniqueness, aromatic uniqueness, molar volume, heat capacity and heat of combustion for hydrocarbons.[24] Vectors which were found important in the multifunctional group problems included: polarizability,[20] dipole moment, [20,23] Van der Waal's a and b constants,[23] density[20] and dipole moment squared.[20]

Tentative values for the carbon number equivalence of various functionalities were assigned in the alcohol and hydrocarbon studies.[21,24]

Boiling point (^{o}K) squared was found the dominant vector for the ethers.[25] Klages[77] has shown the square of the boiling point to be an additive molecular property. Recently, Bach et al,[78] have related the log of the retention

time to the square of the boiling point of a solute, especially for homologous series of solutes.

From Equation 43, we expect the free energy of vaporization, and consequently, the heat of vaporization, and entropy of vaporization, to be possible solute cofactors. In addition, according to the classical equations for solute-solvent interactions[45,79] (see Appendix B), the molar volume, electron polarizability, dipole moment, and the square of the dipole moment should also be possible solute cofactors. The electron polarizability is related to the refractive index, and especially to the molar refraction of the solute. In view of these results and the compatibility of equations 46 and 47 with factor analysis, the results of the target factor analysis studies undertaken by Howery and his coworkers have a solid basis in theory.

Utilizing the most intriguing aspect of target factor analysis, the predictive capability, as described in equation 22, Weiner and Howery,[21,23] were able to predict the retention index for unknown solutes in both the alcohol study and in the study of Rohrschneider's data matrix (a multifunctionality problem) with very good results. Thus, they were able to demonstrate the practicality of the technique quite dramatically.

Selzer and Howery[25] in the ether study demonstrated how combinations of both solute and solvent typical (i.e.,

row and column) vectors gave reproductions with average errors within experimental. The best combination of physical vectors gave average errors of only twice the estimated error.

D. Summary

The theory of gas liquid chromatography is not complete. Explanations of the separation process are inadequate and are theoretically valid only for the simplest of compounds. Some attempts have been made to accomplish some understanding of the interactions involved and their significance. Some advances have been made for the solute part of the process. However, there is practically no information available for the solvent part. In the following chapters, we will attempt to provide some new insight into the solvent involvement in GLC as well as some further insight into the solute cofactors.

CHAPTER V

TARGET FACTOR ANALYSIS OF GAS LIQUID CHROMATOGRAPHIC RETENTION INDICES OF WELL CHARACTERIZED, MONOMERIC STATIONARY PHASES AND A VARIETY OF SOLUTES

A. Introduction

Zielinski and Martire determined the retention of 49 solutes on 7 stationary-phase solvents.[79] Their purpose was to study the effect of the liquid phase functionality on solute retention and selectivity. The liquid phases studied were well defined, monomeric and of similar chain length differing only by a single monofunctional group. The chain lengths were kept constant to eliminate any chain length effects on the retention indices.

They found that the retention index of the normal chain solutes could best be described by the equation

$$I_Z = 100\ n + \Delta I^X_{HDA} + \Delta^X_Y \qquad (50)$$

where I_Z is the solute retention index of liquid phase Z, n is the number of solute carbon atoms, ΔI^X_{HDA} is the contribution of the solute functional group X to the retention index on a straight chain alkane stationary phase solvent HDA, and Δ^X_Y is the retention index dispersion

increment attributed to the solvent functional group Y for a given solute group.

As rationalized in Chapter IV, GLC retention index data should provide valid TFA results. Previous TFA problems were too complicated and poorly defined to identify the solvent cofactors. Zielinski and Martire's data, however, gives us a unique chance to obtain the solvent cofactors. Furthermore, they provide a model (Equation 1) which gives valuable hints for TFA.

The data matrix was obtained from Reference 79. Zielinski and Martire estimate the experimental error to be less than ± 1 RI units. Approximately 12% of the data points were determined by extrapolation and therefore, probably have errors larger than ± 1. The data matrix is given in Table 6 with the 49 solute names. The seven liquid phases and their designations are listed in Table 7.

The solutes belong to three general classifications incorporating 16 branched and normal chain alkanes, 17 alkenes and 16 branched and normal alkyl halides, (chlorides, bromides and iodides).

Table 6 Retention Indices (I) at 45.0 °C

Solute	Solvent						
	HDA	HDE	HDC	HDB	HDI	DOE	DOTE
A. Normal chain							
1. *n*-Pentane	500	500	500	500	500	500	500
2. *n*-Hexane	600	600	600	600	600	600	600
3. *n*-Heptane	700	700	700	700	700	700	700
4. 1-Pentene	481	483	487	488	489	487	490
5. 1-Hexene	582	585	590	591	591	589	591
6. 1-Heptene	681	684	690	691	690	688	690
7. 1-Chloropropane	511	521	546	548	549	537	548
8. 1-Chlorobutane	613	626	649	651	651	638	648
9. 1-Chloropentane	715	726	751	752	752	739	750
10. Bromoethane	486	498	525	528	530	518	528
11. 1-Bromopropane	589	602	629	631	634	617	629
12. 1-Bromobutane	691	704	730	733	735	717	730
13. Iodoethane	585	599	626	632	640	615	631
14. 1-Iodopropane	687	699	726	732	739	715	730
B. Cis/trans alkenes							
15. *trans*-2-Pentene	501	503	506	507	508	506	509
16. *cis*-2-Pentene	505	507	511	513	513	510	514
17. *trans*-2-Hexene	601	605	610	611	612	608	610
18. *cis*-2-Hexene	602	605	610	611	611	608	611
19. *trans*-2-Heptene	697	700	705	705	705	704	705
20. *cis*-2-Heptene	701	704	710	710	711	707	710
21. *trans*-3-Heptene	686	689	692	693	693	693	693
22. *cis*-3-Heptene	688	692	696	697	693	695	696
C. Branched chain							
23. 2-Methylbutane	473	471	472	471	470	474	471
24. 2-Methylpentane	570	569	569	569	567	569	569
25. 3-Methylpentane	583	583	583	583	582	582	583
26. 2,2-Dimethylbutane	535	534	534	534	531	535	535
27. 2,3-Dimethylbutane	565	565	565	565	564	566	566
28. 2-Methylhexane	666	666	666	666	665	667	666
29. 3-Methylhexane	676	675	675	675	674	675	675
30. 3-Ethylpentane	682	683	684	684	684	683	684
31. 2,2-Dimethylpentane	625	625	625	624	622	625	624
32. 2,3-Dimethylpentane	669	669	670	669	670	669	669
33. 2,4-Dimethylpentane	630	630	629	628	626	630	629
34. 3,3-Dimethylpentane	654	655	655	656	655	655	655
35. 2,2,3-Trimethylbutane	636	636	637	636	634	636	637
36. 4-Methyl-1-pentene	549	551	556	556	555	555	557
37. 2-Methyl-2-pentene	579	582	588	589	588	587	588
38. 3,3-Dimethyl-1-butene	508	508	511	512	509	513	514
39. 2-Methyl-1-hexene	676	680	687	687	687	684	686
40. 5-Methyl-1-hexene	649	653	658	658	657	657	658
41. 4,4-Dimethyl-1-pentene	604	607	611	611	608	611	611
42. 2-Chloropropane	468	478	502	504	501	494	503
43. 2-Chlorobutane	577	587	611	612	610	601	609
44. 1-Chloro-2-methylpropane	582	594	616	617	617	607	615
45. 2-Chloro-2-methylpropane	515	524	547	547	543	534	545
46. 2-Bromopropane	548	559	586	588	589	576	585
47. 2-Bromobutane	654	666	692	694	694	679	691
48. 1-Bromo-2-methylpropane	660	671	696	699	699	685	696
49. 2-Bromo-2-methylpropane	591	602	628	629	627	616	625

Table 7. Uniqueness values with correlations for solvents

Solvent Name	Designation	Uniqueness test value	Correlates with solvents[a]
n-heptadecane	HDA	0.65	HDE(.44)
1-hexadecene	HDE	.33	HDA(.44)
1-hexadecyl chloride	HDC	.36	HDB(.26), DOE(.29), DOT (.26)
1-hexadecyl bromide	HDB	.25	HDC(.26),HDI,DOE,DOT
1-hexadecyl iodide	HDI	.90	HDB
di-n-octyl ether	DOE	.30	HDC(.29),HDB,DOT
di-n-octyl thioether	DOT	.23	HDC(.26),HDB,DOE

[a] Other solvents with predicted values greater than 0.15 are listed. Predicted values greater than 0.24 are given in parenthesis.

B. Factor Determination

The first step in a factor analysis is to determine the number of factors that make up the data space. Results for the first five stages of reproduction are listed in Table 8. Both correlation and covariance give the same results.

As shown in Table 8, reproduction of the data matrix using three factors results in a root mean square (rms) error between the predicted matrix and original data matrix of 0.55 RI units with the largest error, 2.95. Only 8.2% ofthe total number of predicted points have errors greater than ± 1 RI units. A three factor model seems likely based on experimental error and consistent with Zielinski and Martire's analysis as embodied in Equation 1.

Table 8 also lists the value of the several functions based on the theory of errors for abstract factor analysis by Malinowski[44]. At a factor space of three, the real error (RE) is estimated at 1.08 RI units, very close to Martire's error estimate. The imbedded error function (IE) shows a slight decrease in going from a two factor space (0.73) to a three factor space (0.71). Theoretically, the IE function should show a minimum at the factor space. If the error is not fairly uniform throughout the data, the IE will not show a minimum. However, if the decrease in the IE function shows a relatively small decrease between two factors then the

Table 8. Factor determination

Number of factors	rms error	Largest error	Percentage errors $>\pm 1$	Percentage errors $>\pm 2$
1	5.99	22.2	98.3	83.7
2	.87	5.1	18.3	2.0
3	.56	2.9	8.2	0.6
4	.35	1.7	1.7	0.0

Number of factors	Real error (RE)	Extracted error (XE)	Imbedded error (IE)	Indicator function (IND)
1	6.94	6.43	2.62	0.193
2	1.36	1.15	.73	.054
3	1.08	.82	.71	.068
4	.88	.57	.66	.097
5	.72	.39	.61	.181

factor space may be within those factors. According to Malinowski's theory our data is a two factor space. The indicator function (IND) also shows a minimum at two factors. However, the RE for two factors is overly large, being 1.36 RI units.

We conclude from the theory of errors that the data matrix is either a two or three factor space. In the following analysis, we will assume a three factor space.

C. Uniqueness Test

The usual second step in a target FA is termed the uniqueness step. A vector consisting of a "1" for the particular designee under test and "0"'s for the other designees is target transformed. Any solute or solvent which exhibit particularly unique interactions should score high on the uniqueness test. If sets of solutes or solvents correlate, i.e., show similar interactions in the data space, they will score mutually high on the uniqueness test for the set. Any solutes or solvents which exhibit similar properties in the factor space will score high on the uniqueness test for their complements.

Listed in Table 9 is the result of the uniqueness test for solvent HDA. The predicted value for HDA of 0.65 indicates some uniqueness. Values for the other solvents are, as expected, low except for the solvent most similar to HDA, HDE (predicted value = 0.44). This test brings out some common property for solvents HDA and HDE.

Table 7 summarizes the results of this step for all the solvents listing all the uniqueness values and the most prominent correlations for each solvents. Solvent HDI, due to the iodide functionality, is the most unique solvent (predicted value, 0.90). HDC correlates highly with HDB and the ethers DOE and DOTE. The only solvents without functional groups (HDA and HDE) also correlate highly.

Table 9. Uniqueness test for solvent HDA

Solvent[a]	Test vector	Predicted vector
HDA	1.0	0.65
HDE	0.0	.44
HDC	.0	-.09
HDB	.0	-.09
HDI	.0	-.01
DOE	.0	.14
DOT	.0	-.04

[a] See Table 7 for solvent designations.

As shown in Table 10, for thirteen representative solutes, all the solutes are relatively non-unique. However, iodoethane and iodopropane, do show some uniqueness (values, .33 and .30 respectively) and correlate with each other indicating the special role of the iodine atom.

Table 10. Uniqueness values for representative solutes[a]

Solute		Uniqueness value
Name	Number	
hexane	2	0.04
1-hexene	5	.02
1-chloropropane	7	.06
2-chloro-2-methyl propane	45	.12
bromoethane	10	.08
1-bromo-2-methyl propane	48	.06
iodoethane	13	.33
1-iodopropane	14	.30
trans-2-pentane	15	.02
cis-2-pentene	16	.02
3-ethylpentane	30	.05
2,4-dimethylpentane	33	.05
3,3-dimethyl-1-butene	34	.05

[a] Solutes iodoethane and 1-iodopropane are the only solutes which correlate

D. Functional Group Effect on Reproduction

One of the questions we can answer with TFA is whether a particular class of solutes or solvents has an effect on the factors in the problem. Is a particular functional group or class associated with one of the factors (or interactions)? To investigate this, we tested the effect on the factor space of removing a particular solvent or solute class.

Table 11 shows the results of this test. Removal of solvent HDA, HDB or HDE one at a time does not result in a significant change in the average rms error of the reproduction or in the percentage of errors exceeding ± 1 RI units. Thus, for the solvents it is apparent that the functional interactions exhibited by solvents HDA, HDE and HDB are not unique but rather are adequately represented in the remaining solvents.

For example, HDC adequately represents the halogen functionality if HCB is removed. Removal of either solvent HDI or DOE does show some significant change, not enough, however, to alter the factor space. For HDI this is probably related to its high uniqueness value. For DOE the reason is less apparent especailly since we should expect DOT to exhibit the same functionality as DOE, though to a lesser degree. The removal of DOE is more important than the removal of the fairly equivalent thioether DOT.

As shown in Table 11 for the solutes, removal of the

Table 11. Effect on reproduction of removal of solute classes and of single solvents, 3 factors

Solute class removed	Solutes remaining	rms error	Percent errors >±1	Largest error
branched chain	22	.54	6.5	2.5
normal chain and cis-trans alkenes	27	.49	5.3	2.4
alkenes	32	.53	7.5	3.1
halides	33	.40	4.3	1.5
chlorides	42	.52	8.1	2.4
bromides	41	.49	7.0	1.8
iodides	47	.55	8.5	2.6
cis-trans alkenes	41	.56	8.4	3.0
cis alkenes	45	.56	8.5	3.1
trans alkenes	45	.55	9.5	3.0

Solvent removed[a]	rms error	Percent errors >±1	Largest error
HDA	.55	8.5	3.0
HDE	.55	8.1	2.8
HDC	.51	6.8	2.2
HDB	.57	10.2	2.9
HDI	.47	5.8	1.8
DOE	.44	4.4	2.4
DOT	.50	6.0	3.1
Complete data matrix	.56	8.2	3.0

[a] See Table 7 for solvent designations.

group of halide-containing-solutes results in a significant change in the factor space. Removal of the branched chain or cis/trans normal alkene pairs does little to alter the number of factors. This supports Zielinski and Martire's conclusions that the functional groups studied here display no selectivity for the separation of cis-trans normal alkenes pairs or for the branched alkanes. In addition, removal of all branched chain solutes shows no significant change in the factor space. Zielinski and Martire also concluded that these monofunctional liquid phases exhibit minimal selectivity toward chain branching in branched alkenes and haloalkanes.

E. Combination of Typical Vectors

The combination step enables us to determine the set of typical solute vectors and the set of typical solvent vectors that best represent the abstract cofactors for this data space. Besides giving us an idea of the most important solutes and solvents from an interaction viewpoint, this step enables us to further extend the results obtained in the uniqueness and functional group analysis steps.

The results for the combination of typical solute and solvent vectors are listed in Table 12. Thus, the combination of solutes pentane, iodopropane and 2,2-dimethylpentane reproduces the original data with an rms error of only 0.65. This compares very favorably with the rms error of 0.56 for the abstract reproduction using three factors (Table 8). The largest error between the reproduced data matrix and the original data matrix is only 2.6. We conclude that the three abstract solute cofactors may be related to a straight chain hydrocarbon (as illustrated by pentane), a halide-containing hydrocarbon (iodopropane) and a branched hydrocarbon (2,2 dimethylpentane).

Listed in Table 13 are the fifteen best solute vector combinations. It is apparent that some solutes can substitute very well for each other. The three best combinations differ only by the substitution of pentane, hexane and heptane for each other. Each of these solutes

Table 12. Summary for solvents and solutes of combinations involving typical and basic vectors, three factors

Combination	rms error[a]	Largest error[a]	Vectors in best set[b]	Correlations from best sets[b]
Typical solvent vectors	0.61	4.3	HDA, HDB, HDI	HDA = HDE
Typical solute vectors	.65	2.6	pentane (or hexane), iodopropane, 2,4-dimethylpentane	pentane = hexane = heptane; iodoethane = iodopropane; 2-methylhexane = 2,2-dimethylpentane
Basic solvent vectors	.66	2.9	UNI, LND, DCV	LND = DX = DCV; DM = DCV (UNI in best 30 sets)
Basic solute vectors	2.74	8.9	NA, CN, DP	DX = DAU = DXU = DEV = DP

[a] Abstract reproduction, 3 factors, gives rms error of 0.55 and largest error of 2.9.

[b] See Tables 7,15 and 17, respectively, for designations.

Table 13. Best combinations of typical solute vectors

Combination[a]			rms error
1	14	31	0.651
2	14	31	.651
3	14	31	.651
13	31	43	.662
14	31	43	.669
1	11	42	.670
2	11	42	.670
3	11	42	.670
13	28	43	.673
18	33	42	.674
11	30	42	.677
11	18	42	.677
1	13	43	.679
2	13	43	.679
3	13	43	.679

[a] See table 6 for solute designations.

represents the associated abstract cofactor equally well. Similarly, by comparing the fourth and fifth combination we can see that iodethane is a valid substitute for iodopropane. The dominant functionality for the cofactor associated with these solutes is the iodine group. These relationships are summarized in the last column of Table 12.

In general, the best solute combination set requires an alkane or alkene, a halide, and a branched solute. Two other observations are worth noting. First, though 2,2-dimethylpentane and 2,4-dimethylpentane are represented in the best combination sets, 2,3-dimethylpentane is not. Secondly, vectors for branched chlorides, such as 2-chloropropane and 2-chlorobutane are more representative cofactors than the normal chain chlorides.

For the solvents, combination of solvents HDA, HDB and HDI reproduces the original data matrix with an rms error of only 0.61 and a largest error of 4.3. As is apparent from Table 14, which lists the best solvent vector combinations, solvents HDA and HDE represent the same cofactor so well that substitution of HDE for HDA gives very similar results. The uniqueness of solvent HDI is apparent in that it is represented in 8 of the 10 best combination sets. These results, then, corroborate the results obtained in the uniqueness tests and functional group effect on reproductions step.

Table 14. Best combinations of typical solvent vectors

Combination[a]			rms error
HDA	HDB	HDI	0.609
NDA	HDC	HDI	.626
HDE	HDB	HDI	.652
HDC	HDC	HDI	.671
HDA	HDI	DOT	.801
HDA	HDC	HDB	.814
HDA	HDI	DOE	.824
HDE	HDI	DOT	.852
HDE	HDC	HDB	.855
HDE	HDI	DOE	.971

[a] See Table 7 for solvent designations.

F. MODEL BUILDING VIA TARGET TESTING

The next step is the target transformation step in which we attempt to identify the abstract factors with physically meaningful vectors. The target transformation capability allows one to test if a particular basic vector represents one of the cofactors in the data space. Good agreement between the predicted and test vectors means the vector most probably represents one of the cofactors of the space. However, one must keep in mind that this agreement is only a mathematical one; we can only infer the true physical meaning of the cofactor.

Tables 15 and 17 list for solutes and solvents, respectively, representative real vectors some of which tested well and some of which did not. Halogen uniqueness, as evidenced for both solutes and solvents in the uniqueness test and in the removal of functional groups procedure of the previous section, should be a good test vector for both the solutes and solvents. For the solutes, the halogen uniqueness vector target transforms very well, indicating that one of the abstract solute cofactors is related to some interaction associated with the halogen-containing solutes. For the solvents, however, halogen uniqueness target transforms poorly, as listed in Table 17.

As listed in Table 15, the solute boiling point vector target transformed only fairly, meaning that some of the predicted data points were poorly predicted.

Table 15.Summary of selected target transformations for solute test vectors

Vector name	Designation	Example values[a]	Points[b]	Qualitative evaluation[c]
boiling point	BP	312,354	45	f
carbon number	CN	ff,ff	6	v
density	DE	1.46,0.62	45	g
dipole moment	DM	2.03,0	31	g
enthalpy of vaporization	EN	ff,7.87	19	g
molar refraction	MR	19,34.6	49	g
molecular weight	MW	109,100	49	g
refractive index	ND	1.42,1.38	45	p
surface tension	ST	ff,18.2	20	f
square BP	SBP	97063,125100	45	v
square DE	SDE	2.13,0.4525	45	g
square EN	SEN	ff,61.97	19	f
square DM	SDM	4.12,0	31	f
square d(boiling point)/dT	SDT	ff..0019	32	f

(continued)

Table 15(continued)

main chain length	MC	2,5	49	f
number atoms	NA	8,23	49	v
number branches	NB	0,2	49	p
number carbon atoms	NC	2,7	49	g
number hydrogen atoms	NH	5,16	49	g
unity	UNI	1,1	49	g
alkene uniqueness	UE	0,0	49	p
halogen uniqueness	UX	1,0	49	v
iodine uniqueness	UI	0,0	49	g
solute group [e]	DX	6.5,0	49	v
ΔI_y [e]	DIY	289,ff	14	p
$\Delta {}^{x}_{y}$ on HDE [e]	DHU	12,0	47	g
$\Delta {}^{x}_{y}$ on halides [e]	DXU	41.0	47	g
$\Delta {}^{x}_{y}$ on ethers [e]	DEU	28,0	47	f

(continued)

Table 15 (continued)

[a] Example test points are given for solutes bromoethane and 2,4-dimethylpentane; ff are free-floated values.

[b] Number of test points on test vector.

[c] v-very good agreement between test and predicted vectors; g-good agreement; f-fair agreement (several points poorly predicted or pattern only predicted); p-poor agreement.

[d] Carbon number vector obtained by assigning values to straight chain alkanes and alkenes (solutes 1-6) and free-floating all other values; see text.

[e] Vectors DX, DIY and the average solute functionality vectors taken from Tables III, IV and V, restpectively, of reference 79.

Table 16. Details of selected target transformations for solute vectors, 3 factors

Solute number[a]	Boiling point squared $\times 10^{-4}$ (SBP)		Alkene uniqueness (UE)		Solute group δ_x (DX)	
	Test[b]	Predicted	Test	Predicted	Test	Predicted
2	11.6	11.7	0.0	0.5	0.0	0.0
5	11.3	11.5	1.0	.4	1.6	1.5
7	10.2	10.4	.0	.1	5.8	6.2
45	10.5	10.4	.0	.1	5.8	5.3
10	9.7	9.9	.0	.0	6.5	7.0
48		13.3	.0	.2	6.5	6.2
13	11.9	11.9	.0	.1	7.1	7.1
14	14.1	13.9	.0	.2	7.1	6.9
15	9.6	9.9	1.0	.3	1.6	1.1
16	9.7	10.0	1.0	.3	1.6	1.3
30	13.4	13.4	.0	.5	.0	.4
33	12.5	12.3	.0	.5	.0	.0
34	9.9	10.0	1.0	.3	1.6	1.1

	Δ_y^x on HDE (DHU)		Carbon number (CN)	
	Test[b]	Predicted	Test[b]	Predicted
2	0.0	0.3	6.0	6.0
5	3.0	2.8	6.0	6.0
7	11.0	10.9		5.9
45	10.0	9.5		5.7
10	12.0	12.3		5.8
48	11.0	11.1		7.5
13		12.8		7.2
14		12.4		8.1
15	3.0	2.0		5.2
16	3.0	2.4		5.2
30	.0	.8		6.9
33	.0	.0		6.2
34	3.0	1.8		6.5

[a] Results for selected solutes shown. See Table 6 for designations.

[b] Blanks are free-floated points.

Table 17. Summary of selected target transformations for solvent vectors

Vector name	Designation	Example values[a]	Points[b]	Qualitative evaluation[c]
boiling point	BP	576,435	6	p
density	DE	0.78,1.12	6	p
dipole moment	DM	0,1.83	4	g
melting point	MP	295,295	5	v[e]
molar refraction	MR	81,91	7	f
molecular weight	MW	241,352	7	g
log refractive index	LND	.157,.170	7	g
reciprocal MW	RMW	.00415,.00283	7	g
square MW	SMW	57831,124150	7	f
unity	UNI	1,1	7	v
ether uniqueness	UET	0,0	7	p
halogen uniqueness	UHA	0,1	7	p
halogen & ether uniqueness	UHE	0,1	7	f
halogen & thioether uniqueness	UHT	0,1	7	f

(continued)

Table 17(continued)

iodine uniqueness	UI	0.1	7	v
ΔI^{x}_{HDA} [d]	DIH	0,386	5	f
solvent group δ [d]	DY	0,7.1	7	v
Δ^{x}_{y} halide [d]	DCV	0,38	7	g
Δ^{x}_{y} alkene [d]	DAV	0,8	7	g

[a] Examples are given for solvents HDA and HDI.

[b] Number of points on test vector.

[c] v-very good agreement between test and predicted vector; g-good agreement; f-fair agreement (several points poorly predicted or pattern only predicted); p-poor agreement.

[d] Average solvent functionality vectors taken from Table V; DIH and DY vectors from Tables III and IV of reference 79 ; see text.

[e] Test vector is a near-unity vector (all points approximately equal), as are the vectors for refractive index, carbon number and number of atoms (not shown in table).

Because removal of branched solutes did not affect the factor space in the previous section, we should not expect a vector testing for branching to test well. As listed in Table 15, a vector testing for branching (designated NB) did test poorly.

Equation 50 provides us with a unique opportunity to test as solute and solvent cofactors the functions described by Martire. The equation can be rewritten

$$I_Z = 100\ (n)_X\ (1)_Y + \Delta I^Y_{HDA} + \Delta^X_Y \qquad (51)$$

where X and Y represent solute and solvent respectively. n, the solute carbon number, is obviously a solute cofactor whose associated solvent cofactor is constant taken as unity. ΔI^X_{HDA}, the contribution of the solute functional group X to the retention index on straight chain alkane solvent HDA, is the product of solute and solvent cofactors: a cofactor associated with the solute functional group and a cofactor associated with the solvent straight chain. Δ^X_Y, the retention index dispersion, is the increment to the retention index attributed to the solvent functional group X. Zielinski and Martire tested the applicability of the substitution:

$$\Delta^X_Y = \delta_X\ \delta_Y \qquad (52)$$

where δ_X and δ_Y measure the interaction strengths of the solutes and solvent group, respectively, and equal the retention index dispersion due to solute group X on solvent

group Y. Thus δ_X would be expected to be a good test vector for the solutes and δ_Y for the solvents.

Therefore, from Equation 51, n, the solute carbon number should represent one of the solute cofactors and target transform very well. A carbon number vector was obtained by assigning carbon number values to the straight chain alkanes and alkenes (solutes 1-6). Ideally only the straight chain alkanes should be assigned carbon numbers on the test vector. However, there are only three n-alkanes in the data set. At least four data points are required in a three factor space to provide a valid target transformation. Howery et al.[24] found the carbon number for n-alkenes equivalent to the corresponding alkane. Therefore, the n-alkenes were added to the test vector. All other data points were free-floated.

The carbon number vector target transformed very well (designated CN in Table 15). Details for this target transformation are given in Table 16 for thirteen selected solutes. Since most of the data points were "free-floated", i.e., left blank, we can assign, using the predicted values for the free floated data points, carbon numbers associated with the functional groups attached to the solutes. Thus a methyl group attached to the 2-position of a solute has an average carbon number value of 0.6; those attached to the 3-position, 0.8. Two methyl groups attached to the same 2-position carbon will therefore have a combined carbon number of 1.2. A methyl group attached to an unsaturated

carbon has a carbon number of 0.9; an ethyl group in the 3-position, 1.9. Similarly, carbon numbers contributions for the halogens are: primary-chlorine (2.9), primary-bromine (3.9), primary-iodine (5.1), secondary-bromine (2.4), and secondary-chlorine (2.4). These results corroborate previous carbon number values found by Howery et al.[24]

Values for the solute vector ΔI^{X}_{HDA} were obtained from Table 3 of Zielinski and Martire's paper. As tabulated in Table 16, this vector (designated DIY) did not target transform well. This is not surprising since this vector is associated with both the solutes and solvents.

Though Δ^{X}_{Y} is associated with both the solutes and solvents, the retention index dispersion, δ_X, was tested as a solute vector using the following procedure. Three vectors were obtained from Martire's data: Δ^{X}_{Y} on alkanes is taken from Table 5 of Reference 79; Δ^{X}_{Y} on halides is an average of the Δ^{X}_{Y} values for the HDC, HDB and HDI solvents; and Δ^{X}_{Y} on ethers is the DOE value.

As tabulated in Table 15, these vectors gave good to fair fits in the target transformation. Details of the target transformation for Δ^{X}_{Y} on alkenes (designated DHU) are given in Table 16 for thirteen selected solutes. The predicted data points are in good agreement with the input data points. Values for the retention index dispersion were free-floated for the iodine-containing solutes, solutes 13 and 14 in the table. The predicted values

for these solutes are quite plausible, since we expect the iodides to exhibit much the same values as the chlorides (11.0) and bromides (12.0). Similarly, for vectors DEU and DXU, the predicted values for the iodides averaged 42.0 and 26.0, respectively.

We expect, however, the solute interaction strength δ_X (designated DX in Table 15) to target transform very well for the solutes. Values for this vector were obtained from Table 4 of Reference 79. Details for the target transformation of δ_X are also given in Table 16, for thirteen selected solutes. Agreement between the input and predicted vectors is very good.

Other vectors which target transformed very well for the solutes include the square of the boiling point (designated SBP, details illustrated in Table 16), density (DE), dipole moment (DM), enthalpy of vaporization (EN), molar refraction (MR), molecular weight (MW), number of atoms (NA), number of carbon atoms (NC) and hydrogren atoms (NH), unity (UNI) and iodine uniqueness (UI). The good fit for the dipole moment, enthalpy and molar refraction vectors is expected from general interaction theory (see Appendix B); molecular weight, atom number and hydrogen number are probably related to the carbon number vector and iodine uniqueness corroborates the high uniqueness found for the iodine containing solutes in the uniqueness test. Boiling point squared (SBP), has been proposed as being directly related to the

retention index[81] and as a primary interaction parameter.[80] Interestingly, this vector has been found to test well on several other gas chromatographic problems.[20-25]

Table 16 also gives details for a solute vector which target transformed poorly: alkene uniqueness (designated DE). Predicted values for this vector do not match the input values. In general, those data points which should predict close to 1.0 average only 0.3; values which should predict 0.0, also average 0.3.

For the solutes, additional vectors which did not test well include: boiling point (^{o}C), melting point (^{o}K and ^{o}C), viscosity, boiling point dependence on pressure (dt/dp), ΔI_{HDA}^{X}, symmetry of molecular structure, number of branched atoms, and the following uniqueness vectors: chlorine, bromine, isopropyl, t-butyl, dimethyl, allylmethyl, double bond and cis-trans.

From the uniqueness test, we expect a vector testing for halogen uniqueness (designated UHA) to target transform well for the solvents. As listed in Table 17, which lists most of the vectors tested for the solvents, a halogen uniqueness vector tested poorly. Thus, halogen uniqueness is not associated with one of the solvent cofactors. However, a vector testing for the iodine uniqueness (UI), target transforms very well. Thus, one of the abstract solvent cofactors may be associated with a unique interaction associated with the solvent iodine functionality.

Similarly, all other vectors we were able to generate from the results of the uniqueness test gave only fair to poor transformations. Details of the ether uniqueness (UET) vector are listed in Table 18. As can be seen, this vector does not give a good fit. However, solvents HDC, HDB, DOE and DOT are predicted high on this vector. This is an indication that this vector is transforming on to a cofactor which exhibits the co-relationships of these solvents. These solvents do correlate on the uniqueness test.

Solvent vectors obtained from Equation 54 were also tested. Values for solvent vector Δ_Y^X for alkenes (designated DEV) were the averaged alkene values from Table 5 of Zielinski and Martire's paper. Solvent vector Δ_Y^X for halide (DXV) was an average of the reported Δ_Y^X for chlorides and bromides in the same table. Solvent interaction vector, δ_Y (designated DY), was obtained from Table 4 of the above reference; Table 3 provided the solvent ΔI_{HDA}^X (DIH) values.

From Equation 51 we expect a unity vector to be the corresponding solvent cofactor for solute cofactor NC (carbon number). Indeed, unity target transforms very well as listed in Table 17. The retention index dispersion vector (DXV and DEV) and the interaction strength of the functional group (DY) also target transformed very well. Details for the solvent interaction strength vector DY are given in Table 18. Agreement between the

Table 18. Details of selected target transformations for solvent vectors, 3 factors

Solvent[a]	Dipole moment (DM)		Ether uniqueness (UET)		ΔI^{x}_{HDA} (DIH)		Solvent group δ_y (DY)	
	Test[b]	Predicted	Test	Predicted	Test	Predicted	Test	Predicted
HDA	0.00	-.02	0.0	0.0	0.	-50.	0.0	-.1
HDE	.51	.54	.0	.1	-19	51	1.6	1.8
HDC		1.90	.0	.5	213	217	5.8	6.0
HDB	1.96	1.95	.0	.4	289	270	6.5	6.5
HDI	1.83	1.83	.0	..0	386	381	7.1	7.1
DOE		1.37	1.0	.5		122	413	4.1
DOTE		1.86	1.0	.4		246	6.2	6.1

[a] See Table 7 for designations.

[b] Blanks are free-floated points.

predicted and input vectors is almost perfect. Details are also given for solvent vector DIH (ΔI^{X}_{HDA}). This vector would seemingly be a very good vector if the predicted values for solvents HDA and HDE were better. ΔI^{X}_{HDA} showed similar results on the solutes; ΔI^{X}_{HDA} is most probably a product term for the solute-solvent interaction of a solute functional group on a straight chain hydrocarbon DHA and is therefore a mixed solute-solvent term. Other solvent vectors which target transformed well include dipole moment (DM), molecular weight (MW) and the log of the solvent refractive index (LND). Melting point apparently was also a very good vector; however, all the data points for this vector were nearly equal. In other words, the melting point vector is nearly equivalent to the unity vector.

As for the solutes, the free-floating feature of the target transformation step gives plausible values for free-floating points of the better solvent vectors. For example, the DM of solvent HDC is predicted to be 1.90, comparable to the actual values of 1.96 and 1.83 for the other halogen solvents HDB and HDI. Other vectors tried but which did not test well for the solvents include: melting point (^{o}K and ^{o}C), refractive index, boiling point (^{o}K and ^{o}C), density, number of carbon atoms and number of atoms. The reciprocal and square of ND and the logarithm of molecular weight, carbon number and atom number were near-unity vectors (all points

about equal) and therefore have no special physical meaning.

For the solvents, 14 vectors were found which tested at least fair; 24 vectors tested at least fair for the solutes. Have we representative cofactors for each of the three abstract factors of our data space? Which ones represent the abstract factors best? Which are equivalent? The combination step provides the answers to these questions.

G. Combination of Basic Vectors

The physical and structural vectors that target transform satisfactorily represent the real, physical analogs of the abstract cofactors. Have we satisfactory analogs for all the abstract cofactors? That is, does our set of target vectors include an analog for all the abstract cofactors? Which set of the target vectors best represents the abstract cofactors or interactions?

Combination of the solute and solvent target vectors enables us to associate the abstract cofactors with physically meaningful vectors. The quality of the combinations obtained provides us with a determination of how well we have defined the abstract cofactors, and whether we have associated real vectors with all the cofactors in the problem.

Tabulated in the bottom row of Table 12 are the results for the combination step for the basic vectors for the solutes and solvents. All the target vectors listed in Tables 15 and 17 which target transformed at least fair were included in the combinations.

The best combination of solvent basic target vectors reproduced the original data matrix with an rms error of 0.66 with a largest error of 2.9. This reproduction is almost as good as the abstract reproduction (0.55, 2.9).

The best real solution consists of unity (UNI), log of refractive index (LRI), and the solute contribution to the retention index for the solvent chloride (DXU). Comparison of the twenty best sets of combinations shows the equivalence of these vectors: DCU with the dipole moment (DM) and the solvent interaction strength (DX) with the log of the refractive index (UND). Unity is the most prominent vector, being present in each of the thirty best combinations. Therefore, unity is definitely one of the solvent cofactors. These results are summarized in the last column of Table 12.

The best combination of solute basic vectors is not expected to reproduce the data matrix as satisfactorily as does the combination of the solute vectors discussed. The best set: atom number (NA), carbon number (NC) and dipole moment (DP) reproduced the original data with an rms error of 2.74 with a largest error of 8.9. Our real vectors are not exact analogs of the abstract cofactors, though very good approximations of them.

Fourteen combination sets reproduced the original data with an rms error less than 4.0. Solute vectors NA and CN are represented in every one of these best combinations. These vectors may represent two of the three solute cofactors. The third cofactor is associated with the third vector of these combinations. These vectors include the solute combination vectors (ΔI_Y, δ_X, and the solute ΔX_Y vectors), and dipole moment (DP).

H. Complete TFA Model

The solutes and solvents involved in this study are not expected to exhibit any strong electron donor or acceptor properties. The molecular interactions involved are primarily Van der Waal's type interactions. As described in Appendix B, the Van der Waal's interactions are usually described by (1) a dispersion term, E_d, related to the solute molar volume and the electron polarizabilities of both solute and solvent, (2) an induced dipole-dipole term, E_i, related to the dipole moment and polarizabilities of the interacting molecules and (3) a dipole-dipole term, E_o, related to the solute and solvent dipole moments. Thus, the total Van der Waal's interaction, E_T, can be described as:

$$E_T = E_d + E_i + E_d \tag{53}$$

Equation 53 can be rewritten as:

$$I_Z = 100\ (n)_x \cdot (1)_y + \Delta^x_{HDA} + \delta_x \cdot \delta_y \tag{54}$$

where subscripts x indicates a solute cofactor and y a solvent cofactor. Δ^x_{HDA} is a combined solute-solvent cofactor term. We have found this to test as a solute cofactor in the target transformation step.

Using target testing, we have related the retention index of solute x on solvent y to a variety of factors, including real dispersive and dipolar interaction terms.

As found in the combination step, the best set of physically meaningful cofactors for the solutes involves NA, CN and DP, nad for the solvents involves UNI, LND and DXV. DXV was found equivalent to DM. In an analogy with equation 53 we can match these solute and solvent cofactors into dispersive, induced dipole and dipole-dipole terms:

$$I_z = k_1 (CN)_x (NC)_y + k_2 (NA)_x (LND)_y + k_3 (DP)_x (DM)_y \quad (55)$$

UNI is equivalent to $(NC)_y$, the solvent chain length (equals 8 for the constant chain length solvents), and k_1, k_2 and k_3 are constants. All terms are products of solute-solvent contributions.

Equation 55 is a complete TFA model consistent with general interaction theory. In the next section we will test the validity of this equation.

I. Predictions Using Key Basic Vectors

To illustrate the applicability of Equation 55 we may solve for constants k_1, k_2, and k_3 using the retention index data for three solute-solvent pairs (three unknowns, three equations) and test the resultant equation on new solute-solvent pairs. Two approaches can be employed: (1) use the best sets of solutes and solvents as found in the combination step for solute and solvent vectors and (2) use regression analysis for all solute-solvent pairs for the entire data set.

As an example of the first approach, we may take the retention equations for three solute-solvent pairs: pentane, on solvent HDA, 2,2-dimethylpentane on HDA and iodopropane on HDC, as follows:

$$I_{\text{pentane, HDA}} = 500 = k_1\ (5) + k_2\ (17)_x\ (.157)_{\text{HDA}} + k_3 (0)_x (0)_y \tag{56}$$

$$I_{\text{dimepentane, HDA}} = 625 = k_1 (6.15)_x + k_2 (23)_x (0.157)_{\text{HDA}} + k_3\ (0)_x (0)_{\text{HDA}} \tag{57}$$

$$I_{\text{iodopropane, HDC}} = 726 = k_1 (8.13)_x + k_2 (11)_x (0.162)_{\text{HDC}} + k_3\ (2.04)\ (1.96) \tag{58}$$

Solving for the constants k_1, k_2 and k_3 we find: $k_1 = 83.732$, $k_2 = 30.475$ and $k_3 = -2.2628$. Alternatively, using a regression analysis of all data points to obtain regression constants k_1, k_2 and k_3, we obtain: $k_1 = 75.807$, $k_2 = 35.819$ and $k_3 = -0.17518$.

Thus Equation 54 becomes:

$$I_z = 83.732\ (CN)_x \cdot (NC)_y + 30.475\ (NA)_x \cdot (LND)_y - 2.2628\ (DP)_x \cdot (DM)_x \quad (59)$$

Testing these results on new solute-solvent pairs gives very good results. For example, the retention index of cis-3-heptene on solvent HDI, using the constants obtained in the first approach, is:

$$I_{x,\ HDI} = 83.732\ (7.096) + 30.745\ (21) \cdot (0.170) + (-2.2628) \cdot (0.356) \cdot (1.83) = 701.5 \text{ RI units} \quad (59A)$$

The reported value is 698 $\pm$ 1 retention index units. Using the regression constants, we obtain 665.8 RI units.

Similarly, for the retention of solute 2-bromo-2-methylpropane on solvent HDI we obtain 622.7 RI units using the constants from the first approach and 584 using regression constants. The reported value is 602 RI units.

Calculating the "k" constants using other sets of vectors gives the same trend in predicted values.

The success of these predictions is a very powerful vindication of our model pairing of cofactors. Table 19 is a tabulation of predicted retention indices for arbitrarily selected data points compared to their expected retention indices. The agreement is highly satisfactory with an average error of 9.8 retention indices. Equation 59 was used.

Table 19. Predicted retention indices for nineteen arbitrarily selected solute-solvent pairs

Solute[a]	Solvent[b]	Predicted[c]	Literature[d]	Error
1	HDA	500	500	0
3	HDC	631	626	5
5	HDA	588	582	6
5	HDI	594	591	3
5	HDE	589	585	4
8	HDC	642	649	-7
9	HDE	746	726	20
19	HDI	688	702	-14
23	DOE	468	474	-12
26	HDA	519	535	-16
28	HDE	666	666	0
33	DOE	629	630	-1
35	HDI	647	634	13
42	HDI	496	501	5
45	HDA	544	515	29
47	HDC	692	681	11
49	HDB	618	629	11
49	HDI	623	602	21

[a] See Table 6 for solute designations.

[b] See Table 7 for solvent designations.

[c] Predicted retention indices using equation 59.

[d] Retention indices from Table 6.

J. SUMMARY

In summary, we have performed a target factor analysis on gas chromatographic retention index data obtained on well defined liquid phases of similar chain length and flexibility. We find that three abstract factors are sufficient to explain the real factor space. Zielinski and Martire also proposed a three term empirical model.

For the solutes, the iodides are the only unique solutes. Combinations of typical solutes vectors show that the solute cofactors consist of vectors associated with an alkane or alkene, a halide and a branched solute. Twenty four real solute vectors target transformed well including vectors associated with atom number, carbon number, boiling point, dipole moment, enthalpy, molar refraction, halogen uniqueness and four vectors obtained from Zielinski and Martire's empirical model. The key physically significant solute cofactors, as found via a combination step, were carbon number, atom number and solute dipole moment.

For the solvents, HDI, the iodide solvent, was most unique. Combination of the solvent vectors show that the best sets include either an alkane or alkene (HDA or HDE), a bromide or chloride (HDB or HDC) and the iodide (HDI). Fourteen real physical and structural vectors target transformed well, including vectors associated with dipole moment, molar refraction, molecular weight,

log of refractive index, unity, iodide uniqueness and four vectors obtained from Zielinski and Martire's tabulation. The best physically significant solvent cofactors were identified, via a combination step, as unity, log of refractive index and Zielinski and Martire's interaction strength associated with dipole moment.

The solute and solvent cofactors are related according to Equation 55, consistent with Zielinski and Martire's equation and interaction theory. A model from target factor analysis is then used to predict the retention index of new solute-solvent pairs.

CHAPTER VI

TARGET FACTOR ANALYSIS OF GAS-LIQUID CHROMATOGRAPHIC RETENTION INDICES OF MONOMERIC SOLVENTS AND SIX SOLUTE SETS

I. INTRODUCTION

A. Introduction

Information on the effect of stationary phases on gas chromatographic retention indices is lacking because of the inherent complexity of the problem. Questions such as "Which phases are redundant?" and "Which phase is most appropriate for a particular problem?" have given rise to a host of proposals including various definitions of the elusive property "polarity." As described in Chapter IV, McReynolds has resorted to ten functional "polarity" probes[55] while others have attempted to define only single polarity probes.

Are McReynolds' ten constants adequate as descriptors of the separation process? Are they redundant? Do the ten solutes selected represent the best set of solutes to be used as functional probes?

From a theoretical viewpoint, how many physical interactions are involved? What are they? Are any of

them unique to a particular solute or solvent?

From a practical standpoint: How does one select a solvent phase? Is there a way of predicting RI data?

McReynolds has compiled a large consistent data base of retention indices at two different temperatures, for over 150 solutes on approximately 200 solvent stationary phases.[85] We have, for the purpose of a TFA analysis, compiled six subsets of this data. Our major purpose is a TFA study of the stationary phases. In addition, we have carried out the analysis for the solutes in these subsets.

Because we are most interested in studying the stationary phases, we have selected eighteen monomeric solvents. We have purposely eliminated all polymers or mixtures. This will allow us to develop test vectors for the solvents.

Retention indices at 120°C for the six problem sets studied were taken from McReynolds' compilation.[80] The data is tabulated in Appendix C. McReynolds' estimates the experimental error at ± 3-± 5 RI units with some points of greater error.[81]

Table 20 is a list of the eighteen liquid phases with their designations and some structural formulas. The solvents include esters (pthalates, adipates, and sebacates), sucroses, amides, fluorine, and a phosphate. Diglycerol, a compound known to GC workers for its unique behavior,[64] is also included in the solvent set.

Table 20: List of solvents

Symbol	Name
A	Bis-(2-ethoxyethyl)pthalate
B	Dibutyltetrachloro pthalate
C	Di-2-ethylhexyl adipate
D	Di-2-ethylhexyl sebacate
E	Diglycerol
F	Diisodecyl pthalate
G	Dioctyl pthalate
H	Dioctyl sebacate
I	Flexol 8N8 $(C_2H_5CH(C_2H_5)COO)_2N(COCH(C_2H_5)C_4H_9)$
J	Hallcomid M18 (N,N-dimethyl stearamide)
K	Hyprose SP-80 (2-hydroxypropyl sucrose)
L	Isooctyldecyl adipate
M	Quadrol (N,N,N',N'-(2-hydroxypropyl)-ethylene diamine)
N	SAIB (Sucrose acetatehexaisobutyrate)
O	SOA (Sucrose octaacetate)
P	TMP Tripelargonate (Celanese Ester No. 9) $(C_2H_5C(CH_2OOCC_8H_{17})_3)$
Q	Tricresol phosphate
R	Zonyl E-7 $\phi(COO(CF_2)_nCH_3)$, n = 3 and 5, mainly

Two sets of solutes include representative solutes from the alcohols, aldehydes, ketones, ethers, esters (acetates and propionates) and hydrocarbons. A basic set (which we designate as MC33) includes 33 solutes and a more complicated set of 53 solutes (desinated as MC53) includes all solutes in MC33 plus 20 branched chain solutes.

The other four solute sets were carbonyl sets: 1) 14 simple ketones (designated KETI) 2) 23 simple, branched and dione ketones (KET2); 3) 16 aldehydes (ALDE) and 4) a combined KET2 and ALDE set (CARB). Some of the questions we hoped to answer with these sets are: Does the addition of branched compounds increase the factor space? How different are the ketone and aldehyde functional groups from a GC viewpoint? How would the results we obtain on these specific solutes compare with previous TFA results on other solutes such as the hydrocarbons, alcohols, and others. Table 21 lists the solutes for all the data sets.

The bulk of the analysis was obtained using the correlation matrix for each problem. However, because of the nature of the data we expect equivalent results for an analysis of the covariance matrix. We have shown that TFA analysis on problem set KET1 using both the covariance and correlation matrices gave equivalent results.

To simplify the presentation, most examples given in this and the following chapters will be for problem

Table 21: List of solutes

Symbol	Solute Name[a]	Symbol	Solute Name[a]
ALDE (aldehyde):			
A2	acetaldehyde	A11	heptanal
A3	propanal	A12	2-ethylhexanal
A4	butanal	A13	acrolein
A5	isobutanal	A14	mathacrolein
A6	pentanal	A15	crotonaldehyde
A7	isopentanal	A16	2-ethyl-2-butenal
A9	2,2-dimethylpropanal	A17	2-ethyl-2-hexenal
A10	hexanal	A18	2,4-hexadienal
KET1 (simple ketone)* and KET2 (expanded ketone):			
K1	acetone*	K15	2,4-dimethyl-3-pentanone*
K2	2-butanone*	K16	2-octanone*
K3	2-pentanone*	K17	2-nonanone*
K4	3-pentanone*	K18	5-nonanone*
K5	3-methyl-2-butanone*	K19	cyclopentanone
K6	2-hexanone*	K20	cyclohexanone
K7	3-hexanone*	K22	5-hexen-2-one
K8	3-methyl-2-pentanone*	K24	4-methyl-3-penten-2-one
K9	2-methyl-2-pentanone*	K25	2,3-butadione
K10	3,3-dimethyl-2-butanone*	K26	2,3-pentadione
K11	2-heptanone	K27	2,4-pentadione
K12	3-heptanone		
CARB (carbonyl):			
All solutes from ALDE and KET2.			

(continued)

Table 21 (continued):

MC33 (McReynolds' straight chain) and
MC53 (straight and branched chain):

OL1	ethanol	ES5	2-methyl-2-butylacetate+
OL2	propanol	ES6	ethylproprionate
OL3	pentanol	ES7	pentylproprionate
OL4	hexanol	ES8	isobutylproprionate+
OL5	octanol	ES9	isopentylproprionate+
OL6	isopropanol+	ET1	ethyl ether
OL7	2-butanol+	ET2	butylethylether
OL8	2-methyl-1-butanol+	ET3	propylether
OL9	2-methyl-1-pentanol+	ET5	pentylether
OL10	2,2-dimethyl-1-propanol+	ET6	isopropylpropylether+
A3	propanal	ET7	isopropylether+
A4	butanal	OX1	ethylene oxide
A5	isobutanal+	OX2	propylene oxide
A6	pentanal	OX3	1,2-butylene oxide
A7	isopentanal+	AR1	benzene
A9	2,2-dimethylpropanal+	AR2	toluene
A11	heptanal	AR3	p-xylene
K2	2-butanone	AR4	ethylbenzene
K6	2-hexanone	AR5	o-xylene+
K8	3-methyl-2-pentanone+	AR6	o-diethylbenzene+
K10	3,3-dimethyl-2-butanone+	AR7	p-diethylbenzene+
K16	2-octanone	AL1	ethane
K28	3-methyl-2-butanone+	AL2	butane
ES1	ethylacetate	AL3	hexane
ES2	pentylacetate	AL4	tetredecane
ES3	hexylacetate	AL5	hexadecane
ES4	isobutylacetate+		

[a] Solutes with superscript, *, in KET1 problem; solutes with superscript, +, in MC53 problem only.

set MC33. Results for all problem sets are given following the MC33 results. The determination of the proper factor space will be illustrated in the concluding half of this chapter. In the following two chapters we will describe the results for the solvents and solutes, respectively.

B. Factor Determination

In Table 22 are listed the results of the factor determination for the six problems. All of the problems are apparently six factor spaces except the KET2 problem (four factors) and the MC53 problem (seven factors). Determination of thecorrect factor space was accomplished by comparing the RMS error, the percentage of errors greater than experimental ($\pm$5 RI) and the largest reproduced errors. For example, we determined the factor space for the ALDE problem set by observing that at six factors the RMS error between the predicted and original data matrix is only 1.50 and only 1.4% of the predicted data matrix points had errors greater than $\pm$5 RI units when compared with the original data matrix. For six factors, the largest error was only 5.7 RI units which is probably within experimental.

The Malinowski functions give nearly the same results as predicted above. Generally, the imbedded error functions (IE) and the indicator function (IND) have minimums at the respective factor space for each problem. However, the IND function on the KET1 problem set shows a minimum at three factors, though the IE gives the correct four factor minimum. The real error (RE) averages 2.4 RI units for the six problem sets, slightly less than the predicted experimental average error of $\pm$3 - $\pm$5 RI units. The RE error for the KET1 problem is significantly higher, which may be related to the failure of the IND function

Table 22. Summary of reproductions for each problem

Problem designation	Number of solutes	Number of factors	rms error	Largest error	Percentage errors > ±5
ALDE	16	6	1.50	5.66	1.4
KET1	14	4	1.73	5.65	2.0
KET2	23	6	1.51	7.61	1.2
CARB	39	6	1.94	9.87	2.7
MC33	33	6	1.94	9.09	3.9
MC53	53	7	1.81	8.61	2.5

to indicate the proper factor space.

Branching increases the factor space at least by one as evidenced in MC53 and KET2. Thus MC53 and KET2 have at least one additional interaction due to branching as compared to MC53 and KET1, respectively, which do not contain any branched solutes.

Apparently, the ketone and aldehyde solutes represented in problem sets ALDE and KET2 exhibit similar interactions. Their combination in the CARB problem retains the same six factor space of the smaller subsets.

Originally the MC33 and MC53 data sets included an extra solute: butyl ether. During the original factor determination, the predicted data point for butyl ether on solvent dioctyl sebacate was found to have the largest predicted error (12 RI units). Increasing the factor space to 9 or 10 factors did not affect substantially the absolute value of this error though all the other data point errors were decreased as expected. Graphing the retention indices of homologous ethers on dioctyl sebacate vs. their carbon numbers should result in a straight line. A graph for the ethers shows the value for butyl ether to deviate by approximately 12 RI units. We conclude the published value for the RI of butyl ether on dioctyl sebacate to be in error by -12 RI units. As a result, to avoid any ambiguity in our problem sets, we removed butyl ether from the problem sets. (Interestingly, the uniqueness test /see below/ did not pick this error up; the uniqueness value of butyl ether is 0.02).

CHAPTER VII

TARGET FACTOR ANALYSIS OF GAS-LIQUID CHROMATOGRAPHIC RETENTION INDICES OF MONOMERIC SOLVENTS AND SIX SOLUTE SETS

II. CHARACTERIZATION OF SOLVENTS

As described in Chapter VI, six solvent-solute sets were derived from McReynolds Compilation[80] and their respective factor spaces determined. In this chapter, we will describe the characterization of the solvent eofactors.

A. Uniqueness

Table 23 summarizes the uniqueness test results for the solvents on the six problem sets. Solvents diglycerol (E) and Zonyl E7 (R) are the most unique for every problem set (averaging .99 and .95, respectively). Diglycerol is a stationary phase long known for its unique behavior,[61,64,82] probably due to its hydrogen bonding ability; Zonyl E7, a fluorine-rich compound, is unique, probably due to the halogen electronegativity.[61]

Solvents Hyprose SP80 (K), quadrol (M) and SOA (O) are moderately unique (averaging .717, .472, and .715, respectively) on all the problem sets, except KET1. Except for SAIB, they are the only solvents which have a β-hydroxy or β-carboxy functional group. The

Table 23. Uniqueness tests for solvents, McReynolds' problems.

	CARB problem, 6 factors		MC33 problem, 6 factors	
Solvent[a]	Uniqueness test value	Correlates[b] with solvents	Uniqueness test value	Correlates[b] with solvents
A	0.36	B(.26),I,J,Q	0.19	B,Q
B	.24	A(.26),I,J,Q	.42	A,G,O,Q(.25)
C	.14		.14	
D	.14		.13	
E	.99		.99	
F	.10		.11	
G	.13		.13	B
H	.14		.13	
I	.17	A,B,J,Q	.12	J
J	.14	A,B,I	.24	I,M
K	.66	M(.42)	.45	M(.33),N,O(.25)
L	.14		.13	
M	.32	K(.42)	.55	J,K(.33),Q
N	.19	O(.31)	.28	K,O,P
O	.83	N(.31)	.61	B,K(.25),N,P
P	.18		.17	N
Q	.18	A,B,I	.28	A,B(.25),M
R	.98		.94	

[a] See Table 20 for solvent designations.

[b] Other solvents with predicted values greater than 0.15 are listed. predicted values greater than 0.24 are given in parenthesis.

(continued)

Table 23 (continued)

Solvent[a]	ALDE problem, 6 factors		MC53 problem, 7 factors	
	Uniqueness test value	Correlates with solvents[b]	Uniqueness test value	Correlates with solvents[b]
A	.37	B(.26),I,J,Q(.25)	.32	M,Φ,Q(.30)
B	.25	A(.26),I,J,Q	.69	K,Q
C	.16	L	.14	
D	.14		.14	
E	.99		.99	
F	.09		.12	
G	.14	N,O	.14	Q
H	.14	L	.14	
I	.15	A,B	.16	J
J	.14	A,B	.26	I,M
K	.53	M(.42),O	.75	B,M(.27)
L	.16	C,H	.14	
M	.56	K(.42)	.56	A,J,K(.27),Q
N	.29	G,O	.28	O(.31)
O	.57	G,K,N(.37)	.74	A,N(.31)
P	.16		.14	
Q	.19	A(.25),B	.33	A(.30),B,G,M
R	.95		.98	

(continued)

Table 23 (continued):

Solvent[a]	KET1 problem, 4 factors: Uniqueness test value	KET1 problem, 4 factors: Correlates with solvents[b]	KET2 problem, 6 factors: Uniqueness test value	KET2 problem, 6 factors: Correlates with solvents[b]
A	.13	K	.23	I,M
B	.08		.10	
C	.09		.15	D,H,L
D	.14		.21	C,H,L
E	.98		.97	
F	.08		.10	
G	.08		.10	
H	.14		.24	C,D,L
I	.09		.20	A,M
J	.09		.14	
K	.44	A,M(.28),O(.28)	.87	M
L	.12		.18	C,D,H
M	.19	K(.28),O	.37	A,I,K
N	.08	R	.17	O(.31)
O	.20	K(.28),M	.82	N(.31)
P	.08		.10	
Q	.10		.10	
R	.88	N	.94	

effect of the two β-carboxy (acetate) groups in solvent SAIB are probably swamped by the six isobutyrate groups. Consequently, though SAIB and SOA are both sucroses they do not share uniqueness. (Wold and Anderson[61] also found a unique correlation for Hyprose SP80 using AFA)

It is apparent that the effect on retention time of the functional groups on the molecule is more important than the underlying carbon skeleton structure. The hydroxypropyl group functionality seems to be more important than sucrose or amine functionality. This is seen even more dramatically in the correlations between the various solvents.

Solvents K and M correlate in every problem set, probably because of their acetate functional groups. However, K, also a sucrose but having no acetate or isobutyrate groups, usually does not correlate with N.

In most problems, except for KET1 and KET2, tricresol phosphate (O) correlates with ethoxyethyl pthalate (P). However, the pthalates do not correlate well among themselves.

Is the uniqueness of a solvent important? How does a particular solvent affect the factor space? Is there a unique interaction associated with a particular solvent? If there is, then we may be able to identify it from the associated solvent molecules' structure. These questions may be answered in the following solvent reproduction test.

B. Effect of Individual Solvents on Reproduction

In this solvent test, a solvent is removed from the data set and its effect on the factor space, as evidenced in the RMS error and percentage of errors greater than experimental, for the data with the seventeen remaining solvents is obtained. If the removal of a particular solvent results in a significant reduction of the factor space, then that solvent contains a relatively unique interaction for the problem set under study.

The results of this test are summarized in Table 24 for the MC33 problem set. No solvent's removal affects the factor space significantly enough to reduce the factor space by one. However, removal of either diglycerol (E), Hyprose SP80 (K), or Zonyl E7 (R) has the greatest effect on the factor space, reducing the percentage errors greater than experimental to an average 1.6% and the RMS error to an average of 1.69 RI units. The result is a clearer and more definite six factor space than was obtained on the original data. These solvents are also the most unique solvents for MC33 as determined in the uniqueness test. Solvents B, M, and O, which are moderately unique in the uniqueness test (see Table 23), also show better improvement in the percentage errors as compared to the non-unique solvents.

Significantly, any correlations derived in the uniqueness test for the solvent problem remain the same on all the new reduced seventeen solvent problem sets.

Table 24. Effect of removal of each solvent vector from problem MC33, 6 factors.

Solvent column vector removed[a]	RMS error	Percentage errors ≥±5	Largest error	Solvent column vector removed[a]	RMS error	Percetage errors≥±5	Largest error
A	1.94	3.6	8.5	J	1.87	3.6	8.7
B	1.77	2.5	7.0	K	1.66	1.6	8.7
C	1.96	3.6	9.3	L	1.95	3.7	9.1
D	1.98	3.6	9.1	M	1.83	2.3	8.4
E	1.71	1.8	9.6	N	1.89	2.1	9.7
F	1.94	3.6	8.9	O	1.88	2.3	8.0
G	1.94	3.0	9.1	P	1.89	2.7	9.3
H	1.96	3.6	9.1	Q	1.89	3.4	7.1
I	1.91	3.4	9.2	R	1.72	1.4	6.9
Complete MC33 problem	1.94	3.9	9.1				

[a] See Table 20 for solvent designations.

For the other problem sets, removal of some solvents does result in reduction of the overall factor space by one factor; other solvents exhibit the same effect as the unique solvents in MC33, reducing the original factor space only slightly.

These results are shown in Table 25. Listed in the first column are those solvents whose removal from the data space affects the reproduction enough to reduce the factor space; these solvents are probably unique representatives of a cofactor in the problem.

The most unique solvents (from the uniqueness test) seem to have the most effect on the solvent factor space. Thus in all the problem sets, solvents E and R, the most unique solvents, have the greatest effect on the solvent space.

Apparently, the unique interactions associated with solvents E and R are not as important in the MC33 problem set probably because of the multiplicity of functional groups for the solvents in MC33. Therefore, the effect of their removal is consequently not as dramatic.

In the second column of Table 25 are listed those solvents whose removal are not as significant to the factor space. Each of these solvents is unique enough to provide a more definite factor determination. Comparison with the uniqueness tables (see Table 23) shows that these solvents are moderately unique in the uniqueness test.

Table 25. Effect of removal of solvent vectors, all problems

Problem	Solvents influencing factor space significantly[a]	slightly[a]
ALDE	E,M,O,R	---
KET1	E,R	I,K,O
KET2	E,K,M,R	A,I
CARB	E,R	K,M
MC33	none	E,K,R
MC53	B,R	E,K,O

[a] See Table 20 for solvent designations.

As found for MC33, the correlations based on the uniqueness test are unaffected by solvent removal for all the problems.

C. Combinations of Typical Solvent Vectors

Combination of the typical solvent columns from the data matrix should provide us with the best set of solvent stationary phases. The best set of stationary phases, if done on a large enough data set, may even provide a basis for a "standard" set of stationary phases. Insight into the identity of the actual interaction factors may then be possible.

Table 26 lists the best sets per factor space for each of the problem sets. For example, the best set of solvent vectors from the original data matrix for problem set ALDE using only one factor, i.e., one vector, is solvent N. The average error between the reproduced matrix and the original data matrix is 35.6 RI units. Using two factor or solvent vectors, solvents E and J provide the best reproduction with an average error of 8.1 RI units. Using six factors, the appropriate factor space for problem ALDE as determined in the factor determination step, solvents A, C, E, K, N and R gave the best combinations with an average error of only 1.7 RI units. The largest error is 9.8 RI units. Compared to the abstract reproduction results as tabulated in Table 22 (average error of 1.5; largest error of 5.8), these results are excellent. We have succeeded in obtaining solvent cofactors related to the abstract cofactors.

Solvents E and R are represented in the best combination set for each problem, another indication of their high

Table26. Summary of best combinations utilizing best solvent vectors from data matrix for each problem, 1-7 factors

Problem	Number of solvent vector cofactors used in the reproduction[a]						
	1	2	3	4	5	6	7
ALDE	35.6 N	8.1 E,J	4.2 D,E,K	3.3 D,E,K,R	2.6 E,I,K,P,R	1.7(9.8) A,C,E,K,N,R	
KET1	31.6 N	4.1 E,F	3.1 E,G,R	1.9(9.5) D,E,M,R			
KET2	31.9 Q	7.3 E,P	5.8 C,E,M	3.6 E,L,M,O	2.2 E,G,K,O,R	1.7(8.4) E,J,K,L,O,R	
CARB	33.8 N	8.6 E,F	5.9 D,E,K	4.7 E,F,K,N	3.0 E,F,K,O,R	2.2(17.2) A,D,E,K,O,R	
MC33	63.9 Q	26.2 G,Q	16.9 F,K,R	5.8 D,E,K,R	3.2 D,E,M,O,R	2.2(16.5) D,E,G,M,O,R	
MC53	59.1 Q	25.9 F,K	15.9 D,K,R	6.3 D,E,K,R	3.5 D,E,M,O,R	2.5 D,E,G,M,O,R	2.0(16.1) A,B,D,E,M,O,R

[a] Upper line - rms error. Largest error in the true factor space given in parenthesis. Lower line - typical solvents in best combinations. See Table 20 for solvent designations.

uniqueness. Thus, these are unique cofactors probably associated with the interactions due to E's hydrogen bonding and R's highly electronegative fluorine atoms.

Comparing the best combination sets for problems MC33 and MC53 shows an interesting correlation. The best combination sets using four, five or six solvent vectors are the same for each problem. MC33 and MC53 share essentially the same factor space. The best combination set using seven solvent vectors for MC53 is similar to the best combination set for MC 33 using six solvent vectors, sharing solvents D, E, M, O and R. Since solvent G (dioctyl phthalate) is similar to solvent A (bis 2-ethoxy-ethylpthalate), also a pthalate, they both probably represent the same cofactor. Thus the seventh cofactor in problem MC53 is probably associated with solvent B (dibutyl tetrachloropthalate), probably due to its ability to form charge-transfer complexes with aromatic hydrocarbons,[83]

There are many combinations of other solvent vectors sets which provide satisfactory reproduction for each problem set. Indeed, if we assume that some of the solvents are redundant, i.e., can represent the same cofactor, then substitution of redundant solvents will not affect the reproduction. For example, there are 18,564 possible combinations of the eighteen solvents in a six factor space. The best thirty combinations all provide excellent reproduction. We can then compare the best thirty combinations to obtain the redundant solvent cofactors. This is illus-

trated in Table 27 which lists the 30 best combinations for MC33.

In the first column of Table 28 are listed the redundant solvent cofactors obtained by comparing the best thirty combinations for each problem set. The most apparent redundancy in every problem set is that of solvents C, D, H and L. Each of these solvents has two C-8 or a C-8 and C-10 alkyl chains attached to carbonyl groups. These seem to be the dominant functionality for those solvents. Solvents F and G also contain C-8 alkyl groups, yet they do not correlate with solvents C, D, H and L. The aromatic ring in solvents F and G, then, probably does induce slightly different interactions. The other correlation among solvents listed in column two of Table 28 are applicable only for the particular problem set. However, many of them vindicate correlations obtained in the uniqueness test. For example, the redundancy of solvents A, B, G and Q for problem MC33 echoes the correlations obtained in the uniqueness test for MC33.

A summary for the better combinations for factors 1—6 in problem MC33 is given in Table 29. The percentage of times a solvent occurred in combinations having an rms error less than an assigned cutoff value for the factor space is listed in this "pattern" table. For the solvent vector combinations, the cutoff value is approximately two times the rms error for an abstract reproduction for the factor space. Thus, for example, the abstract

Table 27. Best combinations of solvent vectors for MC33 problem

Combination[a]	rms error	Combination[a]	rms error
E D G M O R	2.20	G E M N O R	2.34
A D E M O R	2.22	E L M O Q R	2.40
B D E M O R	2.24	B D E N O R	2.43
C E G M O R	2.24	E F M N O R	2.45
D E M O Q R	2.28	C E F M O R	2.46
B C E M O R	2.29	A D E K O R	2.47
A E H M O R	2.30	A D E M N R	2.49
A E C M O R	2.32	D E F M O R	2.50
B E H M O R	2.33	G E M O P R	2.50
A E L M O R	2.34	D E K O Q R	2.50
B E L M O R	2.34	B E M N O R	2.50
G E L M O R	2.35	D E G K O R	2.52
E H M O Q R	2.35	A E H K O R	2.53
G E H M O R	2.36	B C E M N R	2.53
C E M O Q R	2.37	A E M O P R	2.55

[a] See Table 20 for solvent designations.

Table28.Summary of correlations for solvents in each problem

Problem	Correlations based on best combinations[a,b]	Correlations based on percentage tables: Redundant solvents[a]	Correlations based on percentage tables: Cofactor equivalencies[a]
ALDE	C = D = H = L, N = O = Q	N ~ O; A ~ I	E = 3; R = 4, K = 6, R ~ 6
KET1	C = D = H L, N = R, K = M = O, M = Q	N ~ O	R = 3, E = 4
KET2	C = D = H = L, N = O, F = G = J = M	N ~ O	E = 3 & 4 & 5, R = 4, O = 5, H & K = 6
CARB	C = D = H = L, L = P, A = B = J, J = Q	K ~ O	E = 3, N = 4, O & R = 5
MC33	C = D = H = L, A = B = G = Q	J = K = M **C = D = H = L**	E = 4, B & R = 5, N & O = 6
MC53	C = D = H = L, N = O, A = Q, F = G, K = M, G = L	F = G, K = M = O	E = 4, R = 5, B = 7

[a] Solvent designations taken from Table20. Solvents connected by "=" correlate very well; solvents connected by "~" correlate only moderately well.

[b] Based generally on best 30 combinations.

Table 29. Percentage of better reproductions involving typical solvent column vectors from data matrix for MC33 problem

Solvent[a]	Number of solvent vector cofactors used in reproduction					
	1	2	3	4	5	6
A	6	15	19	20	32	31
B	6	7	15	2	40	34
C	6	11	13	13	13	15
D	6	11	13	15	14	17
E	0	13	16	100	100	100
F	6	9	16	15	14	21
G	6	9	17	13	14	24
H	6	11	12	15	14	16
I	6	12	21	21	23	28
J	6	6	21	31	32	29
K	6	14	18	32	31	35
L	6	11	14	15	12	14
M	6	14	19	30	31	39
N	6	14	17	18	4	36
O	6	13	20	21	26	46
P	6	9	12	8	8	16
Q	6	14	18	13	13	19
R	6	7	19	23	81	88
Total number of combinations:	18	153	816	3060	8568	18654
Cutoff for better reproductions (rms)	130	59	36.9	9.91	5.61	3.91
Number reproductions less than cutoff:	17	100	500	200	200	200

(continued)

Table 29 (continued).

rms error, abstract reproduction:	65.6	24.4	15.8	5.2	2.8	1.9
Best combination:						
rms error:	63.9	26.2	16.9	5.8	3.1	2.2
vectors:	Q	G,Q	F,K,R	D,E,K, R	D,E,M, O,R	D,E,G, M,O,R

[a] Solvent designations taken from Table 20.

reproduction using six factors on MC33 has an rms error of 1.90. The cutoff for the six factor percentage column was selected as 3.91. In this case, 500 combinations had a rms less than 3.91 and the percentage of times a given solute occurred in these 500 best combinations is given in that column. Thus, solvent E occurred in all (100%) of the best 500 combinations of solvent vector cofactors for six factors in problem MC33. Solvent R occurred in 88% of the combinations; however, solvent L occurred in only 14%.

For a given number of factors the more important vectors are probably those which have the highest percentages. Thus solvents E and R are probably the most important solvent vectors in MC33. They most likely represent two of the cofactors in the MC33 problem. Solvent L, however, probably does not uniquely represent a solvent cofactor; other solvents can adequately represent the particular cofactor. To a lesser extent solvent O with a percentage of 46% in the six factor space also may represent adequately one of the cofactors.

In a similar type of analysis, Dupuis and Dijkstra[82] also demonstrated the uniqueness of solvent E in a combination of solvents using information theory. Similarly, Wold and Anderson[61], using abstract factor analysis, observed the highly unique character of solvent E.

The data in this table may enable us to associate a particular solvent with the abstract eigenvector.

If a particular solvent increases dramatically in percentage when going from a factor space to the next larger space, that solvent may be associated with the abstract eigenvector added to the space. Solvent E exhibits a marked increase in percentage (13% to 100%) when going from a three factor space to a four factor space. We may associate solvent E with the fourth factor. Similarly solvent L is associated with the fifth factor (23% to 81%). Solvent B may also be associated with the fifth factor (2% to 40%). Solvents N and O are both associated with the sixth factor. Thus, we may conclude that the fourth factor is the unique interaction associated with solvent E, diglycerol, which is probably hydrogen bonding. The fifth factor is probably a halogen-associated interaction (both solvents B and R are halogenated) and the sixth factor is probably associated with acetate groups on a sucrose backbone (i.e., solvents N and O).

The numerical patterns for solvents J, K and M are similar in going from one to six factors. This implies that the solvents J, K and M are equivalent; they behave similarly throughout the entire factor space in MC33. Exchanging these solvents should give similar results in the combination step. Similarly, the same conclusion can be made for solvents C, D, H and L. Also, as determined by comparing the best thirty combinations, they are essentially redundant solvents.

These results are summarized in the last two columns

of Table 28. Redundant solvents, as determined in pattern tables similar to Table 29 for MC33, are listed for each problem in column two; cofactor equivalences are listed in column three. Table 28 thus provides a list of the redundant solvents as determined by comparing the best combinations and analyzing the pattern tables. Substitution of one vector for another should provide similar results in a separation problem.

The uniqueness test, the solvent effect on reproduction and now, the combination of solvent vectors all exhibit the unique cofactor equivalency of solvents E and R. As listed in column three of Table 30, solvents E and R are equivalent to two of the cofactors. We have no other solvents which exhibit nearly as well the interactions associated with these two solvents.

The solvent sets giving the best combinations could make good representatives for "standard" phases. The solvent equivalences listed in Table 28 could provide substitute solvents for some of the "best" solvents.

D. Target Transformation of Basic Vectors

Physical data for the solvents are unfortunately unavailable. An attempt to obtain data from the manufacturers and distributors was fruitless. For example, Supina[84] of Supelco, Inc., advised us that since these compounds are usually of technical grade, physical data, even if available, would be entirely inaccurate. Therefore, we were forced to rely solely on such physical data vectors as molecular weight, molar refraction (calculated from individual group contributions) and the maximum operating temperature as recommended by the various distributors.[85]

Fortunately, we were able to test a host of structural vectors. These include uniqueness vectors testing for aromaticity, carbonyl, ether, phthalate, sucrose, halogen, and hydroxyl functionality, and individual solvent uniqueness. In addition, we devised "scaling" vectors which test for the value of a particular atom or functional group in the solvent, such as atom, carbon, oxygen and halogen atoms and hydroxyl groups.

We are especially interested in the McReynolds constants, based on ten functional probes.[55] As a single measurement of the total polarity, the sum of all ten constants or the first five constants has been proposed. McReynolds also reported the slope 'r' and intercept 'b' for each of the solutes. These fourteen vectors then are most important

in our analysis.

In the target transformation step as well, we can test the correlations obtained in the uniqueness tests. These "correlation" vectors may represent those specific interactions unique to the solvents represented in the vector. Additional vectors may be obtained by optimizing these vectors from their results on the target transformation test.

Listed in Table 30 are some selected target transformation vectors for problem MC33. The vectors with their designations are tabulated together with a qualitative evaluation for each vector. Thus, a molecular weight vector (designation:MW) target transformed well (qualitative evaluation = good); a halogen uniqueness vector (designation: UX) did not transform well (evaluation = poor).

Vectors tested included physical vectors (eg., molecular weight, molar refraction), structural vectors including uniqueness vectors (eg., aromatic uniqueness) and number vectors (i.e., number atoms per molecule) and correlation vectors derived from correlations observed in the uniqueness tests (eg., CV1, CV2). Actual results for five of the test vectors on MC33 are given in Table 31, including vectors which we evaluated as good (MR, CV1), very good (M7), fair (NO), and poor (NS).

As expected, the molar refraction of the solvents (MR) gave good results. Table 31 gives an example

Table 30. Summary of selected target transformations for physically significant solvent test vectors from MC33 problem, 6 factors

Symbol	Name of test vector	Example values[a]	Points[b]	Qualitative evaluation[c]
MW	molecular weight[d]	37,678	18	g
MR	molar refraction[e]	108,141	18	g
Ml	McReynold's constant-benzene	76,344	16	v
OT	maximum recommended operating temperature[g]	ff,ff	11	f
UNI	unity	1,1	18	v
NA	number of atoms[d]	68,85	18	g
NC	" " carbon atoms[d]	22,28	18	g
NH	" " hydrogen atoms	42,38	18	g
NOH	" " hydroxyl groups	0,0	18	f
NO	" " oxygen atoms	4,19	18	f
NS	" " side chains	2,ff	15	p
NX	" " halogen atoms	0,0	18	p
UA	aromatic uniqueness	0,0	18	g

(continued)

Table 30. (Continued)

UC	carbonyl	"	1,1	18	f
UD	dicarbonyl	"	1,ff	16	f
UE	ether linkage	"	0,1	18	f
UG	diglycerol	"	0,0	18	v
UN	nitrogen	"	0,0	18	p
UOH	hydroxyl	"	0,0	18	f
UP	pthalate	"	0,0	18	p
US	sucrose	"	0,1	18	f
UX	halogen	"	0,0	18	p
UZ	Zonyl E-7	"	0,0	18	g
CV1	correlation vector	1^h	0,0	18	g
CV2	"	" 2^h	0,1	18	p

[a] Examples are given for di-2-ethylhexyl adipate (solvent C) and SOA (solvent O); ff are free-floated values.

[b] Number of points on test vector.

(continued)

Table 30. (Continued)

[c] v-very good agreement between test vector and predicted vector; g-good agreement; f-fair agreement (several points poorly predicted or pattern only predicted); p-poor agreement.

[d] For Zonyl E-7, n taken as 4 (see Table 20).

[e] Molar refractions calculated from group contributions or from refractive index data.

[f] All of McReynold's constants, including 'b' and 'r', as well as the reciprocal, square and log of the vectors, tested as very good vectors. Designations for the other McReynold's constants: M2-butanol; M3-pentanone; M4-nitropropane; M5-pyridine; M6-methyl-pentanol; M7-iodobutane; M8-octyne; M9-dioxane; M10-hydrindane; Mb-constant 'b'; Mr-constant 'r'; MF-sum of M1 to M5; MT-sum of M1 to M10.

[g] Taken from references 85 and 87.

[h] See Table 34 for details.

Table 31. Details of selected target transformations for MC33 solvent problem, 6 factors

Solvent[a]	Molar refraction (MR) Test[b]	Molar refraction (MR) Predicted	McReynolds' constant iodobutane Test[b]	McReynolds' constant iodobutane Predicted	No. oxygen (NO) Test	No. oxygen (NO) Predicted	No. of side chains (NS) Test[b]	No. of side chains (NS) Predicted	Correlation vector CV1 Test	Correlation vector CV1 predicted
A	80	96	190	183	6.0	5.8	0.0	-.2	1.0	0.9
B	(96)	86		141	4.0	4.6	.0	-.1	1.0	.9
C	107	140	71	68	4.0	5.2	2.0	1.4	0.0	-.1
D	126	133	68	69	4.0	4.8	2.0	1.4	.0	.1
E	38	39	245	248	5.0	5.1	.0	.0	.0	.0
F	135	109	83	86	4.0	2.4	2.0	1.1	.5	.5
G	113	103	92	96	4.0	2.3	.0	-.9	.5	.6
H	126	131	68	70	4.0	4.7	.0	1.3	.0	.1
I	139	127	98	90	5.0	3.4	3.0	1.4	.0	.2
J	141	132	82	77	1.0	2.4	1.0	1.8	.0	.2
K	204	163	310	287	19.0	16.7		-.8	.0	.1
L	120	135	(72)	69	4.0	4.9	1.0	1.4	.0	.0
M	76	119	208	233	4.0	6.7	.0	.1	1.0	.8
N	192	180	147	152	19.0	14.7		.2	.0	-.4
O	141	160	292	300	19.0	21.5		-2.1	.0	.2
P	(161)	154	77	85	6.0	8.3	1.0	1.1	.0	-.1
Q	103	77	169	159	4.0	2.0	.0	.2	1.0	1.0
R	135	133	146	144	8.0	8.8	.0	.2	.0	.1

[a]See Table 20 for designations

[b]Blank points and known values in parenthesis were free-floated.

target transformation for MR on MC33. Two points were free-floated as evidenced by the parenthesis. Their predicted values are in excellent agreement with the known calculated values, especially considering that the input molar refractions were calculated from group contributions.

The entire set of fourteen vectors from McReynolds' compilation (M1 through M10, Mb, Mr, MF and MT) gave excellent agreement between predicted and tested vectors. The second vector in Table 31 is an example for the iodobutane constant on MC33. Testing of the log, reciprocal and square of each of these fourteen vectors showed these functional forms to give excellent results as well.

For the fourteen McReynolds Constants vectors tested, data points for solvents B and L were free-floated. Apparently, solvent B was not included in McReynold's probe compilation, whereas the identity of solvent L on the compilation is not clear. It is listed as octyl-decyladipate. We were not sure it was unambiguously the same as solvent L (n-octyldecyladipate).

As can be seen in Table 31, the predicted value for solvent L is 69 for the McReynolds iodobutane solute probe (M7). This compares very favorably with the published value for isobutane on octyldecyladipate of 72. In fact, the predicted value for solvent L for all McReynolds constants is in very good agreement ($\pm$7 RI units) with those published for octyldecyladipate. Subsequent to this analysis McReynolds[81] advised us that

Solvent L is indeed the octyldecyladipate listed in McReynold's compilation.

Tables 32 and 33 list the predicted values for the McReynolds constants for solvents B and L on all the problems. The average error for the predicted values on solvent L range from ±2.5 RI units for problem MC33 to ±12.2 RI units for problem ALDE.

The excellent agreement for these constants on solvent L with the published values gives us great confidence in the predicted values for solvent B, though we have no published data to compare them with. However, some predicted values on the MC33 and MC53 data sets are predicted quite differently as compared to predictions derived from the other data sets. This may be due to special interactions between the chlorines of solvent B and the particular probes functionality. For example, constant iodobutane (M7) is predicted higher on MC33 than for the other five problem sets. Probably there is a special interaction between solvent B and iodobutane which is picked up by the increased functionality of the MC33 and MC53 problems.

Other vectors which tested well include uniqueness vectors for solvents diglycerol (UG) and Zonyl E7 (UZ). This is not surprising considering their highly unique characterization in the uniqueness test. In addition, correlation vectors obtained from the uniquenss test (CV1, for example) also tested quite well.

Table 32. Predicted values for the McReynold's vectors for solvent B

Vector[a]	Problem					
	ALDE	KET1	KET2	CARB	MC33	MC53
M1	110	116	113	117	144	157
M2	222	230	235	227	190	225
M3	178	191	186	188	153	192
M4	278	283	282	292	293	283
M5	181	202	194	185	169	283
M6	174	189	189	181	139	167
M7	105	109	109	110	141	158
M8	81	90	87	89	120	121
M9	153	183	171	163	160	232
M10	29	48	42	34	47	92
Mb	.269	.244	.250	.259	.280	.266
Mr	1.84	1.72	1.75	1.79	1.90	1.85

[a] See Table 30 for vector designations.

Table 33. Predicted values for the McReynold's vectors for solvent L

	Problem						
Vector[a]	ALDE	KET1	KET2	CARB	MC33	MC53	Lit.[b]
M1	60	70	76	70	76	78	79
M2	175	158	163	165	173	176	179
M3	109	119	118	115	118	122	119
M4	173	194	192	189	197	197	193
M5	123	116	130	123	127	137	134
M6	136	126	131	130	136	140	141
M7	57	63	68	64	69	71	72
M8	35	45	51	46	50	51	57
M9	100	105	118	11	113	121	119
M10	1	11	9	5	6	11	10
Mr	.287	.284	.285	.283	.285	.283	.284
Mb	1.93	1.93	1.93	1.92	1.93	1.91	1.92
average error	12.2	9.9	4.7	8.2	4.5	2.5	

[a] See Table 30 for vector designations.

[b] From reference 55 .

The oxygen number vector (NO) as illustrated in Table 31 was evaluated as giving only a fair fit. This is because the individual data points were not predicted well, though the general trend of the data for this vector is followed. Thus, those solvents with large values on this vector are generally predicted to be large values.

Other vectors which transformed only fairly included hydroxyl number (NOH), oxygen number (NO), carbonyl uniqueness (UC), ether uniqueness (UE), hydroxyl uniqueness (UOH), sucrose uniqueness (US) and maximum operating temperatures (OT). The maximum operating temperature is our nearest estimation of the boiling point for the solvents. Since sucrose uniqueness (US) tested only fair, we conclude that the sucrose functionality is not a true factor. This was consistent with results from the uniquenss test correlations.

Vectors which tested poorly included side chain number (NS) which is also illustrated in Table 31. It is quite apparent that it gives a very poor fit in the transformation.

Summarized in Table 34 are the vectors which target transformed with best agreement for each of the problem sets. Some vectors target transformed well on all the problem sets. These include the McReynolds' constants vectors (MI-MIO, Mr, Mb, Mf, MT). All the McReynolds' constants represent at least one of the cofactors. Whether all the cofactors of a problem are represented solely by the McReynolds' constants will be determined

Table34.Summary of physical and structural vectors target transforming with best agreement for each problem -solvents

Problem	Test vectors[a]
All 6	M1-M10; MF; MT; Mr; Mb; MR; UEt; UG; UOH; UNI;US; UZ; logs, reciprocals and squares of M1-M10, MF and MT.
ALDE	NC; NOH; UET; CV2; CV4; CV5; CV8; CV13; CV14; CV15;CV16.
KET1	NC; CV2; CV7; CV19.
KET2	MW; NA; NC; NH; NOH; CV1; CV4; CV5; CV7; CV17.
CARB	MW; NA; NH; NO; NOH; UET; CV1; CV3; CV5; CV7; CV18.
MC33	MW; NA; NC; NH; UA; UC; UD;UET; UP;CV1; CV7; CV9; CV10; CV11.
MC53	NA; RMR; UA; UD; UET; UN; UP; UX; CV1; CV5; CV6; CV7; CV13.

(continued)

Table 34. (Continued)

[a] See Table 30 for designations. Points on the correlation vectors (CV) are, in most cases, "0"'s or "1"'s. For each correlation vector, test points having non-zero values are listed below. If the value for a given solvent is other than "1", the value is given in parenthesis. Test points not designated are all "0"'s.

CV1: A,B,F(.5),G(.5),M,Q; CV2: A,K,M,Q; CV3: N,O; CV4: K,M;

CV5: A,B,I,J,Q; CV6: A,K(.5),M,O,Q; CV7: B(-.5),I,J,M,O(-1);

CV8: A,K,M,Q,R(-2); CV9: K,M,N,O; CV10: A(.5),J(.5),K,M,Q(.5);

CV11: A,B,G(.5),O,Q; CV12: G(.5),N,O; CV13: B,K,M; CV14: K,M,O;

CV15: E(.5),N,O; CV16: B(4),I(1.8),M(-3.3),N(.8),R(12);

CV17: B(4),I(.8),M(2.3),N(2.3),R(12); CV18: B(4),I,N,R(12);

CV19: B(4),I(2),N(2),R(12).

in the combination step. However, every one of the constants (MI through MIO) as well as their sums (MF, MT) and slopes (Mr) and intercept (Mb) target transform extremely well on every problem set.

Other vectors which target transform in every problem set include molar refraction (MR), uniqueness vectors for diglycerol (UG), Zonyl E7 (UZ), hydroxyl functionality (UOH) and sucrose functionality (UZ)and a vector involving the recommended maximum operating temperature (OT) for each of the solvents.

Other vectors which tested well in most problems included number of carbons (NC), number of atoms (NA) and correlation vectors CVI, CV5, and CV7. The correlation vectors were constructed originally from the correlations obtained in the uniqueness step. However, they were "optimized", i.e., modified, to provide the best target transformation. The chemical meaning of these correlation vectors is harder to ascertain. For example, vector CVI correlates pthalates (solvents A, B, F, and G) with quadrol (M) and tricresol phosphate (Q).

The molecular weight of the solvents which according to Martire[86] should be a factor was a good test vector (MW) only for problem sets KET2, CARB, and MC33.

E. Combination of Basic Vectors

What is the physical nature of the separation process? Of the interactions involved can any of them be related with a specific solvent? How exclusive are the McReynolds' constants? Are there better solvent probes? Do the constants adequately represent all the interactions that contribute to a chemical separation? Are any redundant?

Combination of the best physical and structural vectors obtained in the target transformation step (listed in Table 34) may provide answers to these questions. Listed in Table 35 are the results of this combination step for all the problems for each factor space.

Not all of the vectors were included in the final combination. Combinations using six factors, for example, of only 25 vectors would involve 177,100 combinations; combinations using six factors on the 37 vectors listed for MC33 (ignoring the functional forms for the McReynolds' constant) would involve 2,324,784 combinations! The computer time involved precluded attempting all the vector combinations. Thus, the following procedure was used: The McReynolds' constant vectors which target transformed best for each problem were included in the combinations. At least six of these vectors were kept in the combinations for each problem, including two or three functional forms of the vectors (e.g., log, reciprocal or square).

A first run was obtained using the above McReynold's vectors and solvent vectors selected from the remaining

Table 35 Summary of best combinations utilizing best physically significant solvent cofactors for each problem, 1-7 factors.

Problem[a]	Number of real vector cofactors used in the reproduction[b] 1	2	3	4	5	6	7	Vectors in ten best combinations
ALDE (21)	83.5 LMF	28.4 LMF,M3	13.5 LMF,UG,M3	9.3 UNI,NX,UG, M3	8.4 LMF,UG,M3, CV5,CV14	5.9(31.2) LMF,Mb,NO, UG,UZ,CV5		LMF,M1,M3,M7,M8,M9,Mb,NO,UG,US, UZ,CV5,CV8,CV14,CV16
KET1 (20)	47.6 LM5	21.6 SM3,LM5	17.3 LM5,NX,SM3	10.2(44.8) LM5,M7,SM3, UG				M1,M3,M5,M7,M10,MF,MT,LM5,LMF, SM3,UG
KET2 (23)	82.1 LMF	22.1 M9,LMF	21.5 LMF,M9,CV4	10.1 LMF,M1,M9, M10	9.1 M1,M5,NOH, UNI,UG	6.0(34.3)[c] M4,Mb,UOH, CV4,CV5,CV17		M1,M4,M10,MF,Mb,Mr,MR,NUH,UOH, UNI,UG,UZ,CV4,CV5,CV17
CARB (21)	82.3 LMF	24.5 LMF,M5	14.5 LMF,M3,UG	14.0 LMF,UG,UOH, M3	9.1 M1,M5,NOH, UG,UNI	6.5(32.3) LMF,M4,UG, UOH,UZ,CV5		M1,M3,M4,M5,M7,M8,M10,Mr,LMF, UG,UOH,UZ,CV3,CV5
MC33 (19)	67.4 LMF	27.8 MF,UNI	19.4 M2,M3,UNI	10.5 M2,M3,UG, UNI	8.6 M1,M2,MF, UG,UNI	6.4(37.) M2,M7,Mb, UG,UNI,CV1		M1,M2,M7,M8,M9,Mr,Mb,UNI,UG,CV1
MC53 (18)	123.0 UNI	30.9 MF,UNI	19.9 M2,M3,UNI	12.1 M2,M3,UG, UNI	10.1 M2,M3,MF, UG,UNI	8.8 M2,M6,M7, UG,UNI,CV1	6.3(30.2) M2,M6,M7,MF, UNI,UZ,CV1	M2,M3,M6,M7,Mr,Mb,MF,SMF,UNI,UG, UZ,CV1,CV5

[a] Lower line - number of real vectors used in final combination step.

[b] Upper line - rms error (largest error in parenthesis for true factor space);
lower line - real vectors occurrring in best combination. See Table 30 for vectors.

[c] For KET2 problem: Vector sets: M4,Mb,CV5,CV17,UOH,UG and M4,Mb,CV5,CV17,UG,CV4 gave the same results.

target vectors to make a total of 18-20 vectors. Solvent physical vectors were deleted after this first run if they were found to be unimportant and new vectors were added to the modified vector set for a second combination run. This was repeated until all the vectors had been tested in the combination step. Column one of Table 35 lists the number of vectors that remained for the final run of combinations for each problem (in parenthesis).

In general, the best combination set for each problem contains two to four of the McReynold's constants, a correlation vector, and at least one uniqueness vector. For example, for problem MC33, the best combination set using six factors, includes three McReynolds' vectors (M2, M7, M6), unity (UNI), diglycerol uniqueness (UG) and correlation vector CV1 and reproduces the original data with an rms error of 6.4 with the largest error 37.0 RI units.

The dominance of the unique interactions associated with solvents E and R is apparent since the uniqueness vectors testing for their interactions, UG and UZ, are major constitents of the best combinations sets for each problem.

Some correlations obtained in the uniqueness test step are apparently associated with at least one of the abstract factors. Also, they are apparently not associated with any of the physical or structural vectors tested. Thus, the correlation vectors are necessary to account for at least one of the cofactors in most of the problems.

Unity is an important vector cofactor for problems MC33 and MC53. This indicates that one of the interactions in these problems is constant and probably depends solely on solute structure, as opposed to solvent structures. This may be the cofactor associated with solute carbon number. Because of the definition of the retention index, we expect the corresponding solvent cofactor to be unity.

We also obtained combinations using only the McReynold's constants (M1-M10) in order to ascertain whether the constants represent all the key cofactors as has been proposed (see Chapter IV). However, the McReynolds' constants by themselves do not provide a satisfactory combination. For example, the best combination of the constants (M1-M10) yielded an rms error of 79.4 on the ALDE problem set. Addition of constants Mr and MB does enhance the reproduction (rms = 9.9 for problem ALDE). Thus as a measure of all the interactions within the separation system the ten functional probes proposed by McReynolds' are inadequate. In fact, many of them are redundant as described below. However, the dominance of the McReynolds' constants in the best combination sets for each problem does indicate that they do represent some of the cofactors. However, any unique interactions inherent in a particular data set must be represented by one of the other vectors (for example, hydrogen bonding by vector UZ).

Those vectors which were represented in the ten best combinations are listed in the last column of Table 35 for

each problem. For example, ten vectors including seven McReynolds' type vectors, unity (UNI), diglycerol uniqueness (UG) and CV1 were the only vectors present in the ten best combinations for problem MC33. These ten, then, are the most significant vectors in MC33.

Column two of Table 36 lists the equivalences of the real vectors obtained by comparing the best thirty combinations for each problem. It is apparent that most of the McReynolds' constants are redundant in most of the problems. Thus it was found that in problem MC33, for example, M1 and M8 were equivalent vectors; either M1 or M8 represented the same cofactor. The aromatic ring in M1 may be slightly equivalent to M8's alkyne bond. Similarly, in MC53, M2 and M6, both alcohols, are found equivalent. In fact, M2 is equivalent to UG, the diglycerol uniqueness, probably due to the -OH functionality.

It must be pointed out that since the best reproductions we have obtained using the solvent physical and structural vectors, though excellent, are not exact reproductions, these correlations among the vectors we observe upon comparison of the best ten combinations may not be exact. We are probably missing a key vector which would best represent one of the solvent cofactors.

Table 37 is a pattern table for some of the vectors used in the best combination for problem MC33. The equivalence of vectors Mb and Mr is apparent from their patterns in going from a factor space of one to a space

Table 36. Summary of correlations for physically significant solvent cofactors in each problem

Problem	Correlations based on best combinations[a]	Correlations based on percentage tables: Redundant vectors[a]	Correlations based on percentage tables: Cofactor equivalencies[a]
ALDE	M1 = M3 = UNI, M1 = M8, M7 = UZ = CV8 = CV14 = CV16, M1 = M9 = NO = US, CV5 = CV14.	Mb = Mr ~ UNI	M3 = 3 & 5, UG = 4, CV5 & CV16 = 6
KET1	M1 = M5 = M7 = M10 = MF = MT, M3 = MF = SM3, M7 = CV19.	LMF = LM5	M3 & UZ & CV18 = 3, UG = 4
KET2	UG = M4 = M10 = CV4, MF = UZ = CV4, Mb = Mr, UG & UOH = CV4 & UOH = UG & CV4.	M8 = M10, NOH = UOH	UNI & LMF = 3, CV17 = 5, M4 & CV4 = 6
CARB	M1 = M5 = M7 = M8 = UZ, LMF = Mr = Mb, M1 = UG, CV3 & UOH = CV5 & LMF.	M4 = M5, M1 = M8, Mb = Mr	M3 & UZ = 3, LMF & CV5 = 6
MC33	M7 = M8 = M9 = MF = UG = CV11, M1 = M7 = M8, Mr = Mb, NA = MH = CV1, Mr & CV1 = LMF & any MC, Mr = Mb = UNI.	M1 = M8, Mb = Mr, UNI ~ M2	M2 & M7 = 4, M1 = 5, UNI = 6
MC53	M3 = M7 = MF, MF = CV6, M2 = M3 = M6 = CV7, Mr = Mb = M6 = CV1, UG = M2.	Mr = Mb, M2 = M6, M2 = UG	M7 & CV7 = 5, CV1 = 6, SMF = 7

[a] See Tables 30 & 34 for designations; MC = general designation for all ten McReynold's constants. Symbol "&" shows vector pairs.

Table 37. Percentage of better reproductions containing specific real solvent cofactors for MC33 problem

Physical vector[a]	Number of real vectors (cofactors) used in reproduction					
	1	2	3	4	5	6
UNI	25	28	40	45	49	89
UG	0	3	21	29	41	40
M1	0	11	15	9	28	34
M2	0	11	24	51	60	81
M3	0	11	22	30	39	46
M7	0	11	2	26	33	23
M8	0	11	8	14	29	28
M9	0	11	22	29	27	32
LM5	25	33	44	39	25	13
Mb	25	28	27	30	39	36
Mr	25	28	29	33	39	41
CV1	0	0	0	0	1	18
CV7	0	0	8	7	18	16

(continued)

Table 37.(continued)

Total number of combinations	19	171	969	3876	11628	27132
Cutoff for better reproductions (rms error)	300.	100.	51.8	33.2	17.4	10.0
Number of reproductions less than cutoff	4	36	100	200	500	811
rms error, abstract reproduction	**65.6**	**24.4**	**15.8**	**5.2**	**2.8**	**1.9**
Best combination :						
rms error	67.4	27.8	19.4	10.5	8.6	6.4
vectors	LM5	UNI,MF	UNI,M2, M3	UG,UNI, M2,M3	UG,UNI, M1,M2,MF	UG,UNI,M2 M7,Mb,CV1

[a] Vector designations taken from Table 30.

of six. There is some correlation in the patterns of vectors UNI and M2. Vector UNI jumps in percentage almost 40% from factor five to six. We conclude that UNI most probably represents cofactor six. Similarly, vector M2 probably represents cofactor four.

These results as well as the results for pattern tables for all the problems are summarized in the last two columns of Table 36. As observed in comparison of the best ten combinations for each problem, many of the McReynolds' constants are indeed redundant.

Summarizing the results of combinations of solvent physical and structural vectors, we find most significantly that the McReynolds' constants do not account for all the interactions in the six problems we have studied. Many of the solute probes used in the derivation of these constants are probably not sufficiently unique to be representative of one cofactor or interaction. However, some of these constants can be associated with specific interaction cofactors.

F. Prediction of new data

The absolute test to determine how well we have succeeded in determining the best combination set that represents the real cofactors for our problems is to use them in predicting retention indices on new solute-solvent pairs. As described in Chapter II, each combination set has a column matrix [C], each column of which is associated with each of the solutes for the problem. The m elements of each column are associated with the m cofactors for the combination. Using Equation 30:

$$d_{xk} = \sum^{m} r_{xm} c_{mk} \qquad (60)$$

where d_{xk} is the predicted retention index for new solvent x on solute k, r_{xm} is the value for cofactor m for the solvent, and c_{mk} is the element of the [C] matrix for cofactor m on solute k, we may predict retention indices for solvents other than those used in the data matrices of each problem if key data r_{xm} are available.

For example, tabulated in Table 38 is the column matrix [C] for the best combination of real solvent cofactors on problem MC33: CV1, UG, UNI, M2, M7 and Mb. The retention index of solute OL1 (ethanol) on a new solvent k would then be:

$$d_{OL1,k} = (CV1)_k\ 9.44 + (UG)_k 124.8 + (UNI)_k 367.4 + (M2)_k 1.16 + (M7)_k(-.36) + (Mb)_k(.93) \qquad (61)$$

Table 38. Column matrix [C] for best combination set, MC33 problem

Key Solvents / Cofactor	Solutes									
	OL1	OL2	OL3	OL4	OL5	A3	A4	A6	A11	
CV1	0.9435E 01	0.5499E 01	0.1322E 02	0.1767E 02	0.2393E 02	0.4326E 02	0.4242E 02	0.4294E 02	0.5101E 02	
UG	0.1248E 03	0.4211E 02	-0.1051E 03	-0.1685E 03	-0.3105E 03	0.3706E 03	0.3183E 03	0.2473E 03	0.1371E 03	
UNI	0.3674E 03	0.4176E 03	0.6567E 03	0.7775E 03	0.9678E 03	0.1430E 04	0.1593E 04	0.1714E 04	0.1969E 04	
M2	0.1164E 01	0.1148E 01	0.1214E 01	0.1226E 01	0.1302E 01	0.2261E 00	0.1839E 00	0.2154E 00	0.2380E 00	
M7	-0.3563E 00	-0.2691E 00	-0.3864E 00	-0.4103E 00	-0.5111E 00	-0.1793E 00	-0.1724E 00	-0.2179E 00	-0.3173E 00	
Mb	0.9322E 00	0.2039E 03	0.1035E 03	0.3726E 02	0.6582E 02	-0.3204E 04	-0.3409E 04	-0.3473E 04	-0.3649E 04	

	K5	K6	K16	ES1	ES2	ES3	ES6	ES7	ET1
CV1	0.4917E 02	0.5712E 02	0.6782E 02	0.3018E 02	0.3964E 02	0.4176E 02	0.2783E 02	0.4149E 02	0.2688E 02
UG	0.3455E 03	0.2285E 03	0.1202E 03	0.3360E 03	0.1551E 03	0.1144E 03	0.2379E 03	0.1088E 03	0.2661E 03
UNI	0.1823E 04	0.2088E 04	0.2349E 04	0.1804E 04	0.2161E 04	0.2291E 04	0.1770E 04	0.2197E 04	0.1062E 04
M2	0.2233E 00	0.2290E 00	0.2554E 00	0.3993E-01	0.8917E-01	0.9481E-01	0.9170E-01	0.7632E-01	0.1827E-01
M7	-0.3504E 00	-0.4691E 00	-0.5655E 00	-0.7349E-01	-0.2173E 00	-0.2402E 00	-0.1015E 00	-0.2046E 00	-0.7609E-01
Mb	-0.4185E 04	-0.4415E 04	-0.4620E 04	-0.4115E 04	-0.4308E 04	-0.4411E 04	-0.3705E 04	-0.4127E 04	-0.1926E 04

	ET2	ET3	ET5	OX1	OX2	OX3	AR1	AR2	AR3
CV1	0.1691E 02	0.1251E 02	0.3174E 02	0.3231E 02	0.3067E 02	0.2942E 02	0.6184E 01	0.1398E 02	0.2374E 02
UG	0.8916E 02	0.8607E 02	-0.3296E 02	0.4147E 03	0.3679E 03	0.2852E 03	0.1477E 03	0.1184E 03	0.9454E 02
UNI	0.1065E 04	0.9816E 03	0.1439E 04	0.1224E 04	0.1449E 04	0.1457E 04	0.9946E 03	0.1153E 04	0.1419E 04
M2	0.8996E-01	0.7553E-01	0.1438E 00	0.1587E 00	0.1194E 00	0.1535E 00	0.1837E-01	0.2217E-01	0.1240E-02
M7	-0.2289E-01	0.2950E-01	-0.1563E 00	0.3941E-01	-0.4811E-01	-0.1889E-01	0.7286E 00	0.6570E 00	0.4863E 00
Mb	-0.1331E 04	-0.1078E 04	-0.1305E 04	-0.2811E 04	-0.3309E 04	-0.3006E 04	-0.1143E 04	-0.1322E 04	-0.1845E 04

	AR4	AL1	AL2	AL3	AL4	AL5
CV1	0.1291E 02	0.1186E-01	0.9539E-01	0.1732E 00	0.9664E-01	0.4740E 00
UG	0.6944E 02	-0.1127E-01	-0.1620E-01	-0.1475E-01	-0.1072E 00	-0.4135E-01
UNI	0.1248E 04	0.2001E 03	0.4003E 03	0.6005E 03	0.1401E 04	0.1601E 04
M2	0.3139E-01	0.9839E-04	0.3352E-03	0.4815E-03	0.6883E-03	0.1303E-02
M7	0.6398E 00	-0.3509E-03	-0.1263E-02	-0.1846E-02	-0.2290E-02	-0.4997E-02
Mb	-0.1341E 04	-0.3098E 00	-0.1175E 01	-0.1825E 01	-0.1989E 01	-0.4949E 01

Thus, for a new solvent, Carbowax 20M, this equation reduces to:

$$d_{OL1,\ Carb.} = (0)\cdot(9.44) + (0)\cdot(124.8) + (1)\cdot(367.4) + (5.36)\cdot(1.16) + (282)\cdot(-.36) + (.2255)\cdot(.93) = 887.8$$

since from Reference 80, M2 =536, M7=282 and Mb=.2255 for Carbowax 20 M. The reported value for the retention index of ethanol on Carbowax 20M is 893.

Listed in Table 39 are the predicted retention index values for nine representative solutes on four new stationary phases for problem MC33 with the reported values. These phases were chosen because they span the gamut of polarities for SE 30 (highly non-polar) to Igepal CO880 (very polar). The results are quite good for most of the solutes considering that we are probably lacking at least one major cofactor in the best combination set. Surprisingly, predictions for the polar solvents, Carbowax 20M and Igepal CO880, are very good for most points, whereas, predictions on the relatively less-polar solvents Apiezon M and SE30 are generally poorer. We can conclude from this that the best combination set adequately spans the dominant interactions involved in the more polar solvents. The non-polar solvents, where we expect the weaker non-polar interactions to be the most dominant, are probably not adequately represented in the best combinations set. The missing cofactor(s) is (are) thus probably associated with the non-polar interactions.

Table 39. Prediction of retention data for nine solutes on four stationary phases for problem MC33

Solute	Calculated and literature retention indices for solvent[a] SE 30		Carbowax 20M		Apiezon M		Igepal C@880	
	Calc.	Lit.	Calc.	Lit.	Calc.	Lit.	Calc.	Lit.
ethanol (OL1)	428	460	888	893	382	412	820	831
pentanal (A6)	858	696	992	984	936	670	838	937
2-hexanone (K6)	997	791	1092	1087	1091	756	1021	1032
pentylacetate (ES2)	1090	842	1184	1177	936	790	1113	1120
pentylpropionate (ES7)	1189	998	1258	1248	1181	954	1189	1191
propyl ether (ET3)	716	687	758	773	679	660	761	761
propylene oxide (OX2)	630	488	760	763	513	461	694	704
benzene (AR1)	720	678	1041	961	698	685	967	906
hexane (AL3)	600	600	600	600	600	600	600	600
average error	115		15		98		22	

[a] Literature values taken from reference 80. Calculated values obtained using equation 60 and [C] matrix from Table 38.

The coefficients of the [C] matrix for the alkane solute AL1 through AL5 are all near zero except for cofactor UNI, which upon observation are equivalent to the carbon number for the solute. Thus, predictions for the alkane solutes are perfect (illustrated in Table 40 for solute hexane). All other solutes, however, have at least four significant coefficients in the [C] matrix thus requiring at least four cofactors.

CHAPTER VIII

TARGET FACTOR ANALYSIS OF GAS-LIQUID CHROMATOGRAPHIC RETENTION INDICES OF MONOMERIC SOLVENTS AND SIX SOLUTE SETS-- III. CHARACTERIZATION OF SOLUTES

As described in Chapter VI, six solvent-solute sets were derived from McReynolds'[80] compilation and their respective factor spaces determined. In Chapter VII, TFA was used to characterize the solvents with very good results as evidenced in the satisfactory prediction of new solvent-solute retention indices. In this chapter, we will describe the characterization of the solute cofactors.

A. Uniqueness Test--Solutes:

Tables 40 and 41 summarize the results of the uniqueness test for the solutes for the six problems. For the ALDE problem, the most unique solutes were acetaldehyde (designated A2, uniqueness value=.913), 2,2-dimethylpropanol (A9, .705) and 2,4-hexadienal (A18, .727). These results are not surprising: A2 is the smallest solute with the aldehyde functionality, A9 probably shows the most steric hindranee and A18 is the only diene. Comparing the most correlated solutes, we notice correlations

Table 40. Uniqueness tests for solutes in MC33 and MC53 problems.

Solute[a]		MC33 problem, 6 factor		MC53 problem, 7 factor	
Symbol	Name	Uniqueness test value	Correlates with solutes[b]	Uniqueness test value	Correlates with solutes[b]
OL1	ethanol	.29	OL2 (.26),OL3	.17	
OL2	propanol	.24	OL1(.26),OL3,OL4	.13	
OL3	pentanol	.20	OL1,OL2,OL4,OL5	.12	
OL4	hexanol	.22	OL2,OL3,)L5(.26)	.15	OL5
OL5	octanol	.36	OL3,OL4(.26)	.28	OL4
OL6	isopropanol*			.13	
OL7	2-butanol*			.12	
OL8	2-methyl-1-butanol*			.11	
OL9	2-methyl-1-pentanol*			.11	
OL10	2,2-dimethyl-1-propanol*			.16	
A3	propanol	.17	OX1	.18	
A4	butanal	.10		.11	
A5	isobutanal*			.07	
A6	pentanal	.06		.07	
A7	isopentanal*			.04	
A9	2,2-dimethylpropanal*			.05	
A11	heptanal	.11	K16	.10	
K2	2-butanone	.12		.08	
K6	2-hexanone	.14	K16	.09	
K8	3-methyl-2-pentanone*			.12	
K10	3,3-dimethyl-2-butanone*			.13	
K16	2-octanone	.30	A11,K6,ET5	.23	
K28	3-methyl-2-butanone*			.09	
AC1	ethyl acetate	.22	AC2,PR1	.14	
AC2	pentyl acetate	.19	AC1,AC3,PR1,PR2	.10	
AC3	hexyl acetate	.23	AC2,PR1,PR2	.12	
AC4	isobytyl acetate*			.16	
AC5	2-methyl-2-butyl acetate*			.10	
PR1	ethyl propionate	.19	AC1, AC2,AC3	.12	

(continued)

Table 40. (continued)

PR2	pentyl propionate	.17	AC2,AC3	.12	
PR3	isobutyl propionate*			.08	
PR4	isopentyl propionate*			.09	
ET1	ethyl ether	.12		.08	
ET2	butyl ethyl ether	.03		.03	
ET3	propyl ether	.03		.04	
ET5	pentyl ether	.22	K16	.20	
ET6	isopropyl propyl ether*			.12	
ET7	isopropyl ether*			.06	
OX1	ethylene oxide	.19	A3	.15	
OX2	propylene oxide	.12		.09	
OX3	1,2-butylene oxide	.07		.05	
AR1	benzene	.28	AR2(.25),AR3,AR4(.25)	.17	AR2
AR2	toluene	.26	AR1(.25),AR3,AR4(.25)	.17	AR1,AR3,AR5
AR3	p-xylene	.22	AR1,AR2,AR4	.27	AR2,AR5(.28)
AR4	ethylbenzene	.24	AR1(.25),AR2(.25),AR3	.14	
AR5	o-xylene*			.29	AR2,AR3(.28)
AR6	o-diethylbenzene*			.22	AR7
AR7	p-diethylbenzene*			.24	AR6
AL1	ethane	.01		.01	
AL2	butane	.03		.03	
AL3	hexane	.06	AL5	.06	AL5
AL4	tetradecane	.35	AL5(.40)	.33	AL5(.38)
AL5	hexadecane	.45	AL3,AL4(140)	.43	AL3,AL5(.38)

[a] Solutes with superscript,*, in MC53 problem only.

[b] Other solutes with predicted values greater than 0.15 are listed. Predicted values greater than 0.25 are given in parenthesis.

Table 41. Uniqueness tests for solutes in ALDE, KET2 and CARB problems

Solute[a]		Uniqueness test value	Correlates with solutes[b]	Uniqueness test value	Correlates with solutes[b]
Symbol	Name	CARB problem, 6 factor		ALDE problem, 6 factor	
A2	acetaldehyde	.51	A13,K1	.91	A3
A3	propanal	.24	A4	.39	A2,A6,A7,A11
A4	butanal	.11	A3	.15	A13
A5	isobutanal	.08		.20	A9(.25)
A6	pentanal	.12		.16	A3,A11
A7	isopentanal	.09		.27	A3,A9
A9	2,2-dimethyl propanal	.13		.71	A5,(.25),A7,A15
A10	hexanal	.10		.14	A11
A11	heptanal	.16		.31	A3,A6,A10,A12,A17
A12	2-ethyl hexanal	.17	K18	.49	A11,A16,A17(.27)
A13	acrolein	.17	A2	.45	A4,A14(.26)A15
A14	methacrolein	.08		.19	A13(.26)
A15	crotonaldehyde	.12	A18	.38	A9,A13,A18(.28)
A16	2-ethyl-2-butenal	.08		.18	A12,A18
A17	2-ethyl-2-hexenal	.11		.33	A11,A12 (.27),A18
A18	2,4-hexadional	.39	A14,K20	.73	A15,A16,A17
				KET2 problem, 6 factor	
K1	acetone*	.20	A2	.44	K2,K27(.28)
K2	2-butanone*	.09		.14	K1
K3	2-pentanone*	.08		.10	
K4	3-pentanone*	.05		.18	K12
K5	3-methyl-2-butanone*	.10		.12	K10
K6	2-hexanone*	.05		.07	
K7	3-hexanone*	.04		.10	K18
K8	3-methyl-2-pentanone*	.09		.11	K10
K9	2-methyl-2-pentanone*	.10		.12	K10
K10	3,3-dimethyl-2-butanone*	.20		.34	K5,K8,K9

(continued)

Table 41. (Continued)

K11	2-heptanone	.06		.15	K16,K17
K12	3-heptanone	.07		.24	K4,K17(.26)
K15	2,4-dimethyl-3-pentanone*	.10		.13	
K16	2-octanone*	.11		.32	K11,K17
K17	2-nonanone*	.17		.25	K11,K16,K19
K18	5-nonanone*	.18	A12	.40	K7,K12(.26)K17
K19	cyclopentanone	.25	K20(.25)	.34	K20(.36),K24
K20	cyclohexanone	.29	A18,K19(.25)	.45	K19(.36),K24
K22	5-hexen-2-one	.09		.15	K27
K24	4-methyl-3-penten-2-one	.09		.32	K19,K20
K25	2,3-butadione	.47	K26(.41)	.50	K26(.45)
K26	2,9-pentadione	.39	K25(.41)	.46	K25(.45)
K27	2,4-pentadione	.11		.58	K1(.28),K22

[a]Solutes with superscript,*, also in KET1 problem.

[b]Other solutes with predicted values greater than 0.15 are listed. Predicted values greater than 0.25 are given in parenthesis.

based on branching (for example, solutes A5, A7 and A9) and general molecule similarity (A13 and A14). Unsaturation does not seem to be important: 2-ethyl-hexanal (A12) and 2 ethyl-2-hexenal (A17) correlate very well.

For problem KET2, the unique solutes include acetone (K1), cyclohexanone (K20), and the diones (K25, K26, K27). Correlations seem to be based on branching (for example K5, K8, K9 and K10), cyclo uniqueness (K19 and K20), and adjacent dione functionality (K25 and K26). 2,4-pentadione (K2) does not correlate with the other diones, probably because the two carboxy groups are not on adjacent carbons. Rather, acetone (K1) correlates well with 2,4-pentadione (K27); both K1 and K27 have no keto-groups.

Problem KET1 showed similar uniqueness values (results not shown).

Problem CARB contains two unique solutes, acetaldehyde (A1, .50) and 2,3-butadione (K25, .422). As with KET2 problem, major correlations involve the two cyclo solutes (K19 and K20) and the adjacent diones, K25, K26.

For the MC33 and MC53 problems, the only unique solutes are the alkanes AL4 and AL5 (.430), probably because they serve as the upper anchors in having the greatest retention index for the data.

The larger MC53 solute set hides some of the minor correlations apparent in the smaller MC33 problem set.

The alcohols OL1 through OL5 correlate well on MC53, though on MC33 only the higher molecular weight alcohols OL4 and OL5 correlate. Thus only the major correlations among the alcohols (OL4 and OL5) is apparent. The esters correlate well only on MC33 whereas the aromatics correlate on both MC33 and MC53.

In summary, the uniqueness test on the solutes has revealed those unique solutes for each of the problems and those solutes which correlate well. Some of these correlations can be used in the target transformation step as vectors to test whether the correlated solutes represent one of the interactions or cofactors in the problem. Thus, a vector testing for adjacent diones on KET2 may test well. The identity of the cofactors may be deduced from these correlations. For example, one of the cofactors in the MC33 problem may be associated with aromatic character.

B. Solutes: Effect of Functionality on Reproduction

One of the reasons for selecting the MC33 and MC53 problem sets was to study the effect of each of the solute classes on the factor space. Table 42 summarizes the results of these analysis. Each of the individual solute functional groups in sets MC33 and 53 were factor analyzed separately on the entire 18 solvent group set. In all cases, the MC53 factor space is always larger for a solute class than the space for MC33 by one factor. This is probably due to a "branching" factor.

The individual solute classes are simpler systems as compared to the original problem sets. Each solute class alone required either a 2 (on MC33) or 3 (on MC53) factor space with the 18 solvents to adequately reproduce the data matrix. As expected, the straight chain alkane solutes require only one factor. Addition of the five straight chain alkanes to each of the individual solute classes provides a 3 (on MC33) or 4 (on MC53) factor space.

Summarized in Table 43 are the results when only one solute class is removed from the MC33 problem set. Removal of any single solute class does not have a significant effect on the factor space. Thus it is not readily apparent that there are any unique interactions or cofactors associated with any single solute class. However, as we shall see in the target transformation step, there are

Table 42. Summary of reproductions for subsets of solutes for MC33 and MC53 problems

Solute group	Designations[a]	Number of solutes MC33	Number of solutes MC53	Number of factors[b] MC33	Number of factors[b] MC53	Largest errors[c] MC33	Largest errors[c] MC53	Percentage errors >5[c] MC33	Percentage errors >5[c] MC53
alcohol	OL1-OL10	5	10	2	3(4)	6.1	8.9(6.4)	1.1	5.6(0.6)
aldehyde	A3-A11	4	7	2	3	4.5	8.0		3.2
ketone	K2-K28	3	6	d	3	d	4.1	d	0.0
branched	all-MC33	-	20	-	6	-	11.3	-	3.3
ester	ES1-ES9	5	9	2,3	3,4	9.3,3.7	13.1,4.6	7.8,0.0	3.7,0.0
ether	ET1-ET7	4	6	c	3	c	9.5	c	2.8
oxide	OX1-OX3	3	3	c	c	c	c	c	c
aromatic	AR1-AR7	4	7	c	3	c	4.7	c	0.0
alkane	AL1-AL5	5	5	1	1	0.0	0.0	0.0	0.0
alkanes and	alcohols	10	15	3	4	5.6	9.0	0.6	3.3
" "	aldehydes	9	12	3	4	4.2	5.0	0.0	0.0
" "	ketones	8	11	3	4	2.5	4.1	0.0	0.0
" "	esters	10	14	3	3(4)	6.3	10.6(5.8)	1.1	5.6(0.8)
" "	ether	9	11	3	4	8.6	1.5	3.1	7.6
" "	oxide	8	8	3(4)	4	11.1(.4)	0.4	3.5(.0)	0.0
" "	aromatics	9	12	3(4)	4	10.9(2.8)	4.4	3.1(.0)	0.0

[a] See Table 21 for solute designations.
[b] Two values given where factor space not clear.
[c] Second values for the larger factor space.
[d] Number of solutes insufficient for factor space.

Table 43. Effect of removal of each solute group from MC33 problem, 6 factors

Solute group removed	Designations[a]	Number of solutes left	rms error	Percentage errors > ± 5	Largest error
alcohol	OL1-OL5	28	1.72	2.2	9.0
aldehyde	A3-A6	29	1.75	2.9	8.2
ketone	K2-K16	30	1.77	2.4	8.6
ester	ES1-ES7	28	1.87	2.2	8.6
ether	ET1-ET5	29	1.67	2.3	8.7
oxide	OX1-OX3	30	1.79	1.9	8.2
aromatic	AR1-AR4	29	1.68	2.5	6.1
alkane	AL1-AL5	28	2.00	3.0	6.8
Complete MC33 problem		33	1.76	2.7	8.5

[a] See Table 21 for solute designations.

unique interactions associated with the carbonyls, alcohols, esters, and aromatics.

C. Combinations of Typical Solute Vectors

Combination of the original solute column vectors can provide insight into the key interactions and the most representative solutes for a particular data set. Solutes which are quite similar (in an interaction sense) will behave similarly and consequently should be good substitutes for each other.

Not all of the solutes were used in combination for the CARB, MC33 and MC53 problems. The number of calculations were much too large, using the computer facilities available. For example, combination, using six factors, of the 33 solutes in MC33 would involve 1,107,568 combinations; using seven factors on the 53 solutes of MC53: 154,143,080; and six factors on the 37 solutes of CARB: 2,324,784. Therefore, combinations of solute vector cofactors for CARB, MC33 and MC53 were calculated only for 20, 18 and 18 solute vectors respectively, as explained in footnote of Table 44.

Tabulated in Table 44 are the best combinations of the solute vector cofactors for each of the problem sets per factor space. Thus, for example, the best combination for problems KET1, KET2, and CARB, using only one factor, is provided by solute K6 with a rms error between the predicted and original data matrices of 32.3, 32.3, and 34.6, respectively. For MC33 in 6 factors, two alcohols (OL1, OL5), an aldehyde (AL1),

Table44.Summary of combinations utilizing best solute vector cofactors for each problem, factors 1-7

Problem[a]	Number of solute vector cofactors used in the reproduction[b,c]						
	1	2	3	4	5	6	7
ALDE (16)	38.5 A6	7.8 A13,A17	4.3 A4,A12,A18	3.3 A4,A9,A11,A18	2.4 A2,A4,A9, A17,A18	1.7(8.8) A2,A7,A9, A12,A13,A18	
KET1 (14)	32.2 K6	4.1 K2,K17	2.9 K2,K10,K17	1.9(11.8) K2,K10,K16,K18			
KET2 (23)	32.3 K6	7.1 K2,K17	5.4 K2,K17,K25	3.4 K5,K17,K19,K25	3.2 K2,K10,K18, K20,K25	1.8(11.) K1,K10,K16, K18,K19,K25	
CARB (20)	34.6 K6.	8.7 A17,K1	6.3 A4,K18,K19	4.8 A4,K18,K20,K25	3.0 A4,A18,K10, K18,K26	2.3(12.5) A4,A13,K10, K18,K20,K26	
MC33 (18)	65.8 A11	26.0 ET5,OX3	16.4 OL1,ES6.,ET5	5.5 OL5,ES7 ,OX1, AL1	3.2 OL1,OL5,K2, AR1,AL1	2.3(10.3) OL1,OL5,A11, AC1,AR2,AL5	
MC53 (18)	60.6 A11	26.6 K2,ET5	15.3 OL1,ES6 ,ET5	6.2 OL1,OL5, ES7, AL1	3.6 OL1,AC3,OX1, AR2,AL1	2.7 OL1,OL5,A11, AC1,AR1,AL1	2.4(15.8) OL1,OL5,A11, K16,AC1,AR2, AL1

(continued)

Table 44. (continued)

[a] Lower line - number of solute vectors used in the reproduction. The following twenty selected solutes were used in combination for the CARB problem: A2,A4, A7,A9,A11,A12,A13,A17,A18,K1,K10,K11,K12,K16,K18,K19,K20,K25,K26,K27. The following eighteen selected solutes were used in combination for the MC33 and MC53 problems: OL1,OL5,A3,A11,K2,K16,ES1,ES3,ES6,ES7,ET1,ET5,OX1,OX3,AR1,AR2, AL1,AL5.

[b] Upper line - rms error; lower line - solute vector cofactors occurring in best combination. Largest error for the best combination in the true factor space for each problem is given in parenthesis. See Table 21 for solute designations.

and acetate (AC1), an aromatic (AR2), and an alkane (AL5) provide the best combination with an rms of 2.3. The largest absolute error is only 10.3. MC53 which has a 7 factor space differs in the best combination using 7 factors from the 6 factor MC33 combination by the addition of a ketone (K16). We expect MC33 and MC53 to be similar. Indeed, using 6 factors, MC53 has the same solute set as MC33, differing only in the substitution of AR1 for AR2, solutes which we shall later see are equivalent.

Thus, though no branching solutes were included in the MC53 combination step, the results for 7 factors are quite satisfying. The best combination set is almost identical to the best set of problem MC33 and reproduces the original data with an rms error of only 2.4.

If we compare the solutes represented in the best 20 or 30 combinations for the nth factor (as determined in the reproduction step of a problem) we can determine which solutes are good substitues for each other. For example, Table 45 lists the best twenty combinations for the MC33 problem. Note that solutes AL1 and AL5 are equivalent. The equivalency of AC1 and AC3 is also apparent.

The second column of Table 46 summarizes these correlations for all the problems. Thus the alkanes, as a group, are found equivalent in problems MC33 and MC53, as are the acetates and aromatics. For problem ALDE, aldehydes with methyl groups as well as the straight chain aldehydes correlate well. For KET1 and KET2, ketones with

Table 45. Best combinations of solute vectors, MC33 problem

Combinations set[a]						rms error
OL1	OL5	A11	ES1	AR2	AL5	2.29
OL1	OL5	A11	ES1	AR2	AL1	2.29
OL1	OL5	A11	ES1	AR1	AL5	2.30
OL1	OL5	A11	ES1	AR1	AL1	2.30
OL1	OL5	A11	ES3	AR2	AL1	2.30
OL1	OL5	A11	ES3	AR2	AL5	2.30
OL5	A11	ES1	OX1	AR2	AL1	2.30
OL5	A11	ES1	OX1	AR2	AL5	2.30
OL1	OL5	A11	ES3	AR1	AL1	2.31
OL1	OL5	A11	ES3	AR1	AL5	2.31
OL1	OL5	A11	ES7	AR2	AL1	2.31
OL1	OL5	A11	ES7	AR2	AL5	2.31
OL1	OL5	A11	ES7	AR1	AL1	2.33
OL1	OL5	A11	ES7	AR1	AL5	2.33
OL1	A11	ES1	OX1	AR2	AL5	2.33
OL1	A11	ES1	OX1	AR2	AL1	2.33
OL1	OL5	A11	ES7	AR2	AL1	2.34
OL1	OL5	A11	ES7	AR2	AL5	2.34
OL5	A11	ES1	OX1	AR1	AL1	2.34
OL5	A11	ES1	OX1	AR1	AL5	2.34
OL1	OL5	A11	ES6	AR1	AL5	2.35
OL1	OL5	A11	ES6	AR1	AL1	2.35
OL5	A11	ES3	OX1	AR2	AL1	2.35
OL5	A11	ES3	OX1	AR2	AL5	2.35
OL1	A11	ES1	OX1	AR1	AL1	2.37
OL1	A11	ES1	OX1	AR1	AL5	2.37
OL1	A11	ES3	OX1	AR2	AL1	2.38
OL1	A11	ES3	OX1	AR2	AL5	2.38
OL5	A11	ES6	OX1	AR2	AL1	2.39
OL5	A11	ES6	OX1	AR2	AL5	2.39
OL5	ES3	ET5	OX1	AR2	AL5	2.41

[a] See Table 21 for solute designations.

Table 46. Summary of correlations for solutes in each problem

Problem	Correlations based on best combinations[a]	Correlations based on percentage tables: Redundant solutes[a]	Correlations based on percentage tables: Cofactor equivalencies[a]
ALDE	A4=A13; A5=A7=A9; A11=A12=A17; A3=A4=A6=A10	A4=A6; A5=A11~A17	A15 & A16 & A18=3; A9=4; A2=5; A7 & A13 & A14=6
KET1	K2=K3=K4; K6=K16=K17; K5=K9=K10=K15; K5=K18	K4=K7; K15=K17=K18	K1 & K2=2; K10=3; K4 & K5=3; K15 & K18=4; K16~4
KET2	K1=K27; K5=K10; K8=K10; K19=K20; K25=K26	K2=K5=K8; K7=K12; K11=K16~K17; K25=K26=K27; K19~K20	K22 & K25 & K26=3; K19 & K20=4; K4 & K18=5 K1=6; K27=3 & 5 & 6
CARB	A4=A7; A2=A13; A12=K18; K1=K10; K19=K20; K25=K26	A2=A4; K11=K18; K25=K26	K25 & K26=4; K1 & K10=5; A18=3; A7 & A13=6
MC33	AL1=AL5; ES1=ES3; AR1=AR2; ES3=ES6=ES7=OX1; OX1=OL1=OL5	OL1~OL5; AR1=AR2; AL1=AL5	OL3=3; OL=4; ET1~4; AR1 & AR2=5
MC53	AL1=AL5; AR1=AR2; ES3=ES7; ES1=ES6; K16=AR1; OL1=OL5=A3 A11=A3;	OL1=OL5; AR1=AR2; A11=AL5; ES1 ES3; ES6~ES7	OL1 & OL5=3 & 4; AL1 & AL2=4; AR1 & AR2=5; ES1=6; A3 & K16=7

[a] Solute designations taken from Table 21. Correlated solutes related by an "=" for good correlations; "~" for weak correlations. See footnotes, Table 44, for solutes used in combinations for CARB, MC33 and MC53.

Table 46. Summary of correlations for solutes in each problem

Problem	Correlations based on best combinations[a]	Correlations based on percentage tables: Redundant solutes[a]	Correlations based on percentage tables: Cofactor equivalencies[a]
ALDE	A4=A13; A5=A7=A9; A11=A12=A17; A3=A4=A6=A10	A4=A6; A5=A11~A17	A15 & A16 & A18=3; A9=4; A2=5; A7 & A13 & A14=6
KET1	K2=K3=K4; K6=K16=K17; K5=K9=K10=K15; K5=K18	K4=K7; K15=K17=K18	K1 & K2=2; K10=3; K4 & K5=3; K15 & K18=4; K16~4
KET2	K1=K27; K5=K10; K8=K10; K19=K20; K25=K26	K2=K5=K8; K7=K12; K11=K16~K17; K25=K26=K27; K19~K20	K22 & K25 & K26=3; K19 & K20=4; K4 & K18=5 K1=6; K27=3 & 5 & 6
CARB	A4=A7; A2=A13; A12=K18; K1=K10; K19=K20; K25=K26	A2=A4; K11=K18; K25=K26	K25 & K26=4; K1 & K10=5; A18=3; A7 & A13=6
MC33	AL1=AL5; ES1=ES3; AR1=AR2; ES3=ES6=ES7=OX1; OX1=OL1=OL5	OL1~OL5; AR1=AR2; AL1=AL5	OL3=3; OL=4; ET1~4; AR1 & AR2=5
MC53	AL1=AL5; AR1=AR2; ES3=ES7; ES1=ES6; K16=AR1; OL1=OL5=A3 A11=A3;	OL1=OL5; AR1=AR2; A11=AL5; ES1 ES3; ES6~ES7	OL1 & OL5=3 & 4; AL1 & AL2=4; AR1 & AR2=5; ES1=6; A3 & K16=7

[a] Solute designations taken from Table 21. Correlated solutes related by an "=" for good correlations; "~" for weak correlations. See footnotes, Table 44, for solutes used in combinations for CARB, MC33 and MC53.

Table 47. Percentage of better reproductions involving selected specific solute column vectors from data matrix for MC33 problem.

Solute[a]	Number of solute vector cofactors used in reproduction					
	1	2	3	4	5	6
OL1	0	11	31	45	59	59
OL5	6	4	15	47	58	59
A3	6	13	18	18	12	18
A11	6	11	18	18	19	28
K2	6	10	23	15	25	19
K16	6	11	9	13	17	24
ES1	6	10	23	23	11	30
ES3	6	10	20	15	20	30
ES6	6	14	19	19	9	24
ES7	6	11	22	18	21	28
ET1	6	5	0	36	20	19
ET5	6	15	24	32	25	32
OX1	0	13	13	22	26	24
OX3	6	12	15	19	19	14
AR1	6	10	7	0	45	49
AR2	6	11	11	0	46	50
AL1	6	15	18	29	34	46
AL5	6	14	18	29	34	46
Total number of combinations:	15	153	816	3060	8568	18564
Cutoff for better reproductions (rms)	120	50.2	29.2	13.2	5.69	3.51
Number reproductions less than cutoff:	15	100	200	500	500	1000

(continued)

Table 47. (Continued)

rms eror, abstract reproduction:	63.6	23.7	15.0	5.15	2.77	1.90
Best combination:						
rms error:	65.8	26.0	16.6	5.50	3.16	2.29
vectors[a]:	All	ET5,OX3	OL1,PR1, ET5	OL5,PR2, OX1,AL1	OL1,OL5, K2,AR1, AL1	OL1,OL5, A11,ES1, AR2,AL5

[a] Solute designations taken from Table 21.

are relatively important. Since they have the highest percentage of occurrence in the six factor space of any of the solutes, we conclude that, of the solute types, the alcohols are most likely uniquely representative of one of the interactions or factors in the MC33 problem. To a lesser extent, the same conclusion may be determined for the aromatics, AR1 and AR2, and for the alkanes, AL1 and AL2. Comparing with the best combination sets (Table 45) will show these three functional groups to be well represented.

We may also speculate on the association of a particular solute and the abstract eigenvectors. When going from a factor space to the next larger one, if a particular solute increases dramatically in percentage, it may be associated with the abstract eigenvector added to the space. AR1 and AR2 exhibit a marked increase in percentage (45 percentage points) in going from a four factor to a five factor space. Thus the aromatics may be associated with the fifth factor or eigenvalue. Comparing with the best combination per factor space in Table 47, we also notice that the aromatics are first represented in the fifth factor. For the other solutes we conclude that OL1 associates with the third cofactor and OL5 with the fourth. Ethyl-ether (ET1) seems to associate, as well, with the fourth cofactor. However, in going to the next higher factor space, it loses its importance, falling from 36 to 20 percent.

The numerical pattern for solutes AR1 and AR2 is quite similar in going from one to six factors. This means that solutes AR1 and AR2 are equivalent. Exchanging these solutes in a combination will give equivalent results, as found in the correlations we observed when comparing the best combinations. The alcohols OL1 and OL5 also show similar patterns as do the alkanes AL1 and AL5.

These results are summarized in columns 2 and 3 of Table 46 for all the problem sets. Solutes are correlated with both abstract cofactors and with other near-equivalent solutes.

The most unique solutes, as found in the uniqueness step, are those which exhibit the best equivalences with the abstract cofactors. From Table 46 we can compare the most unique solutes with the cofactor equivalencies as found in the pattern tables. Thus, for problem ALDE, the three most unique solutes A2, A9 and A18 are good representatives of the fifth, fourth and third abstract cofactors, respectively. For problem KET2, K25, K26, and K27 are all equivalent to the third cofactor and from the uniqueness test are also equivalent to each other. K25 and K26 correlated, as well, in the best 20 combinations.

D. Target Transformation of Solute Vectors

In the target transformation step we attempt to associate the abstract cofactors with physically meaningful vectors. Comparison of the predicted and test vector provides the means of determining whether a test vector is a basic cofactor.

For the solutes, as described in Appendix B, we expect molar volume, enthalpy of vaporization, molar refraction, boiling point, dipole moment and polarizability to provide good test vectors.

Summarized in Table 48 are all the test vectors which target transformed with good agreement for the six problem sets.

Designations and example values for 23 representative vectors which tested from poor to excellent are listed in Table 49.

Test vectors include physical data, such as can be found in the literature and chemical handbooks for the various solutes. The free-floating feature of TFA enables us to leave blank any unknown data points. The target transformation feature of TFA will predict the missing data points quite reasonably if the test vector transforms well. Details are provided in Table 50 of four selected target transformation on the CARB problem for fourteen selected solutes. Of these, three (MV, XM and VP)

Table 48. Summary of physical and structural vectors target transforming with best agreement for each problem

Problem	Test vectors[a]
All 6[b]	BP, BPK, C, DH, EDH, MW, NC, NCM, SBPK, X, XM.
ALDE	DT, K, LVP, MP, MPK, MR, MV, **ND, UD, UNI, CV%, CV6, CV7.**
KET1	K, KP, LVP, MP, MPK, MR, MV, SYM, UAS, UNI, USC, VP, CV2, CV3, CV4.
KET2	KP, LRI, LVP, MP, MPK, MR, MV, NO, UAD, UAS, UCY, UD, **UDI, UNI, USC,** VP, **VMH, CV1, CV8, CV9, CV10, CV11, CV12.**
CARB	K, **MD, MP, MPK, MR, MV, NO, UAD, UAS, UCY, UDI, UK, UNI, VMH, VP.**
MC33	DIP, DK, K, MA, NH, ST, UAR, UCA, UE, UOH, VA, VB, VIS, EDM, CV13.
MC53	DIP, DK, K, MA, NH, ST, UAR, UOH, VA, VB, EDM, CV13.

[a] See Table49 for most of the designations.

Other designations are: C: number of total carbon atoms; DIP: dipole moment of functional group; DK: dielectric constant; DT: d(BP)/dT; KP: position of keto group; LRI: log of refractive index; LVP: log of vapor pressure; MA: main chain length; **MD: molar dispersion; ND: number**

(continued)

Table 48. (Continued)

double bonds; NH: number hydrogen atoms; ST: surface tension; SYM: symmetry uniqueness; VA: Van der Waal's constant "a"; VB: Van der Waal's constant "b"; VIS: viscosity; EDM: electric dipole moment.

The following are uniqueness test vectors: UAD; adjacent dione; UAR: aromatic; UAS: adjacent side chain; UCA: carbonyl; UD: double bond; UDI: dione; UE: ester; UK: ketone; UOH: alcohol.

Points on the following correlation vectors (CV) are in most cases "0"'s or "1"'s. For each correlation vector, test points having non-zero values are given below. If the value for a given solute is other than "1", the value is given in parenthesis. (Test points not designated are all "0"'s.) CV1: K1, K27; CV2: K1,K2,K3(.5),K7,K10(-.8),K18; CV3: K1,K2,K3,K4,K5(.5), K7,K18; CV4: K4,K7,K15,K17,K18(2.); CV5: A5,A7,A9,A15; CV6:A12,A16,A17,A18; CV7: A10,A11,A12,A13(-.5),A16,A17,A18; CV8: K1,K2(.5),K27; CV9: K19,K20, K24,K27(-2.); CV10: K19,K20,K24; CV11: K1,K24(-.5),K27; CV12: K11,K16,K17; CV13: AL4,AL5.

[b] Vectors testing well in each of the six problems.

Table 49. Summary of selected target transformations for physically significant test vectors for CARB problem, 6 factors.

Symbol	Name of test vector	Example values[a]	Points[b]	Qualitative evaluation[c]
BP	boiling point (C)	128,173	36	g
BPK	boiling point (K)	401,446	36	v
DH	heat of combustion	ff,1205	15	v
EDH	exponent DH	ff,.13	15	v
K	volumetric susceptibility[e] $\times 10^{-6}$	ff,ff	19	f
LDH	log DH	3.08,ff	15	g
MP	melting point (C)	-56,-16	29	f
MPK	melting point (K)	217,257	29	g
MR	molar refraction	30,39	24	v
MV	molar volume ml/mole	124,137	24	v
MW	molecular weight	100,128	39	v
NC[d]	number carbon atoms	6,8	14	v
NCM	number carbon atoms, main chain	6,8	39	g
NO	number oxygen atoms	1,1	39	f

(continued)

SBPK	square BPK	160920,198920	35	v
UNI	unity	1,1	39	v
UCY	cyclic uniqueness	0,0	39	f
UK	ketone "	0,1	39	g
USC	side chain "	0,0	39	p
X	specific susceptibility[e] $\times 10^{-6}$	.693,ff	20	g
XM	molar susceptibility[e] $\times 10^{-6}$	69.4,ff	20	v
VP	vapor pressure	ff,ff	10	g
VMH	molar heat of vaporization	ff,11649	25	g

[a] Examples are given for hexanal (solute A10) and octanone (solute K16); ff are free-floated values.

[b] Number of points on test vector.

[c] v-very good agreement between test vector and predicted vector; g-good agreement; f-fair agreement (several points poorly predicted or pattern only predicted); p-poor agreement.

[d] Points correspond only to straight molecules. All other points were free-floated.

[e] Diamagnetic susceptibilities from reference 88 .

Table 50. Details of selected target transformations for CARB problem, 6 factors

Solute[a]	Molar volume (MV) (ml/mole) Test[b]	Molar volume (MV) (ml/mole) Predicted	Molar susceptibility (XM) $\times 10^{-6}$ Test[b]	Molar susceptibility (XM) $\times 10^{-6}$ Predicted	Side chain uniqueness (USC) Test	Side chain uniqueness (USC) Predicted	Vapor pressure (VP) (lb/in^2) Test[b,c]	Vapor pressure (VP) (lb/in^2) Predicted
A2	57.1	55.9	22	20	0.0	0.3		8.6
A4	90.5	91.4	46	47	.0	.2		2.1
A7	108.4	103.2	58	56	1.0	.3		.1
A9	109.6	110.7		60	1.0	.7		1.0
A10	123.7	125.1	69	69	.0	.2		2.1
A12		157.5		87	1.0	.5		-.4
K1	74.0	75.0	34	36	.0	.5	4.3	4.1
K4	106.4	108.0	58	58	.0	.4	.9	.6
K6	124.2	124.3	69	68	.0	.5	.3	.2
K10	125.0	126.1	70	69	1.0	.8	.6	.6
K16	157.5	155.7		87	.0	.5		-.6
K19		97.2	52	51	.0	.0		-.4
K25		76.5		33	.0	.1		-1.8
K27		104.9		50	.0	.1		-2.1

[a] See Table 21 for designations.
[b] Blank points are free-floated points.
[c] Test valid; six other points not shown.

target transform well, one (USC) did not target transform.

Structural vectors based on our chemical knowledge and intuition can be constructed. Thus, a vector testing for unsaturation which we can call an unsaturation or double bond uniqueness test vector would consist of values of "1.0" for those solutes which are unsaturated and of "0.0" for the saturated solutes. Those solutes whose designations we would be unsure of, could be free-floated. For example, on an unsaturation uniqueness vector, an aromatic solute would be free-floated, the predicted vector then assigning the proper "unsaturation" value to the aromatic.

"Correlation" vectors representing the correlations obtained in the uniqueness tests may also be constructed. In these vectors, we can assign partial values, such as +0.5 or -0.5, to those solutes which do not correlate as strongly, or correlate negatively as compared to the other solutes in the vector. Sometimes the correlation vector can be optimized by modifying it after observing its performance in the target transformation. However, modifying it too much would lose its physical significance by optimizing onto one of the abstract factors.

Only those correlation vectors which target transformed well are listed in Table 48. However, all of the apparent correlations observed in the uniqueness tests were tested.

As expected, boiling point (BP), enthalpy of vaporization (DH) (for those data sets available), and NC (carbon number

based on straight chain alkanes) generally tested well. Interestingly, the boiling point squared (SBPK) was a very good vector. This vector has been found to be a good vector for most solute GC data sets which were factor analyzed.[25]

Table 51 is a list of the tested target vectors which did not target transform well. Some vectors resemble a unity vector in that all data points are equivalent. They are indicated in Table 51 by a "u".

Three carbon number vectors were tested. On each problem set, all three target transformed well. Two of the vectors tested for the total number of carbon atoms in a solute (NC) or in the main chain of the solute (NCM). The third vector tested for the carbon number (C), in which only values for the straight chain alkanes were used. Values for the other solutes were free-floated. This would allow us to assign a carbon number value (on a retention index scale) to the various functional groups and branches of the free-floated solutes.

A target vector must have at least n-data points in the n-factor space it is tested in. Preferably, there should be more than n+1 data points. The MC53 problem set is a 7-factor space problem; however, there are only 5 straight chain alkanes. Consequently for set MC33, values for the aldehydes A5, A7, A9 obtained from the carbon number (C) target transformation vector from problem set ALDE were input for those solutes.

Table 51. Solute vectors which did not target transform well

Problem	Vectors[a]
ALDE	RI (u); SG (u); SR (u); RIN (u); SD (u); NB; NBC.
KET1	DEN (u); RI (u);dBP/dP (u); SG(u);RIN (u); SD (u); VMH; VPB.
KET2	dBP/dP(u); K (u); NB; RI; RIN; SD (u); SG; SR; SYM; UD; VPB.
CARB	dBP/dP(u);DEN (u); KP; NB; NBC; ND; RI; RIN; SD (u); SG; SR (u); SYM; UD; USC; VPB.
MC33	MP; DEN; DT.
MC53	SG.

[a] u=unity vector. See Table 49for most vector designations. Other designations are: DEN: density; dBP/dP: boiling point dependence on pressure; NB: number branches; NBC: number branched carbons;RI: refractive index; RIN: refractive intercept; SD: specific dispersion; SG: specific gravity; SR: specific refraction; VPB: constant "b" vapor pressure equation (ref. 88).

The carbon number values for the oxides and aromatics of the main chain length vector of MC33 and MC53 were free-floated. Predicted values for the -C-C- ring averaged 4.6 on MC53 and 4.2 on MC33. Predicted values for the benzene ring averaged 13.5 for MC53 and 10.8 for MC33. Thus the pi-system seems to add approximately six carbon atoms.

Molar refraction (MR) and molar volume (MV) target transformed well on all the carbonyl sets (data was not available for problems MC33 and MC53). According to Equations 83 and 84, solute molar volume and molar refraction are theoretical factors. Dipole moment, predicted in Equation 85 as a possible factor, also tested well in the MC33 and MC53 problem sets where the data was available. Van der Waals constants "a" and "b" also tested well for sets MC33 and MC53, the only problems where they could be tested due to data availability.

Correlation vectors that tested well are also listed. It is hard to rationalize these vectors chemically. However, some of these vectors such as CV5 and CV10 may represent molecular weight correlations.

Uniqueness test vectors for aromatics (UAR) and alcohols (NOH) tested well in MC33 and MC53. A uniqueness vector for ester also tested well in MC33. These vectors indicate that a unique interaction is associated with the alcohols, aromatics and for MC33, with the esters.

E. Combinations of Basic Solute Vectors

In the combination step we can determine 1) which of those vectors which target transformed well are redundant (i.e., represent the same abstract cofactor) 2) whether we have adequately determined the entire cofactor space, i.e., have a basic vector for each of the n abstract cofactors and 3) which set of vectors represents best the cofactor space.

Because the total number of vectors which tested well in the target transformation step was so large for each problem, it was impossible to obtain the combination of all the vectors due to limited computer time. Thus, the combination of the physically significant vectors was obtained using the same method as previously described for the solvent physical vectors. A combination of a preliminary selected set of the most promising vectors (as determined by their performance in the target transformation) was first obtained. Those vectors which are found in the preliminary step as insignificant were then removed and other new vectors were then added to the vector set, for a second modified combination set. For some problems this was repeated as many as six times. Thus the final combination vector set, though containing only 17 to 19 vectors, was the result of the combination on all the good vectors.

Table 52 provides a summary of the results for the six problems. In every problem, SBPK, the square of the boiling point, is the best vector combination for the one factor space. However, its signficance diminishes in the higher

Table 52. Summary of combinations utilizing best physically significant solute cofactors for each problem, factors 1-7.

Problem[a]	Number of real vector cofactors used in the reproduction[b] 1	2	3	4	5	6	7	Vectors in ten best combinations[a]
ALDE (17)	48.3 SBPK	16.3 SBPK, EDH	7.2 BPK, MW, NC	6.3 DT, MR, NC, XM	4.3 EDH, DH, K, MR, X	2.5 (11.2) DH, EDH, K, NC, X, XM		DH, DT, EDH, K, MR, MV, NC, X, XM
KET1 (17)	39.7 SBPK	4.5 SBPK, NC	3.7 NCM, NC, SBPK	2.8 (20.6) CV2, NC, SBPK, XM				CV2, CV4, KP, LVP, MR, MV, NC, SBPK, XM
KET2 (17)	41.1 SBPK	25.4 LRI, SBPK	19.1 DH, NC, SBPK	10.9 BP, BPK, MV, VMH	9.2 BPK, DH, LRI, MR, NC	6.8 (47.5) BPK, KP, LRI, LVP, NC, VP		BPK, DH, KP, LRI, LVP, MB, MV, NC, SBPK, UNI, VA, VP
CARB (19)	43.2 SBPK	19.2 DH, SBPK	15.2 BPK, DH, SBPK	13.8 BPK, DH, SBP, VMH	11.7 DH, K, NC, VP, XM	9.8 (106) DH, K, NC, UK, VP, XM		BPK, DH, K, MR, MV, MW, NC, SBPK, UK, UNI, VA, VP, XM
MC33 (18)	120. SBPK	81.2 DIP, SBPK	30.7 DIP, NCM, UOH	23.6 DIP, NCM, UOH, X	17.3 DIP, MA, UOH, X, VB	11.0 (81.9) C, DIP, ST, UOH, X, VB		C, DH, DIP, K, MA, NCM, ST, UOH, VA, VB, X
MC53 (17)	111. SBPK	69.7 BPK, NC	62.6 BPK, NC, UOH	37.0 EDH, K, VA, VB	19.4 K, MA, NC, DIP, UOH	16.4 DK, K, NC, ST, UOH, XM	14. (167) DIP, K, NC, NCM, ST, UOH, XM	DK, DIP, K, MA, NC, NCM, ST, UOH, VA, VB, XM

[a] Lower line - number of real vectors used in the final combination step.

[b] Upper line - rms error; lower line - real vectors occurring in best combination. Largest error for the best combination of the true factor space for each problem is given in parenthesis. See Table 48 for vector designations.

factor spaces.

The best combinations give quite good results for problems ALDE, KET1, and KET2. For problems CARB, MC33 and MC53 the results are not as satisfactory (rms error = 9.8, 11.0 and 14.0, respectively). In the latter three problems, we are obviously missing a representative for at least one of the abstract cofactors in our best combination set. Thus, our vector combinations do not fully represent the entire cofactor space.

In general, the best combination set consists of a carbon number vector (NC or C), a vector which can be associated with vaporization (BP, SBPK, EDH, DH, VP, LVP or VB), a vector related to magnetic susceptibility (VM, K or X) and vectors which may relate to specific interactions (CV2, UK, WH, ST and DIP). Objective comparison of those vectors which are dominant in the best combinations of one problem with those in another problem are not possible because many of the vectors were not available in all the problems. For example, dipole moment (DIP) data was available only for problems MC33 and MC53.

Similarly, in comparing the best combination sets for these problems with previous factor analysis solutions on other problems, we must keep in mind that the same set of vectors was not used in determining the best combination. Also, since the solvents and solutes comprising the other data sets were different, there are other dominant or specific interactions.

Four previous studies have, using factor analysis on gas chromatographic data, determined the best combination of physically significant vectors to reproduce the original data matrix. Weiner and Parcher[22] for thirteen butenoic acid esters in a four factor space found UNI, BP, VA and C or MR as the best (±5 RI units) combination set. Howery[24], for twenty five hydrocarbons (six factors), found C, UA, MV or DH, UE and two specific structural vectors to provide the best (±8 RI units) combination. Weiner and Howery,[21] for twenty five alcohols (5 factors), found UNI, MW, MR, C and BPK gave the best (±8.4 RI units) combination and Selzer and Howery,[25] for eighteen ethers (six factors), found C, NA, SBPK, square of C and two specific structural vectors to give the best (±5.3 RI units) combination.

In general, in each of these studies, a carbon number vector (C), a vector associated with the bulk molecule (dispersion forces) (MV, MR, MW, NA) and a vector associated with vaporization (BP, SBP, MVH) are important cofactors. Specific interaction vectors (UA, UE) and specific structural vectors represent the other major cofactors. These results, then, have also been duplicated in the six problems in this study, as described above.

Tables 53 and 54 give the pattern tables for the CARB and MC33 problems. Thus, we see that SBPK is a significant vector in the CARB problem, equivalent probably to the second or third factors. In the MC33 prob-

Table 53. Percentage of better reproductions involving specific real solute cofactors for CARB problem, 6 factors.

Real Vector[a]	Number of real vector cofactors used in reproduction					
	1	2	3	4	5	6
BPK	9	31	22	23	24	23
DH	0	6	16	17	61	48
EDH	0	14	16	19	21	28
K	9	6	13	17	16	54
MR	9	8	16	21	21	23
MV	9	6	17	21	21	26
MW	9	11	11	20	19	30
NC	9	11	37	36	42	37
NO	0	3	5	14	17	17
SBPK	9	42	71	81	74	78
UNI	9	8	18	16	19	15
UK	0	0	0	3	44	63
VMH	0	3	2	14	15	24
VP	0	3	3	15	16	40
XM	0	6	11	13	17	16
X	9	3	16	18	18	28
C	9	14	14	18	16	13
Total number of combinations:	19	171	969	3876	11628	27125
Cutoff for better reproductions (rms error):	165.	70.	30.1	21.6	19.7	15.0
Number reproductions less than cutoff:	11	36	100	500	1000	1000

(continued)

Table 53 (continued)

rms error, abstract reproduction:	33.3	7.7	5.7	4.3	2.6	1.9
Best combination:						
rms error:	43.2	19.2	15.2	13.8	11.7	9.76
vectors[a]:	SBPK	DH, SBPK	BPK,DH, SBPK,	BPK,DH, SBPK,V	DH,K, NC,VP XM	DH,K, NC,UK, VP,XM

[a] Vector designations taken from Table 48.

Table 54. Percentages of better reproductions involving specific selected solute cofactors for MC33 problem, 6 factor

Real vector[a]	Number of real vectors used in reproduction					
	1	2	3	4	5	6
DIP	0	20	46	66	71	89
K	6	10	6	25	33	31
NC	6	10	26	19	27	27
NCM	6	12	17	25	41	40
SBPK	6	24	57	18	0	0
ST	6	26	0	45	47	49
UE	0	0	3	2	0	50
UOH	0	8	31	44	74	100
VA	6	16	14	33	42	38
VB	6	10	11	20	30	25
X	6	10	3	20	21	43
Total number combinations	18	153	816	3060	8568	18654
Cutoff for better reproductions (rms error)	300.	210.	100.	52.4	26.5	17.9
Number of reproductions less than cutoff	10	51	35	100	200	200
rms error, abstract reproduction	63.6	23.7	15.	5.2	2.8	1.9

[a] Vector designations taken from Table 48 .

lem, however, SBPK loses its significance totally in the higher factor spaces (at 5 and 6 factors its percentage is 0.0).

The pattern for the vectors MD, MV, and MR in problem CARB are obviously similar, indicating the equivalence of these vectors. We do expect the molar volume and molar refraction (really a relative measurement of molar volume) to be equivalent.

Table 55 summarizes the correlations obtained on all the problem sets from the pattern tables, as well as correlations obtained from comparison of the best 30 combinations. Vectors, MV and MR, as expected, are equivalent for those solute sets where they were tested (ALDE, KET1, KET2, CARB). They were not tested in MC33 and MC53 because of lack of data.

Table 55. Summary of correlations for physically significant solute cofactors for each problem

Problem	Correlations based on best combinations[a]	Correlations based on percentage tables: Redundant vectors[a]	Correlations based on percentage tables: Cofactor equivalencies[a]
ALDE	MR=MV; DT=X=XM	EDH~X; MW~C; MR~MV	SBP=2 & 3; NC=4; EDH & K & X=5; DT=6
KET1	CV2=CV4=KP; NS=NCM; MR=MV; LVP=XM	CV2=CV4; C=NCM; MW=MR	UNI & NC=2; SBPK=3
KET2	MV=MR=VA=BPK; DH=XM; LRI & LVP=SBP & MV	MV~MR; LRI~NC	SBPK & BPK=2; BP=5; KP & NC & VP=6
CARB	MR=MV; DH=XM; BPK=UNI; BPK & VA=NC & XM=NC & DH	EDH=MW; MV=MR=MD	SBP=2; SBP & NC=3; DH & UK=5; K & VP=6
MC33	C=NCM=MA=DH; UE=X; UAR=VA	NC~VB; NCM~VA; DK~DM	ROH & DIP=3; DIP & ST & VIS=4; UOH=5; UOH & UE=6
MC53	MA=NCM; K=VB; DK=DIP	VA=VB	VA & VB & NC=4; UOH & DIP & ST=5; XM=6; DK & NCM & ST=7

[a] See Table 48 for designations. Correlated solutes related by an "=" for good correlations; "~" for weak correlations.

F. Prediction of New Solute Data

As described for the solvents, the definitive test of our factor model is whether our sets of real cofactors can be used to predict new solute data. For the solutes, then

$$d_{xk} = \sum^{m} r_{xm} \quad c_{mk} \tag{62}$$

where d_{xk} is the predicted retention index for new solute x on solvent k, r_{xm} is the value for key, basic cofactor m for the solute, x, as determined in the combination portion of the analysis, and c_{mk} is the element of the [C] matrix for solute cofactor m on solvent k. Using this equation we may predict the retention index for solutes other than those used in the data matrix.

Tabulated in Table 56 is the column matrix [C] for the best combination of real solute cofactors on problem MC33: -X, ROH, DIP, VB, ST, C. Thus, for example, the retention index for a new solute x on solvent G (dioctyl pthalate) is:

$$\begin{aligned} d_{x,G} = \; & (-X)_x \cdot (-260.9) + (UOH)_x \cdot (190.1) + (C)_x \cdot (49.) + \\ & (DIP) \cdot (63.6) + (VB)_x \cdot (2101.) + (ST)_x \cdot (8.72) \end{aligned} \tag{63}$$

Thus for a new solute, butanol:

$$\begin{aligned} d_{butanol,G} = \; & (.7627)\ (+260.9) + (1)(190.1) + (4)\ (49.) + \\ & (.66)\ (63.6) + (.1143)\ (2101 + (24.6)\ (8.72) = 748 \end{aligned} \tag{63a}$$

Table 56. Column matrix C for best combination set, MC33 problem.

Cofactor[a]	Solvents[b]								
	A	B	C	D	E	F	G	H	I
-X	-407.6	-249.3	-257.1	-235.4	79.7	-263.2	-260.9	-232.4	-285.2
UOH	-263.6	164.6	193.8	187.1	461.8	184.2	190.1	185.2	251.9
C	31.8	57.9	45.9	50.1	-33.7	48.3	49.0	49.8	44.3
DIP	91.1	65.5	47.6	47.5	209.6	59.4	63.6	47.6	63.3
VB	2839.	1668.	2095.	1928.	2763.	2127.	2101.	1918.	2318.
ST	14.2	9.5	9.7	8.9	14.3	8.8	8.7	9.0	9.3

	J	K	L	M	N	O	P	Q	R
-X	-202.7	-565.1	-250.5	-466.8	-499.7	-679.8	-306.3	-321.3	-748.1
UOH	286.1	503.9	186.4	476.9	257.0	331.6	191.7	251.6	159.8
C	52.6	6.4	46.8	22.9	23.1	4.3	48.4	40.3	24.0
DIP	64.0	129.0	47.1	127.4	59.4	100.4	42.3	89.2	80.5
VB	1928.	3782.	2042.	3416.	3149.	3855.	2102.	2455.	5209.
ST	6.4	20.9	9.6	14.2	18.4	26.3	11.1	11.1	25.6

[a] See Table 48 for cofactor designations.
[b] See Table 20 for solvent designations.

The reported value for the retention index of butanol on solvent G is 782. We do not expect very good agreement between the predicted and reported values, since our factor model is not complete, missing at least one factor for MC33.

Table 57 compares the calculated and reported values of five solutes, butanol, butyl acetate, ethyl benzene, ethyl butyrate and heptane, on four non-key solvents: E, G, N and R. These solutes were chosen to represent each of the major solute classes in MC33. The solvents include the two unique solvents E and R and two non-unique solutes G and N. It is obvious that the better predictions are obtained for the non-unique solvents. Our basic solute cofactor set fails in accounting for the unique interactions of the solvents E and R.

The uniqueness of solvents E and R is apparent in the [C] matrix of Table 56. The factor loading for cofactor C (carbon number) is negative for both solvents, whereas, for all other solvents it is positive. Solvent E, in addition, also loads on cofactor X (magnetic susceptibility) in the opposite direction than all the other solvents. Thus, an analysis of the solute [C] matrix provides us with some important solvent information.

Our ability to predict solute retention data is limited by the lack of data for the solutes on the real physical cofactors. We may be able to compensate by estimating the value for a real solute cofactor for a particular solute.

Table 57. Prediction of retention data for five new solutes on four stationary phases for problem MC33

	Calculated and literature retention indices for solvents[a]							
	E		G		N		R	
Solute	calc.	lit.	calc.	lit.	calc.	lit.	calc.	lit.
butanol	1054	1443	748	782	880	933	852	941
butylacetate	947	1179	841	869	943	983	1086	1153
ethylbenzene	579	1158	1065	947	1016	1024	793	1070
ethylbutyrate	1198	1136	922	848	1016	954	1202	1106
heptane	659	700	719	700	717	700	732	700
average error	261		52		36		112	

[a] Literature values taken from reference 80 . Calculated values obtained using equation 62 and C matrix from Table 56. See Table 20 for solvent designations.

Thus, the value for the surface tension (cofactor ST) for butyl acetate was estimated at 24 dynes/cm since ethyl acetate is reported to have a surface tension of 23.9 dynes/cm. The carbon number (cofactor C) for ethyl benzene was estimated at 12.8 since the benzene ring was found equivalent to 10.8 carbons in the target transformation step. Obviously, these estimates by themselves introduce large errors into the predictions.

We are fortunate that data is available on many solutes for the real solute cofactors found in MC33. The calculated retention data do not agree satisfactorily with reported retention data, but this is expected since our best cofactor model for MC33 does not reproduce our original data matrix satisfactorily (as shown in Table 52).

We had hoped to demonstrate the accuracy of the prediction equation using the solute cofactors for problem ALDE, which gave excellent reproduction of the original data: EDH, DH, K, NC, X and XM. However, data is not available for these solute cofactors for many aldehydes other than those used in the ALDE matrix. Solute cofactor data was found for only four solutes: benzaldehyde, cinnamic aldehyde, p-hydroxybenzaldehyde and formaldehyde. Reported retention data is available only for formaldehyde. However, formaldehyde is expected to be unique due to its small molar volume and molecular weight. Thus we do not expect good agreement between predicted and reported retention data for this compound. The remaining three solutes

are aromatic, a functionality not represented in ALDE. We would therefore not expect the ALDE solute cofactors to adequately represent any aromatic interaction cofactor and consequently, the predicted retention data for these solutes is probably not accurate. Indeed, predicted retention indices average 1000 RI units greater than expected.

G. Summary

In summary, we have completed a target factor analysis of six subsets of solute-solvent retention index data from McReynolds' compilation utilizing 18 selected solvents on various solutes including four sets comprising the carbonyl functionalities, aldehyde and ketones, and two sets with representatives of several classes of compounds. The factor space was found to include four to seven factors as corroborated by Malinowski's criteria.

The effect of solvent uniqueness was tested using two different techniques:

1) by a target transformation uniqueness test, and
2) by a solvent functionality (effect on reproduction) test

The major functionality of the solvent molecule was found more significant than the underlying molecular skeletal structure. As expected from previous work, solvents diglycerol and Zonyl E7 were found to be highly unique.

These uniqueness results were corroborated in the combination of typical solvents step. Thus diglycerol and Zonyl E7 are present in every best combination set. Pattern tables associate them with the third to fifth abstract solvent cofactors.

The combination of solvent step provides a means for suggesting the best set of solvents which comprise a key solvent set. From the correlations obtained in the

uniqueness and combination steps, satisfactory substitutes for the various solvents may be obtained.

Associating the abstract solvent cofactors with physically significant cofactors in the target transformation step, we find that the McReynolds' constants do represent many of the solvent cofactors. However, of greater significance, as illustrated in the combination step, they are not sufficient to span the entire factor space. Other significant vectors include those predicted by theory, i.e., molar refraction and molecular weight, and structural vectors and correlation vectors obtained in the uniqueness tests.

Combinations of the physical vectors show the McReynolds' constants to represent two to three of the abstract cofactors. The cofactors associated with the unique interactions of diglycerol and Zonyl E7 also are unique and extremely important as illustrated in the pattern tables. The best combination sets were fairly satisfactory. However, for most of the problem sets, at least one major cofactor is probably missing from our final compilation.

The best combination sets provide surprisingly good predictions of new retention index data for new solutes on the solvent set. The results are especially good for the polar solvents indicating that we have good representative cofactors for the polar interactions.

Similarly for the solutes, the uniqueness of the individual solutes and classes of solute compounds was

explored. Solutes correlate, as expected, according to the dominant functionality of their structure. The aldehyde and ketone functionalities are found to exhibit essentially the same type of interaction.

Despite computer limitations, a method is described that gives satisfactory results on the combination of solutes cofactors. Correlation obtained from the combination step are also easily understood from a chemical standpoint.

As predicted theoretically (see Appendix B), molar volume or refraction, enthalpy of vaporization or boiling point, dipole moment and polarizability should target transform to abstract cofactors, as verified in the target transformation step. Three carbon number vectors, including one used to provide a carbon number scale relative to the n-alkanes, also target transform well. Typical carbon numbers of 4.4 are assigned to an oxide ring (-C(O)-C-) and 12.4 to benzene.

Combinations of the real physically significant vectors yield excellent reproduction results for most of the problem sets. In general, the best combination set consists of a carbon scale vector, vectors associated with vaporization and magnetic susceptibility, and a vector relating to specific interactions for a functional group. Previous studies of other solute cofactors using target factor analyses obtained similar results.

Examples are also given illustrating the predictive capabilitity of TFA using solute vector cofactors.

CHAPTER IX

TARGET FACTOR ANALYSIS OF GAS-SOLID-CHROMATOGRAPHIC RETENTION INDICES INVOLVING IONIC SORBENTS

A. Introduction

Gas liquid chromatography (GLC) developed much more rapidly than its predecessor, gas solid chromatography (GSC). For many years the utility of GSC was restricted to separations of hydrocarbons and of low boiling gases. Recently, however, the applicability of GSC has been extended through the development of many new sorbents.

Of these new support materials, some of the most useful are the macroreticular ion exchange resins. The macroreticular resins are mechanically and thermally stable under conditions normally used in GSC. The cation exchange resins can be readily converted to many metal ionic forms without the anionic polymer matrix being greatly changed in the process. The macroreticular resins should therefore be suitable for studying the effect of the metal cation on the retention behavior of compounds in GSC.

Hirsch et al[89] reported the retention indices of twenty one alkane and aromatic compounds on a macroreticular cation exchange resin, Amberlyst A15, using 15 metal

ionic forms at 160°C. More recently, he has furnished us with more complete data obtained at 180°C.[90] The retention indices of the n-alkanes were curve fitted and are therefore not in exact multiples of 100 as should be by definition. The data is reported in Table 58. The experimental error is reported to be ±3-±4 RI units. The data contain the retention indices of 21 normal, branched, cyclic and olefinic alkanes and aromatic solutes on 9 ionic forms of the Amberlyst A15 resin. In addition, the retention index of the 21 solutes on Chromosorb P, are also included. Chromosorb P is a relatively inert diatometous support prepared from Sil-O-Cel C22 firebrick. It has a relatively hard surface and is useful in GLC because it has a high capacity to hold liquid phases (up to 30% loading). IN GSC, its utility is in separation of hydrocarbons and other low polarity compounds. Being a relatively inert support, Chromosorb P should provide a good test of the Amberlyst A15 backbone's activity in the same data base. Both Amberlyst A15 and Chromosorb P have high surface areas with high pore volume. In the following, we shall designate each of the ionic forms of the resin by the respective atomic symbol, i.e., the symbol Ag will stand for the Ag^+-ionic form. Chromosorb P will hereafter be designated CP.

A simple analyses of the original data in Table 58 leads us to the conclusions in the following paragraphs.

The branched alkanes are retained for shorter times

Table58. Gas-solid-chromatographic retention indices at 180°C[90].

Solute		Sorbent[a]										
No.	Name	Li+	Na+	K+	Rb+	Cs+	Ag+	Mg++	Ca++	Sr++	Ba++	CP
1	n-pentane	501	500	499	500	498	500	505	501	502	502	500
2	n-hexane	598	599	601	600	601	600	598	600	600	598	600
3	n-heptane	702	701	702	700	702	699	699	699	699	698	701
4	n-octane	799	800	798	800	801	800	799	799	799	803	799
5	cyclopentane	515	500	505	508	506	500	500	496	502	498	544
6	cyclohexane	609	605	605	599	602	603	585	580	583	595	648
7	cycloheptane	723	726	721	720	729	722	701	698	702	702	776
8	cyclooctane	832	838	831	834	841	833	811	809	812	815	895
9	methylcyclopentane	593	583	596	585	595	587	575	570	573	576	620
10	2,4-dimethyl-pentane	656	647	662	653	664	662					663
11	2,2,4-trimethyl-pentane	727	724	739	734	742	742	692	692	699	705	743
12	1-hexene	662	661	653	651	642	1197	650	675	671	660	592
13	cyclohexene	672	669	663	661	659	1208	645	681	686		653
14	benzene	741	756	804	854	894	1088	702	827	875	919	638

(continued)

Table 58. (continued)

15	chlorobenzene	941	952	998	1069	1098	1070					798
16	ethylbenzene	937	941	962	991	1015	1330					830
17	toluene	860	863	892	925	951	1220	815	1002	1055	1095	
18	o-xylene	964	962	985	1002	1033	1375					
19	1-heptene	757	756	743	742	736		713	765	763	758	691
20	3-heptene	728	733	731	729	728						694
21	1-octene	850	853	835	842	836						791

[a] Cationic forms, Amberlyst A15; CP - Chromosorb P

than the straight chain alkanes. The retention is shortest for the divalent resin forms. The cyclic alkanes show slightly longer retention with the univalent ion forms and shorter retention times with the divalent cationic forms of the resin. Thus, the effect of higher cationic charge is shorter retention times. Most likely, the non-polar unsaturated compounds are repelled by the ionic groups and attracted by the hydrocarbon matrix of the resin. Thus, greater ionic charge results in more repulsion and shorter retention times.

The olefins are retained longer than the n-alkanes; the greatest retentions are for the smaller ions. As ionic size increases, the retention decreases. This is probably a result of an ion-induced dipole interactions between the stationary phase and the olefin. The silver form is unusual in that it forms stable complexes with olefins. Thus, the retention times of those olefins reported on Ag are extremely long.

The aromatics also show an increased retention with cation size probably due to electrostatic interactions. The adsorption of the aromatics on Ag also shows evidence for silver complex formation with the aromatic π-system.

Chlorobenzene is the most polar of the solutes. For example, ethylbenzene is retained less than chlorobenzene on all the ionic forms except Ag. The permanent dipole moment of chlorobenzene strengthens its adsorption. The electron density on the benzene ring in ethylbenzene is

much greater than in chlorobenzene where the electron withdrawing halogen atom is present. Thus, ethylbenzene is retained longer on Ag; it forms a more stable complex with silver than the electron deficient chlorobenzene ring.

Unfortunately, the data matrix in Table 58 is not complete. For example, the retention index on Ag is not reported for the heptenes and 1-octene. Consequently, a TFA on the entire matrix is not possible. (As explained in Chapter III, no data point in the original data matrix may be left blank). However, subsets of the original data matrix may be studied. The results of the reproduction and uniqueness steps for these subsets will be discussed and a complete TFA will be shown for the most representative subset.

Ten subsets were studied, three including only the monovalent cation resin forms. In each case, the purpose was to include as many of the solutes and solvents as possible. Unique Chromosorb P column was kept in seven of the subsets. One of our purposes in this study was to study the effect of this column on the factor space.

B. Reproduction and Uniqueness Test:

12 X 10 Set

Table 59 gives a summary of the reproduction for the subset called 12X10 (12 solutes, 10 solvents). Problem 12X10 includes four straight chain alkanes (solutes in Table numbered 1-4), five cyclic alkanes solutes (5-9), a branched alkane (solute 11) and two normal alkenes (solutes 12 and 19). The sorbents included the following nine ionic resin forms: Li, Na, K, Rb, Cs, Mg, Ca, Sr, Ba and the Chromosorb P column.

At four factors only ten data points (8.3%) are predicted with errors greater than ± 4 RI units. This compares to 32 points (26.7%) for three factors. The Real Error for four factors is predicted to be 2.16, whereas for three factors its value is 4.0. Experimental error is estimated by Hirsch to be less than $\pm$ 4 RI units. Thus the Real Error also substantiates a four factor space. The Indicator function, as well, shows a minimum at four factors. We conclude that problem 12X10 is a four factor space.

Table 60 gives a summary of the solute uniqueness test for problem set 12X10. Solutes 2,2,4-trimethylpentane and 1-heptene were the most unique (.90 and .83, respectively). The reason for the uniqueness of 2,2,4-trimethylpentane is probably because it is the only branched straight-chain alkane. The cause for the

Table59.Factor determination, 12x10 problem

Factor	rms error	Largest error	Percentage error > ±4	Real error	Imbedded error	Indicator (IND)
1	12.96	71.4	63.3	15.7	4.95	0.193
2	4.99	21.6	32.5	5.8	2.58	.090
3	3.44	9.9	26.7	4.0	2.21	.082
4	1.94	7.7	8.4	2.6	1.67	.073
5	1.44	5.0	3.4	2.1	1.50	.085

Table 60. Uniqueness values and correlations for solutes, 12x10 problem

Solute :	Uniqueness test value	Correlates with solutes[a]
n-pentane	0.17	2,3,4(.22)
n-hexane	.16	1,3,4(.21)
n-heptane	.20	1,2,4(.23),12
n-octane	.28	1(.22),2(.21),3(.23),12
cyclopentane	.18	6,7,8(.20)
cyclohexane	.22	5,7(.23),8(.26)
cycloheptane	.25	5,6(.23),8(.28)
cyclooctane	.32	5(.20),6(.26),7(.28)
methylcyclo-pentane	.12	11(.21)
2,2,4-trimethyl-pentane	.90	9(.21)
1-hexene	.39	3,4,19(.31)
1-heptene	.83	12(.31)

[a] Other solutes with predicted values greater than 0.15 are listed. Predicted values greater than 0.20 are given in parenthesis. Solutes 5 and 11 correlate negatively (-.18).

uniqueness of heptene is less apparent.

Also listed in Table 60 are the correlations obtained from the uniqueness test. The n-alkanes correlate well with each other as do the simple cyclic compounds (solutes 5-8) and the alkenes (solutes 12 and 19 respectively). Solutes 2,2,4-trimethylpentane and methylcyclopentane also correlate well.

Table 61 presents a summary of the sorbent uniqueness test on subset 12X10. As expected sorbent CP is extremely unique (uniqueness value .98) being the only non-ion exchange stationary phase. Mg also shows high uniqueness probably due to the small charge to volume ratio of the Mg^{+2} ion. Li and Na correlate highly together as do K, Rb and Cs and the divalent Ca, Sr and Ba.

Table 61. Uniqueness values and correlations for sorbents, 12x10 problem

Sorbent[a]	Uniqueness test value	Correlates with sorbents[a,b]
Li	0.21	Na(.21), Mg(-.18)
Na	.22	Li(.21), Mg(-.19)
K	.35	Rb(.27),Cs(.35)
Rb	.21	K(.27),Cs(.27)
Cs	.37	K(.35),Rb(.27)
Mg	.88	Li(-.18),Na(-.19)
Ca	.35	Sr(.30),Ba(.24)
Sr	.27	Ca(.30),Ba(.21)
Ba	.18	Ca(.24),Sr(.21)
CP	.97	

[a] Ionic resin forms are identified by their respective cation symbol; CP is Chromosorb P sorbent.

[b] Sorbents with predicted values greater than |0.15| are listed.

Reproduction and Uniqueness Test: All Subsets

Results for the reproduction step for all subsets are summarized in Table 62. The solutes and solvents in each subset are also identified.

Since the original data is incomplete, comparison of the effect on the same data of removal of a particular solute or solute group (or sorbent or sorbent group) is not always possible. However, notwithstanding this limitation, some important conclusions may be drawn from the data in Table 62, as well as from results on the uniqueness test (not shown).

For example, comparison of the factor size of subsets 13X9 (13 solutes, 9 sorbents) and 12X9 shows that the removal of solute benzene, the only aromatic in set 13X9, results in reduction of the factor space by one. Thus benzene represents a unique interaction. The fairly high uniqueness value of benzene (ranging between .45-.99) in all other sets where it is present further substantiates this conclusion. Chlorobenzene, which we expect to be the most unique aromatic, is indeed the most unique solute (uniqueness values 0.48-0.51) in the subsets where it is present (18X6, 19X6, 21X5).

For the sorbents, Mg, in every case, is very unique (.57-.97). Removal of Mg from the 13X10Ba sets (13 solutes, 10 solvents, contains Ba) set to produce set 13X9Ba reduces

Table 62. Reproduction summary for 13 subsets

Problem	Solutes Number	Solutes Designees [a]	Sorbents Number	Sorbents Designees [b]	Factor size
21x5	21	(1-21)	5	(Li-Cs)	3
18x6	18	(1-9,11-14)	6	(Li-Ag)	3,4
19x6	19	(1-16,19-21)	6	(Li-Cs,CP)	3,4
13x8	13	(1-9,11-14)	8	(Li-Cs,Mg,Ca,Sr)	4
13x9Mg	13	(1-9,11-14)	9	(Li-Cs,Mg,Ca,Sr,CP)	5
13x10Ag	13	(1-9,11-14)	10	(Li-Ag,Mg,Ca,Sr,CP)	5
13x10Ba	13	(1-9,11,12,14,19)	10	(Li-Cs,Mg,Ca,Sr,Ba,CP)	5,6
13x9	13	(1-9,11,12,14,19)	9	(Li-Cs,Ca,Sr,Ba,CP)	4,5
12x9P	12	(1-9,11,12,19)	9	(Li-Cs,Ca,Sr,Ba,CP)	4
12x10	12	(1-9,11,12,19)	10	(Li-Cs,Mg,Ca,Sr,Ba,CP)	4
12x9	12	(1-9,11,12,19)	9	(Li-Cs,Mg,Ca,Sr,Ba)	3

[a] See Table 59 for solute designations.

[b] See Table 61 for sorbent designations.

the factor space from 5 to 4, further substantiating its uniqueness.

The uniqueness value of Ag is extremely high (.99) in both subsets (13X10Ag, 18X6) where it is present. This would indicate a unique interaction associated with Ag is expected. However, comparison of sets 13X9Mg and 13X11Ag does not show the expected change in the factor space with removal of Ag.

When sorbent CP is removed from problem 12X10 to form 12X9, reproduction of the original data requires only three factors. In addition, removal of CP in problem 13X8 as compared to 13X9Mg, and in 21X5 compared to 19X6, show a drop in the factor space in each case. The average uniqueness value of 0.98 for CP also indicates its unique interactions. Dispersion (London) forces would be expected to predominate due to CP's lack of anionic groups. This further substantiates that one of the sorbent cofactors in problem 12X10 is associated with CP.

For the solutes the following correlations are quite evident in most of the problems: a)normal alkanes (solutes 1-4); b) cyclic alkanes (5-8); c) alkenes (12.19.21); d) aromatics (14-18); e) branched alkanes (9-11); f) cyclohexene and 2,2,4-trimethylpentane (6,11); and g) 1-hexene and cyclohexene (12,13).

For the sorbents, the univalent solvents seem to correlate according to ionic size. The smaller counter ions Li and Na correlate well together, as do the larger

counter ions, K, Rb, and Cs. The divalent forms (excepting the unique Mg) Ca, Sr, and Ba also correlate well together.

Summarizing the results of the reproduction and uniqueness tests of all the problem sets, we find: 1) for the solutes: unique interactions are associated with the aromatics, branched methyl groups and alkenes; 2) for the sorbents: Mg and Ag and P are quite unique and the rest of the sorbents correlate according to ionic size and charge.

C. Combination Step:

12X10 Problem- Using Typical Vectors

Because we wanted to study the gas-solid interactions for the most complete, yet simplest problem, we chose to continue the target factor analysis on problem 12X10.

Due to unique complexing interaction of the Ag^+ ion, we chose not to include Ag; sorbent CP was retained so that we could study how it differs from the other GSC stationary phases and consequently what it might tell us about the interactions involved with the anionic resin backbone of Amberlyst A15.

As discussed above, problem 12X10 is a four factor space; i.e., four interactions are involved between the solutes and sorbents. The combination step in the TFA may give us thebest sets of solutes and solvents which best represent these four factors. Their identity may provide some insight into the physical meaning of these interactions.

For the solutes, the best combination set involved the combination of hexane, cyclohexane, 2,2,4-trimethylpentane and hexene. Thus, each of the different kinds of solutes in the data are represented. The rms error was 2.08 with the largest error being only 10.7. This compares very favorably with the abstract reproduction error of 1.94 obtained for four factors in the reproduction step.

Comparison of the best fifteen combination sets shows correlations of the normal-alkanes (solutes 1-4) and the cyclo-alkanes (solutes 6-8). These correlations were predicted in the uniqueness tests.

For the sorbents the best combination set included the univalent and divalent K and Ca, and the two unique sorbents Mg and PP. The rms error was 2.10 with the largest error of 8.0. Correlations obtained by comparing the best thirty combinations sets are the same as obtained in the uniqueness steps. Two sets of univalent cation resin forms: Li and Na; and K, Rb, and Cs, respectively, correlate according to cation size, and all the divalent forms correlate.

The pattern tables for combination of sorbents and solutes are listed in Tables 63 and 64. For four factors, sorbent CP is represented in 100 percent of the 30 best combinations with average errors less than 4.1. Mg is present in 60 percent of these combinations. As is evident from Table 63, Mg may represent the third factor while sorbent CP is perhaps the second factor.

Similarly, for the solutes, 1-hexene is present in 69% of the best combinations and is most likely the fourth factor. The branched alkane, 2,2,4-trimethyl-pentane, is present in 60% of the best combinations and is most likely the third factor.

Correlations obtained from the pattern tables for both the solutes and solvents are similar to those obtained

Table 63. Percentage of better reproductions involving sorbent column vectors from 12x10 problem

Sorbent	Number of sorbent vector cofactors used in reproduction 1	2	3	4
Li	11.1	6.7	18.2	36.7
Na	11.1	6.7	18.2	33.3
K	11.1	20.0	18.2	30.0
Rb	11.1	26.7	18.2	30.0
Cs	11.1	20.0	18.2	23.3
Mg	11.1	0.0	54.5	60.0
Ca	11.1	20.0	30.3	33.3
Sr	11.1	20.0	33.3	33.3
Ba	11.1	13.3	21.2	10.0
CP	0.0	46.7	66.7	100.0
Total number of combinations	9	45	120	210

(continued)

Table 63. (continued)

Cutoff for better reproductions (RMS error)	25.	8.0	6.8	4.1
Number of reproductions less than cutoff	9	15	33	30
RMS error, abstract reproduction	13.0	5.0	3.4	2.0
Best combination:				
RMS error	13.4	5.5	4.1	2.1
Vectors	Rb	Sr, CP	K, Ca, CP	K, Mg, Ca, CP

Table 64. Percentage of better reproductions involving solute column vectors from 12x10 problem

Solute No[a]	Number of solute vector cofactors used in reproduction 1	2	3	4
1	10.	8.3	32.	43.2
2	10.	16.7	21.	21.6
3	10.	20.8	24.	31.1
4	10.	16.7	22.	14.9
5	10.	0.0	15.	21.6
6	10.	20.8	17.	24.3
7	10.	25.0	20.	18.9
8	10.	25.0	19.	21.6
9	10.	20.8	21.	31.1
11	10.	0.0	37.	59.5
12	0.	25.0	39.	48.6
19	0.	16.7	33.	68.9

(continued)

Table 64. (continued)

Total number of combinations	12	66	220	495
Cutoff for better reproductions	20.	7.	7.	4.
Number of reproductions less than cutoff	10	24	100	74
RMS error, abstract reproduction	12.96	4.99	3.44	1.97
Best combination:				
RMS error	13.6	5.2	4.2	2.1
Vector No[a]	3	8,12	1,7,12	3,7,11,12

[a] See Table 59 for solute designations.

on the uniqueness steps of the factor analysis (eg., cyclic alkanes correlate for the solutes).

Thus from the combination of solutes and solvent vectors we find that the unique entities are represented in the best combinations, together with a representative of each of the different types of compounds. However, we would like to be able to identify each of the factors with physically meaningful vectors. This can be achieved in the target transformations step.

D. Target Transformation

A total of 95 physical and chemical vectors were tested (47 for the solutes, 48 for the sorbents) as well as their squares, logs and reciprocals. Seventeen (7 for the solutes; 10 for the sorbents) were chemical vectors including unity and correlation vectors derived from the uniqueness tests.

SORBENTS

Table 65 lists all the vectors which target transformed well for the sorbents. No specific data vectors (such as water content) were available for the various cationic forms of the resins. However, since we are measuring the effect of cation substitution on the resin we would expect that cation data vectors (such as enthalpy of cation hydration) would test well. In general, this is the case. The effect of the resin backbone on the cation is probably nearly constant for all cationic forms.

Thus, as expected, ionic charge (oxidation number) is an excellent vector, as are enthalpy and free energy of cation hydration. Interestingly, the predicted value for the ionic charge of the free-floated sorbent CP is -1.0. This may relate to the presence of the -Si-O- functionality in CP. However, the ionic size vectors (ionic radius, volume, molar refraction) only tested fairly.

Patterns for some of the vectors seem to be duplicative. For example, R and Pauling's electronegativity nearly

Table 65. Summary of selected target transformations for physically significant sorbent test vectors for 12x10 problem, four factors

Test vector	Example values[a]	Points[b]	Qualitative evaluation[c]	Value for CP[d]	Functional forms[e]
absolute electron affinity	118,ff	5	f	34.9	(s)
electron affinity (ev)	.47,6.11	6	f	-45.0	r
electronegativity (Pauling)	.93,1.0	9	g	1.9	r
enthalpy of cation formation	-57.,-130	9	f	110.2	
enthalpy of cation hydration	106,399	9	f	-156.1	(r,1)
entropy of cation hydration	26,61	9	g	30.7	(r)
free energy, cation formation	-63,-132	9	f	118.1	
free energy, cation hydration	98,381	9	g	-165.3	(r,1)
free energy, ion solvation	90,ff	5	f	40.1	
hydration number	1.5,5.2	7	f	7.04	
hydrated radii	1.86,1.06	9	g	.43	r
ionic charge (valence)	1,2	9	v	-1.	
ionic charge/radius[96]	1.03,2.02	9	g	3.38	r,s,(1)

(continued)

Table 65 (continued)

ionic mobility[95]	43.5,ff	5	f	34.9	(s)
ionic molar refraction[103]	0.9,1.2	8	f	-1.72	
ionic potential (polarizing strength)[104]	1.0,1.8	9	f	6.39	r
ionic stability ratio[104]	2.44,5.82	9	g	-1.12	r,s,l
ionic volume	-6.6,-28.6	9	f	6.34	
ionic radius, crystal	.97,.99	9	f	-1.8	
R[f]	14.2,22.6	8	g	-5.6	r,s
R'= R/charge	14.2,12.3	8	g	19.3	
softness parameter (σ_p)[98]	.21,.18	9	g	.33	
" " (σ_A)[98]	.93,.09	9	g	-.10	
Li(1),Na(1),Mg(-1)[g]	1,0	10	g	--	
K(1),Rb(1),Cs(1)[g]	0,0	10	g	--	
Li(1),Na(1),Ca(1),Sr(1),Ba(1)[g]	1,1	10	g	--	
unity	1,1	10	v	--	

(continued)

Table 65. (continued)

[a] Example test points are given for Na and Ca sorbents.

[b] Number of points on test vector.

[c] v=very good agreement between test vector and predicted vector; g=good agreement; f=fair agreement (several points poorly predicted or pattern only predicted).

[d] Predicted values for free-floated Chromosorb P values.

[e] Other functional forms of the vectors which target transformed well: r=reciprocal; s=square; l=log. Forms in parenthesis target transformed only fair.

[f] $R = IP(r_m + r_x)/ne^2$ where IP is the ionization potential for cation m, r_m and r_x are the cation and anion radii, respectively, n is the charge on m and e is the charge of the electron.[101]

[g] Correlation vectors based on results from the uniqueness test. All other values "0", except CP which is free-floated.

duplicate mathematically the unity vector, and R^e and Pearson's softness parameter seem to duplicate the ionic charge vector.

Beside uniqueness vectors for the Mg and P sorbents, three correlation vectors tested well. They are correlations of a) K, Rb, and Cs; b)Li, Na, Ca, Sr and Ba; and c) Li and Na correlated negatively with Mg. Thus, correlations derived from the uniqueness test do indeed target transform well.

Unity is a very good vector. It is probably an overall cofactor associated with the interaction of the solutes with the anionic resin matrix. (As we shall see in the combination of physical vectors, unity is associated with the first factor).

Most of the vectors which did not test well involved atomic data (as opposed to cationic data). Unsuccessfully tested atomic parameters included atomic weight, equivalent weight, diatomic bond strength and molecular polarizability. Other vectors which tested poorly included: crystal bond length, second ionization potential, magnetic susceptibility of the chloride, enthalpy of formation, entropy of cation formation, cube of the (Jenny and Pallman) ionic radius, equivalent conductance, and polarizability.

Target transformations on the data set with sorbent CP removed (set 12X9) (a three factor space) gives essentially the same results as for the 12X10 problem. This further substantiates the high uniqueness of Sorbent CP.

Column 5 of Table 65 lists the predicted values for

the free-floated Chromosorb CP data points on each of the target transformation vectors. For example, the predicted value for the ionic charge of Chromosorb P is -1.0, as opposed to the input values for all the other sorbents which are all positive. This significantly demonstrates the unique character of sorbent Chromosorb P. Chromosorb P is predicted to be opposite in character to the other sorbents.

Similarly, the predicted values for Chromosorb P are significantly different for the following vectors: free energies and enthalpies of cation formation and hydration, electron affinity, hydrated radii, ionic molar refraction and stability ratio, R, and softness parameter σ_A. Since these vectors have no physical meaning for CP, the free-floated values only serve to remind us that CP is unique.

SOLUTES

Table 66 lists all the vectors which target transformed well for the solutes. As expected (see Appendix B) molar refraction, molar volume, enthalpy of vaporization, and molecular weight are good vectors. Two carbon number vectors also target transformed well. One vector (NC) tested for the total number of carbons in a solute; another vector (CN) free-floated the data points for the cyclic and branched solutes (solutes 5-10). Results from the latter vector further verify the results obtained in the

Table 66 Summary of selected target transformations for physically significant solute vectors for 12x10 problem, four factors

Test vector	Example values[a]	Points[b]	Qualitative evaluation[c]
boiling point (^{o}K)[d]	341,372	12	u
carbon number CN	6,ff	6	v
carbon number NC	6,8	12	v
critical volume[d]	.37,.47	12	g
critical temperature (^{o}C)[d]	234,ff	9	g
enthalpy of combustion	99,ff	10	g
enthalpy of vaporization[d]	6.9,7.4	12	g
$d(\Delta H_v)/dt$	-14,ff	6	g
entropy of vaporization[d]	92,101	10	g
melting point (^{o}K)	178,166	11	f
molar refraction	299,393	9	v
molar volume	132,166	9	v
molecular weight	86,114	12	g

(continued)

Table 66.(continued)

surface tension[d]	18,18	12	g
viscosity	.30,ff	6	g
cyclic uniqueness	0,0	12	f
alkene uniqueness	0,0	12	g
unbranched cyclic uniqueness	0,0	12	f
n-alkane + hexene uniqueness	0,0	12	v

[a] Example values are given for hexane (solute 2) and 2,2,4-trimethylpentane (solute 11).

[b] Number of points on test vector.

[c] v=very good agreement between test vector and predicted vector; g=good agreement; f=fair agreement; (several points poorly predicted or pattern only predicted); u=equivalent to unity vector.

[d] Square of the vector target transformed successfully. For critical temperature and entropy of vaporization the agreement is only fair.

gas-liquid chromatography problems of the previous chapters. Each branched methyl group is found equivalent to 0.8 carbon numbers and a cyclic structure adds 0.6 to the carbon number. Thus, a solute such as 2,2,4-trimethylpentane is predicted to have a carbon number of 6.8 (3 x 0.6 for the three methyl groups plus 5.0 from the pentane chain).

As found in several factor analyses of retention indices, the square of the boiling point (^{o}K) target transformed very well. Thus, this cofactor is significant for both gas-liquid and gas-solid chromatographic systems. However, as we shall see and as noted in many of the GLC problems, its importance in an overall model is shown to be minimal on the combination step for GSC.

Structural and correlation vectors which tested well included test vectors for unsaturation and cyclic uniqueness. A vector testing for cyclic uniqueness, with solute 9, the branched cyclic compound, free floated also tested well.

The following test vectors did not target transform well: boiling and melting points (all ^{o}C), critical temperature, critical pressure, dielectric constant, critical density, critical value of PV/RT, density, refractive index, free energy of formation, log of free energy of formation, vapor pressure, log of vapor pressure, specific dispersion, pressure dependence of the boiling point, alkane uniqueness, and branching uniqueness. Vectors testing for entropy of vaporization, boiling point (^{o}K), specific refraction, and refractive intercept were too much like the unity vector to furnish new information.

E. Combination of Basic Vectors

SORBENTS

The best combination of sorbent vectors gave quite satisfying results. An rms error of 2.4 and a largest error of 10.0 were obtained with the four vectors: unity, free energy of ionic solvation, electron affinity and R. The free energy of solvation correlates well with ionic mobility in the best combinations and in pattern tables; this vector is probably equivalent to the fourth factor.

Unity is the significant factor (present in 100% of the best combinations, all factors) and appears to be equivalent to factor one. It accounts for the major amount of variance in the data. This suggests that the major interaction for the system is a common constant property due to adsorption on the resin backbone.

SOLUTES

The best combination of four solute vectors (without the Van der Waals vectors) reproduced the original data matrix with an rms error of 6.0; the largest error was 18.0. The best set included the carbon number vector (CN), alkene uniqueness, critical volume and temperature dependence of the enthalpy of vaporization. Comparing the best 20 combinations, the temperature dependence vector is found also to be equivalent to alkane uniqueness or alkane + hexane uniqueness.

The carbon number vector CN is the significant vector for the best 100 combinations, being represented in 82 per cent for 4 factors. This is expected from the definition of the retention index. The dominance of the vector shows that the major interaction is probably a dispersive type related to molecular size.

The molar refraction and molar volume vectors correlate well together in the best combinations. This has been found as well in the GLC problems (as explained in Chapter V and VIII).

Pattern tables for the best solute combinations show no definite results. Carbon number CN and the square of the enthalpy are likely equivalent to factor three.

F. Prediction of Data

As described in Equation 30, the best set of cofactors r_{xm} can be used to predict new data. Using the column matrix cofactors calculated from the TFA combination step with that set in the factor space, we can predict the retention index, I_{xk}, of new solute x on an original sorbent k from the equation:

$$I_{xk} = \sum_{1}^{4} r_{xm} \; c_{mk} \qquad (64)$$

The only sorbent not included in our test set is the Ag form of the resin. This was due to the unique complexation of the Ag^+ ion with some of the solutes, a totally new interaction that we would not expect to be spanned by our best set of four factors. Thus predictions on Ag would not be expected to be satisfactory and indeed attempts to do so were unsuccessful.

However, predictions using the solute vectors to obtain new retention index data for solutes on our original sorbent set do give excellent results. Inserting the values of the vectors; carbon number, alkene uniqueness, critical volume and temperature dependence of the enthalpy for any new solute, into the above equation together with the column-designee matrix loadings associated with those vectors should give us the retention index for the solute.

For example, the retention index of 2,4-dimethylpentene on Na is 647. We can estimate the carbon number for this solute using the rules formulated earlier. Each methyl group is assigned a value of 0.8 and the pentyl skeleton has a value of 5.0. Thus the carbon number of this solute is 6.6. It is obviously not an alkene and therefore the alkene value is 0.0. The values for the critical volume can be estimated from available data for similar solutes by noting that each methyl group contributes approximately 0.5 to the value. The critical volume of 2,2,4-trimethylpentane is 0.47 l/m and therefore the value for our solute is estimated as 0.41 l/m. Similarly the temperature dependence of enthalpy is estimated as -13.7 cal/^{o}m. In Table 67, is listed the calculated column matrix. The coefficients in matrix[C] for Na are 72.39, 70.81, -3.08 and 313.4.

Inserting these values into equation 64 gives the predicted value I*, for the retention index:

$$I^* = (72.39)\cdot(6.6)+(70.8)\cdot(0.0)+(=3.08)\cdot(-13.7)+ (313.4)\cdot(0.41) = 648.0 \quad (65)$$

well within experimental error of the reported value 647.

Similarly, for this same solute on the unique sorbent CP the equation becomes:

$$I^* = (101.6)(6.6)+(107)(.0)+(.355)(-13.7)+(10.07)(.41)=661.6 \quad (65A)$$

The reported value is 663.

Table 67. Column matrix for best combination of basic sorbent vectors

	Sorbent[a]									
Factor	Li	Na	K	Rb	Cs	Mg	Ca	Sr	Ba	PP
1	72.89	72.39	70.34	69.82	73.13	55.37	57.73	60.03	62.93	101.6
2	69.55	70.81	46.03	51.07	35.52	61.71	105.8	98.21	86.71	-107.4
3	-3.02	-3.08	-1.17	-2.25	-.654	-9.37	-8.40	-7.48	-6.25	.355
4	311.1	313.4	437.2	396.8	414.3	328.9	319.6	323.3	327.8	-10.1

[a] See Table 9 for sorbent designations.

Predictions for an alkene solute, 1-actene, are almost as good. We can estimate, as above, values for the enthalpy temperature dependence and critical volume for solute 1—octene. The critical volume, for example, is estimated by comparing the free-floated values for solutes 1-hexene (0.37) and 1-heptene (0.44) obtained in the target transformation step. Addition of one methylene group to 1-hexene adds 0.07 units to get the critical volume of 1-heptene. Thus, the critical volume of 1-octene can be estimated as 0.44 + 0.07 or 0.51 units. Similarly, the temperature dependence of enthalpy for 1-octene is estimated as -10.6. The carbon number is 8.0 and the alkene uniqueness is obviously 1.0.

Predicting then the retention index for 1-octene on Na, we find:

$$I^* = (72.39)\cdot(8.0) + (70.81)\cdot(1.0) + (-3.08)\cdot(-10.6) \quad (313.4)\cdot(0.51) = 841.8 \tag{66}$$

The column designee coefficients are the same as those above. The reported value is 853.

Similarly, for sorbent CP:

$$I^* = (101.6)\cdot(8.0) + (-10.7)\cdot(1.0) + (0.355)\cdot(-10.6) + (-10.07)\cdot(0.51) = 793 \tag{66a}$$

The reported value is 791.

Thus, our designations for the physical meaning of the cofactors are quite valid and useful for prediction.

G. SUMMARY

In summary, target factor analysis has provided us with the number and identity of the solute and solvent parameters in a gas-solid chromatography problem. Reproduction and uniqueness tests show the expected unique interactions of the Ag+ ion due to complex formation, as well as the extra interaction associated with aromatic compounds.

For a representative 12X10 set, four factors are sufficient to span the data space. Combination of vectors from the data matrix associates these four factors with representatives of each of the solute or sorbent classes. The solutes are associated with normal, cyclic and unsaturated and branched contributions. The sorbents are associated with univalent and divalent cation resin forms as well as the two unique sorbents: Mg and Chromosorb P.

Parameters which tested for the sorbents included Pauling electronegativity, entropy and free energy of cation hydration, ionic charge, hydrated radii, ionic charge/radius, and Pearson's softness parameter. Chemical vectors which gave good transformations included Mg and CP sorbent uniqueness vectors and two correlation vectors obtained from the uniqueness test.

Molar refraction, molar volume, enthalpy of vaporization and boiling point squared were good solute

physical vectors. Chemical solute vectors which transformed well included two carbon number vectors and unsaturation and cyclic uniqueness vectors.

The best set of physical vectors found in the final combination step may be associated with theoretical interaction terms as described in Appendix B. Thus for the solutes, vectors for carbon number and critical volume may be associated with dispersion terms; alkene uniqueness and enthalpy temperature dependence with dipole-dipole interaction terms. Similarly, for the solvents, the dominant vector, unity, accounts for adsorption by the resin itself, whereas the other vectors account for interaction of the counter ions with the solutes.

As a final test of our solution, very good predictions for new solutes on two of the original sorbents in our sample problem are obtained.

Appendix A

TARGET FACTOR ANALYSIS OF ION EXCHANGE EQUILIBRIUM CONSTANTS

Introduction

Ion exchange resins are widely used in research and industry. However, the theory of ion exchange interactions still requires an all-encompassing treatment. What factors are responsible for the resin's preference for one ion over another? In this appendix we will discuss the target factor analysis of a data matrix containing ion exchange equilibrium constants for 20 cations on Dowex 50 resin.

Theory

A typical cation exchange may be represented by the following equation:

$$A^+ + B^+R^- \rightleftharpoons B^+ + A^+R^- \qquad (67)$$

where cations A^+ and B^+ in solution are exchanged with counter ions in the resin, B^+R^- and A^+R^-.[31]

The equilibrium between ion exchangers and solutions has been the subject of numerous experimental and theoretical investigations. Various models of the exchange have been proposed,[31] each model satisfactory in

some respects, but no model adequately explaining the ion exchanger process in all aspects.

Ion exchange equilibrium is attained when an ion exchanger is placed in an electrolyte solution containing a counter ion which is different from that in the ion exchanger. The selectivity of an ion exchanger describes the preference of the ion exchanger for one counter ion over another counter ion. Quantities such as separation factors, selectivity coefficients and distribution coefficients have been used to quantify the ion exchange equilibrium.

Theoretically, the thermodynamic equilibrium constant, K_B^A, is used. This is defined by:

$$\Delta G^{\circ} = -RT \ln K_B^A \tag{68}$$

ΔG° is the standard free energy change for the ion exchanger. If electrolyte sorption and changes in swelling of the ion exchange resin are neglibible, the thermodynamic equilibrium constant can be related to the activities of the ions by:

$$K_B^A = \frac{\bar{a}_A^{|Z_B|} \; a_B^{|Z_A|}}{\bar{a}_B^{|Z_A|} \; a_A^{|Z_B|}} \tag{69}$$

where a_A, a_B, $\bar{a}_A$, and $\bar{a}_B$ are the activities of counter ions A and B in the solution and within the ion exchange resin, respectively, and Z_A and Z_B are the

charges of the respective counter ions.[31]

Factors associated with selectivity may include: counter ion charge, ionic solvation, swelling pressure, sieve action and specific interactions such as ion pair formation, electrostatic attractions and London interactions.

We may define individual hypothetical equilibrium constants K_A and K_B for each counter ion to represent the free energy change for each counter ions interaction with the ion exchanger resin. For $A \longrightarrow AR$,

$$\Delta G^{o}_{A} = -RT \ln K_A \tag{70}$$

and for $B \longrightarrow BR$,

$$\Delta G^{o}_{B} = -RT \ln K_B \tag{71}$$

Thus for the equilibrium reaction $A + BR \longrightarrow AR + BR$:

$$\Delta G^{A}_{B} = -RT \ln K^{A}_{B} \tag{72}$$

and

$$\ln K^{A}_{B} = \ln K_A - \ln K_B \tag{73}$$

where

$$K^{A}_{B} = \frac{K_A}{K_B} \tag{74}$$

Thus the rational equilibrium constant for the ion exchange process can be rewritten in terms of the quotient of the individual equilibrium constants for cations A and B.

The equilibrium constant for any counter ion exchange can be calculated from the equilibrium constants for

the exchange of that counter ion with a third counter ion. For example, the equilibrium constant for the exchange of H and Li, K_H^{Li}, can be written in terms of the equilibrium constants for the exchanges of Li and Na (K_H^{Na} and K_{Li}^{Na}).

$$K_H^{Li} = \frac{K_H^{Li}}{K_{Li}^{Na}} \tag{75}$$

In general,

$$K_A^B = \frac{K_A^C}{K_B^C} \tag{76}$$

This "triangle" rule allows us to set up an ion selectivity where the exchange of all ions is expressed relative to one of the ions. The affinity of a lithium ion to a resin has been arbitrarily assigned a value of unity.[96] All other ion affinities are then expressed relative to lithium according to the "triangle" rule.

Data

Bonner and Smith[91] experimentally determined the rational equilibrium constants for twenty univalent and divalent cations on Dowex 50 X4, X8, and X16 resins. These three resins differ only in the degree of divinylbenzene content: 4, 8 and 16 percent, respectively.

The equilibrium constants were corrected for the ionic strengths of the solutions. The data were reported

in the form of a selectivity scale with the affinity of each ion for the resins based on a lithium affinity of unity. The data are tabulated in Table 68.

The data were converted to a data matrix of equilibrium constants for the 20 ions using the triangle rule. Thus three 20 X 20 data matrices were generated for the three Dowex 50 resin forms. The data matrix for the X8 form of the resin is tabulated in Table 69.

The data matrices were then subjected to target factor analyses. The data were transformed to the log since the log of the equilibrium constant is related to the free energy of the ion exchange (Eq. 68).

Results and Discussion

Results were equivalent for all three resin forms. From Equation 73 we expect the data to be a two factor space, each factor then being related to the individual ionic equilibrium constants K_A and K_B. Mathematically, the factor space must be equal to two since the data was generated using the triangle rule as above. As anticipated, with two factors the data matrix is reproduced exactly.

Since each column of the data matrix is equivalent to a selectivity scale of all ions relative to the column ion, we expect that all possible combinations of the original columns using two columns should reproduce the

Table 68. Selectivity Scale for Twenty Cations on Dowex 50[a]

Cation	4% DVB	8% DVB	16% DVB
Li	1.00	1.00	1.00
H	1.32	1.27	1.47
Na	1.58	1.98	2.37
NH_4	1.90	2.55	3.34
K	2.27	2.90	4.50
Rb	2.46	3.16	4.62
Cs	2.67	3.25	4.66
Ag	4.73	8.51	22.9
Tl	6.71	12.4	28.5
UO_2	2.36	2.45	3.34
Mg	2.95	3.29	3.51
Zn	3.13	3.47	3.78
Co	3.23	3.74	3.81
Cu	3.29	3.85	4.46
Cd	3.37	3.88	4.95
Ni	3.45	3.93	4.06
Ca	4.15	5.16	7.27
Sr	4.70	6.51	10.1
Pb	6.56	9.91	18.0
Ba	7.47	11.5	20.8

[a]Data from reference 91.

Table 69. Log of rational equilibrium constants for twenty cations on Dowex 50X8

	H	Li	Na	K	Rb	Cs	Ag
H	0.0	0.1038E 00	-0.1931E 00	-0.3585E 00	-0.3958E 00	-0.4078E 00	-0.8268E 00
Li	-0.1040E 00	0.0	-0.2967E 00	-0.4622E 00	-0.5003E 00	-0.5114E 00	-0.9281E 00
Na	0.1928E 00	0.2967E 00	0.0	-0.1656E 00	-0.2027E 00	-0.2154E 00	-0.6326E 00
K	0.3585E 00	0.4624E 00	0.1658E 00	0.0	-0.3716E-01	-0.4964E-01	-0.4672E 00
Rb	0.3959E 00	0.4997E 00	0.2030E 00	0.3743E-01	0.0	-0.1233E-01	-0.4295E 00
Cs	0.4081E 00	0.5119E 00	0.2151E 00	0.4961E-01	0.1199E-01	0.0	-0.4179E 00
Ag	0.8256E 00	0.9294E 00	0.6328E 00	0.4670E 00	0.4298E 00	0.4175E 00	0.0
Tl	0.9896E 00	0.1093E 01	0.7968E 00	0.6310E 00	0.5937E 00	0.5815E 00	0.1641E 00
NH_4	0.3028E 00	0.4065E 00	0.1099E 00	-0.5601E-01	-0.9313E-01	-0.1051E 00	-0.5229E 00
Mg	0.4135E 00	0.5172E 00	0.2206E 00	0.5461E-01	0.1745E-01	0.5180E-02	-0.4123E 00
Ca	0.6068E 00	0.7126E 00	0.4160E 00	0.2502E 00	0.2130E 00	0.2009E 00	-0.2168E 00
Sr	0.7098E 00	0.8136E 00	0.5169E 00	0.3512E 00	0.3139E 00	0.3017E 00	-0.1158E 00
Ba	0.9569E 00	0.1061E 01	0.7640E 00	0.5984E 00	0.5610E 00	0.5488E 00	0.1313E 00
Co	0.4691E 00	0.5729E 00	0.2762E 00	0.1106E 00	0.7335E-01	0.6107E-01	-0.3565E 00
Ni	0.4905E 00	0.5944E 00	0.2978E 00	0.1319E 00	0.9482E-01	0.8243E-01	-0.3354E 00
Cu	0.4816E 00	0.5855E 00	0.2887E 00	0.1232E 00	0.8565E-01	0.7372E-01	-0.3439E 00
Zn	0.4365E 00	0.5403E 00	0.2438E 00	0.7809E-01	0.4060E-01	0.2857E-01	-0.3893E 00
Cd	0.4850E 00	0.5888E 00	0.2923E 00	0.1265E 00	0.8920E-01	0.7700E-01	-0.3410E 00
Pb	0.6923E 00	0.9961E 00	0.6994E 00	0.5336E 00	0.4964E 00	0.4842E 00	0.6670E-01
UO_2	0.2853E 00	0.3892E 00	0.9237E-01	-0.7314E-01	-0.1107E 00	-0.1226E 00	-0.5406E 00

(continued)

Table 69 (continued)

	Tl	NH_4	Mg	Ca	Sr	Ba	Co
H	-0.9914E 00	-0.3026E 00	-0.4134E 00	-0.6091E 00	-0.7100E 00	-0.9586E 00	-0.4685E 00
Li	-0.1092E 01	-0.4067E 00	-0.5171E 00	-0.7122E 00	-0.8125E 00	-0.1060E 01	-0.5735E 00
Na	-0.7959E 00	-0.1101E 00	-0.2204E 00	-0.4157E 00	-0.5171E 00	-0.7645E 00	-0.2765E 00
K	-0.6308E 00	0.5576E-01	-0.5502E-01	-0.2503E 00	-0.3516E 00	-0.5986E 00	-0.1107E 00
Rb	-0.5935E 00	0.9307E-01	-0.1773E-01	-0.2132E 00	-0.3143E 00	-0.5607E 00	-0.7314E-01
Cs	-0.5817E 00	0.1055E 00	-0.5243E-02	-0.2007E 00	-0.3019E 00	-0.5482E 00	-0.6098E-01
Ag	-0.1643E 00	0.5228E 00	0.4123E 00	0.2167E 00	0.1159E 00	-0.1314E 00	0.3566E 00
Tl	0.0	0.6869E 00	0.5762E 00	0.3808E 00	0.2799E 00	0.3262E-01	0.5206E 00
NH_4	-0.6661E 00	0.0	-0.1107E 00	-0.3063E 00	-0.4067E 00	-0.6536E 00	-0.1662E 00
Mg	-0.5768E 00	0.1106E 00	0.0	-0.1952E 00	-0.2967E 00	-0.5436E 00	-0.5552E-01
Ca	-0.3809E 00	0.3062E 00	0.1953E 00	0.0	-0.1007E 00	-0.3478E 00	0.1399E 00
Sr	-0.2798E 00	0.4071E 00	0.2964E 00	0.1011E 00	0.0	-0.2472E 00	0.2408E 00
Ba	-0.3292E-01	0.6542E 00	0.5434E 00	0.3481E 00	0.2472E 00	0.0	0.4878E 00
Co	-0.5200E 00	0.1664E 00	0.5576E-01	-0.1397E 00	-0.2403E 00	-0.4681E 00	0.0
Ni	-0.4989E 00	0.1878E 00	0.7737E-01	-0.1180E 00	-0.2190E 00	-0.4660E 00	0.2160E-01
Cu	-0.5086E 00	0.1790E 00	0.6819E-01	-0.1273E 00	-0.2284E 00	-0.4750E 00	0.1242E-01
Zn	-0.5528E 00	0.1339E 00	0.2325E-01	-0.1726E 00	-0.2733E 00	-0.5200E 00	-0.3245E-01
Cd	-0.5045E 00	0.1824E 00	0.7151E-01	-0.1238E 00	-0.2248E 00	-0.4724E 00	0.1578E-01
Pb	-0.9745E-01	0.5895E 00	0.4789E 00	0.2835E 00	0.1824E 00	-0.6449E-01	0.4232E 00
UO_2	-0.7033E 00	-0.1726E-01	-0.1278E 00	-0.3233E 00	-0.4248E 00	-0.6716E 00	-0.1838E 00

(continued)

Table 69.(continued)

	Ni	Cu	Zn	Cd	Pb	UO_2
H	-0.4908E 00	-0.4815E 00	-0.4365E 00	-0.4855E 00	-0.8928E 00	-0.2857E 00
Li	-0.5952E 00	-0.5850E 00	-0.5406E 00	-0.5884E 00	-0.9957E 00	-0.3893E 00
Na	-0.2976E 00	-0.2890E 00	-0.2434E 00	-0.2924E 00	-0.6990E 00	-0.9259E-01
K	-0.1319E 00	-0.1232E 00	-0.7779E-01	-0.1267E 00	-0.5331E 00	0.7335E-01
Rb	-0.9474E-01	-0.8566E-01	-0.4048E-01	-0.8936E-01	-0.4962E 00	0.1106E 00
Cs	-0.8249E-01	-0.7366E-01	-0.2826E-01	-0.7676E-01	-0.4841E 00	0.1229E 00
Ag	0.3351E 00	0.3440E 00	0.3692E 00	0.3406E 00	-0.6651E-01	0.5402E 00
Tl	0.4990E 00	0.5080E 00	0.5530E 00	0.5046E 00	0.9726E-01	0.7042E 00
NH_4	-0.1878E 00	-0.1791E 00	-0.1337E 00	-0.1824E 00	-0.5901E 00	0.1745E-01
Mg	-0.7727E-01	-0.6803E-01	-0.2319E-01	-0.7160E-01	-0.4789E 00	0.1281E 00
Ca	0.1183E 00	0.1271E 00	0.1723E 00	0.1239E 00	-0.2832E 00	0.3235E 00
Sr	0.2191E 00	0.2281E 00	0.2732E 00	0.2248E 00	-0.1824E 00	0.4244E 00
Ba	0.4663E 00	0.4752E 00	0.5204E 00	0.4719E 00	0.6446E-01	0.6715E 00
Co	-0.2136E-01	-0.1278E-01	0.3262E-01	-0.1592E-01	-0.4237E 00	0.1838E 00
Ni	0.0	0.9026E-02	0.5423E-01	0.5609E-02	-0.4012E 00	0.2052E 00
Cu	-0.8774E-02	0.0	0.4532E-01	-0.3486E-02	-0.4112E 00	0.1962E 00
Zn	-0.5404E-01	-0.4528E-01	0.0	-0.4866E-01	-0.4559E 00	0.1511E 00
Cd	-0.5683E-02	0.3460E-02	0.4844E-01	0.0	-0.4067E 00	0.1998E 00
Pb	0.4017E 00	0.4106E 00	0.4558E 00	0.4072E 00	0.0	0.6069E 00
UO_2	-0.2055E 00	-0.1965E 00	-0.1512E 00	-0.2000E 00	-0.6073E 00	0.0

original data matrix exactly. All possible combinations of typical vectors had root mean squared errors less than two percent (this error is probably due to round-off errors).

No cations were truly unique as determined in a uniqueness test. Hydrogen (uniqueness value=0.19), lithium (0.26), thallium (0.21), barium (0.19) and lead (0.16) exhibited slight uniqueness. Two sets of cations seemed to correlate in this test: hydrogen-lithium and silver-thallium-barium-lead.

We have in this system a unique opportunity to study the factor(s) directly related to an ion's affinity to a resin (K_A or K_B). Therefore, target transformation can provide us with an ideal tool to determine the real physical factor related to the individual ionic equilibrium constants.

Eleven target vectors target transformed with good fits. They are listed in Table 70. Equation 73 can be rewritten:

$$\ln K_B^A = (\text{unity})_B \ln K_A - \ln K_B (\text{unity})_A \quad (77)$$

where unity is the cofactor for each ionic equilibrium constant. Thus, the unity vector should best represent one of the cofactors. Eight of the vectors which target transformed well are related to ion-water interactions. A correlation vector obtained from the uniqueness test also transformed well.

Table 70. Vectors which target transformed well on Dowex 50X8 data matrix

Designation	Description	Typical values[a]
CV	correlation vector[b]	-1,1
DGI	free energy of solvation [94]	-273,ff
DHC	enthalpy of solvation - calculated [94]	-277,ff
DHI	enthalpy of solvation [94]	-146,ff
ED	energy of dissociation [95]	73,ff
HR	hydrated radii [94]	3.4,
HRP	hydrated radii - Pallman [96]	7.3,ff
IC	limited ionic conductance [97]	38.6,ff
IM	ionic mobility [95]	33.5,ff
UNI	unity	1,1
WU	resin water uptake [91]	211,122

[a] Values given for Li and Ba resin forms; free-floated values are designated by ff.

[b] CV obtained from uniqueness test: Values for H,Li =-1.0; Tl,NH_4,Ba,Pb = 1.0; all other resin ionic forms = 0.0.

Most of the parameters tested poorly, as shown in Table 71, including: ionic volume and charge, hydration number, ionization potential and softness. Many of these vectors are related to ionic hydration and would be expected to be important in the dehydration of an ion in the ion exchanger.

The effect of ionic charge was also tested by factor analyzing the univalent and divalent portions of the data matrices individually (i.e., 9 X 9 and 11 X 11 matrices). The results were equivalent to the factor analyses results on the total 20 X 20 matrix, indicating that ionic charge is not a factor.

Combination of the eleven target vectors which transformed satisfactorily resulted in 17 combinations with RMS errors less than 2%. They are listed in Table 72.

As expected from Equation 77, unity is represented in the four best combinations. However, almost as valid reproductions are obtained without unity as a cofactor. We can rewrite Equation 73 in the form:

$$\ln K_B^A = k_A F_A + k_B F_B \qquad (78)$$

where cofactor F_A and F_B are related to the equilibrium constant K_B^A by factor loadings k_A and k_B. Thus, for example, the physically meaningful analog of Equation 82 would be:

$$\ln K_B^A = k_A \, (E_{dissociation})_A + k_B \, (\text{hydrated ratio}) \qquad (79)$$

Table 71. Vectors which did not target transform successfully for the ion exchange data matrix

Vector type	Vector description [a]
Elemental	atomic weight[88]; atomic number[88]; atomic radii[95]; boiling point[88]; density[88]; energy of dissociation[95]; entropy at 298°K[88]; entropy of melting[88]; entropy of vaporization[88]; heat capacity[88]; heat of melting[88]; heat of formation[88]; heat of vaporization[88]; heat of vaporization constants[88]; magnetic susceptibility[88]; polarizability[98]; molecular bond length[88]; bond strength[88]; molecular polarizability[98].
Ionic	crystal radii (3)[88,99,100]; charge/volume[101]; charge/radii[96]; desolvation energy[98]; Edward's α and β[98]; E^{o}[95]; electron affinity (3)[102,107]; electron density[104], -average[104]; electronegativity (Pauling and from stability ratios) (6)[104,103,95]; -cation[104]; E_K[105]; E_{KH}[105]; entropy[100]; equivalent conductance (2)[88,106,97]; E_m[102]; excess polarizing strength[109]; free energy of sublimation[103]; formation-enthalpy, entropy and free energy (4)[88,103]; hydration - enthalpy, entropy and free energy (12)[107,97,102,95,99]; hydration number (6)[94,31,95,108]; hard and soft acidity[98]; ionic B_n[107]; idealized energy density[94]; ionization potential -

Table 71. (continued)

first, second and third (in volts and kcal/gm) (7)[88,104,103]; -average[104]; K_d[96]; magnetic moment[103]; magnetic susceptibility of chloride[88]; maximum electron storage density[94]; metal class[101]; mobility[95]; molar refraction-Fajans[103]; molar volume - conventional[100], ionic[100], crystalline[100] and ionic crystalline[100]; NMR ionic coefficients (4)[97]; partial molar heat capacities; polarizability[102,98,103]; potential[102,103]; radii -hydrated[102,103], -unhydrated[95]; R' [101]; R" [101]; relative affinities of ligand atoms (4)[98]; size[96]; softness parameters (3)[98]; solvent isotope effects in hydration (3)[97]; stability ratio[104]; S_x(cosphere numbers) (2)[97]; thermodynamic potential[94]; viscosity[100]; water uptake of cationic resin[96].

[a]Values in parenthesis are total number of different vectors with same designation.

Table 72. Best combination sets of physically significant vectors for Dowex 50X8

Combination[a]		rms error ($\times 10^{+3}$)	Combination[a]		rms error ($\times 10^{+3}$)
ED	UNI	3.66	DGI	IM	6.52
DGI	UNI	4.14	DHC	DHI	6.57
DHC	UNI	4.27	DHC	IM	6.58
HR	UNI	4.80	ED	IC	7.19
ED	HR	5.35	DGI	HR	7.33
DHI	ED	5.70	ED	IM	7.38
DGI	IC	6.32	DHC	HR	7.49
DHC	IC	6.38	DHI	UNI	8.61
DGI	DHI	6.41			

[a] See Table 71 for vector designations.

if we use the fifth best combination set. Using TFA, we can express the solution in many different, though mathematically equivalent, ways.

Thus, the combination step obtains the best representative vectors of K_A and K_B and also the best combination of real vectors that represent the rational equilibrium constant for the entire ion exchange equilibrium.

Summary

Target factor analyses of a data matrix containing rational equilibrium constants for twenty univalent and divalent cations has succeeded in obtaining real physically meaningful vectors that best represent the rational equilibrium constants. In addition, we have succeeded in separating the equilibrium constant into a sum of two ionic equilibrium constants each related to the free energy change for an ion in the ion exchanger and free (solvated) form. These include the energy of dissociation for an ion, hydrated ionic radii, and enthalpy change of ionic hydration.

Appendix B

INTERMOLECULAR FORCES

Equilibrium distribution in chromatography is determined in large measure by the forces between molecules in each phase--the molecular interactions. These interactions are interrelated to the chemical structures of the interacting molecules.

The net energy of interaction, E_{ij}, between 2 adjacent non-bonded atoms, i and j, is the sum of both attractive and repulsive forces:[45,92]

$$E_{ij} = \frac{A}{r^{12}} - \frac{\sum B}{r^{Z}} \qquad (80)$$

where r is the distance separating the nucleii of i and j and A and B are constants for a particular i and j pair. This equation identifies one major repulsive interaction A/r^{12} and several attractive interactions. The equilibrium separation, r^{e}, is determined mainly by the repulsive term. The interactions at equilibrium can be regarded mainly as a function of the attractive term.

The attractive term includes

a) Dispersion or London forces, $(E_{ij})_d$, expressed by

$$(E_{ij})_d = \frac{-3}{2} \frac{I_i I_j}{I_i + I_j} \frac{\alpha_i \alpha_j}{r^6} \qquad (81)$$

I_i and I_j refer to the first ionization potentials of atoms i and j and α_i and α_j are their respective polarizabilities.

To a good approximation

$$(E_{ij})_d = \frac{-3}{8} \frac{Ii\ \alpha_i^2}{r_i^6} \frac{Ij\ \alpha_j}{r_j^6} \qquad (82)$$

The total dispersion energy, E_d, can be shown[45] by integrating over all volume elements and all atoms, to be:

$$E_d = C''\ \bar{V}_i\ (\alpha_e{}^v)_i\ (\alpha_e{}^v)_j \qquad (83)$$

where C" is a constant, $\bar{V}_i$ is the molar volume of compound i and $\alpha_e{}^v$ is the electron polarability per unit volume.

$\alpha_e{}^v$ is related to the refractive index n of the compound by the Lorentz-Lorentz Equation:

$$\alpha_e^v = \frac{3}{4} \pi N \frac{n^2 - 1}{n^2 + 2} \qquad (84)$$

where N is Avogodros' number. Consequently, α is also related to the molar refraction.

According to the Hildebrand rule, ΔH_v, the heat of vaporization, closely parallels T_b, the boiling point. Thus, a plot of T_b or ΔH_v vs. $\bar{V}_i$ would show good correlation for non-polar and slightly polar compounds. T_b correlates better however with $\bar{V}_i \left(\frac{n^2-1}{n^2+1}\right)^2$ (see Reference 45, pp 37-39).

b) Electrostatic Dipole Interactions - Keesom forces:

Where both molecules have permanent dipoles, the average net energy $(E_{ij})_o$ from dipole orientation is:

$$(E_{ij})_o = \frac{2\bar{\mu}_i^2 \ \bar{\mu}_j^2}{3kTr^6} \qquad (85)$$

where $\bar{\mu}_i$ and $\bar{\mu}_j$ are the permanent dipole moments of adjacent molecules, k is the Boltzman constant, and T is the absolute temperature. Equation 85 is derived for a dilute gas. In the liquid phase, however, some observers suggest that $(E_{ij})_o$ is directly propertional to $\bar{\mu}_i \cdot \bar{\mu}_j$.[45]

Interatomic distances do not vary greatly for different pairs of adjacent atoms. Therefore to a good approximation:

$$r = (r_i + r_j) \approx 2 \ (r_i \ r_j)^{1/2} \qquad (86)$$

Applying (7) to (6):

$$(E_{ij})_o = \frac{1}{3} \frac{1}{kT} \ \frac{\bar{\mu}_i^2}{r_i^3} \cdot \frac{\bar{\mu}_i^2}{r_j^3} \qquad (87)$$

c) Electrostatic Dipole--Induced Dipole Interaction (Debye Interactions):

A molecular with a permanent dipole moment will induce temporary dipoles in adjacent molecules. The attractive energy $(E_{ij})_i$ from these dipole-induced dipole is:

$$(E_{ij})_i = \frac{\bar{\mu}_i^2 \ \alpha_j}{r^6} \qquad (88)$$

When both molecules possess dipole moments the total induction interaction is:

$$(E_{ij})_i = \frac{\bar{\mu}_i^2 \cdot \alpha_j + \bar{\mu}_j^2 \alpha_i}{r^6} \qquad (89)$$

Applying (86) to (88) and (89):

$$(E_{ij})_o = \frac{-1}{2} \frac{\bar{\mu}_i^2}{r_i^3} \frac{\alpha_j}{r_j^3} \qquad (90)$$

$$(E_{ij})_o = \frac{-1}{2} \frac{\bar{\mu}^2 i}{r^3 i} \frac{\alpha j}{r_j^3} - \frac{1}{2} \frac{\bar{\mu}_j{}^2}{r_j^3} \frac{\alpha_i}{r_i^3} \qquad (91)$$

Because the intermolecular interactions actually apply to individual functional groups of a molecule the overall molecular dipole moment may not correlate well with the interactions. Equations 88-91 actually apply to individual bond dipoles rather than to the molecular dipole moment, which may be the aggregate of dipoles which cancel each other out.

d) Hydrogen Bonding or Electron Donor-acceptor Reactions

The interaction energy, E_{ab}, of various acid-base interactions can be expressed as

$$E_{ab} = E_a^* E_b^* + C_a^* C_b^* \qquad (92)$$

where E_a^* and E_b^* measure the hard (electrostatic) acid and base strengths of a given molecule and $Ca^* + C_b^*$ are corresponding measures of soft (covalent) acid and base

strengths. Pearson[93] has provided a qualitative theory of hard-soft interactions. Hydrogen bonding is a relative hard acid-base interaction. Selectivity in ion exchange seem to be governed by the criterion of hard vs. soft interactions.[45,91]

The total energy of interaction, E_{ij}, of a molecule, i, with a surrounding phase, j, is the sum of all possible interactions.

$$E_{ij} = E_d + E_o + E_i + E_{ab} \tag{93}$$

From Equations 82, 83, 87, 90, 91 and 92, E_{ij} can be divided into a solute-solvent product form suitable to a TFA treatment. Thus,

$$E_{ij} = C'' \bar{V}_i (\alpha^V_e)_i \cdot (\alpha^V_e)_j + \frac{1}{3kT} \frac{\bar{\mu}_i^2}{r_i^3} \cdot \frac{\bar{\mu}_j^2}{r_j^3} + \frac{-1}{2} \frac{\bar{U}_i^2}{r_i^3} \cdot \frac{\alpha_j}{r_j^3} + E_a^* \cdot E_b^* + C_a^* \cdot C_b^* \tag{94}$$

As shown, the main factors which should control the interaction energy include: molar refraction, molecular size and polarizability of solute (dispersion interaction), molecular polarity of solute and solvent (electrostatic interactions) and the electron donor acceptor properties of solutes and solvents. These vectors should provide suitable target transformation vectors in a real system.

A. Data for MC33 and MC53 problems

Solutes[a]	Solvents			
	A	B	C	D
OL1	0.7320E 03	0.6490E 03	0.5560E 03	0.5400E 03
OL2	0.8410E 03	0.6590E 03	0.6670E 03	0.6510E 03
OL3	0.1062E 04	0.8750E 03	0.8780E 03	0.8640E 03
OL4	0.1171E 04	0.9790E 03	0.9780E 03	0.9650E 03
OL5	0.1381E 04	0.1184E 04	0.1182E 04	0.1168E 04
OL6	0.7540E 03	0.5890E 03	0.5960E 03	0.5810E 03
OL7	0.8580E 03	0.6980E 03	0.7050E 03	0.6920E 03
OL8	0.1023E 04	0.8460E 03	0.8510E 03	0.8380E 03
OL9	0.1126E 04	0.9450E 03	0.9450E 03	0.9310E 03
OL10	0.9360E 03	0.7690E 03	0.7810E 03	0.7710E 03
A3	0.7160E 03	0.5830E 03	0.5580E 03	0.5470E 03
A4	0.8120E 03	0.6830E 03	0.6520E 03	0.6410E 03
A6	0.9200E 03	0.7870E 03	0.7560E 03	0.7460E 03
A11	0.1129E 04	0.9920E 03	0.9600E 03	0.9480E 03
A5	0.7630E 03	0.6440E 03	0.6170E 03	0.6060E 03
A7	0.8690E 03	0.7460E 03	0.7190E 03	0.7070E 03
A9	0.7770E 03	0.6750E 03	0.6510E 03	0.6420E 03
K2	0.8320E 03	0.6910E 03	0.6570E 03	0.6450E 03
K6	0.1027E 04	0.8840E 03	0.8480E 03	0.8370E 03
K16	0.1239E 04	0.1092E 04	0.1046E 04	0.1034E 04
K28	0.8800E 03	0.7500E 03	0.7160E 03	0.7060E 03
K8	0.9730E 03	0.8440E 03	0.8090E 03	0.7980E 03
K10	0.9180E 03	0.7980E 03	0.7630E 03	0.7530E 03
ES1	0.8010E 03	0.6760E 03	0.6470E 03	0.6340E 03
ES2	0.1110E 04	0.9760E 03	0.9440E 03	0.9330E 03
ES3	0.1214E 04	0.1079E 04	0.1043E 04	0.1030E 04
ES4	0.9480E 03	0.8290E 03	0.8050E 03	0.7930E 03
ES5	0.9610E 03	0.8570E 03	0.8300E 03	0.8210E 03
ES6	0.8830E 03	0.7590E 03	0.7360E 03	0.7240E 03
ES7	0.1191E 04	0.1061E 04	0.1029E 04	0.1018E 04
ES8	0.1036E 04	0.9170E 03	0.8910E 03	0.8800E 03
ES9	0.1150E 04	0.1025E 04	0.9930E 03	0.9830E 03
ET1	0.5970E 03	0.5410E 03	0.5190E 03	0.5150E 03
ET2	0.7730E 03	0.7200E 03	0.7050E 03	0.7020E 03
ET3	0.7560E 03	0.7050E 03	0.6950E 03	0.6920E 03
ET5	0.1164E 04	0.1105E 04	0.1084E 04	0.1080E 04
ET6	0.7120E 03	0.6610E 03	0.6550E 03	0.6490E 03
ET7	0.6630E 03	0.6180E 03	0.6070E 03	0.6010E 03
OX1	0.6120E 03	0.5000E 03	0.4660E 03	0.4550E 03
OX2	0.6800E 03	0.5640E 03	0.5300E 03	0.5250E 03
OX3	0.7760E 03	0.6690E 03	0.6360E 03	0.6270E 03
AR1	0.8560E 03	0.7850E 03	0.7280E 03	0.7240E 03
AR2	0.9600E 03	0.8940E 03	0.8290E 03	0.8280E 03
AR3	0.1064E 04	0.1002E 04	0.9290E 03	0.9270E 03
AR4	0.1054E 04	0.9760E 03	0.9180E 03	0.9150E 03
AR5	0.1107E 04	0.1037E 04	0.9580E 03	0.9550E 03
AR6	0.1263E 04	0.1169E 04	0.1119E 04	0.1115E 04
AR7	0.1256E 04	0.1158E 04	0.1115E 04	0.1112E 04
AL1	0.2000E 03	0.2000E 03	0.2000E 03	0.2000E 03
AL2	0.4000E 03	0.4000E 03	0.4000E 03	0.4000E 03
AL3	0.6000E 03	0.6000E 03	0.6000E 03	0.6000E 03
AL4	0.1400E 04	0.1400E 04	0.1400E 04	0.1400E 04
AL5	0.1600E 04	0.1600E 04	0.1600E 04	0.1600E 04

[a] See Tables 20 and 21 for solvent and solute designations.

(continued)

	E	F	G	H
OL1	0.1368E 04	0.5470E 03	0.5630E 03	0.5390E 03
OL2	0.1397E 04	0.6630E 03	0.6720E 03	0.6510E 03
OL3	0.1489E 04	0.8780E 03	0.8900E 03	0.8620E 03
OL4	0.1534E 04	0.9840E 03	0.9950E 03	0.9630E 03
OL5	0.1629E 04	0.1189E 04	0.1202E 04	0.1167E 04
OL6	0.1307E 04	0.5910E 03	0.6020E 03	0.5810E 03
OL7	0.1336E 04	0.7010E 03	0.7120E 03	0.6920E 03
OL8	0.1438E 04	0.8490E 03	0.8590E 03	0.8370E 03
OL9	0.1473E 04	0.9470E 03	0.9560E 03	0.9300E 03
OL10	0.1329E 04	0.7740E 03	0.7820E 03	0.7710E 03
A3	0.1120E 04	0.5730E 03	0.5840E 03	0.5480E 03
A4	0.1146E 04	0.6730E 03	0.6830E 03	0.6400E 03
A6	0.1195E 04	0.7790E 03	0.7880E 03	0.7470E 03
A11	0.1289E 04	0.9820E 03	0.9930E 03	0.9500E 03
A5	0.1078E 04	0.6360E 03	0.6460E 03	0.6070E 03
A7	0.1130E 04	0.7370E 03	0.7470E 03	0.7080E 03
A9	0.1006E 04	0.6670E 03	0.6740E 03	0.6410E 03
K2	0.1192E 04	0.6810E 03	0.6910E 03	0.6450E 03
K6	0.1257E 04	0.8740E 03	0.8830E 03	0.8360E 03
K16	0.1356E 04	0.1082E 04	0.1091E 04	0.1032E 04
K28	0.1174E 04	0.7380E 03	0.7490E 03	0.7070E 03
K9	0.1210E 04	0.8330E 03	0.8420E 03	0.7990E 03
K10	0.1150E 04	0.7860E 03	0.7920E 03	0.7540E 03
ES1	0.1098E 04	0.6590E 03	0.6680E 03	0.6340E 03
ES2	0.1231E 04	0.9640E 03	0.9730E 03	0.9320E 03
ES3	0.1294E 04	0.1067E 04	0.1075E 04	0.1029E 04
ES4	0.1126E 04	0.8170E 03	0.8230E 03	0.7940E 03
ES5	0.1099E 04	0.8470E 03	0.8540E 03	0.8200E 03
ES6	0.1108E 04	0.7490E 03	0.7560E 03	0.7250E 03
ES7	0.1260E 04	0.1054E 04	0.1061E 04	0.1017E 04
ES8	0.1146E 04	0.9090E 03	0.9160E 03	0.8800E 03
ES9	0.1218E 04	0.1017E 04	0.1024E 04	0.9810E 03
ET1	0.8280E 03	0.5250E 03	0.5270E 03	0.5160E 03
ET2	0.8800E 03	0.7140E 03	0.7180E 03	0.7020E 03
ET3	0.8600E 03	0.7020E 03	0.7070E 03	0.6920E 03
ET5	0.1152E 04	0.1102E 04	0.1108E 04	0.1078E 04
ET6	0.8220E 03	0.6600E 03	0.6650E 03	0.6490E 03
ET7	0.7680E 03	0.6090E 03	0.6170E 03	0.6030E 03
OX1	0.1056E 04	0.4780E 03	0.4870E 03	0.4600E 03
OX2	0.1055E 04	0.5520E 03	0.5580E 03	0.5270E 03
OX3	0.1092E 04	0.6540E 03	0.6610E 03	0.6270E 03
AR1	0.1044E 04	0.7390E 03	0.7500E 03	0.7240E 03
AR2	0.1113E 04	0.8480E 03	0.8600E 03	0.8270E 03
AR3	0.1162E 04	0.9530E 03	0.9640E 03	0.9280E 03
AR4	0.1158E 04	0.9410E 03	0.9470E 03	0.9160E 03
AR5	0.1217E 04	0.9840E 03	0.9910E 03	0.9550E 03
AR6	0.1302E 04	0.1136E 04	0.1146E 04	0.1116E 04
AR7	0.1269E 04	0.1136E 04	0.1145E 04	0.1113E 04
AL1	0.2000E 03	0.2000E 03	0.2000E 03	0.2000E 03
AL2	0.4000E 03	0.4000E 03	0.4000E 03	0.4000E 03
AL3	0.6000E 03	0.6000E 03	0.6000E 03	0.6000E 03
AL4	0.1400E 04	0.1400E 04	0.1400E 04	0.1400E 04
AL5	0.1600E 04	0.1600E 04	0.1600E 04	0.1600E 04

	I	J	K	L
OL1	0.6250E 03	0.6380E 03	0.1104E 04	0.5450E 03
OL2	0.7350E 03	0.7460E 03	0.1295E 04	0.6570E 03
OL3	0.9540E 03	0.9650E 03	0.1435E 04	0.8700E 03
OL4	0.1061E 04	0.1069E 04	0.1542E 04	0.9700E 03
OL5	0.1276E 04	0.1275E 04	0.1752E 04	0.1174E 04
OL6	0.6570E 03	0.6650E 03	0.1086E 04	0.5870E 03
OL7	0.7630E 03	0.7710E 03	0.1183E 04	0.6960E 03
OL8	0.9240E 03	0.9360E 03	0.1374E 04	0.8410E 03
OL9	0.1025E 04	0.1033E 04	0.1467E 04	0.9360E 03
OL10	0.8500E 03	0.8630E 03	0.1246E 04	0.7740E 03
A3	0.5920E 03	0.5610E 03	0.8870E 03	0.5560E 03
A4	0.6910E 03	0.6620E 03	0.9750E 03	0.6430E 03
A6	0.7970E 03	0.7680E 03	0.1093E 04	0.7540E 03
A11	0.1006E 04	0.9690E 03	0.1295E 04	0.9530E 03
A5	0.6530E 03	0.6230E 03	0.9160E 03	0.6130E 03
A7	0.7560E 03	0.7260E 03	0.1041E 04	0.7160E 03
A9	0.6840E 03	0.6530E 03	0.9160E 03	0.6480E 03
K2	0.6930E 03	0.6670E 03	0.1012E 04	0.6530E 03
K6	0.8940E 03	0.8600E 03	0.1214E 04	0.8440E 03
K16	0.1103E 04	0.1064E 04	0.1445E 04	0.1041E 04
K28	0.7590E 03	0.7250E 03	0.1072E 04	0.7150E 03
K9	0.8530E 03	0.8180E 03	0.1164E 04	0.8060E 03
K10	0.7970E 03	0.7740E 03	0.1086E 04	0.7620E 03
ES1	0.6760E 03	0.6380E 03	0.9710E 03	0.6450E 03
ES2	0.9820E 03	0.9440E 03	0.1278E 04	0.9400E 03
ES3	0.1085E 04	0.1049E 04	0.1379E 04	0.1039E 04
ES4	0.8350E 03	0.7960E 03	0.1113E 04	0.8030E 03
ES5	0.8630E 03	0.8260E 03	0.1113E 04	0.8280E 03
ES6	0.7640E 03	0.7300E 03	0.1056E 04	0.7320E 03
ES7	0.1070E 04	0.1023E 04	0.1341E 04	0.1024E 04
ES8	0.9260E 03	0.8880E 03	0.1133E 04	0.8880E 03
ES9	0.1033E 04	0.9950E 03	0.1296E 04	0.9900E 03
ET1	0.5330E 03	0.5170E 03	0.6690E 03	0.5200E 03
ET2	0.7160E 03	0.7030E 03	0.8858E 03	0.7050E 03
ET3	0.7040E 03	0.6910E 03	0.8830E 03	0.6940E 03
ET5	0.1104E 04	0.1091E 04	0.1223E 04	0.1082E 04
ET6	0.6610E 03	0.6490E 03	0.7420E 03	0.6520E 03
ET7	0.6150E 03	0.6040E 03	0.6930E 03	0.6040E 03
OX1	0.4970E 03	0.4680E 03	0.7870E 03	0.4660E 03
OX2	0.5670E 03	0.5340E 03	0.8440E 03	0.5300E 03
OX3	0.6660E 03	0.6360E 03	0.9550E 03	0.6340E 03
AR1	0.7490E 03	0.7300E 03	0.9810E 03	0.7260E 03
AR2	0.8550E 03	0.8340E 03	0.1092E 04	0.8290E 03
AR3	0.9560E 03	0.9360E 03	0.1165E 04	0.9300E 03
AR4	0.9460E 03	0.9240E 03	0.1173E 04	0.9170E 03
AR5	0.9870E 03	0.9650E 03	0.1243E 04	0.9560E 03
AR6	0.1143E 04	0.1117E 04	0.1391E 04	0.1116E 04
AR7	0.1137E 04	0.1116E 04	0.1374E 04	0.1111E 04
AL1	0.2000E 03	0.2000E 03	0.2000E 03	0.2000E 03
AL2	0.4000E 03	0.4000E 03	0.4000E 03	0.4000E 03
AL3	0.6000E 03	0.6000E 03	0.6000E 03	0.6000E 03
AL4	0.1400E 04	0.1400E 04	0.1400E 04	0.1400E 04
AL5	0.1600E 04	0.1600E 04	0.1600E 04	0.1600E 04

	M	N	O	P
OL1	0.9880E 03	0.7120E 03	0.9300E 03	0.5490E 03
OL2	0.1097E 04	0.8220E 03	0.1041E 04	0.6600E 03
OL3	0.1334E 04	0.1041E 04	0.1263E 04	0.8780E 03
OL4	0.1445E 04	0.1149E 04	0.1371E 04	0.9820E 03
OL5	0.1680E 04	0.1362E 04	0.1585E 04	0.1187E 04
OL6	0.9800E 03	0.7370E 03	0.9350E 03	0.5870E 03
OL7	0.1085E 04	0.8470E 03	0.1047E 04	0.7000E 03
OL8	0.1280E 04	0.1005E 04	0.1223E 04	0.8490E 03
OL9	0.1379E 04	0.1108E 04	0.1328E 04	0.9520E 03
OL10	0.1163E 04	0.9270E 03	0.1126E 04	0.7800E 03
A3	0.8180E 03	0.6740E 03	0.8660E 03	0.5480E 03
A4	0.9090E 03	0.7810E 03	0.9650E 03	0.6500E 03
A6	0.1026E 04	0.8900E 03	0.1077E 04	0.7590E 03
A11	0.1240E 04	0.1092E 04	0.1281E 04	0.9610E 03
A5	0.8570E 03	0.7400E 03	0.9140E 03	0.6150E 03
A7	0.9770E 03	0.8440E 03	0.1022E 04	0.7180E 03
A9	0.8520E 03	0.7600E 03	0.9120E 03	0.6470E 03
K2	0.9460E 03	0.8020E 03	0.1001E 04	0.6560E 03
K6	0.1144E 04	0.9910E 03	0.1192E 04	0.8510E 03
K16	0.1360E 04	0.1197E 04	0.1397E 04	0.1056E 04
K28	0.9920E 03	0.8540E 03	0.1042E 04	0.7160E 03
K8	0.1092E 04	0.9460E 03	0.1133E 04	0.8100E 03
K10	0.1022E 04	0.8990E 03	0.1073E 04	0.7680E 03
ES1	0.8820E 03	0.7820E 03	0.9730E 03	0.6440E 03
ES2	0.1201E 04	0.1091E 04	0.1277E 04	0.9510E 03
ES3	0.1312E 04	0.1196E 04	0.1384E 04	0.1056E 04
ES4	0.1043E 04	0.9370E 03	0.1113E 04	0.8040E 03
ES5	0.1049E 04	0.9500E 03	0.1103E 04	0.8290E 03
ES6	0.9710E 03	0.8660E 03	0.1045E 04	0.7340E 03
ES7	0.1280E 04	0.1171E 04	0.1347E 04	0.1039E 04
ES8	0.1124E 04	0.1015E 04	0.1185E 04	0.8920E 03
ES9	0.1242E 04	0.1128E 04	0.1302E 04	0.1004E 04
ET1	0.6240E 03	0.5690E 03	0.6610E 03	0.5070E 03
ET2	0.8170E 03	0.7580E 03	0.8380E 03	0.7010E 03
ET3	0.7920E 03	0.7420E 03	0.8140E 03	0.6920E 03
ET5	0.1221E 04	0.1142E 04	0.1215E 04	0.1086E 04
ET6	0.7190E 03	0.6930E 03	0.7590E 03	0.6460E 03
ET7	0.6690E 03	0.6410E 03	0.6990E 03	0.5980E 03
OX1	0.6950E 03	0.5740E 03	0.7600E 03	0.4550E 03
OX2	0.7700E 03	0.6590E 03	0.8370E 03	0.5380E 03
OX3	0.8700E 03	0.7540E 03	0.9310E 03	0.6350E 03
AR1	0.9150E 03	0.8290E 03	0.9920E 03	0.7340E 03
AR2	0.1023E 04	0.9290E 03	0.1096E 04	0.8400E 03
AR3	0.1129E 04	0.1029E 04	0.1193E 04	0.9470E 03
AR4	0.1116E 04	0.1024E 04	0.1189E 04	0.9330E 03
AR5	0.1175E 04	0.1069E 04	0.1251E 04	0.9740E 03
AR6	0.1319E 04	0.1233E 04	0.1401E 04	0.1129E 04
AR7	0.1312E 04	0.1219E 04	0.1381E 04	0.1124E 04
AL1	0.2000E 03	0.2000E 03	0.2000E 03	0.2000E 03
AL2	0.4000E 03	0.4000E 03	0.4000E 03	0.4000E 03
AL3	0.6000E 03	0.6000E 03	0.6000E 03	0.6000E 03
AL4	0.1400E 04	0.1400E 04	0.1400E 04	0.1400E 04
AL5	0.1600E 04	0.1600E 04	0.1600E 04	0.1600E 04

	Q	R
OL1	0.6890E 03	0.7300E 03
OL2	0.7950E 03	0.8280E 03
OL3	0.1017E 04	0.1053E 04
OL4	0.1126E 04	0.1160E 04
OL5	0.1335E 04	0.1367E 04
OL6	0.7130E 03	0.7660E 03
OL7	0.8170E 03	0.8770E 03
OL8	0.9810E 03	0.1017E 04
OL9	0.1084E 04	0.1129E 04
OL10	0.8950E 03	0.9350E 03
A3	0.6770E 03	0.8150E 03
A4	0.7730E 03	0.9210E 03
A6	0.8810E 03	0.1034E 04
A11	0.1086E 04	0.1246E 04
A5	0.7270E 03	0.8840E 03
A7	0.8290E 03	0.9910E 03
A9	0.7430E 03	0.9210E 03
K2	0.7870E 03	0.9800E 03
K6	0.9800E 03	0.1182E 04
K16	0.1188E 04	0.1396E 04
K28	0.8370E 03	0.1042E 04
K8	0.9330E 03	0.1140E 04
K10	0.8780E 03	0.1096E 04
ES1	0.7540E 03	0.9490E 03
ES2	0.1065E 04	0.1262E 04
ES3	0.1164E 04	0.1369E 04
ES4	0.9020E 03	0.1110E 04
ES5	0.9200E 03	0.1110E 04
ES6	0.8350E 03	0.1014E 04
ES7	0.1146E 04	0.1331E 04
ES8	0.9890E 03	0.1187E 04
ES9	0.1102E 04	0.1300E 04
ET1	0.5820E 03	0.6580E 03
ET2	0.7600E 03	0.8160E 03
ET3	0.7440E 03	0.7870E 03
ET5	0.1147E 04	0.1193E 04
ET6	0.6940E 03	0.7440E 03
ET7	0.6440E 03	0.7050E 03
OX1	0.5800E 03	0.7010E 03
OX2	0.6430E 03	0.7940E 03
OX3	0.7440E 03	0.8820E 03
AR1	0.8240E 03	0.8670E 03
AR2	0.9330E 03	0.9810E 03
AR3	0.1029E 04	0.1108E 04
AR4	0.1023E 04	0.1070E 04
AR5	0.1072E 04	0.1154E 04
AR6	0.1234E 04	0.1287E 04
AR7	0.1224E 04	0.1270E 04
AL1	0.2000E 03	0.2000E 03
AL2	0.4000E 03	0.4000E 03
AL3	0.6000E 03	0.6000E 03
AL4	0.1400E 04	0.1400E 04
AL5	0.1600E 04	0.1600E 04

B. Data for ALDE, KET1 and KET2 problems

Solutes[a]	Solvents[a]			
	A	B	C	D
A2	0.6190E 03	0.4850E 03	0.4380E 03	0.4250E 03
A3	0.7160E 03	0.5830E 03	0.5580E 03	0.5470E 03
A4	0.8120E 03	0.6830E 03	0.6520E 03	0.6410E 03
A5	0.7630E 03	0.6440E 03	0.6170E 03	0.6060E 03
A6	0.9200E 03	0.7870E 03	0.7560E 03	0.7460E 03
A7	0.8690E 03	0.7460E 03	0.7190E 03	0.7070E 03
A9	0.7770E 03	0.6750E 03	0.6510E 03	0.6420E 03
A10	0.1026E 04	0.8920E 03	0.8580E 03	0.8480E 03
A11	0.1129E 04	0.9920E 03	0.9600E 03	0.9480E 03
A12	0.1164E 04	0.1038E 04	0.9990E 03	0.9900E 03
A13	0.7470E 03	0.6000E 03	0.5590E 03	0.5480E 03
A14	0.8010E 03	0.6750E 03	0.6350E 03	0.6250E 03
A15	0.9500E 03	0.7790E 03	0.7320E 03	0.7210E 03
A16	0.1085E 04	0.9330E 03	0.8830E 03	0.8720E 03
A17	0.1254E 04	0.1104E 04	0.1063E 04	0.1053E 04
A18	0.1283E 04	0.1091E 04	0.1022E 04	0.1010E 04
K1	0.7400E 03	0.5920E 03	0.5570E 03	0.5420E 03
K2	0.8320E 03	0.6910E 03	0.6570E 03	0.6450E 03
K3	0.9170E 03	0.7770E 03	0.7470E 03	0.7350E 03
K4	0.9200E 03	0.7800E 03	0.7500E 03	0.7420E 03
K5	0.8800E 03	0.7500E 03	0.7160E 03	0.7060E 03
K6	0.1027E 04	0.8840E 03	0.8480E 03	0.8370E 03
K7	0.1003E 04	0.8680E 03	0.8350E 03	0.8250E 03
K8	0.9730E 03	0.8440E 03	0.8090E 03	0.7980E 03
K9	0.9630E 03	0.8310E 03	0.8000E 03	0.7880E 03
K10	0.9180E 03	0.7980E 03	0.7630E 03	0.7530E 03
K11	0.1130E 04	0.9860E 03	0.9470E 03	0.9340E 03
K12	0.1109E 04	0.9710E 03	0.9410E 03	0.9320E 03
K15	0.9790E 03	0.8670E 03	0.8340E 03	0.8270E 03
K16	0.1239E 04	0.1092E 04	0.1046E 04	0.1034E 04
K17	0.1345E 04	0.1193E 04	0.1151E 04	0.1138E 04
K18	0.1299E 04	0.1159E 04	0.1126E 04	0.1117E 04
K19	0.1099E 04	0.9150E 03	0.8680E 03	0.8540E 03
K20	0.1213E 04	0.1027E 04	0.9700E 03	0.9600E 03
K22	0.1052E 04	0.8910E 03	0.8490E 03	0.8360E 03
K24	0.1063E 04	0.9100E 03	0.8660E 03	0.8550E 03
K25	0.8540E 03	0.7060E 03	0.6810E 03	0.6600E 03
K26	0.9450E 03	0.7980E 03	0.7750E 03	0.7560E 03
K27	0.1114E 04	0.9300E 03	0.8780E 03	0.8600E 03

	E	F	G	H
A2	0.1108E 04	0.4530E 03	0.4570E 03	0.4270E 03
A3	0.1120E 04	0.5730E 03	0.5840E 03	0.5480E 03
A4	0.1146E 04	0.6730E 03	0.6830E 03	0.6400E 03
A5	0.1078E 04	0.6360E 03	0.6460E 03	0.6070E 03
A6	0.1195E 04	0.7790E 03	0.7880E 03	0.7470E 03
A7	0.1130E 04	0.7370E 03	0.7470E 03	0.7080E 03
A9	0.1006E 04	0.6670E 03	0.6740E 03	0.6410E 03
A10	0.1241E 04	0.8860E 03	0.8960E 03	0.8460E 03
A11	0.1289E 04	0.9820E 03	0.9930E 03	0.9500E 03
A12	0.1237E 04	0.1031E 04	0.1040E 04	0.9900E 03
A13	0.1152E 04	0.5830E 03	0.5930E 03	0.5490E 03
A14	0.1124E 04	0.6560E 03	0.6680E 03	0.6250E 03
A15	0.1329E 04	0.7670E 03	0.7790E 03	0.7220E 03
A16	0.1321E 04	0.9200E 03	0.9320E 03	0.8730E 03
A17	0.1372E 04	0.1092E 04	0.1103E 04	0.1055E 04
A18	0.1579E 04	0.1059E 04	0.1072E 04	0.1009E 04
K1	0.1180E 04	0.5780E 03	0.5890E 03	0.5440E 03
K2	0.1192E 04	0.6810E 03	0.6910E 03	0.6450E 03
K3	0.1209E 04	0.7690E 03	0.7790E 03	0.7360E 03
K4	0.1198E 04	0.7730E 03	0.7820E 03	0.7410E 03
K5	0.1174E 04	0.7380E 03	0.7490E 03	0.7070E 03
K6	0.1257E 04	0.8740E 03	0.8830E 03	0.8360E 03
K7	0.1215E 04	0.8600E 03	0.8700E 03	0.8260E 03
K8	0.1210E 04	0.8330E 03	0.8420E 03	0.7990E 03
K9	0.1192E 04	0.8220E 03	0.8290E 03	0.7900E 03
K10	0.1150E 04	0.7860E 03	0.7920E 03	0.7540E 03
K11	0.1303E 04	0.9770E 03	0.9840E 03	0.9340E 03
K12	0.1263E 04	0.9660E 03	0.9750E 03	0.9330E 03
K15	0.1127E 04	0.8570E 03	0.8650E 03	0.8260E 03
K16	0.1356E 04	0.1082E 04	0.1091E 04	0.1032E 04
K17	0.1409E 04	0.1180E 04	0.1192E 04	0.1138E 04
K18	0.1350E 04	0.1157E 04	0.1166E 04	0.1115E 04
K19	0.1497E 04	0.9030E 03	0.9160E 03	0.8550E 03
K20	0.1565E 04	0.1013E 04	0.1026E 04	0.9600E 03
K22	0.1318E 04	0.8720E 03	0.8840E 03	0.8350E 03
K24	0.1342E 04	0.8900E 03	0.9010E 03	0.8570E 03
K25	0.1223E 04	0.6840E 03	0.6940E 03	0.6610E 03
K26	0.1238E 04	0.7830E 03	0.7930E 03	0.7560E 03
K27	0.1535E 04	0.9090E 03	0.9220E 03	0.8600E 03

	I	J	K	L
A2	0.4880E 03	0.4570E 03	0.7860E 03	0.4350E 03
A3	0.5920E 03	0.5610E 03	0.8870E 03	0.5560E 03
A4	0.6910E 03	0.6620E 03	0.9750E 03	0.6480E 03
A5	0.6530E 03	0.6230E 03	0.9160E 03	0.6130E 03
A6	0.7970E 03	0.7680E 03	0.1093E 04	0.7540E 03
A7	0.7560E 03	0.7260E 03	0.1041E 04	0.7160E 03
A9	0.6840E 03	0.6530E 03	0.9160E 03	0.6480E 03
A10	0.9050E 03	0.8720E 03	0.1195E 04	0.8540E 03
A11	0.1006E 04	0.9690E 03	0.1298E 04	0.9530E 03
A12	0.1049E 04	0.1016E 04	0.1302E 04	0.9970E 03
A13	0.6060E 03	0.5760E 03	0.9200E 03	0.5560E 03
A14	0.6750E 03	0.6470E 03	0.9660E 03	0.6310E 03
A15	0.7920E 03	0.7640E 03	0.1176E 04	0.7280E 03
A16	0.9390E 03	0.9020E 03	0.1281E 04	0.8810E 03
A17	0.1113E 04	0.1077E 04	0.1437E 04	0.1058E 04
A18	0.1092E 04	0.1052E 04	0.1561E 04	0.1018E 04
K1	0.6010E 03	0.5680E 03	0.9440E 03	0.5530E 03
K2	0.6930E 03	0.6670E 03	0.1032E 04	0.6530E 03
K3	0.7820E 03	0.7570E 03	0.1112E 04	0.7430E 03
K4	0.7910E 03	0.7620E 03	0.1104E 04	0.7460E 03
K5	0.7530E 03	0.7250E 03	0.1072E 04	0.7150E 03
K6	0.8940E 03	0.8600E 03	0.1214E 04	0.8440E 03
K7	0.8790E 03	0.8500E 03	0.1183E 04	0.8340E 03
K8	0.8330E 03	0.8180E 03	0.1164E 04	0.8060E 03
K9	0.8330E 03	0.8110E 03	0.1140E 04	0.7970E 03
K10	0.7970E 03	0.7740E 03	0.1086E 04	0.7620E 03
K11	0.1000E 04	0.9630E 03	0.1322E 04	0.9430E 03
K12	0.9850E 03	0.9510E 03	0.1281E 04	0.9370E 03
K15	0.8760E 03	0.8460E 03	0.1129E 04	0.8330E 03
K16	0.1103E 04	0.1063E 04	0.1425E 04	0.1041E 04
K17	0.1205E 04	0.1156E 04	0.1521E 04	0.1144E 04
K18	0.1177E 04	0.1138E 04	0.1450E 04	0.1120E 04
K19	0.9490E 03	0.8830E 03	0.1382E 04	0.8660E 03
K20	0.1040E 04	0.9900E 03	0.1506E 04	0.9680E 03
K22	0.8970E 03	0.8620E 03	0.1247E 04	0.8440E 03
K24	0.9030E 03	0.8730E 03	0.1294E 04	0.8630E 03
K25	0.7110E 03	0.6760E 03	0.1036E 04	0.6710E 03
K26	0.8000E 03	0.7720E 03	0.1123E 04	0.7670E 03
K27	0.9360E 03	0.8990E 03	0.1350E 04	0.8700E 03

	M	N	O	P
A2	0.7060E 03	0.5560E 03	0.7530E 03	0.4250E 03
A3	0.8180E 03	0.6740E 03	0.8660E 03	0.5480E 03
A4	0.9090E 03	0.7810E 03	0.9650E 03	0.6500E 03
A5	0.8570E 03	0.7400E 03	0.9140E 03	0.6150E 03
A6	0.1026E 04	0.8900E 03	0.1077E 04	0.7590E 03
A7	0.9770E 03	0.8440E 03	0.1022E 04	0.7180E 03
A9	0.8520E 03	0.7600E 03	0.9120E 03	0.6470E 03
A10	0.1135E 04	0.9910E 03	0.1179E 04	0.8620E 03
A11	0.1240E 04	0.1092E 04	0.1281E 04	0.9610E 03
A12	0.1254E 04	0.1128E 04	0.1299E 04	0.1005E 04
A13	0.8340E 03	0.7030E 03	0.9110E 03	0.5530E 03
A14	0.8840E 03	0.7650E 03	0.9550E 03	0.6310E 03
A15	0.1071E 04	0.9020E 03	0.1139E 04	0.7350E 03
A16	0.1197E 04	0.1036E 04	0.1250E 04	0.8870E 03
A17	0.1366E 04	0.1208E 04	0.1409E 04	0.1059E 04
A18	0.1445E 04	0.1218E 04	0.1500E 04	0.1021E 04
K1	0.8560E 03	0.7020E 03	0.9090E 03	0.5510E 03
K2	0.9460E 03	0.8020E 03	0.1001E 04	0.6560E 03
K3	0.1034E 04	0.8880E 03	0.1083E 04	0.7480E 03
K4	0.1029E 04	0.8840E 03	0.1077E 04	0.7510E 03
K5	0.9920E 03	0.8540E 03	0.1042E 04	0.7160E 03
K6	0.1144E 04	0.9910E 03	0.1192E 04	0.8510E 03
K7	0.1110E 04	0.9700E 03	0.1156E 04	0.8380E 03
K8	0.1092E 04	0.9460E 03	0.1133E 04	0.8100E 03
K9	0.1073E 04	0.9340E 03	0.1117E 04	0.8010E 03
K10	0.1022E 04	0.8990E 03	0.1073E 04	0.7680E 03
K11	0.1254E 04	0.1095E 04	0.1293E 04	0.9520E 03
K12	0.1223E 04	0.1069E 04	0.1257E 04	0.9380E 03
K15	0.1081E 04	0.9550E 03	0.1110E 04	0.8390E 03
K16	0.1360E 04	0.1197E 04	0.1397E 04	0.1056E 04
K17	0.1464E 04	0.1297E 04	0.1494E 04	0.1151E 04
K18	0.1408E 04	0.1256E 04	0.1437E 04	0.1127E 04
K19	0.1249E 04	0.1059E 04	0.1326E 04	0.8790E 03
K20	0.1382E 04	0.1169E 04	0.1433E 04	0.9840E 03
K22	0.1161E 04	0.1009E 04	0.1241E 04	0.8510E 03
K24	0.1184E 04	0.1016E 04	0.1234E 04	0.8660E 03
K25	0.9380E 03	0.8420E 03	0.1094E 04	0.6720E 03
K26	0.1044E 04	0.9340E 03	0.1178E 04	0.7710E 03
K27	0.1251E 04	0.1062E 04	0.1340E 04	0.8780E 03

	Q	R
A2	0.5730E 03	0.7190E 03
A3	0.6770E 03	0.8150E 03
A4	0.7730E 03	0.9210E 03
A5	0.7270E 03	0.8840E 03
A6	0.8810E 03	0.1034E 04
A7	0.8290E 03	0.9910E 03
A9	0.7430E 03	0.9210E 03
A10	0.9860E 03	0.1146E 04
A11	0.1086E 04	0.1246E 04
A12	0.1126E 04	0.1293E 04
A13	0.6950E 03	0.8380E 03
A14	0.7590E 03	0.9070E 03
A15	0.9020E 03	0.1088E 04
A16	0.1041E 04	0.1214E 04
A17	0.1209E 04	0.1395E 04
A18	0.1221E 04	0.1427E 04
K1	0.6900E 03	0.8950E 03
K2	0.7870E 03	0.9800E 03
K3	0.8750E 03	0.1070E 04
K4	0.8800E 03	0.1056E 04
K5	0.8370E 03	0.1042E 04
K6	0.9800E 03	0.1182E 04
K7	0.9610E 03	0.1141E 04
K8	0.9330E 03	0.1140E 04
K9	0.9170E 03	0.1125E 04
K10	0.8780E 03	0.1096E 04
K11	0.1085E 04	0.1286E 04
K12	0.1068E 04	0.1250E 04
K15	0.9440E 03	0.1132E 04
K16	0.1188E 04	0.1396E 04
K17	0.1293E 04	0.1501E 04
K18	0.1253E 04	0.1443E 04
K19	0.1048E 04	0.1257E 04
K20	0.1163E 04	0.1379E 04
K22	0.9950E 03	0.1185E 04
K24	0.1009E 04	0.1219E 04
K25	0.7950E 03	0.9890E 03
K26	0.8930E 03	0.1083E 04
K27	0.1046E 04	0.1257E 04

BIBLIOGRAPHY

1. R. B. Catell, "Factor Analysis, an Introduction and Manual for the Psychologist and Social Scientist", Harper and Bros., N.Y., 1952 .

2. H. H. Harmon, "Modern Factor Analysis", 2nd Ed., U. of Chicago Press, Chicago, 1975 .

3. R. J. Rummell, "Applied Factor Analysis", Northwestern Univ. Press, Evanston, Ill., 1970 .

4. E. R. Malinowski, Ph.D. Thesis, Stevens Institute of Technology, Hoboken, N.J., Dissertation Abstracts, 23, Abstract 62-2027, 1963 .

5. P.T. Funke, E.R. Malinowski, D.F. Martire, L.Z. Pellara, Sep. Sci., 1, 661 (1966).

6. P. H. Weiner, E. R. Malinowski, J. Phys. Chem., 74, 4537 (1970).

7. E. R. Malinowski, P. H. Weiner, "A Computer Program for Factor Analysis of Physical Properties", U.S. Army Research Office (Durham) Contract No. DA-31-124-ARO-D-90, Project No. CRD-AP-3930-C.

8. E. R. Malinowski, D. G. Howery, P.H. Weiner, J. M. Soroka, P.T. Funke, R.B. Selzer, A. Levinstone, "FACTANAL--Target Transformation Factor Analysis", Program 320, Quantum Chemistry Program Exchange, Indiana University, Bloomington, Ind., 1976 .

9. D. G. Howery, Amer. Lab., 8(2), 14 (1976) and D.G. Howery in "Statistics", R.F.Hirsch (ed.), Franklin Institute Press, Philadelphia, 1978, p. 185.

10. P. H. Weiner, Chemtech, 7, 321 (1977).

11. D. G. Howery, E. R. Malinowski, "Factor Analysis in Chemistry", John Wiley and Sons, N. Y., 1979.

12. P. H. Weiner, J. Amer. Chem. Soc., 95, 5845 (1973).

13. J. H. Kindsvater, P.H. Weiner, T. J. Klingen, Anal. Chem., 46, 982 (1974).

14. P.H. Weiner, E. R. Malinowski, J. Phys. Chem., 75, 1207 (1971).

15. P. H. Weiner, E.R. Malinowski, Ibid., 75, 3160 (1971).

16. M.L. Weiner, P. H. Weiner, J. Med. Chem., 16, 655 (1973).

17. D. G. Howery, Bull. Chem. Soc., Japan, 45, 2643 (1972).

18. E. R. Malinowski, M. McCue, Anal, Chem., 49, 284 (1977).

19. A. R. Levinstone, Special Report, Stevens Inst. of Tech., Hoboken, N.J., 1969.

20. P. H. Weiner, D. G. Howery, Anal. Chem., 44, 1189 (1972).

21. P. H. Weiner, C.J. Dack, D.G. Howery, J. Chromatogr., 69, 249 (1972).

22. P.H. Weiner, J. F. Parcher, J. Chromatogr. Sci., 10, 612 (1972).

23. D.G. Howery, P.H. Weiner, J. S. Blinder, Ibid., 12, 366 (1974).

24. D. G. Howery, Anal. Chem., 46, 829 (1974).

25. R. B. Selzer, D.G. Howery, J. Chromatogr., 115, 139 (1975).

26. P. H. Weiner, J.F. Parcher, Anal. Chem., 45, 302 (1973).

27. P. H. Weiner, D. G. Howery, Can. J. Chem., 50, 448 (1972).

28. D. R. Burgard, S. P. Perone, J. C. Wubers, Anal. Chem., 49, 1445 (1977).

29. J. E. Davis, A. Shepard, N. Stanford, L. B. Rogers, Ibid., 46, 821 (1974).

30. R. W. Rozett, E. M. Peterson, Ibid., 47, 1301 (1975); R. W. Rozett, E. M. Peterson, Ibid., 47, 2377 (1975); R. W. Rozett, E. M. Peterson, Ibid., 48, 817 (1976).

31. Helferrich, "Ion Exchange", McGraw-Hill Book Co., N. Y., 1962 .

32. D. L. Deuwer, H. Freiser, Anal. Chem., 49, 1940 (1977).

33. L. B. Justice, T. L. Isenhour, Ibid., 47, 2286 (1975).

34. G. L. Ritter, S. R. Lowry, T. L. Isenhour, C. L. Wilkens, Ibid., 48, 591 (1976).

35. S. C. Elliot, N. A. Hartman, S. J. Hawkes, Ibid., 43, 1938 (1971).

36. Z. Z. Hugus, A. A. El Awady, J. Phys. Chem., 75, 2454 (1971).

37. J. J. Kankare, Anal. Chem., 42, 1322 (1970).

38. M. R. Bacon, G. E. Maciel, J. Amer. Chem. Soc., 95, 2413 (1973).

39. R. N. Carey, S. Wold, J. O. Westgard, Anal. Chem., 47, 1824 (1975).

40. G. Wernimont, Ibid, 39, 554 (1966).

41. J. I. Bulmer, H. F. Shurvell, J. Phys. Chem., 77, 256 (1973).

42. D. L. Duewer, B. R. Kowalski, J. L. Fasching, Anal. Chem., 48, 2002 (1976).

43. O. Exner, Coll. Czechoslov. Chem. Comm., 31, 3222 (1966).

44. E. R. Malinowski, Anal Chem., 49, 606 (1977) and 49, 612 (1977).

45. B. L. Karger, L. R. Snyder, C. Horvath, "An Introduction to Separation Science", John Wiley and Sons, N.Y., 1973.

46. J. M. Miller, "Separation Methods in Chemical Analysis", John Wiley and Sons, N.Y., 1975.

47. H. B. Littlewood,"Gas Chromatography", 2nd Ed., Academic Press, N.Y., 1970.

48. E. Kovats in "Advances in Chromatography", Vol. 1, J. C. Giddings, and R.A. Keller, Eds., Marcel Dekker, N.Y., 1973.

49. E. Kovats, A.I.A. Kellemans, Anal. Chem., 36, 31A (1964) and L. S. Eitre, Chromatographia, 6, 489 (1973).

50. A. Hartkopf, J. Chromatogr. Sci., 12, 113 (1974).

51. L. Rohrschneider, J. Chromatogr. 22, 6 (1966).

52. R. A. Keller, J. Chromatogr. Sci., 11, 49 (1973) and R. A. Keller, B. L. Karger, L. Snyder in "Gas Chromatography 1970", R. Stock and G. Perry, eds., Institute of Petroleum, London, 1971 ; B. L. Karger, L. R. Snyder, C. Eon, Anal. Chem., 50, 126 (1978).

53. L. Rohrschneider in "Advances in Chromatography", Vol. IV, J. C. Giddings, and R. A. Keller, eds., Marcel Decker, N.Y., 1968 .

54. W. R. Supina in "Recent Advances in Gas Chromatography", I. Domsky and J. A. Perey, eds., Marcel Decker, Inc., N.Y., 1971 .

55. W. O. McReynolds, J. Chromatogr. Sci., 8, 685 (1970).

56. L. Kaplan, C. Szita, J. Takacs, G. Taijan, J. Chromatogr., 65, 115 (1972).

57. Applied Science Laboratories, Inc., Catalog 20, 1977, State College, PA., 16801.

58. S. J. Preston, J. Chromatogr. Sci., 8, 18A (1970), and S. Hawkes, D. Grossman, A. Hartkopf, T. Isenhour, J. Leary, J. Parcher, S. Wold, J. Yancey, Ibid., 13, 115 (1975).

59. D. L. Massart, H. Declercq, Anal. Chem., 46, 1988 (1974), and references therein.

60. A. Hartkopf, S. Grunfeld, R. Delumyea, J. Chromatogr. Sci., 12, 119 (1974).

61. S. Wold, K. Anderson, J. Chromatogr., 80, 43 (1973).

62. J. Chastriette, J. Chromatogr. Sci., 14, 357 (1976).

63. D. H. McClosky, S. J. Hawkes, Ibid., 13, 1 (1973).

64. A. Eskes, F. Dupuis, A. Djikstra, H. Declercq, D. Massart, Anal. Chem., 47, 2168 (1975).

65. S. Wold, J. Chromatogr. Sci., 13, 525 (1975).

66. S. R. Lowry, S. Tsuge, J. S. Leary, T. L. Isenhour, Ibid., 12, 124 (1974); J.S. Leary,et al., Ibid.,11,201 (1973)

67. D. L. Massart, P. Sanders, Ibid., 12, 617 (1974).

68. J. Novak, J. Ruzickova, S. Wicar, J. Janak, Anal. Chem., 45 1365 (1973) and J. Novak, J. Chromatogr. 78, 269 (1973).

69. T. H. Risby, P. C. Jurs, B. L. Reinbold. J. Chromatogr. 99, 173 (1974).

70. C. E. Figgins, T. H. Risby, P. C. Jurs, J. Chromatogr. Sci., 14, 453 (1976).

71. S. R. Lowry, H. B. Woodruff, T. L. Isenhour, J. Chromatogr. Sci., 14, 129 (1970).

72. J. Takacs, C. Szita, G. Taijan, J. Chromatogr., 56, 1 (1971); J. Takacs, G. Talas, I. Bernatti, G. Czako, Ibid., 67, 205 (1972); Z. Szentirmai, G. Taijan, J. Takacs, Ibid., 73, 11 (1973); and G. Taijan, J. Chromatogr. Sci, 14, 309 (1976).

73. G. Castello, G. D'Amato, J. Chromatogr., 58, 127 (1971).

74. G. Castello, M. Lunardelli, M. Berg, Ibid., 76, 31, (1973).

75. E. A. Streuller, M. Orloff, Ibid., 101, 23 (1974).

76. J. R. Chretieen, J. E. Dubois, Anal. Chem., 49, 747 (1977).

77. F. Klages, Ber. DT. Chem. Ges., 76, 788 (1943).

78. R. W. Bach, F. Dotsch, H. A. Friedrichs, L. Marx, Chromatographia, 4, 459 (1971).

79. W. L. Zielinski, D. Martire, Anal. Chem. 48, 1111 (1976).

80. W. O. McReynolds "Gas Chromatographic Retention Data", Preston Technical Abstracts Co. Evanston, Ill., 1966.

81. W. O. McReynolds, Personal Communications, 1975.

82. F. Dupuis, A. Dijkstra, Anal. Chem., 47, 379 (1975).

83. S. H. Langer, H. Purnell, J. Phys. Chem., 70, 904 (1966).

84. W. Supina, Personal Communication, 1976.

85. Altech Associates, Catalog 19, 1975, Arlington Hts., Ill., 60004 and Supelco, Inc., 1974 Catalog, Bellafonte, PA., 16828.

86. D. Martire, Anal. Chem., 46, 626 (1974).

87. S. J. Hawkes, E. F. Mooney, Anal Chem., 36, 1473 (1964).

88. "Handbook of Chemistry and Physics", R. B. Weast, ed., 53rd edi., Chemical Rubber Co. Press, Cleveland, 1972.

89. R. F. Hirsch , H. C. Stober, M. Kowblansky, F. N. Hubner, A. W. O'Connell, Anal. Chem., 45, 2100 (1973).

90. R. Hirsch, Personal Communication, 1977.

91. O. D. Bonner, L. L. Smith, J. Phys. Chem., 61, 326 (1957).

92. D. M. Young, A. D. Crowell, "Physical Adsorption of Gases", Butterworths, London, 1962 and A. D. Crowell in "The Solid Gas Interface" Vol. 1, A. E. Flood, ed., M. Decker, N.Y., 1966 .

93. R. G. Pearson, J. Am. Chem. Soc., 85, 3533 (1963).

94. J. O. M. Bockris and A. K. N. Reddy, "Modern Electrochemistry", Vol. II, Plenum Press, N.Y., 1970.

95. F. A. Cotton and G. Wilkinson, "Advanced Inorganic Chemistry", 3rd ed., Interscience Publishers, N.Y., 1972.

96. R. Kunin, "Ion Exchange Resins", R. E. Krieger Publishing Co., N.Y., 1972.

97. F. Franks, ed., "Water, A Comprehensive Treatise", Vol. III, Plenum Press, N.Y., 1973.

98. R. G. Pearson, ed., "Hard Acid and Soft Bases", Dowden Hutchinson and Ross, Inc., N.Y., 1973.

99. D. F. C. Morris in "Structure and Bonding", Vol. IV, Springer-Verlag, N.Y., 1968.

100. R. A. Horne, ed., "Water and Aqueous Solutions; Structure, Thermodynamics and Transport Processes", Wiley Interscience, N.Y., 1972.

101. F. P. Dwyer, D. P. Miller, "Chelating Agents and Metal Chelates", Academic Press, N.Y., 1967.

102. O. K. Rice, "Electronic Structure and Chemical Bonding", 1st edition, McGraw-Hill, N.Y. 1940.

103. A. A. Grinberg, "An Introduction to the Chemistry of Complex Compounds", translated by J. R. Leach, Pergamon Press, London, 1962.

104. R. T. Sanderson, "Chemical Periodicity", Reinhold, New York, 1960.

105. G. Zundel, "Hydration and Intermolecular Interaction: Infra-red Investigations with Polyelectrolyte Membranes", Academic Press, N.Y. 1969.

106. E. Marinsky, ed., "Ion Exchange" Vol. I, M. Dekker, N.Y., 1966.

107. A. K. Covington, T. Dickenson, "Physical Chemistry of Organic Solvent Systems", Plenum Press, London, 1973.

108. E. Gleuckauf and G. P. Kitt, Proc. Royal Soc., London (Ser. A), 228, 322 (1955).