EXAM: A Two-State Thermodynamic Analysis Program



William H. Kirchhoff



National Institute of Standards and Technology

he National Institute of Standards and Technology was established in 1988 by Congress to "assist industry in the development of technology . . . needed to improve product quality, to modernize manufacturing processes, to ensure product reliability . . . and to facilitate rapid commercialization . . . of products based on new scientific discoveries."

NIST, originally founded as the National Bureau of Standards in 1901, works to strengthen U.S. industry's competitiveness; advance science and engineering; and improve public health, safety, and the environment. One of the agency's basic functions is to develop, maintain, and retain custody of the national standards of measurement, and provide the means and methods for comparing standards used in science, engineering, manufacturing, commerce, industry, and education with the standards adopted or recognized by the Federal Government.

As an agency of the U.S. Commerce Department's Technology Administration, NIST conducts basic and applied research in the physical sciences and engineering and performs related services. The Institute does generic and precompetitive work on new and advanced technologies. NIST's research facilities are located at Gaithersburg, MD 20899, and at Boulder, CO 80303. Major technical operating units and their principal activities are listed below. For more information contact the Public Inquiries Desk, 301-975-3058.

Technology Services

- Manufacturing Technology Centers Program
- Standards Services
- Technology Commercialization
- Measurement Services
- Technology Evaluation and Assessment
- Information Services

Electronics and Electrical Engineering Laboratory

- Microelectronics
- · Law Enforcement Standards
- Electricity
- Semiconductor Electronics
- Electromagnetic Fields¹
- Electromagnetic Technology¹

Chemical Science and Technology Laboratory

- Biotechnology
- Chemical Engineering¹
- · Chemical Kinetics and Thermodynamics
- Inorganic Analytical Research
- Organic Analytical Research
- Process Measurements
- Surface and Microanalysis Science
- Thermophysics²

Physics Laboratory

- Electron and Optical Physics
- Atomic Physics
- Molecular Physics
- Radiometric Physics
- Quantum Metrology
- Ionizing Radiation
- Time and Frequency¹
- Quantum Physics¹

Manufacturing Engineering Laboratory

- Precision Engineering
- Automated Production Technology
- Robot Systems
- Factory Automation
- Fabrication Technology

Materials Science and Engineering Laboratory

- Intelligent Processing of Materials
- Ceramics
- · Materials Reliability¹
- Polymers
- Metallurgy
- Reactor Radiation

Building and Fire Research Laboratory

- Structures
- Building Materials
- Building Environment
- Fire Science and Engineering
- Fire Measurement and Research

Computer Systems Laboratory

- Information Systems Engineering
- Systems and Software Technology
- Computer Security
- Systems and Network Architecture
- Advanced Systems

Computing and Applied Mathematics Laboratory

- Applied and Computational Mathematics²
- Statistical Engineering²
- Scientific Computing Environments²
- Computer Services²
- Computer Systems and Communications²
- Information Systems

¹At Boulder, CO 80303.

²Some elements at Boulder, CO 80303.

EXAM: A Two-State Thermodynamic Analysis Program

William H. Kirchhoff

Chemical Sciences Division Office of Basic Energy Sciences U.S. Department of Energy

and

Biotechnology Division Chemical Science and Technology Laboratory National Institute of Standards and Technology Gaithersburg, MD 20899

March 1993



U.S. Department of Commerce Ronald H. Brown, Secretary National Institute of Standards and Technology

Raymond G. Kammer, Acting Director

National Institute of Standards and Technology Technical Note 1401 Natl. Inst. Stand. Technol. Tech. Note 1401 110 pages (Mar. 1993) CODEN: NTNOEF U.S. Government Printing Office Washington: 1993

For sale by the Superintendent of Documents U.S. Government Printing Office Washington, DC 20402–9325 *EXAM* is a computer program written for the analysis of data from calorimetric or spectroscopic data on the denaturation of proteins, oligonucleotides, lipids and other macromolecules in terms of a two-state model of thermodynamic equilibrium. For $A \rightarrow nB$ and $nA \rightarrow B$ transitions, data are fit by an iterative, non-linear least squares calculation to two-state equations of the form:

$$\frac{dQ}{dT} = (1-\alpha) [B_a + B'_a (T-T_0)] + N \left(\frac{\Delta H(T)^2}{RT^2}\right) F(\alpha) + \alpha [B_b + B'_b (T-T_0)]$$

for calorimetric data and

$$I(T) = (1 - \alpha) [\epsilon_A + \epsilon_A'(T - T_0)] + \alpha [\epsilon_B + \epsilon_B'(T - T_0)]$$

for spectroscopic data. In these equations, α is the extent of reaction and is defined as the ratio of the number of moles of B at a particular temperature to the number of moles of B when the material is completely denatured. α can be expressed in terms of the equilibrium coefficient K by

$$\frac{K}{K_0} = \frac{(2\alpha)^n}{2(1-\alpha)} = e^{-\frac{\Delta C}{RT}}$$

for $A \rightarrow nB$ and

$$\frac{K}{K_0} = \frac{2\alpha}{(2(1-\alpha))^n} = e^{-\frac{\Delta G}{RT}}$$

for $nA \rightarrow B$. K_0 is the equilibrium constant at the reference temperature T_0 where $\alpha = 1/2$. The two-state model in *EXAM* includes the temperature dependence of the transition enthalpy ΔH , heat capacity, ΔC_p , and Gibbs free energy ΔG in the form of power series expansions in $(T-T_0)$:

$$\Delta H(T) = \Delta H_0 + \Delta C_{p0} (T - T_0) + \frac{1}{2} \Delta C'_{p0} (T - T_0)^2 + \dots$$

$$\Delta C_p(T) = \Delta C_{p0} + \Delta C'_{p0} (T - T_0) + \dots$$

and

$$-\frac{\Delta G}{T} = \left(\frac{\Delta H_0}{T_0}\right) \left(\frac{T-T_0}{T}\right) + \Delta C_{p0} \left(\ln\frac{T}{T_0} - \frac{T-T_0}{T}\right) + \frac{1}{2} \Delta C_{p0}' \left(\frac{(T+T_0)(T-T_0)}{T} - 2T_0 \ln\frac{T}{T_0}\right)$$

where the subscript 0 refers to the reference state defined by T_0 . In the equation for the calorimetric data, N is the number of moles of material under investigation in the calorimeter and is determined by the analysis. If the weight of material in the calorimeter is known, the molecular weight can be determined. In performing the linearized least squares analysis, analytical derivatives are used.

Acknowledgment

Many people have contributed to the development of this program by using early versions and communicating their problems and requests for additional features. Particular thanks are due to Dr. Ann Ginsburg, Dr. Frederick Schwarz, Dr. Andrew Shrake, Dr. Robert Goldberg, and Professors Gordon Kresheck, Phoebe K. Dea, and C. S. Raman. The encouragement and support of Dr. Stanley Abramowitz throughout this process is also greatly appreciated. The impetus for the development of this program came while the author was a guest worker in the laboratory of Dr. Ira Levin in the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) at NIH and his hospitality and enthusiasm will always be warmly remembered and appreciated. Finally, particular gratitude must be expressed to Dr. Philip Ross, also of NIDDK, for many, many suggestions and criticisms and for his uncanny knack of knowing just when the author's efforts were flagging and needed a boost.

Table of Contents

Abstract ii	ii
Acknowledgment	v
Introduction	1
Computer Requirements	1
The Two-State Model	2 8 8 9
The Least Squares Fit 1 Linearization 1 Initial Estimates 1 Tests for Convergence and Divergence 1 Systematic Errors - Estimation of σ from Third Differences 1 Parameter Constraints 1	0 .0 .2 .2 .3
Data Files, Data and Units 1 The Independent Variable X 1 The Dependent Variable Y 1 The Independent Variable t 1 Data files consisting of one variable only 1 Data consisting of two or more variables 1 Headers and Footers 1 Long Files 1	5 .5 .6 .7 .7 .7
The EXAM Program1Program and Documentation Conventions1Installing the Program1When the Program Aborts1Starting the Program2Starting the Program2The (PAR).EXM File2Entering Data2Previewing Data - Headers and Footers2Data Files2Data Files2Data Interpretation2Data Entry from the Keyboard2The Main Menu2Control of Graphics and Other Program Parameters2Data Transformations2Tian Transformation3Dividing by the Scan Rate: Y \rightarrow Y/(dX/dt)	88990011235667801

Reverse the order of X Y relative to t	33
Compress (average) Data	33
Replace Y with $(2Y)^{n/2}(1-Y)$	33
Replace Y with $s(Y + \Delta Y)$ Replace X with $s(X + \Delta X)$	34
Replace Y with Y^n Replace X with X^n	
Replace V with 1/V Replace X with 1/X	34
Replace V with $Ln(V)$ Replace X with $Ln(X)$	35
$\begin{array}{llllllllllllllllllllllllllllllllllll$	35
Replace Y with $\Sigma X C n Y^n$ Replace X with $\Sigma X C n Y^n$	00
Curdeniag of Date	JJ 26
	30
Scaling of Display Axes	30
	0.0
Analysis Options	38
Analysis Options Menu	38
Summary of Analysis Options and Commands	38
Principal Displays	38
Additional Graphs	39
Data Editing, Selection, or Manipulation	39
Data Analyses	40
Other Menu Options:	41
Analysis Options and Commands - Detailed Descriptions	42
Principal Displays: [F1] [F2] [Shift-F2] [F9] [Tab]	42
Cursors and Markers: $[\rightarrow]$ [Ctrl \rightarrow] [\leftarrow] [Ctrl \leftarrow] [\uparrow] [\downarrow] [Home] [End] [F5]	
[Ctr]-F5]	43
Full-Screen Cursor: [Ctrl-Home]	45
Zooming and Unzooming: [Shift-F5] [Alt-F5]	46
Data Deactivation, Selection, and Editing: [Del] [Ins] [Alt-F1] [Shift-F1]	46
Data Transformations [Ctrl-F1]	49
Baseline Drawing: [F6] [F8]	49
Curve Fitting: [F7] [Shift-F7]	10
General Information selecting data ranges bottom line information	40
Power Sories Fit:	40
Fypopoptial Fit:	50
DSC and SDEC: The Two State Analysis of Calerimetric and Spectroscenic Date	01
Losst Squares Devender Penerta: [Ctr] F7] [Alt F7]	. 00
Source and Pocolling Fitting Functional [Alt F2] [Ctr] F2]	00
Saving and Drawing Crophs of Etting Equations, [Alt F9]	00
Drasing and Drawing Graphs of Fitting Functions: [F0], [Alt-F0]	60
Residuals $[F9]$ [Snift-F9] [Ctrl-F9] [Alt-F9]	61
Extrapolation and Interpolation: [Shift-F8]	63
Confidence Intervals: [Ctrl-F8]	63
Area Calculation: [Shift-F6]	63
Peak Height Estimation: [Alt-F6]	64
Reference Graphs: [Ctrl-F4] [Alt-F4] [F4] [Shift-F4]	65
Drawing data file graphs on the current display: [F3]	67
Saving Displays as Files: [Shift-F3]	68
Printing the Display [F10]	69
Printing Labels and Messages on the Display [Ctrl-F10]	70
Warning beeps	70

Exam	ples	71
1.	A simple example of model errors	71
2.	The order of performing the Tian transformation and dividing by the scan rate doesn't	
	matter	81
3.	Even with an apparent random scatter in the residuals, the model errors are greater	
	than suggested by the standard deviation of the parameters	84
4.	ΔH from the two-state analysis differs slightly from ΔH_{vH} calculated from a van't Hoff	
	analysis.	87
5.	The analysis of overlapping profiles.	92

EXAM

A Two-State Thermodynamic Analysis Program

Introduction

The program EXAM has been developed to assist the interpretation and analysis of differential scanning calorimetry or spectroscopy data in terms of a two-state model of chemical equilibrium. The program can also be used as a general purpose, data analysis program. Data in the form of X,Y pairs can be plotted and analyzed individually as functions of time or Y can be analyzed as a function of X. The graph of Y vs X (or X vs t, or Y vs t) or its first derivative (calculated from finite differences between adjacent data) or its integral (calculated by summing adjacent trapezoidal areas) can be plotted on the screen. Graph segments can be fit by least squares to a polynomial of order 0 to 8, an exponential function, or a thermodynamic two-state model. The area of the graph above a baseline can be calculated. Peaks and valleys can be located and their distances from a baseline calculated. The residuals, that is, the differences between the data and a baseline or a fitting function, can be displayed. Calculated graphs can be saved as separate files. Various transformations of the data can be performed. Data from different files can be compared. If the computer is equipped with a compatible printer (such as an IBM graphics printer), displayed graphs may be printed. The program is menu driven and graphically oriented. After a preliminary session of reading data files or entering data from the keyboard, a graph of the data is displayed which is then manipulated and analyzed using function and keypad keys. The full menu of options can be displayed by pressing the [?] key.

Computer Requirements

The program has been written in Fortran 77 for the IBM XT, AT, and System 2 series of personal computers and compatibles equipped with a mathematics coprocessor (e.g., Intel 8087, 80287, 80387, etc.) with 640 K of available memory operating in DOS 3.1 or equivalent. If the program is run on computers without a mathematics coprocessor, the computer may lock up and, if so, it must be turned off and on to be restarted.

The two-state model is a thermodynamic equilibrium model in which the components of a system exist in either of two states, designated A and B, that can be fully described by an equilibrium constant K. In this section the analytical expressions will be derived for describing calorimetric or spectroscopic temperature profiles in terms of the two-state model parameters:

- ΔH the molar enthalpy of the transition from one state to the other,
- ΔC_n the change in heat capacity at constant pressure,
- $\Delta C'_{n}$ the temperature dependence of the change in heat capacity,
- T_0 the temperature at the midpoint of the transition,
- N the number of moles of material being analyzed, and
- n the stoichiometry of the transition.

The derivation will draw upon simple ideas found in most elementary physical chemistry textbooks. It is being included with this documentation to provide the reader with a convenient reference for understanding how the EXAM program works. At the end of this section the comparison of the approach used in EXAM will be compared with other common techniques appearing in the literature.

The equilibrium constant K for a process in which one mole of species A dissociates into n moles of species B, $A \rightarrow nB$, as the temperature is increased is given by

$$K = \frac{N_B^n}{N_A}.$$

where N_A and N_B are the number of moles of species A and B in the calorimeter cell. For spectroscopic measurements, the conventional expression for the equilibrium constant K is in terms of concentrations. In the derivations to follow, N_A and N_B can be assumed to have either meaning, concentrations or total numbers, without necessarily distinguishing which. This is because the derivation considers only the ratios of equilibrium constants so that the mass or volume of solvent will cancel.

Similarly, for a transition involving polymerization, $nA \rightarrow B$,

$$K = \frac{N_B}{N_A^n}.$$

In the following discussion, the $A \rightarrow nB$ case will be emphasized. Corresponding expressions for the $nA \rightarrow B$ case will be given where appropriate.

Note that not all of the fragments need be identical. That is, the situation where $A \neq B + B' + B'' + \cdots$, will display the same calorimetric or spectroscopic profile as one where all the fragments are identical. This is because $N_B = N_{B'} = N_{B'} = \cdots$ so that

$$K = \frac{N_B \cdot N_{B'} \cdot N_{B''} \cdot \cdots}{N_A} = \frac{N_B^n}{N_A}$$

The only factor of consequence is the number n of product species. Situations of overlapping, stepwise transformations, $A \neq B \neq C \neq \cdots$, which have been studied by Friere and Biltonen [1], are not treated in this version of the program.

The two-state model provides a functional dependence of the measured heat capacity on temperature of the following form:

$$\frac{dQ}{dT} = N_A C_{pA}(T) + \Delta H(T) \left(\frac{1}{n} \frac{\partial N_B}{\partial T} \right)_p + N_B C_{pB}(T)$$

$$= N_A C_{pA}(T) - \Delta H(T) \left(\frac{\partial N_A}{\partial T} \right)_p + N_B C_{pB}(T),$$
(1)

where dQ/dT is the differential heat added to or evolved by the sample. $C_{pA}(T)$ and $C_{pB}(T)$ are the heat capacities of species A and B and, as indicated, are functions of temperature. ΔH is the enthalpy change for the reaction *per mole of oligomer*, i.e.,

$$\Delta H(T) = nH_B(T) - H_A(T) \quad for \quad A \to nB$$

or

$$\Delta H(T) = H_B(T) - nH_A(T) \quad for \quad nA \to B,$$

and, as indicated, is also a function of temperature.

The extent of reaction α is defined as

$$\alpha = \frac{N_B}{N_{\infty}} = 1 - \frac{N_A}{N_0}$$

where N_0 is the total number of moles of A at low temperature where $N_B=0$ and N_{∞} is the total number of moles of B at high temperature where $N_A=0$. α thus approaches 0 at low temperatures and 1 at high temperatures. For $A \to nB$, $N_{\infty} = nN_0$ and for $nA \to B$, $N_0 = nN_{\infty}$. In the discussion and equations that follow N, without an identifying subscript of 0 or ∞ , is the number of moles of pure oligomer. That is, $N = N_0$ for $A \to nB$ and $N = N_{\infty}$ for $nA \to B$.

Using $N_B = n(N_0 - N_A)$, the equilibrium constant can be expressed in terms of N_0 and α as

$$K = \frac{(nN_0)^n \alpha^n}{N_0(1-\alpha)}$$

so that

$$\frac{1}{K}\left(\frac{\partial K}{\partial T}\right)_{p} = \left(\frac{n}{\alpha} + \frac{1}{1-\alpha}\right)\left(\frac{\partial \alpha}{\partial T}\right)_{p}$$

The differential change in N_A and N_B can then be expressed in terms of α and the equilibrium constant as:

$$-\frac{dN_A}{dT} = \frac{1}{n} \frac{dN_B}{dT} = N_0 \left(\frac{\partial \alpha}{\partial T}\right)_p = \frac{N_0}{K} \left(\frac{\partial K}{\partial T}\right)_p \left(\frac{\alpha (1-\alpha)}{n(1-\alpha)+\alpha}\right) = \frac{N_0}{K} \left(\frac{\partial K}{\partial T}\right)_p F(\alpha).$$
(2)

The temperature dependence of K can be obtained from

$$K = K_0 e^{-\frac{\Delta G}{RT}}$$

where K_0 is the equilibrium constant at some specified reference state to be defined below and ΔG is the Gibbs free energy change from the reference state to the state defined by K. The temperature dependence of ΔG can be obtained from the Gibbs-Helmholtz equation:

$$\left(\frac{\partial(\Delta G/T)}{\partial T}\right)_{p} = -\frac{\Delta H}{T^{2}}$$

so that

$$\frac{1}{K} \left(\frac{\partial K}{\partial T} \right)_{p} = \frac{1}{R} \left(\frac{\partial \left(-\Delta G/T \right)}{\partial T} \right)_{p} = \frac{\Delta H}{RT^{2}}.$$
(3)

Substituting eqs. (2) and (3) into (1) gives for $A \rightarrow nB$

$$\frac{dQ}{dT} = N(1-\alpha)C_{pA} + N\left[\frac{(\Delta H)^2}{RT^2}\right]F(\alpha) + \alpha nNC_{pB}$$
(4a)

$$F(\alpha) = \frac{\alpha(1-\alpha)}{n(1-\alpha) + \alpha}$$
(5a)

and for $nA \rightarrow B$

$$\frac{dQ}{dT} = N(1-\alpha)nC_{pA} + N\left[\frac{(\Delta H)^2}{RT^2}\right]F(\alpha) + \alpha NC_{pB}$$
(4b)

$$F(\alpha) = \frac{(1-\alpha)\alpha}{n\alpha + (1-\alpha)}.$$
 (5b)

To evaluate eq. (4a) or (4b) explicit expressions for the dependence of ΔH , ΔG and α on T are required. Also, the reference state must be defined.

The extent of reaction α is related to the temperature through the equilibrium constant:

$$K = K_0 e^{-\frac{\Delta G}{RT}} = \frac{N_B^n}{N_A} = \frac{(\alpha N_{\infty})^n}{(1-\alpha)N_0}.$$

The reference state N_{A0} , N_{B0} , T_0 will be defined to be the state where $\alpha = 1/2$, i.e., the midpoint of the transition. This state is frequently denoted by the subscript 1/2 or m for midpoint, i.e., $T_{1/2}$ or T_m .

Substituting $\alpha = 1/2$ to evaluate K_0 then gives

$$\frac{K}{K_0} = \frac{(2\alpha)^n}{2(1-\alpha)} = e^{-\frac{\Delta G}{RT}}$$
and, for nA \rightarrow B,
(6a)

$$\frac{K}{K_0} = \frac{2\alpha}{(2(1-\alpha))^n} = e^{-\frac{\Delta G}{RT}}.$$
(6b)

Following the approach of Clarke and Glew [2], the temperature dependence of ΔH and the heat capacity change ΔC_p can each be expressed as a power series about $T=T_0$, the temperature of the reference state,

$$\Delta H(T) = \Delta H_0 + \Delta C_{p0} (T - T_0) + \frac{1}{2} \Delta C'_{p0} (T - T_0)^2 + \cdots$$

$$\Delta C_p(T) = \Delta C_{p0} + \Delta C'_{p0} (T - T_0) + \cdots$$
(7)

where $\Delta H_0 = n H_{B0}$ - H_{A0} is the enthalpy change at the reference temperature T_0 ,

$$\Delta C_{p0} = \left(\frac{\partial \Delta H}{\partial T}\right)_{p,T=T_0} = nC_{pB0} - C_{pA0}$$

is the heat capacity change per mole of oligomer at the reference temperature T_0 and

$$\Delta C_{p0}' = \left(\frac{\partial \Delta C_p}{\partial T}\right)_{p, T=T_0}$$

Substituting eq. (7) into the Gibbs-Helmholtz equation and integrating between T_0 and T gives the desired expression for the temperature dependence of $\Delta G/T$,

$$-\frac{\Delta G}{T} = \left(\frac{\Delta H_0}{T_0}\right) \left(\frac{T - T_0}{T}\right) + \Delta C_{p0} \left(\ln \frac{T}{T_0} - \frac{T - T_0}{T}\right) + \frac{1}{2} \Delta C_{p0}' \left(\frac{(T + T_0)(T - T_0)}{T} - 2T_0 \ln \frac{T}{T_0}\right).$$
(8)

To give a better sense of the order of magnitude of each of the terms in this expression, they can be expanded as power series in $T-T_0$ to give

$$\left(\ln\frac{T}{T_0} - \frac{T - T_0}{T}\right) \approx \frac{1}{2} \left(\frac{T - T_0}{T_0}\right)^2 - \frac{2}{3} \left(\frac{T - T_0}{T_0}\right)^3$$
and

$$\left(\frac{(T+T_0)(T-T_0)}{T} - 2T_0 \ln \frac{T}{T_0}\right) \approx \frac{1}{3}T_0 \left(\frac{T-T_0}{T_0}\right)$$

so that

$$-\frac{\Delta G}{T} \approx \left[\frac{\Delta H_0}{T} + \frac{1}{2}\Delta C_{p0}\left(\frac{T-T_0}{T_0}\right) + \left(\frac{1}{6}\Delta C'_{p0}T_0 - \frac{2}{3}\Delta C_{p0}\right)\left(\frac{T-T_0}{T_0}\right)^2\right]\left(\frac{T-T_0}{T_0}\right). \tag{9}$$

The value of dQ/dT at any given temperature is calculated from eq. (4a) or (4b) by evaluating $\Delta G/T$ from eq. (8), α from eq. (6a) or (6b), F(α) from eq. (5a) or (5b), and $\Delta H(T)$ and $\Delta C_p(T)$ from eq. (7). This procedure is straightforward except for the calculation of α from 6a or 6b when n > 1. When n > 1, α must be obtained by Newton's method of successive approximations. This is done for $A \rightarrow nB$ by first rewriting eq. (6a) as

$$\frac{\alpha}{(1-\alpha)^{\frac{1}{n}}} = \frac{2^{\frac{1}{n}}e^{-\Delta G}}{2} = k.$$

Defining α_i as the ith approximation to α , $k_i = k(\alpha_i)$, and $k'_i = dk_i/d\alpha_i$, then α_{i+1} is given by:

$$\boldsymbol{\alpha}_{i+1} = \boldsymbol{\alpha}_i + \frac{k-k_i}{k_i'} = \boldsymbol{\alpha}_i + \left(\frac{k}{k_i}-1\right)nF(\boldsymbol{\alpha}_i)$$

where $F(\alpha_i)$ is given by eq. (5a). If we take the initial value of α_0 as

$$\alpha_0 = 1 - \frac{1}{n+k^n} \quad \text{for } k \ge 1$$
$$= k \qquad \text{for } k < 1$$

then the iterative process converges in less than 10 iterations near $\alpha = 0.5$. As α approaches 0 or 1, the convergence is much faster. The case of $nA \rightarrow B$ is performed with the identical calculation but with α replaced by 1- α and ΔG by - ΔG .

In practice, C_{pA} , C_{pB} , C'_{pA} , and C'_{pB} cannot be reliably obtained from DSC measurements of dilute solutions of proteins, lipids, or nucleic acids for various reasons such as lack of sample purity and imprecision in matching sample and reference cells, although under carefully controlled situations, ΔC_p and $\Delta C'_p$ may be. To reflect this, eq. (4a) or (4b) is rewritten as:

$$\frac{dQ}{dT} = (1-\alpha) \left[B_a + B'_a (T-T_0) \right] + N \left(\frac{\Delta H(T)^2}{RT^2} \right) F(\alpha) + \alpha \left[B_b + B'_b (T-T_0) \right].$$
(10)

Equation (10) is the equation that is fit by the EXAM program to calorimetric temperature profiles. $B_a + B'_a(T-T_0)$ is the pre-transition asymptotic baseline. $B_b + B'_b(T-T_0)$ is the post-transition asymptotic baseline. $\Delta B = B_b - B_a$ can interpreted as $N\Delta C_p$ and $\Delta B' = B'_b - B'_a$ as $N\Delta C'_p$ by EXAM in a variety of ways as described in the section on the least squares fit. The sum of the first and third terms in eq. (10) is the profile baseline and the factors (1- α) and α in the first and third terms provide the sigmoidal transition of the baseline from state A to the baseline of state B. *Even in the absence of interpretability of the parameters from a two-state analysis of the data, the sigmoidal baseline of the analysis can be used as a good estimate of the true baseline.*

Although eq. (10) and its derivation may seem complicated, the general functional form of dQ/dT is fairly simple. The shape is determined by the dependence of α on temperature. For n=1, that is, for A \rightarrow B transitions, α has the particularly simple form:

$$\alpha = \frac{1}{1 + e^{-z}}$$

where, to first order in $(T-T_0)$,

$$z = \frac{\Delta H_0}{RT_0^2} \left(T - T_0\right).$$

Also, for n=1, $F(\alpha)=\alpha(1-\alpha)$. α is seen to approach 0 exponentially as $T \rightarrow -\infty$ and 1 as $T \rightarrow +\infty$. 1- α has the opposite behavior and their product gives the characteristic bell shape of the DSC profile. 99% of the profile occurs between $T_0 \pm 5.3 RT_0^2 / \Delta H$. For $T_0=50^\circ$ and $\Delta H = 100 \text{ kJmol}^{-1}$, this represents a range of $\pm 46 \text{ }^\circ\text{C}$.

No mention has been made of thermal lag effects which can distort the profile shape of eq. (10). Depending on the scan rate, these effects can be profound or negligible. Under favorable circumstances, the distorted profile can be adjusted by use of the Tian equation [3].

$$Y_{adj} = Y + \tau \frac{dY}{dt}$$

where τ is the instrument time constant, t is the time, and Y is the experimentally observed dQ/dT. The EXAM program includes a provision for applying the Tian equation to account for thermal lag.

The area above the baseline is the heat required to convert the material in the calorimetric cell from state A to state B. Dividing by the mass of the material in the cell gives the specific heat per gram of material. This specific heat is generally referred to as ΔH_{cal} . The van't Hoff enthalpy, ΔH_{vH} , is the specific heat per mole and is obtained from van't Hoff plots generated from the development curve, i.e., from the integral of the calorimetic profile. It is generally accepted that α , the extent of reaction, can be calculated directly from the profile by dividing the integral of the profile as a function of temperature by the integral of the entire profile. a thus defined will vary sigmoidally from 0 to 1 and will be nearly equal to the true value of α defined as N_{B}/N_{ω} .[†] The van't Hoff plot is a plot of the logarithm of the equilibrium constant K/K₀ (from α and eq. (6a) or (6b)) vs 1/T. This graph should be a straight line with a slope of $\Delta H_{vH}/R$ and an intercept of $\Delta H_{vH}/RT_0$. From the ratio of $\Delta H_{vH}/\Delta H_{cal}$, the molecular weight of the material in the calorimetric cell can be calculated. In the EXAM program, the distinction between the calorimetric enthalpy and the van't Hoff enthalpy is not made. Inspection of eqs. (6a), (6b), and (8) shows that the van't Hoff enthalpy is just the value of ΔH appearing in the two state model equations and is the value that will be obtained when a calorimetric profile is fit to eq. (10). Note that the number of moles of material in the cell, N, is one of the parameters obtained from a least squares fit of the calorimetric data to eq. (10). From the number of moles N and the weight of material in the cell, the molecular weight can be obtained just as it can be from $\Delta H_{vH} / \Delta H_{cal}$.

It can be seen from eq. (9) that if ΔH and ΔC_p are both positive as they usually are, there may be a temperature $T_L < T_0$ where the ΔH and ΔC_p terms will cancel resulting in $\Delta G=0$. A second exotherm will be seen centered at T_L . This has been generally ascribed to a so-called "cold denaturation." Below T_L , the B state is again more stable than the A state. The reality of this presumes that the truncated power series expansion of $\Delta H(T)$ as indicated in eq. (7) is valid at temperatures T_L far removed from T_0 . Without any evidence that this is so, the appearance of a "cold denaturation" from eq. (10) should be viewed as a numerical artifact resulting from an overextrapolation of the linear form of ΔH . One practical consequence of this however will be the lack of convergence from time to time of α to its low temperature limit of 0. If ΔC_p is large enough relative to ΔH , α may not reach 0 over the temperature span of the experiment. This will generally signal the appearance of a "cold denaturation" at a relatively high temperature (but still less than T_0).

[†] It will not be exactly so in all cases for the following reason. The integral above the baseline is the integral of the second term of eq. (10) and also the integral of the second term of eq. (1). As can be seen from eq. (1), if ΔH is not a function of T then the integral of the profile, the integral of the second term of eq. (1), will be equal to N ΔH . If ΔH is a function of temperature and if n = 1, this will still be so because the heat capacity term, $\Delta C_p(T-T_0)$, is antisymmetric about T_0 . However, for n \neq 1 and ΔH a function of T, the integral of the profile will not be exactly equal to N ΔH and the value of α calculated from the integral of the profile will therefore not be exactly equal to the true value of α defined as $\alpha = N_B/N_{e}$.

Spectroscopic Data

The two state model can also be used to extract ΔH and T_0 from temperature profiles of spectroscopic data when the spectra of the two states are different. The observed spectral density (or intensity) at a particular temperature will be a superposition of the spectral density of the two states weighted by their relative mole fractions, viz.,

$$I(T) = \varepsilon_A N_A + \varepsilon_B N_B$$

where ε_A and ε_B are the molar extinction coefficients of pure species A and B, I(T) is the measured spectral intensity, and N_A and N_B are the number of moles of species A and B in the observation volume. From the definition of α and allowing ε_A and ε_B to be linear functions of temperature, the spectral intensity as a function of temperature is given by

$$I(T) = (1 - \alpha) [\varepsilon_A + \varepsilon'_A (T - T_0)] + \alpha [\varepsilon_B + \varepsilon'_B (T - T_0)]$$
⁽¹¹⁾

It can been seen that eq. (11) is identical to eq. (10) with the explicit value of N in the middle term of eq. (10) set equal to 0. In the EXAM program, both calorimetric and spectroscopic data are analyzed with the same equations. However, for spectroscopic data, the baseline coefficients B_a , B'_a , B_b , and B'_b are not interpreted as thermodynamic quantities. Rather $B_a + B'_a(T-T_0)$ and $B_b + B'_b(T-T_0)$ are simply baseline terms representing the spectroscopic intensities of pure state A and pure state B. The values of the thermodynamic quantities ΔH and T_0 are determined from the temperature dependence of α .

Comparison to Other Methods

A rich literature exists on the application of the two-state model to a variety of cooperative phenomena. A thorough review of this literature is beyond the scope of this documentation. Nevertheless, a few of the more commonly referenced approaches should help in comparing the results from EXAM with results from other approaches. Sturtevant and his colleagues have examined a large number of proteins and lipid assemblies [4]. In their approach, the development curve represented by the integral of the profile is used to determine an appropriate baseline. The profile above this baseline is then analyzed by a least squares process involving the parameters $T_{1/2}$ (= T_0), $\Delta h_{1/2}$ (the profile height at $T=T_{1/2}$), and $\beta=\Delta H_{vH}/\Delta h_{cal}$. Δh_{cal} is the area under the calorimetric profile divided by the mass of material undergoing denaturation. β should therefore be equal to the molecular weight, MW. Deviations of β from MW are a measure of intermolecular and intramolecular cooperativity. Sturtevant also defines $\Delta H_{cal} = \Delta h_{cal} \times MW$. If N_{EXAM} represents the value of N returned by EXAM then the mass of material undergoing denaturation divided by N_{EXAM} should be the apparent or effective molecular weight of the material. If N_{true} is the true number of moles of material, then the mass of material will be $N_{true} \times MW$. β should therefore be compared to $(N_{EXAM}/N_{true}) \times MW$, $\Delta H_{vH} = \beta \Delta h_{cal}$ with ΔH_{EXAM} , and ΔH_{cal} with $(N_{EXAM}/N_{true}) \times \Delta H_{EXAM}$.

The approach by Sturtevant is similar to that of Privalov [5] who defined an effective enthalpy $\Delta H_{eff} = 4RT_0^2\Delta h_{1/2}/Area$. The analysis performed by EXAM will give approximately (but not exactly because of the temperature dependence of ΔH) the same values of ΔH from the peak height and from the area/N_{EXAM}. ΔH_{EXAM} most closely approximates ΔH_{eff} . Privalov defines the cooperativity of the transition as $\Delta H/\Delta H_{eff}$ where $\Delta H = \Delta H_{cal} = (Profile Area)/N_{true}$, consistent with Sturtevant's definition of ΔH_{cal} . The cooperativity defined as $\Delta H/\Delta H_{eff}$ should be compared with N_{EXAM}/N_{true}.

The principal difference therefore between the approaches of Sturtevant et al. and Privalov et al. and that of EXAM is the following. Whereas the other authors define two measures of ΔH , ΔH_{vH} and ΔH_{cal} , and attempt to interpret their differences in terms of perturbations to the two state-model, EXAM uses one value of ΔH which closely approximates ΔH_{vH} and a value of the number of moles, N_{EXAM} , whose departure from N_{true} is to be interpreted in terms of perturbations to the two state model. ΔH_{EXAM}

model dependent however whereas $\Delta H_{cal} = (N_{EXAM}/N_{true}) \times \Delta H_{EXAM}$ will be less so.

As already mentioned, Friere and Biltonen [1], use a statistical mechanical approach to extract the partition coefficient from the profile in such a way that a sum of profiles resulting from a stepwise process, $A \neq B \neq C \neq \cdots$, can be deconvoluted and the individual ΔH 's for each step determined. In this case ΔH corresponds to the value of ΔH returned by EXAM. This method requires that the baseline be flat and the technique of Sturtevant mentioned above of using the integral of the profile to construct a development curve for the baseline can be used to provide a suitably flat baseline. Alternatively, EXAM can be used to generate a sigmoidal baseline which will approximate the baseline obtained by integration of the profile. This sigmoidal baseline can then be subtracted from the profile and the difference can be analyzed using the approach of Friere and Biltonen.

Becktel and Schellman [6] define temperatures T_g , T_h , and T_s where $\Delta G=0$, $\Delta H=0$, and $\Delta S=0$ respectively. T_g corresponds to T_0 returned by EXAM. The values of ΔG , ΔH , and ΔS as functions of temperature can be obtained once ΔC_p and ΔH are known at any convenient reference temperature. Becktel and Schellman place particular emphasis on so-called stability curves, plots of ΔG vs T, the maximum of which corresponds to T_s . Becktel and Schellman discuss in detail the implications of the stability curves, their dependence on ΔC_p , the reliability of their determination, and the significance of cold denaturation.

References

- [1] Friere, E. and Biltonen, R.L., *Biopolymers*, 17 (1978) 463.
- [2] Clarke, E.C.W. and Glew, D.N., Trans. Faraday Soc., 62 (1966) 539.
- [3] Calvet, E., Recent Progress in Microcalorimetry. In *Experimental Thermochemistry*, vol. II, ed. Skinner, H.A., New York: Interscience Publishers, (1962) 385-410.
- [4] Sturtevant, J.M., Ann. Rev. Phys. Chem. 38 (1987) 463.
- [5] Privalov, P.L., Adv. Protein Chem., 33 (1979) 167.
- [6] Becktel, W.J. and Schellman, J.A., *Biopolymers* 26 (1987) 1859.

Linearization

The two-state thermodynamic parameters are obtained from a least squares fit of eq. (10) to calorimetric temperature profiles or eq. (11) to spectroscopic temperature profiles. Since in either case the data are non-linear functions of some of the model parameters, the least squares fit must be conducted iteratively. The iterative procedure is based on a Taylor's Series expansion of the function around the current values of the parameters, cutting the expansion off at the first term, and performing a linear least squares fit of the corrections to the coefficients to the values of Y(obs) - Y(calculated) where Y = dQ/dT in eq. (10) or I(T) in eq. (11). That is, if C_i^n are the values of the model parameters following the nth iteration, then the corrections to the values of these parameters, δ_i^n , are obtained from a linear least squares fit of:

$$Y(obs) - Y(C_i^n) = \sum_i \frac{\partial Y}{\partial C_i} \Big|_{C_i = C_i^n} \delta_i^n$$
(12)

and the new coefficients C_i^{n+1} are calculated from $C_i^{n+1} = C_i^n + \delta_i^n$.

In EXAM all of the derivatives $\partial Y/\partial C_i$ are calculated analytically. For example, if Y = dQ/dT, then

$$\frac{\partial Y}{\partial \Delta H_0} = -[B_a + B_a'(T - T_0)] \frac{\partial \alpha}{\partial \Delta H_0} + \frac{2N\Delta H(T)F(\alpha)}{RT^2} + \frac{N\Delta H(T)^2}{RT^2} \left(\frac{\partial F(\alpha)}{\partial \Delta H_0}\right) + [B_b + B_b'(T - T_0)] \frac{\partial \alpha}{\partial \Delta H_0}$$

where

$$\Delta H(T) = \Delta H_0 + \Delta C_{p0} (T - T_0) + \frac{1}{2} \Delta C'_{p0} (T - T_0)^2$$

$$\frac{\partial F(\alpha)}{\partial \Delta H_0} = \frac{n(1-\alpha)^2 - \alpha^2}{(n(1-\alpha) + \alpha)^2} \frac{\partial \alpha}{\partial \Delta H_0}$$
$$\frac{\partial \alpha}{\partial \Delta H_0} = F(\alpha) \frac{1}{K} \frac{\partial K}{\partial \Delta H_0} = F(\alpha) \frac{\partial}{\partial \Delta H_0} \left(-\frac{\Delta G}{RT}\right) = F(\alpha) \frac{T-T_0}{RTT_0}$$

The remaining derivatives, $\partial Y/\partial \Delta C_p$, $\partial Y/\partial \Delta C'_p$, $\partial Y/\partial N$, and $\partial Y/\partial T_0$ are evaluated in a similar fashion. The derivatives with respect to the linear terms, $\partial Y/\partial B_a$, $\partial Y/\partial B'_a$, $\partial Y/\partial B_b$, and $\partial Y/\partial B'_b$, are just the coefficients of B_a , B'_a , B_b , and B'_b in eq. (10) except when $\Delta B = B_b - B_a$ is being interpreted as $N\Delta C_p$, and/or $\Delta B' = B'_b - B'_a$ is being interpreted as $N\Delta C'_p$. These special constraints will be discussed at the end of this section.

Initial Estimates

The initial estimates of the model parameters are made in the following way. The low and high temperature asymptotic baselines are determined from a linear least squares fit of those portions of the profile far removed the two-state transition under analysis. The segments of the profile used for determining the asymptotic baseline are under the control of the user. If the profile consists only of a portion of a two-state transition profile, the baseline can be set by the user or can be assumed to be a constant, linear baseline (which is equivalent to assuming that $N\Delta C_p = 0$ and $N\Delta C'_p = 0$).

The initial estimates of ΔH_0 , N, and T_0 are determined from the temperature of the peak maximum T_m , the area of the profile above a stepped baseline formed by extrapolating the asymptotic baselines to T_m , and the height of the profile at the peak maximum. The height is calculated from a point midway between the the extrapolated baselines at T_m . The area is calculated as the sum of the two half-areas as indicated in the figure below:



From T_m , the height h, and the area, A = A(low) + A(high), the initial estimates of ΔH_0 , N, and T_0 are calculated from:

$$A = N \Delta H$$

$$h = \frac{N \Delta H^2}{R T_m^2 (\sqrt{n} + 1)^2}$$

and

$$\frac{\Delta H(T_m - T_0)}{RT_m T_0} = \ln \left(\frac{K_m}{K_0}\right)$$

where the value of the equilibrium constant at the peak maximum, K_m , is calculated from eq. (6a) or (6b) using the value of α_m at the peak maximum which is approximated by setting $dF(\alpha)/d\alpha = 0$ (ignoring the slowly varying explicit dependence of dQ/dT on T in eq. (10)) giving:

$$\alpha_m \approx \frac{\sqrt{n}}{1+\sqrt{n}}$$
.

If ΔC_p or $\Delta C'_p$ is being varied in the analysis, it is given a starting value of $(B_b-B_a)/N$ or $(B'_b-B'_a)/N$ respectively.

The temperature profiles of spectroscopic data display a sigmoidal shape and the initial estimates of ΔH and T_0 are obtained from the midpoint of the transition region and the slope of the sigmoid at the midpoint determined from the the two points nearest the midpoint. As with the calorimetric data, the pre- and post-transition baseline parameters, B_a , B'_a , B_b , and B'_b , are first estimated by linear least

squares fits of data far from the transition interval. The temperature T_m nearest the midpoint is the one for which Y is closest to the average of the two baselines: $[B_a + B_b + (B'_a + B'_b)(T-X_0)]/2$ where X_0 is some arbitrary origin of the temperature scale. Defining the line connecting the two data nearest T_m as

$$Y = a + b(T - T_m)$$

then the initial estimates of T_0 and ΔH are given by

$$T_0 = T_m + \frac{B_a + B_b - 2a}{2b - B_a' - B_b'}$$

and

$$\Delta H = \frac{RT_0^2(n+1)(2b - B_a' - B_b')}{B_b - B_a}$$

After the initial estimates of the values of the parameters are calculated, they are printed on the screen and may be altered by entering values from the keyboard. Also, parameters may be selected to be held fixed, constrained to be functions of each other, or varied freely in the least squares fit.

Tests for Convergence and Divergence

The least squares fit of eq. (10) or eq. (11) follows the iterative procedure of eq. (12) until it converges, is interrupted by the user, or reaches a maximum number of iterations. Several different tests are applied to determine whether the iterative procedure has converged. If the correction δ_i^n to each and every parameter i at the end of the nth iteration is less than 10% of the standard deviation for that parameter returned by the least squares fit then the procedure is deemed to have converged. A second test is based on the size of the corrections compared to the values of the parameters themselves. If the absolute value of δ_n^i/C_n^i for each and every parameter being varied is less than one part in 10⁶, then the procedure is deemed to have converged. This latter test is most useful when calculated, "errorless" data are being fit. Finally, the standard deviation σ^0 returned by the linearized least squares fit is compared to the standard deviation σ calculated directly from Y(obs) - Y(calc) before the linearized least squares fit is performed. If the value of $(\sigma - \sigma^0)/\sigma < .005$, that is, if the standard deviation has not decreased significantly, the procedure is assumed to have converged. If any of the three convergence tests is successful, the iterative procedure is ended. On the other hand, if the absolute value of the correction to and the standard deviation of each of two or more parameters increases from one iteration to the next, that is, if $|\delta_i^n| > |\delta_i^{n-1}|$ AND $\sigma_i^n > \sigma_i^{n-1}$ for two or more parameters, the procedure may be diverging and the user is prompted to intervene.

Systematic Errors - Estimation of σ from Third Differences

Because systematic errors arising from the approximate nature of the two-state model and/or experimental sources typically overwhelm random measurement errors, the standard deviations of the parameters are not provided automatically with the values of the parameters. Rather, the values of the standard deviations of the parameters can only be reviewed while a graph of the residuals, $R_i = Y_i(obs) - Y_i(calc)$, is displayed. If no systematic error is apparent, the standard deviations of the parameters may have some significance in assessing the true values of the parameters and their confidence intervals. This will seldom happen however and the standard deviations of the parameters will be of little use in the interpretation of the results of the least squares analysis.

As a means of signaling the presence of systematic errors, the overall precision of the data is estimated from third differences. If first differences are defined as $d_i^1 = Y_i - Y_{i-1}$, then third differences are

defined as $d_i^3 = d_i^2 - d_{i-1}^2 = Y_i - 3Y_{i-1} + 3Y_{i-2} - Y_{i-3}$. The estimate of the standard deviation of the data from third differences is obtained from

$$\sigma^2 \approx \frac{1}{20(n-3)} \sum_{i=4}^n \left(d_i^3 \right)^2.$$

If the standard deviation of the fit is larger than the estimate from third differences, then some systematic error is likely. In reporting the analyses of least squares fits in EXAM, the value of σ estimated from third differences is always reported. The extent to which σ from third differences approximates random measurement error depends upon the relative magnitude of the random measurement error and the functional change in Y among neighboring points. For most of the data encountered, the third differences appear as a random scatter of points.

The best approach for assessing systematic error is through comparison of the value of N calculated by EXAM with the known number of moles of material in the calorimeter cell.

Parameter Constraints

Several options are offered for conducting two-state analyses. For DSC data, some of these options result in constraints being placed on the parameters. These constraints, which apply <u>only</u> to DSC data, add considerable complexity to the least squares analysis. Specific situations will now be described in some detail primarily for the benefit of those who wish to interpret and use the source code for the EXAM program.

The asymptotic baseline parameters and transition parameters may be evaluated independently (the program default option) or simultaneously. Because of the approximate nature of the two state model, deviations from the two state model are typically an order of magnitude greater than the random errors in the data as can be easily seen by displaying the residuals R_i . The systematic nature of departures from ideal, two-state behavior can result in large distortions in the values of the parameters when the asymptotic and transition parameters are varied simultaneously. When all parameters are varied simultaneously, the quality of the least squares fit, as judged by the standard deviation, improves but the interpretability of the parameters, as judged by the agreement between N and the known amount of material in the DSC cell, can deteriorate. The optimum way in which an analysis should be conducted depends on the system being studied. Consequently, flexibility in the conduct of the analysis is offered.

When the baseline parameters, B_a , B'_a , B_b , B'_b , are held fixed while any of the transition parameters, ΔH_0 , T_0 , N, ΔC_p , $\Delta C'_p$, is varied, B_a and B_b must be corrected following each iteration for shifts in the value of T_0 . For example, the pre-transition baseline is defined as $B_a + B'_a(T-T_0)$. If T_0^n is the value of T_0 following the nth iteration, then before the next iteration is performed, B_a must be shifted to $B_a^n = B_a^{n-1} + B'_a(T_0^n-T_0^{n-1})$. Similar statements apply to B_b . This is important only if ΔC_p is being interpreted as $\Delta B/N$ or $\Delta C'_p$ as $\Delta B'/N$, but the correction is always applied regardless.

When the baseline parameters **are held fixed** and ΔC_p is varied, ΔC_p is determined solely from the shape of the transition profile and not from $(B_b-B_a)/N$. When ΔC_p is constrained in the least squares fit by $\Delta C_p = \Delta B/N = (B_b-B_a)/N$, ΔC_p is calculated from $(B_b-B_a)/N$ following each iteration and is held fixed at this value on the next iteration. The value reported by the program is the value calculated from $(B_b-B_a)/N$ following the last iteration. Similar statements apply to $\Delta C'_p$ when it is constrained to be $(B'_b-B'_a)/N$. This is true regardless of whether the baseline parameters are being varied simultaneously with the transition parameters or not.

When the baseline parameters are constrained, e.g., $B_b = B_a + N\Delta C_p$, then $B_a + N\Delta C_p$ is substituted for B_b in eq. (10), and the baselines thus contribute explicitly to the values of N and ΔC_p . The derivatives with respect to N and ΔC_p must include the additional contributions from their appearance in the baseline.

When the baseline parameters and transition parameters are varied simultaneously further complexities can arise. Attempts to vary B_a , B_b , and ΔC_p independently[†] will usually result in a diverging iterative process. Similarly, attempts to vary B'_a , B'_b , and $\Delta C'_p$ independently will also usually diverge. Therefore, and because the concept of two independent values of the heat capacity does not make much sense, when the baseline and transition parameters are varied simultaneously, at least one of B_a , B_b , and ΔC_p must be held fixed at some value or one of the three must be constrained by the relationship $\Delta C_p = (B_b - B_a)/N$. Similarly, at least one of B'_a , B'_b , and $\Delta C'_p = (B_b - B_a)/N$. Therefore, the only way to obtain a value of ΔC_p that is independent of the value of $(B_b - B_a)/N$ is to hold the baseline parameters fixed and vary ΔC_p in the least squares fit.

An additional complexity can arise when the baseline parameters and transition parameters are varied simultaneously because of the appearance of T_0 in the equations for the asymptotic baseline, viz., $B_a + B'_a(T-T_0)$ and $B_b + B'_b(T-T_0)$. Consider the baseline $Y = B_a + B'_a(T-T_0)$ alone. The three parameters B_a , B'_a , and T_0 cannot be obtained from a least squares fit because $\partial Y/\partial T_0$ and $\partial Y/\partial B_a$ are not independent. One is a constant multiple of the other. Attempts to evaluate B_a , B'_a , and T_0 will result in division by 0. Therefore, $\partial Y/\partial T_0$ contains contributions from the baseline terms only when one of B_a or B_b is constrained by $B_b = B_a + N\Delta C_p$ and/or one of B'_a or B'_b is constrained by $B'_b = B'_a + \Delta C'_p$. Under these circumstances the baseline parameters and T_0 are correlated, but, for the example of $B_b = B_a + N\Delta C_p$, $\partial Y/\partial B_a$, $\partial Y/\partial N$, $\partial Y/\partial \Delta C_p$, and $\partial Y/\partial T_0$ are no longer linearly dependent.

In practice, it makes no difference in the outcome of the analysis which of B_a , B_b , and ΔC_p is constrained by the relationship $\Delta C_p = (B_b - B_a)/N$ or which of B'_a , B'_b , and $\Delta C'_p$ is constrained by $\Delta C'_p = (B'_b - B'_a)/N$. The choice is left as a matter of convenience.

Most of the above complexities do not arise in the analysis of spectroscopic data because the baseline parameters have no thermodynamic significance. Hence, only the corrections of B_a and B_b arising from shifts in T_0 with each iteration are made.

[†] This would, in effect, define two independent values of ΔC_p , one derived from the temperature dependence of ΔH and the other from the difference in the asymptotic baselines at T_0 .

The EXAM program is designed to read calorimetric or spectroscopic data files consisting of one to ten ASCII (alphanumeric characters) fields per record (entries or values per line) with each record in the file separated by a carriage return or a carriage return plus a line feed. Fields in each record are assumed to be separated by blank spaces or commas. Other separators can be used but the reading of data will be slow as each record is individually interpreted. Floating point numbers must use the conventional form 0.1234E+05 rather than 0.1234+05.

In order to accommodate both calorimetric and spectroscopic data, the following conventions are adopted.

Three variables X, Y, and t are defined by the program though their identification with specific fields in the data file is under the control of the user.

The Independent Variable X

In any two-state analysis that is performed on the data, X is interpreted as the temperature. For other types of analyses, such as a linear least squares fit of one variable as a power series of another, X is merely the independent variable. Whenever X is to be interpreted as the temperature, its units must be degrees centigrade. If the temperature values are in other units, they must first be converted to degrees centigrade using the appropriate data transformation functions incorporated in the program.

The Dependent Variable Y

For calorimetric data, Y may be, for example, in units of voltage, power, or heat capacity. The units of Y will determine the units of some but not all of the parameters returned by the twostate analysis. One may think of Y in volts (or millivolts, or microvolts) as being related to Y in watts (or milliwatts or microwatts) by means of a scale factor, the calibration constant of the cell. Similarly, Y in watts is related to Y in joules/K by another scale factor, the scan rate. If these scale factors are not functions of temperature, the two-state analysis can be performed on Y expressed in any units but at a sacrifice of the interpretibility of N, the number of moles of material.

The units of ΔH returned by the two-state analysis are always kilojoules/mole or kilocalories/mole.

The units of ΔC_p are always kJ K⁻¹ mol⁻¹ or kcal K⁻¹ mol⁻¹.

The units of $\Delta C'_p$ are always kJ K⁻² mol⁻¹ or kcal K⁻² mol⁻¹.

The choice of calories or joules is under the control of the user with joules as the program default.

However, the units of N, the number of moles of material in the calorimeter cell, are determined by the units of X and Y. The relationship of the units of N to those of X and Y can most readily be seen from integration of eq. (10) ignoring the temperature dependence of ΔH :

$N\Delta H = \int Y dX$

so that Units of N = (Units of Y)×(Units of X)/(Units of Δ H):

Units of Y	<u>Units of N=XY/∆H</u>
microjoules/° (microcalories/°)	nanomoles
millijoules/° (millicalories/°)	micromoles
joules/° (calories/°)	millimoles

and so on. If the values of Y are in units other than multiples of J K^{-1} (or cal K^{-1}) they must first be converted to multiples of J K^{-1} (or cal K^{-1}) by suitable transformations incorporated into the program if the value of N is to be interpreted in terms of the number of moles of material contained in the cell.

The question of the units of N and ΔH appears to be a difficult one for some people to grasp. It can most readily be understood by multiplying both sides of eq. (10) by a constant scale factor S. If Y=SdQ/dT then baseline parameters returned by the least squares analysis will be SB_a, SB'_a, etc., and the value of N will be SN. The values of ΔH , ΔC_p , and $\Delta C'_p$ will be unaltered since their functional dependence on Y is unchanged. Their values are determined by the shape of the DSC profile as well as its magnitude and their units depend only on the units of the gas constant R in eq. (10) and in the definition of the equilibrium constant. The value of R can be selected to be in units of kJ K⁻¹ mol⁻¹.

If either the scan rate or the calibration constant is a function of temperature (as they usually are), the two-state analysis of Y in any units other than heat capacity will give inaccurate results. Two transformations are provided to convert Y into units of heat capacity from units of volts or watts. Both transformations involve calibration constants which are functions of temperature. Y in units of volts can be converted to Y in units of watts by means of a calibration constant C(T) that can be expressed as a power series in T:

$$Y_{Watts} = Y_{Volts} \times C(T) = Y_{Volts} \sum_{i=0}^{n} C_i T^i.$$

Y in units of watts (joules sec⁻¹) can be converted to Y in units of joules K^{-1} by dividing by a scan rate s(t) = dT/dt expressed as a power series in time t:

$$Y_{JK^{-1}} = \frac{Y_{watts}}{s(t)} \qquad T = \sum_{i=0}^{3} C_i t^i \qquad s(t) = \frac{dT}{dt} = \sum_{i=1}^{3} i C_i t^{i-1}.$$

For *spectroscopic data*, Y can be the spectroscopic response variable or can be defined by the user as the ratio or difference of two spectroscopic measurements as a function of temperature.

The Independent Variable t

The third variable is a second independent variable t, intended primarily to serve as a time scale for the calculation of the scan rate and for correcting the data for an instrument time constant. If a time axis is not necessary, the t variable can be set equal to the index of the record in the file. The "time scale" can also be used as a basis for graphing Y and X scales independently.

Data files consisting of one variable only

If a data file contains one entry per record, it is assigned to the Y variable. The program next generates a t (time) scale using the index of each record and an incremental time step Δt provided by the user. If the incremental step is unity, the t scale is simply the index of the record. Following the generation of a t scale, the program generates an X (temperature) scale as a polynomial function of

16

the t scale using coefficients supplied by the user and stored in a parameters file ({PAR}.EXM) used by the program.

Data consisting of two or more variables

If a data file contains more than one entry per record, the assignment of the entries to the X, Y, or t variables is controlled by the user. Any of the entries, or the difference or ratio of any two of the entries can be assigned to the Y variable. Any other entry can be assigned to the X and t scales. If X or t is undefined by the data and so indicated by the user, their values are generated as described in the preceding paragraph on data files with one entry per record.

Headers and Footers

The first line of each data file is tested as a title. If the first line contains no letters (other than E), it is assumed to be data and the title for the data is taken to be the file name. If the first line does contain letters, it is assumed to consist of a title for the data and the first 30 characters are used as a title in the graphical displays of the data.

Many files contain header information in which various operational, date and time information describing the data to follow are contained in the first few lines of the file. These can be identified and ignored by using a file preview option. Similar information following the data, footers, can be identified by the program if the number of entries per record is less than the number of entries per record in the data that precede the footers. If, on the other hand, footers contain the same or greater number of entries per record and are purely numerical entries, they are interpreted as data, but may be subsequently deleted by the user from the graphical display (see use of [Shift-F1] in the later section describing analysis options and commands.)

Long Files

Files of any length will be analyzed, but only 1400 records can be analyzed directly. If a file contains more than 1400 records the program will average the data in pairs. If the file is still too long, the data will be averaged in groups of 4 or 8 or...etc., until the entire data file has been read. The averaging maintains the scale of all variables and their transformations. Distortions in the data caused by the averaging can be tested by performing additional averaging and testing for significant differences in the analyses.

Program and Documentation Conventions

The program prompts for information to be entered from the keyboard and/or waits for presses of function, number, or cursor pad keys in response to menu options. When values of data are entered from the keyboard, entry must be followed with [Enter]. When selecting from a list of options, pressing the selected key alone will usually initiate the appropriate response without pressing the [Enter] key though this is not possible for menus with more than 10 numbered options. Numbers generally should be entered from the main keyboard since the number pad is reserved for cursor moves. When responding to prompts for input from the keyboard, pressing the [Enter] key alone selects the default value. The default value is always printed on the screen in square brackets following the prompt.

In order to signal that the program has responded to a key press, a response message is always printed before the requested operation is performed. When the operation is completed, the response message is usually erased. For very slow operations, an attempt is made to estimate the time the operation will take and to include this in the response message. On newer computers, some of the operations are completed so quickly that little time is available to read the response message before the operation is finished and the message is changed. No critical information is contained in these response messages and they can be ignored.

Frequently, pressing a second key while the program is responding to the first key press leaves the second and subsequent key presses in a buffer. The program will respond to these in turn. As an inevitable side effect, holding a key depressed can fill the keyboard buffer with a string of repeated entries. A few entries for which this can be extremely annoying have been modified. However, the user is encouraged to be decisive but brief when pressing the keys.

Throughout this documentation and in the prompts issued by the program, keyboard keys are identified by enclosing their representations in square brackets such as [1], [F1], [Esc], etc. These keys are the standard IBM PC keys (e.g., [Enter]) which may have different designations on other keyboards (e.g., [Return]). Program prompts calling for the entry of file names and data are described using as examples names and numbers enclosed in angle brackets such as, <filename> or <1.23456>.

All requests for single letter responses, e.g., X, Y, or t in the data definition prompt described below, are case *insensitive*. That is capital or lower case letters may be entered.

Operations can be cancelled by pressing the **[Esc]** key in response to prompts for additional information required by those operations. The **[Esc]** key is also the exit key. By repeatedly pressing **[Esc]** in response to prompts, the program will eventually end.

The program was written, with a few significant exceptions, in Fortran 77 and compiled with a program from Lahey Computer Systems, Inc. In addition to the standard Fortran 77 language, the program uses special functions provided with the compiler. In particular, the program uses a special function for calling IBM BIOS interrupt services for reading the keyboard, printing characters to the screen and drawing graphs on the screen.

Installing the Program

The program can be run from a single diskette drive, but performance is improved if the program is installed on a fixed disk.

1. From diskette.

The computer should be booted and operating in IBM DOS or equivalent. When the booting is complete, insert the program diskette and type EXAM. When the EXAM title and prompt for data file appear (see below), remove the program diskette and insert a diskette containing data. Proceed from there as directed by the program.

2. From Fixed Disk.

Select or create an appropriate directory where the program will reside. Copy to that directory the files *EXAM.EXE*, *(PAR).EXM*, and *F77L.EER*. For proper handling of error conditions, the *F77L.EER* file should be in the same directory as the *EXAM.EXE* file. If a path has been established with the PATH command (either directly or by the *AUTOEXEC.BAT* file) to the directory where *EXAM.EXE* resides, EXAM can be run from any directory. The PATH command and the use of the *AUTOEXEC.BAT* file are described in the IBM DOS manual. The program has been written assuming a default directory named *EXAM*, but this is not necessary for proper functioning of the program.

When the Program Aborts

Occasionally the program may abort as a result of numerical error such as attempting to divide by zero. The file *F77L.EER* that accompanies the program was provided with the Lahey compiler and contains error messages that will be printed to the screen and which may help the user to understand what went wrong.

If the graphics screen is active when the program aborts, the computer will remain in graphics mode. To return to the normal text mode, type the appropriate MODE command for the resident display device (see the MODE command in the DOS manual.) or reboot the computer by pressing the Ctrl-Alt-Del key combination or rerun EXAM and exit naturally. To exit the program naturally, press the [Esc] key in response to program prompts until the program ends.

Starting the Program

The program is run by typing EXAM[Enter] and begins with a title page and an introductory session of prompts for data entry and analysis options.



The dateline is updated each time a new feature is added to the program and serves to identify the current version. The hidden menu of function key definitions that can be displayed with [?] will be discussed below but can be printed on the screen at almost any time. Pressing any other key will clear the screen and prompt for data entry.

The {PAR}.EXM File

EXAM uses a parameter file, *{PAR}.EXM*, to store default values of operating parameters defined by the user during the course of running EXAM. If the program cannot find a file titled *(PAR).EXM* in the default directory (the directory that was current when EXAM was started) the following prompt is issued before the title page is displayed:

If [1] is selected, program default values are assigned to the parameters. Different versions of the $\langle PAR \rangle$. *EXM* file can be stored in different directories allowing unique default options to be defined for individual directories.

Entering Data

EXAM prompts for data entry with the following message:

```
Enter name of data file
   [Enter] alone prompts for data entry from the keyboard
   <Filename>/P allows previewing of data file
   RUN <command> executes DOS command
   [Esc] to cancel
Entry: __
```

The name of the data file is entered **including the path** if the data file is not in the current default directory (the directory that is current when EXAM is started). If the file cannot be found, the message

```
COULD NOT FIND DATA FILE: <FILENAME>
```

is printed and the data entry prompt is repeated. If the file cannot be interpreted, for example when a file other than a data file is inadvertently requested, warning messages and a prompt to end data entry appear.

The RUN <command> option is offered primarily to allow recall of the names of data files without leaving EXAM. For example, entering "Run Dir *.dat/w" in response to the data entry prompt will result in a display in the wide format of all the files in the default directory with the extension name "dat". Any command can be executed with the RUN <command> option provided the command can be found and the program that executes the command can fit into the available memory. Available memory means in addition to that used by EXAM and any other programs that may be running. After the command has been executed, the data entry prompt is repeated.

Following entry of the name of the data file, the reading and interpreting of the data begins. The first line of the data file is tested as a title. If the first line contains no letters (other than E), it is assumed to be data and the title for the data is taken to be the file name. If the first line does contain letters other than E, it is assumed to consist of a title for the data and the first 30 characters are used as a title in the graphical displays of the data.

If the data file contains field separators other than commas and spaces, each record is read, character by character, and numerical entries are distinguished from non-numerical entries. If this happens, the reading of data will be very slow and a warning message to that effect is displayed on the screen:

```
File may contain unusual field separators. Interpretation will be slow.
```

Previewing Data - Headers and Footers

If <file name>/p is entered in response to the data entry prompt, the first 15 records of the file are displayed on the screen and the first record to be interpreted as data can be selected.

```
The following are the lines (records) of data in <filename>
 Enter the line number of the first line (record) of data to be analyzed.
                                          [Esc] returns to the beginning of
 [Enter] alone previews the next lines.
 the file or, if there, aborts the preview process.
 1: 00
 2: "AMP.GAIN, SAMPLE RATE, AND FILE LENGTH (SEC) ARE AS FOLLOWS:"
 3: 20
 4: 2
 5: 200
 6: 1
 7:0
 8: 0
 9: "EVEN VALUES=A-B, ODD VALUES=B; UNITS ARE MICROJOULES/SECOND"
10: 4.549561, -5.430908
11: 4.545899, -4.694824
12: 4.544678,-2.8125
13: 4.544678, -2.513428
14: 4.542237,-2.059326
15: 4.541016, -.3442383
Enter Selection ([Esc] to cancel, [Enter] for more): _
```

All records preceding the selected record are ignored. This display is uninterpreted, other than locating the ends of records by interpreting carriage returns (^M or hex OD) or carriage return - line feed combinations (^M^J or hex OD OA) as end of record indicators. The selected record may be made the default starting record for reading subsequent data files during the EXAM session by so indicating when prompted. Once EXAM is exited, the default starting record always reverts to the first record in the file. It is not possible to save a title from the data file for titling results in EXAM when a starting record is selected in the preview mode. A default title consisting of the file name is assigned by EXAM. The default first record can be changed at any time by previewing a file.

Many files contain header information in which various operational, date and time information describing the data to follow are contained in the first few records of the file. Previewing files allows header information to be identified and ignored by EXAM. Similar information following the data, footers, can be identified by the program if the number of entries per record is less than the number of entries per record in the data that precede the footers. If, on the other hand, footers contain the same or a greater number of entries per record and are purely numerical entries, they are interpreted as data. They may be deleted later using the data editing functions described elsewhere (See [Shift-F1] in the section describing analysis options.)

Long Data Files

Files of any length will be analyzed, but only 1400 values of X and Y (and t) can be analyzed directly. If the file contains more than 1400 data records, the program will average the data in pairs and display the averaged data for analysis. If the file is still too long, the data will be averaged in groups of 4, 8, ... etc., until the entire data file has been read. The averaging maintains the scale of all variables and their transformations. If the data have been averaged, a message of the following form is printed to the screen before analysis is continued:

Number of data in file <filename>, 5235, exceeds allowable limit. Consequently, each point in the displayed data is an average of 4 points in the original file. All data transformations are scaled as required. The last 3 points in the file have been ignored.

The averaging process may cause some distortion in the profile. This can be tested by additional averaging (see section on Data Transformations) and comparing analyses performed before and after the additional averaging.

Data Interpretation

The section on Data Files, Data, and Units describes in general terms how calorimetric, spectroscopic, or other data are assigned to dependent and independent variables. Because EXAM has been designed to accept as wide a range of data types as possible, the program must be instructed on how the data are to be interpreted.

Following the completion of the reading of data file, the Data Definition Prompt appears:

```
The reading of 81 data records took 21 secs.
<filename> contains 9 data fields. Program variables are defined as:
    X (temperature) = Field 1
    Y (response) = Field 2
    t (time or index) = Field 3
Select: [Enter] to continue
    [X,x], [Y,y], or [T,t] to redefine variable
    [R,r] to review beginning lines of data
    [Esc] to abandon this data file
```

The prompt contains the length of time required to read the data file and the number of data fields contained in the data records. If this number is less than the true number of data fields, then header information may have been interpreted as data. This can be verified by selecting [R] in response to the Data Definition Prompt. If header information has been incorrectly identified as data, the data interpretation should be abandoned with [Esc], the data file previewed (see above), and the header information disregarded.

The above example of the Data Definition Prompt contains the default data definitions assumed by EXAM for records containing three or more fields prior to interpreting the first data file. Once data definitions are selected, they become the default definitions for subsequent files unless the structure of the data changes (e.g., a change in the number of fields per record) whereupon the program defaults are invoked. For every file read, the Data Definition Prompt is offered.

Selecting [X], [Y], or [t] in response to the Data Definition Prompt results in additional prompts for definitions of X, Y and t such as:

Which field (1 - 5, 0 if none) contains the values of X? [1] _

The program offers additional options of generating values of X, Y or t from functions of the data indices or data fields. These options are offered if a [0], signifying none, is entered in response to the prompt to identify which data field contains a particular variable.

Entering [0] in response to the prompt to select a field for the independent variable X results in the following prompt:

```
Define the X (temperature) variable as
    1: X = t
    2: X = ∑Cnt<sup>n</sup>
Selection: [1] __
```

If [2] is selected, an additional menu is offered with subsequent prompts for the entry of data:

```
Defining X = \SigmaCnt<sup>n</sup>

C0 = 0.0

C1 = 1.0

C2 = 0.0

C3 = 0.0

C4 = 0.0

Select index, 0-4, of any coefficient to be changed [none] _

<Enter new value for C1 [1]: 12.3456>

<Do you wish to save these values? Y/N [N] _>
```

The value of any coefficient may be changed by selecting the coefficient index and entering the value when prompted. The new values may be saved as defaults in the *{PAR}.EXM* file.

The Y (response) variable can be set equal to any data field or to the ratio of or the difference between two data fields. Entering [0] in response to the prompt to select the field containing the values of Y results in the following prompt:

Selecting a ratio or a difference results in an additional series of prompts to identify the A and B fields:

24

```
To set Y = <A> - <B>:
Which field (1-9, 0 if none) is A? [4] 2
Which field is B? [5] _
```

The t (time or index) variable can be set equal to the index of each datum or to a multiple of that index. Entering [0] in response to the prompt to select a field for t results in the following prompt:

Define t = $\Sigma \Delta t$ by entering the value of the increment Δt [10]: _

The time interval, Δt becomes a default for the program and is stored in the (PAR). EXM file.

As each variable is defined, the Data Definition Prompt is refreshed, for example:

```
X (temperature) = \Sigma Cnt^n
Y (response) = \langle 2 \rangle - \langle 3 \rangle
t (time or index) = \Sigma \Delta t, \Delta t = 10.0
```

In addition to the options offered by the Data Definition Prompt, the X and Y variables can be replaced by functions of themselves using a variety of functions given in the Data Transformation Menu described below.

Uninterpretable Data

Interpretation of data in non-standard formats will be attempted. However, if data in the requested file cannot be interpreted, the program responds with various warning messages. If some of the data can be interpreted, any or possibly all of the following messages are printed:

Analysis can continue by responding affirmatively to prompts. These messages most commonly result when the data file contains headers, footers or both.

If the program is unable to interpret any data in the file, the following message results:



and the program returns to the Data Entry Prompt.

Data Entry from the Keyboard

Pressing the [Enter] key alone in response to the Data Entry Prompt clears the screen and produces the Keyboard Entry Prompt:



Each value of X and Y is entered in turn at the cursor prompt. If entries extend beyond the bottom of the screen, the data column scrolls. Errors can be corrected using the cursor arrows to select the entry, re-entering the data or deleting the data with the **[Del]** key. New values can be inserted with the **[Ins]** key. The default value for X or Y is the value of the index N. If **[Esc]** is pressed, the keyboard entry is aborted and the user is again prompted to enter the name of a data file. When keyboard entry is complete, pressing **[End]** will accept the entered data and the program resumes with the presentation of the main menu.

The Main Menu

Once the data have been read or entered and interpreted, the display and analysis begin with the Main Menu:

MAIN MENUData file: <filename>Title line: <FILE: filename> or <title>The following options are available:1: Analyze Y vs X5: Analyze a new file2: Analyze X vs t6: Change title line3: Analyze Y vs t7: Execute DOS command4: Transform Data8: Program EnvironmentENTER NUMBER or [Esc] to exit program [1] __

Enter the number corresponding to the analysis or the operation to be performed. If [ENTER] alone
is depressed, the default value of [1], Y vs X, will be assumed. The analysis options of Y vs t and X vs t allow the independent analysis of X or Y in terms of a second independent variable t. This variable, usually considered to be time or the index value of each datum, is set by the user when defining the data.

Option 4 allows the user to perform certain transformations on the data. We will skip this for the moment but return to it in detail following description of the remaining options.

Option 5 allows a new data file to be loaded, replacing the current data file.

Option 6 allows a title to be defined for subsequent data displays. This title will appear on the top line of the graphical displays (when help and propmpts are not displayed).

Option 7 allows DOS commands in response to the prompt:



to be performed. Other programs can be run (if there is sufficient room in memory for them) without exiting completely from EXAM. Following completion of each command, the above prompt is repeated until [Enter] alone is pressed, whereupon the main menu is redisplayed.

Control of Graphics and Other Program Parameters

Option 8 allows for limited control of the program environment by offering several options. The first of these relate to the type of display:

```
What kind of display do you have?
     1. CGA
     2. EGA
     3. VGA
     4. MCGA
     5. MDA
Enter number [3]:
Enter number of horizontal pixels in display [640]: _
Enter number of vertical pixels in display [350]:
Enter maximum number of display pages for graphics display [2]:
Are you using a color monitor? Y/N [Y] _
Enter color value, 1-15,
   for fitting functions [10]: _
   for reference curves [3]: _
   for markers [14]:
   for residuals and inactive data points [12]:
```

Through BIOS interrupt calls the program attempts to identify the type of graphics display available, but for some non-standard operating systems it may be necessary to define the graphics parameters. In general, other than the selection of colors for particular features, these should not be changed unless warranted by graphics display problems. The response to the questions will determine which of the BIOS graphics and text modes will be used by the program. If a color display is selected, the text mode will be 3, otherwise 2. If CGA is selected, the graphics mode will be 6; if EGA or VGA, the graphics mode will be 16 for color display or 15 for monochrome; if MCGA, the graphics mode will be 18 for color and 17 for monochrome. The default values for the number of horizontal and vertical pixels are those defined by the standard BIOS graphics modes. The default number of graphics display pages is also determined by appropriate BIOS interrupt calls. If difficulties are encountered when displaying the analysis options menu from the graphics display, the number of graphics display pages should be set to 1.

If graphics colors are available, the color values specified are the usual definitions: 1: blue, 2: green, 3: cyan, 4: red, 5: magenta, 6: yellow, 7: white, and 8 through 15 the same sequence with enhanced intensity.

After defining the display parameters, additional options are offered:

```
Enter the number of blank lines to be reserved below graph [0]: _
Enter printer number LPT# [1]: _
Beep for graphics prompts? Y/N [Y] _
Pause at each iteration in iterative fits? Y/N [N] _
The units of the gas constant determine the units of ΔH, ΔCp, and ΔCp'.
Select units for the gas constant
1. R = 0.00831451 kJ/°mol
2. R = 0.00198722 kcal/°mol
Enter number [1]: _
```

Blank lines can be left at the bottom of the display for writing text when the graphs are to be printed as permanent records. The default is none. The default printer setting is LPT1. Normally, error messages and prompts for input while displaying graphs will be accompanied by a beep. This can be turned off. When performing iterative least squares fits, the program does not pause at the end of each iteration unless signalled to do so by pressing a key. For modern personal computers, this can be difficult to accomplish within the short time required by the computer to process each iteration. Therefore, the option of stopping at the end of each iteration is offered. Finally, the selection of the units reported for ΔH , ΔC_p , and $\Delta C'_p$ in kilojoules or kilocalories is offered.

The selections for the program environment are saved in the (PAR).EXM file and become the default values until they are changed.

Data Transformations

Commonly, primary data from a calorimetric or spectroscopic measurement are not appropriate for analysis in terms of a two-state model without first transforming the data. For example, the data from a calorimetric measurement may be in the form of voltages which must first be transformed to power values using a calibration constant (which may be a function of temperature) and then to heat capacity values by dividing by the scan rate. Data transformation options have been included in the EXAM program for this purpose. Although motivated primarily for common situations encountered in calorimetry, the range of data transformations included will allow a variety of general purpose data transformations. Selecting [4] from the Main Menu or pressing [Ctrl-F1] while in graphics display (bypassing return to the Main Menu) presents the Data Transformation Menu:

[Enter] alone to end and resume analysis Select Transformation: 7: $\mathbf{Y} \rightarrow \mathbf{s} (\mathbf{Y} + \Delta \mathbf{Y})$ 13: $X \rightarrow g(X+\Delta X)$ 1: Tian Transformation 8: $\mathbf{Y} \rightarrow \mathbf{Y}^{n}$ 14: $X \rightarrow X^n$ 2: $Y \rightarrow Y/(dX/dt)$ (t=N Δt) 9: $\mathbf{Y} \rightarrow \mathbf{1/Y}$ 15: $X \rightarrow 1/X$ 3: Interchange X and Y 4: Reverse order of X,Y 10: $Y \rightarrow Ln(Y)$ 16: $X \rightarrow Ln(X)$ 17: $X \rightarrow Exp(X)$ 11: $Y \rightarrow Exp(Y)$ 5: Compress (average) Data **12:** $\mathbf{Y} \rightarrow \mathbf{Y} \cdot \Sigma \mathbf{Cn} \mathbf{X}^{n}$ **18:** $X \rightarrow \Sigma Cnt^n$ 6: $Y \rightarrow (2Y)^n / 2(1-Y)$ [Esc] to cancel all transformations performed in this session Enter Selection ([Enter] alone to end, [Esc] to cancel): _

The Transformation Menu serves as the prompt for selecting the transformations to be performed. The selection is made by entering the corresponding number (and pressing [Enter] following entry of the number). After entering the number of the transformation to be performed, additional information may be requested before execution. Pressing [Esc] when prompted to enter additional information will cancel the transformation.

Each call for the Transformation Menu is treated as an entry into a transformation session in which a sequence of transformations is performed. After each transformation is performed a list of all transformations performed on the current data is printed on the bottom of the screen, e.g.:

```
Transformations performed: Y/(dX/dt) Y/1000 s(Y+\Delta Y) (2Y)^{2}(1-Y) s(X+\Delta X) 1/X Y/1000 Ln(Y)
```

Those transformations performed during the current session are highlighted in the list. These can be undone by selecting [Esc] in response to the Transformation Menu prompt. Those transformations performed in a prior session cannot be undone. On exit from the session, the data are replaced by their transformed values.

If any transformation is repeated during a data transformation session a warning message to that effect is printed:

You have already performed this transfomation. Do you wish to repeat it? Y/N [N] $_$

If, following completion of a transformation session (even if no transformation is performed), the resulting values of Y are too large or too small to be displayed conveniently in a fixed field format, the values may be scaled by a factor of 10^3 or 10^{-3} in response to one or the other of two prompts:

Maximum value of Y is greater than 10,000 Do you wish to rescale Y values by .001

Maximum value of Y is less than .001 Do you wish to rescale Y values by 1000 If rescaling is not selected, the left hand margin of the display may be quite large or the axis limits may be displayed in floating point notation. Note that rescaling the values of Y will have no effect on the values of ΔH , T_{c} , ΔC_{p} and $\Delta C'_{p}$ obtained from a two-state analysis, but **will scale** N and the baseline parameters B_{a} , B'_{a} , B_{b} , and B'_{b}

For the most part, the transformations are self-evident. A few require explanation.

1. Tian Transformation^{\dagger}

The Tian transformation allows a means for cancelling the effect of an instrumental, exponential response in the measured values of Y:

 $Y = Y + \tau (dY/dt).$

The Tian transformation requires a time scale. If a time scale is not included with the data, it is generated from the time increment per datum (entered by the user) and the index value of each datum in the file. If the t scale is the index scale (implying a time increment Δt of unity) a prompt will be issued to allow a time increment to be entered.

On performing the Tian transformation for the first time, an explanatory message is printed on the screen.

Transform Y according to the Tian Equation: $Y \rightarrow Y + \tau(dY/dt)$ where t is the time and τ is the instrument time constant. If a time scale is not included in the data file, it is calculated from $t = I\Delta t$ where Δt is the time increment per datum and I is the index of each datum in the data file. In calculating dY/dt, Y is expressed as a power series in t, fit to M points placed symmetrically about each value of Y(I),t(I) and the derivative dY/dt is evaluated from the coefficients of the fit.

If an attempt is made to perform the Tian transformation more than once on a set of data, a warning message is printed.

The Tian transformation also requires input of an instrument time constant τ . The value is entered from the keyboard in response to the prompt:

Enter value for time constant τ (in secs) [100]: _

The default value is stored in the file *(PAR).EXM* and is recalled each time EXAM is run. Each new value entered becomes the subsequent default value.

The derivative dY_j/dt required by the Tian transformation is calculated for each value of Y_j from the coefficients of a power series of order p in the time t obtained by a least squares fit to m values of Y placed symmetrically about Y_j :

[†] Calvet, E., Recent Progress in Microcalorimetry. In *Experimental Thermochemistry*, vol. II, ed. Skinner, H.A., New York: Interscience Publishers, (1962) 385-410.

$$Y_n = \sum_{i=0}^{p} C_{ji} t_n^i \quad \text{for } n = j - \frac{m}{2} , \quad n = j - \frac{m}{2} + 1 , \quad \dots \quad n = j + \frac{m}{2}$$

$$\frac{dI_j}{dt} = \sum_{i=1}^{j} i C_{ji} t_j^{i-1}.$$

The number of data m in the segment and the order p of the power series are under the control of the user. For the jth value, Y_j , dY/dt is obtained from a fit of the data from j-m/2 to j+m/2. For the first or last m/2 values, a single power series is used to calculate dY_j/dt for all the values from j = 1 to m/2 or from j = n-m/2 to n.

For the value of dY/dt to be as accurate as possible, the number of data above and below the value being transformed should be equal in the polynomial estimation of dY/dt. Therefore, the number of data used to calculate dY/dt is always odd. If an even number of points is entered for m, it is increased by 1. Also, the estimate of dY/dt is more accurate for p = 1 or 3 than for p = 2 (because of the nodal structure of the residues of the least squares fit).

The values for p and m are entered from the keyboard in response to the prompts:

The default value of m is 2p+1.

The values of p and m determine the length of time required to perform the transformation, the noise in the transformed values, and the distortion due to inaccuracy of the power series spline. Small values of p and m will result in noisy data (when τ is larger than Δt , $Y+\tau\Delta Y/\Delta t$ can be very noisy), but the transformation will be performed rapidly with minimal distortion. For large values of p and m, the transformation may take several minutes. Therefore, an estimate is made of the time required and printed on the screen:



Very large values of m may result in a distortion of the profile shape when the data segment cannot be fit accurately with a power series. This can be tested by displaying Y vs t, zooming to particular ranges of the data, fitting the data to a polynomial, and investigating the residuals.

2. Dividing by the Scan Rate: $Y \rightarrow Y/(dX/dt)$

For DSC data, this transformation allows for the conversion of power values to heat capacity by calculating the temperature scan rate, dX/dt, and dividing the power values by the temperature scan rate (e.g., μ W/(K sec⁻¹) = μ J K⁻¹). This transformation requires a time scale. If a time scale is not included with the data, it is generated from the time increment per datum (entered by the user) and the index value of each datum in the file. If the t scale is the index scale (implying a time increment Δ t of unity) a prompt will be issued to allow a time increment to be entered.

The first time this transformation is selected, the following explanatory prompt is printed:

Transform power values to heat capacity values by dividing each value of the power by the scan rate: Y = Y/(dX/dt), where X is the temperature and t is time. The scan rate, dX/dt, is calculated as a function of t by fitting X to a cubic power series in t. The scan rate at each time t is calculated from the analytical derivative of the cubic polynomial. If a time scale is not included with the data, it is generated from t=N Δ t, where Δ t is the time increment per datum. ([Esc] to cancel)

The derivatives dX/dt are calculated analytically from the coefficients of a fit of the values of X to a cubic polynomial in t. Note that the scan rate can be a function of time and, hence, temperature.

If markers are active when this transformation is performed, the fit of X to t may be conducted over the range between the markers by entering Y at the prompt:

Calculate dX/dt from data between markers only? Y/N [N] _

If [Ctrl-F1] is pressed to enter the transformation session from a graphical display and if a zoomed plot is active when [Ctrl-F1] is pressed, the plot is first unzoomed and active markers are placed at the limits of the zoomed plot. Also, if markers were active from the present or a previous file and this transformation was selected from the Main Menu, the markers will be presumed to be still active. In both cases the above prompt is printed.

The transformation is initiated by calculating the scan rate as a function of time (and hence temperature) from a cubic power series in t. Since this may require several seconds on slow computers, a message is printed to the screen:

Calculating T = Σ Cntⁿ for n = $0 \rightarrow 3$ Please wait....

Following completion of the cubic fit, the final option is offered:

Replace X values with their smoothed values? Y/N [N] _

If yes, each X value will be replaced by its value calculated from the least squares fit. If the fit covered only the range of data between markers, the smoothed values of X outside this range will be extrapolated from the fit within the range. After the transformation has been completed, the values of the scan rate at the beginning and end of the scan are as printed on the top of the Transformation Menu screen:

NOTE: Scan rate at 19.3° = 0.355°/min and at 69.9° = 0.357°/min.

Following the least squares fit of X to a cubic polynomial in t, the standard deviation of the fit is compared with an estimate of the standard deviation obtained from third differences. (See the earlier

section on *The Least Squares Fit* for an explanation of third differences.) If systematic trends or discontinuities exist in the dependence of X on t, the two estimates of the standard deviation will not agree. If the standard deviation of the cubic fit is more than a factor of three larger than the estimate of the standard deviation from third differences, a warning message and prompt similar to the following are printed:

```
CAUTION! \sigma(T) from third differences = 0.00199°

\sigma(T) from cubic fit = 0.0179°

Measures of \sigma(T) do not agree! Advise analyzing T vs t,

performing cubic fit, and inspecting residuals.

Continue with the transformation? Y/N [N] __
```

giving the user the opportunity to inspect the values of X before performing the transformation.

As with the Tian transformation, an attempt to perform this transformation more than once on a data file will result in a warning message.

3. Interchange X and Y

This can be used to apply transformation 12 to Y or transformation 18 to X.

4. Reverse the order of X,Y relative to t

If the values of X run from high values to low, some of the analysis procedures become unstable or confusing (e.g., the cursor moves are opposite to intuition). Although this will be spotted by the program and a sorting option offered, the sorting procedure is especially slow when the data are in the reverse order. This option greatly speeds the ordering operation. In addition, transformation 15, $X \rightarrow 1/X$, automatically reverses the order of the transformed X axis and the user may wish to transform it back to the order conforming to the original order of X. The t scale is unchanged by this transformation.

5. Compress (average) Data

If files containing more than 1400 records are read, the data are averaged pairwise with $Y_i = (y_{2i-1} + y_{2i})/2$, $X_i = (x_{2i-1} + x_{2i})/2$, and $T_i = (t_{2i-1} + t_{2i})/2$. The pairwise averaging continues as necessary to read the entire data file. Since the averaging may distort the shape of the data profile, this transformation is offered in order to test for possible distortion. The transformation is identical to the averaging procedure used for long files. Performing the averaging again and comparing the analysis before and after data compression will give a measure of the dependence of profile parameters on the averaging process. In addition, this transformation may be used to reduce the number of data and hence speed analysis time when performing preliminary analyses of data.

6: Replace Y with (2Y)ⁿ/2(1-Y)

This is the functional dependence of the equilibrium constant K on the extent of transition α given by eq. (6a) or (6b) in the section on *The Two-State Model* (with Y replacing α in the equation). If a negative value of n is entered, Y is replaced with $2Y/(2-2Y)^n$, the appropriate form for $nA \rightarrow B$. If

Y = 1, the denominator vanishes and the transformed value of Y becomes ∞ . Consequently, corresponding values of X, Y, and t are dropped from the data. The prompt for entering the value of n contains the explicit reminder of the difference between the two cases A \rightarrow nB and nA \rightarrow B and the intended use of this transformation for calculating K from α :

Replace Y, the extent of reaction, with Keq, the equilibrium constant, For A-mB, Keq = (2Y)ⁿ/2(1-Y), enter a positive value for n. For nA-B, Keq = 2Y/(2(1-Y))ⁿ, enter a negative value for n. Enter value for n [1]: _

Note: In the list of transformations performed appearing at the bottom of the Transformation Menu screen and in the "results" file (see [Alt-F7] in the section on Analysis Options), only the $A \rightarrow nB$ form for this transformation, namely $(2Y)^n/2(1-Y)$, is given.

Since the intended use of this transformation is to calculate the equilibrium constant K from the extent of transition α , a few words on the calculation of α are warranted. For calorimetric data, the extent of reaction can be calculated from the integral of the transition profile over the baseline. The preferred baseline is one derived from a two-state analysis of the profile. The integral is generated from the graph of Y with [Shift-F2] and Y is permanently replaced with the integral using [Alt-F1]. Finally, α is obtained by normalizing Y in the Transformation Menu with 7: Y \rightarrow s(Y + Δ Y) using the defaults Y_{min} and $1/(Y_{max}-Y_{min})$ for Δ Y and s respectively. For spectroscopic data, one need only normalize Y to obtain α .

7. Replace Y with $s(Y + \Delta Y)$ 13. Replace X with $s(X + \Delta X)$

The default values for the increments ΔY and ΔX are the minimum values of Y and X respectively. The default value for the scale factor s is the range in Y or X if the default value of ΔY or ΔX is selected. Otherwise the default value is 1. Thus, for example, if the default values are selected for the transformation of Y, the resulting graph will be normalized with Y ranging from a minimum value of 0.0 to a maximum value of 1.0. If, for the transformation $X \rightarrow s(X + \Delta X)$, s is negative, the ordering of the data is reversed (with respect to the independent variable t) so that the new values of X are in increasing order. The markers (which are identified by their index values) are also reversed to preserve the correspondence of markers with data.

8. Replace Y with Yⁿ 14. Replace X with Xⁿ

Only integral values of n are allowed.

9. Replace Y with 1/Y 15. Replace X with 1/X

If a value of 0 is encountered in this transformation, the values of X, Y, and t are dropped from the set. For the transformation in X, the ordering of the X,Y values with respect to the independent variable t are reversed so that the new values of X will be in increasing order and maintain correspondance with the values of Y. If markers are active, they are also reversed so that they will continue to correspond to the original values of X. If the inverse values of X are less than .01, an option of rescaling by a factor of 1000 is offered.

10. Replace Y with Ln(Y) 16. Replace X with Ln(X)

Values of X or Y less than or equal to 0 result in those values (and their corresponding values of Y or X and t) being dropped from the data set.

11. Replace Y with Exp(Y) 17. Replace X with Exp(X)

Values exceeding the maximum or minimum value for a floating point number are replaced with exp(87) or exp(-87).

12. Replace Y with Y \cdot \Sigma Cn X^n 18. Replace X with \Sigma Cnt^n

Transformation 12, allows Y to be multiplied by a calibration factor which itself is a function of temperature. For example, if the values of Y are voltage differences between reference and sample cells, the calibration function, $C(T) = \Sigma C_n T^n$, can be obtained from a calibration curve of Power/Voltage = C(T). This transformation can also be used to transform Y from, for example, calories/min to Joules/sec. The values of C_n may be edited if this transformation is selected and the new values may be stored in the file (PAR).EXM to be recalled each time EXAM is run.

Transformation 18, replacing X with a power series function of time, is the identical function offered when only values of Y are entered as data. In effect, the temperature scale is generated from a known scan rate function. The values of C_n may be edited if this transformation is selected and the new values may be stored in the file *(PAR).EXM* to be recalled each time EXAM is run.

The values of C_n for X and for Y are separate and independent

When either of these transformations is called, a parameter editing menu is offered, for example, for transformation 12:

```
Replace Y with Y \cdot \Sigma CnX^n

C0 = 1.0
C1 = 0.0
C2 = 0.0
C3 = 0.0
C4 = 0.0
Select index, 0-4, of any coefficient to be changed [none] ____
```

As coefficients are selected, the values are entered in response to a prompt

```
Enter new value for C0 [1]: _
```

Following entry, the parameter editing menu is again offered with changed values highlighted. If **[Enter]** alone is pressed, the parameter editing session ends with the prompt

Do you wish to save these values? Y/N [N]

If yes, the new values of the parameters are saved in the (PAR).EXM file.

Ordering of Data

For some of the analyses in EXAM to be conducted correctly, the data must be in order of increasing X. Before the analysis session begins, the ordering of X is checked and if X is not in increasing order, the following warning message and prompt are printed on the screen:



If the analysis is continued without ordering of the data, some of the functions may not work properly or the program may abort. Care should be taken, but under some circumstances it may be desireable not to reorder the data. If the data are in reverse order, i.e., in order of decreasing X, it may be considerably quicker to reverse the order of the data by selecting this transformation from the Data Transformation Menu. Note that the t scale is not changed by the reordering. This is not a problem if the t scale is simply the index of the data. However, one should be aware that the correspondence between X (and also Y) and t is changed by the reordering of the data.

Scaling of Display Axes

Following selection of the analysis to be performed, X vs Y, X vs t, or Y vs t, the option of autoscaling the axes is offered:



The axis limits for autoscaling are determined from the lowest and highest values of X and Y in the file with a sufficient margin to display all the data. If autoscaling is not selected, entry of the minimum and maximum values of X and Y for the plot will be prompted at appropriate times. Values larger than or less than the range in the data can be entered.

This completes the description of the Main Menu. After selection of the analysis to be performed, Y vs X, X vs t, or Y vs t, and whether or not autoscaling is to be invoked, the selected graph appears on the screen. From this point on, the program is controlled with function and cursor keys alone or in combination with the [Shift], [Ctrl], and [Alt] keys through a passive menu that can be displayed

by pressing [?/]. The top line of the display contains summary help and prompt information. The bottom line of the display contains marker and cursor information.

Analysis Options Menu

Once a graph of the data is displayed, the program is controlled with function and cursor keys alone and in combination with the [Shift], [Ctrl], and [Alt] keys. The top line of the display contains summary help and prompt information. The bottom line of the display contains marker and cursor information. The Analysis Options Menu can be displayed at any time by pressing the [?/] key:

Principal Displays	Data Editing, Selection, or Manipulation
[F1] Data [F2] Derivative [Shift-F2] Integral [F9] Residuals [Tab] Redraw	<pre>[Del],[Ins] Deactivate or Reactivate data [Ctrl-F1] Transform data [-] Invert Y [Shift-F1] Delete, replace or insert data [Alt-F1] Replace data with display [F5], [Ctrl-F5] Turn markers on/off, or move [Shift(Alt)-F5] Zoom or (stepwise) unzoom</pre>
Additional Displays	Data Analysis
[F8] Calculated [F4] Reference [F3] Other File [Shift-F9] Residuals [Shift-F4] Resid (Ref)	<pre>[F7],[Shift-F7] Least squares fit to f(X), f(X-Xo) [F6] Draw baseline [Shift-F6] Calculate area [Alt-F6] Locate peak or valley, calculate value [(Ctrl,Alt)-F8] (Confidence intervals, asymptotes) [Shift-F8] Calculate Y ± σ(Y) from X</pre>
Other: [Ctrl(Alt)-F9] Ca [Shift-F3] Save [(Alt,Ctrl)-F3] [Ctrl-F4] Save [Alt-F4] Save [(Ctrl,Alt)-F7] [Ctrl-Home] Full X,Y cura	alculate (+deactivate outliers) & display residuals display or calculation to a file (Save in memory, Recall) fitting function or add display to Reference or recall Reference or add graph of fitting function to Reference Parameters from least squares fit \rightarrow (screen, file) sor [F10] Print screen [Ctrl-F10] Text on screen

Summary of Analysis Options and Commands

The Analysis Options Menu is summarized below. Following this summary each of the menu options will be described in detail.

Principal Displays

All graphs are based on the selected data set: Y vs X, Y vs t, or X vs t. Least squares fits can be performed on any of the principal displays, including the residuals display.

[F1] Display selected graph: Y vs X, Y vs t, or X vs t.
[F2] Display first derivative as calculated from first differences.
[Shift-F2] Display the integral of the data.
[F9] Display the residuals, e.g., Y(obs) - Y(calc) vs X, following a least squares fit.

[Tab] Redraw graph with a solid line. If already a solid line clear the screen and redraw the graph with points.

Additional Graphs

[F8]	Draw or erase on the current display a calculated line from a least squares fit.
[F4]	Draw the reference graph on the current display.
[F3]	Draw the graph of another file on the current display.
[Shift-F9]	Draw the residuals on the current display.
[Shift-F4]	Subtract the Reference Graph from the display, save as a graph of residuals, and then draw it.

Data Editing, Selection, or Manipulation

[←,→]	Arrow keys move the data cursor, a short vertical bar, from one data point to the next.
[Ctrl-←,→]	"Fast" data cursor move (moves cursor \sqrt{N} points, where N is the number of points displayed.)
[↑,↓]	Move the data cursor to the highest or lowest point on the display.
[F5]	Place or erase a marker at the current data cursor position. Markers are used for selecting the range of a graph for zooming and for least squares fits. Markers appear as vertical lines from the data points to the X axis.
[Ctrl-F5]	Turn both markers on or off at their current positions.
[Shift-F5]	Zoom to display data between and including markers. Zoomed plots can be repeatedly zoomed to smaller and smaller regions of the graph. If no markers are displayed, restore full display (unzoom).
[Alt-F5]	Stepwise unzoom from a repeated zoom regardless of the status of the markers. The program remembers up to 10 sequential zooms.
[Home]	Move data cursor to next lowest marker. If at lowest marker, move cursor to first data position. If at first data position, move to highest marker.
[End]	Move data cursor to last data point. If there, erase cursor and markers.

[Ctrl-Home]	Replace the data cursor with a full screen cursor in the form of a cross-hair which can be moved to any point within the displayed axes. The full screen cursor position is not limited to the values of the data and can be used to measure or position any point on the display and to draw straight lines between cursor positions. The full screen cursor routine is also invoked automatically for positioning graphs on an existing display or for editing data.
[Del]	Deactivate the datum at the data cursor or, if both markers are displayed, all data points between markers. Deactivated data points are not included in least squares fits.
[Ins]	Reactivate datum at the data cursor or all points between markers.
[Ctrl-F1]	Transform data using various options offered by the Transformation Menu. This is identical to selection of Option 4 in the Main Menu.
[•]	Invert Y, i.e., Y -▶ -Y.
[Shift-F1]	Delete, replace or insert data.
[Alt-F1]	Replace X and Y with the current display. Deactivated data and data beyond the display range are lost.

Data Analyses

[F7]	Perform a least squares fit of the data between the markers in the current display to a selected function: A two-state calorimetry model, a two-state spectroscopy model, a polynomial, or an exponential function. Follow prompts for selection of function. If either of the two-state functions is selected, a menu of parameters is offered from which to choose those to be varied or to be held fixed. Starting values for the parameters can be entered in this menu.
[Shift-F7]	Similar to $[F7]$ only the least squares fit is to $f(X-X_0)$ rather than $f(X)$.
[Alt-F3]	Save a fitting function in memory as a reference for future recall. Associated statistics are not saved.
[Ctrl-F3]	Recall a fitting function. This replaces the current fitting function but without associated statistical information.
[F6]	Calculate and draw a linear baseline between two markers or between a marker and a cursor.
[Shift-F6]	Calculate the area "under" the display and "above" a baseline if one is presently displayed.
[Alt-F6]	Locate a peak or a valley between two markers and calculate the value of Y above or below a displayed baseline. If no baseline is displayed, the absolute value of Y at the peak or valley is calculated.

[Ctrl-F8]	Draw confidence intervals of a fitting function on the current display,
[Alt-F8]	Draw or erase the asymptotic limits of DSC or spectroscopic functions.
[Shift-F8]	Based on a least squares fit, calculate the value of $Y \pm \sigma(Y)$ from a value of X supplied by the user. $\sigma(Y)$ is the standard deviation of the calculated value of Y based on the statistics of the least squares fit.

Other Menu Options:

- [Ctrl-F9] Recalculate the residuals for the principal data Y (not the integral of Y or the derivative of Y or the residuals of Y themselves) using the currently active function (most recent least squares fit or function recalled with [Ctrl-F3])
- [Alt-F9] This performs the same operation as [Ctrl-F9] but in addition, for DSC functions, deactivates points whose calculated values are more than two standard deviations away from the baseline. [Ctrl-F9] and [Alt-F9] are used primarily to analyze overlapping two-state transition profiles.
- [Shift-F3] Save the current display or fitting function graph to a file.
- [Ctrl-F4] Save the current display as a reference graph which can later be drawn by pressing [F4]. If a reference graph already exists, the new graph can replace it, the reference and display can be exchanged, or the display can be added to, subtracted from, or ratioed with the reference.
- [Alt-F4] Save a graph of a fitting function or a baseline as a reference graph. The graph to be saved can also be added to or subtracted from an existing reference graph.
- [Ctrl-F7] Print the values of the parameters from the most recent least squares fit on the current display.
- [Alt-F7] Save the values of the parameters from the most recent least squares fit in a file.
- [F10] Print the current display. This option works on only a limited number of printers.
- [Ctrl-F10] Allow text entered from the keyboard to be written on the current display.

Analysis Options and Commands - Detailed Descriptions

Principal Displays: [F1] [F2] [Shift-F2] [F9] [Tab]

Four graphs are available and are selected by pressing the function keys [F1], [F2], [Shift-F2], or [F9]. In the following discussion, the displays are occasionally identified by the function keys that produce them. Thus [F1], [F2], [Shift-F2] and [F9] can refer either to the keys or, in italics without boldface emphasis, to the corresponding graphs themselves. [Tab] connects the displayed points with solid lines. If [Tab] is repeated, the screen is cleared and the graph is redrawn, point by point.

[F1] displays the graph of primary data: Y vs X, X vs t, or Y vs t, depending on the selection from the Main Menu. [F1] is the default display following the main menu. If [F1] is pressed a second time, the points of the display are connected by straight line segments. If pressed a third time, the screen is cleared and the original [F1] graph is again displayed. That is, repeated pressing of [F1] behaves like [Tab].

[F2] displays the first derivative of the *current display*. The derivative is calculated from the averages of ratios of differences between adjacent values:

$$\left(\frac{dY}{dX}\right)_{i} = \frac{1}{2}\left(\frac{y_{i+1}-y_{i}}{x_{i+1}-x_{i}} + \frac{y_{i}-y_{i-1}}{x_{i}-x_{i-1}}\right).$$

This method of estimating the first derivative maintains the X scale. The first and last points of the derivative graph are determined by extrapolation. If the data spacing is small compared with the noise in the data, the first derivative plot may appear as scattered points. If **[F2]** is pressed a second time, the derivative of the current display, that is, the derivative of the derivative is displayed.

[Shift-F2] displays the integral of the *current display*. The integral is formed by summing the areas of adjacent trapezoids relative to the X-axis or to a baseline.

$$I_{i} = \left(\int_{0}^{x} y dx\right)_{i} = \sum_{j=2}^{i} \frac{1}{2} (y_{j} - b_{j} + y_{j-1} - b_{j-1}) (x_{j} - x_{j-1}) \quad \text{for } i = 2 \cdots N \qquad I_{1} = 0$$

where b_j is the value of the baseline at the *jth* point. If neither a baseline and nor a least squares fit line is displayed on the screen, two options are offered:

If a least squares fit line or a baseline is displayed on the screen, these options are not offered. The integral is evaluated above the line, that is, the integral of $(Y-Y_{calc})dX$ is calculated and displayed. If the data are fit to a calorimetric, two-state model, both a function line and a baseline are displayed and the data are integrated above the baseline. If the integral of YdX or (Y-Y(1))dX is desired instead of $(Y-Y_{calc})dX$, then the calculated line must first be erased with [F8].

If [Shift-F2] is pressed a second time, the integral of the current display, that is, the integral of the integral will be displayed.

Note that [F2] and [Shift-F2] are not reciprocal. This is because of approximations implicit in estimating derivatives and integrals with finite elements. From a comparison of the expressions for estimating the derivative and the integral, it can be easily shown that

$$\left(\frac{d}{dx}\int_{0}^{x} y dx\right)_{i} = \frac{1}{4}y_{i-1} + \frac{1}{2}y_{i} + \frac{1}{4}y_{i+1}.$$

Thus the graph resulting from [Shift-F2]+[F2] only approximates [F1]. Also, as a result of the averaging implicit in the formulas for the derivative and the integral, the scatter of data in [Shift-F2]+[F2] will be approximately 60% that of the data in [F1].

[F9] displays the residuals graph formed by subtracting a fitting or reference function from a selected display: *[F1], [F2], [Shift-F2],* or *[F9].* A fitting function is obtained by performing a least squares fit or a baseline fit (see **[F7]** and **[F6]** below). A reference function is obtained by saving a fitting function with **[Alt-F3]** and recalling it with **[Ctrl-F3]**. The fitting or reference function can always be identified by pressing **[Ctrl-F7]**. When EXAM is first run and before any fitting function has been defined, the fitting function is the nul function, i.e., $Y_i(calc) = 0$. When **[F9]** is first pressed following a least squares fit, a baseline fit, or recall of a reference function, the residuals are calculated by subtracting from the displayed values of Y the values of Y calculated from the fitting function. Once the residuals have been calculated they can only be changed by performing another fit, reading a new data file, recalling a reference graph, or selecting **[Ctrl-F9]** or **[Alt-F9]**. If, for example, a new data file is read, replacing the current data, pressing **[F9]** will cause the residuals to be calculated for the current display using the most recently derived fitting function.

If **[F9]** is pressed a second time and no intervening least squares fits have been performed, the data points will be connected with straight line segments. If pressed a third time, the screen will be erased and the pointwise residual graph redrawn. Repeated pressings of **[F9]** without intervening least squares fits or recall of reference functions behave like repeated pressings of **[Tab]**.

If a least squares fit of a residual graph itself is performed, pressing **[F9]** will calculate and display the residuals of the fit of the residuals, **replacing** the residuals from the initial least squares fit.

Cursors and Markers: $[\rightarrow]$ [Ctrl \rightarrow] [\leftarrow] [Ctrl \leftarrow] [\uparrow] [\downarrow] [Home] [End] [F5] [Ctrl-F5]

A data cursor and two markers are used to identify specific data points or data ranges for zooming, fitting, area calculations, and data deactivation and activation. The data cursor appears as a short vertical line segment intersecting each data point. Markers appear as vertical lines extending down from the data to the X axis. The data cursor is moved left by $[\leftarrow]$ and $[Ctrl \leftarrow]$ and right by $[\rightarrow]$ and $[Ctrl \rightarrow]$. $[\leftarrow]$ and $[\rightarrow]$ move the data cursor from one point to the next whereas $[Ctrl \leftarrow]$ and $[Ctrl \leftarrow]$

The following paragraphs describe how the data cursor and markers react to various keys and key combinations. Though the description may appear on first reading to be complicated, the keys and their responses have been designed to be intuitive.

If no data cursor is displayed, pressing $[\rightarrow]$ or $[\leftarrow]$ will place one on the leftmost or rightmost point of the display respectively. The $[\uparrow]$ and $[\downarrow]$ keys move the cursor to the highest and lowest points on the display respectively. If a cursor move past the leftmost or rightmost point of the display is attempted, the cursor wraps to the other end of the display.

The values of X and Y (or X and t, or Y and t) at the data cursor are printed on the bottom line of the display. If a calculated curve is *displayed*, the calculated value of Y (in parentheses) is also printed on the bottom line of the display. If a least squares fit has been performed and if the resulting residuals, *[F9]*, are displayed, the number in parentheses following the value of Y (the residual) is the value of the residual divided by the standard deviation of the fit. That is, it is the number of standard deviations by which Y(obs) differs from its expected value of Y(calc). While the cursor key(s) remain depressed, the cursor information at the bottom of the screen is in general *not* renewed. It is always renewed when the cursor key is released. This is done to move the data cursor more rapidly, particularly for older computers, though this will depend on the speed of the computer and the time delay for testing the keyboard.

If the data cursor is on the leftmost point of the display and no markers are active, pressing [Home] will turn the cursor off. If it is on any point other than the leftmost point of the display, pressing [Home] will move the cursor to the leftmost point on the display. If the cursor is on the rightmost point of the display, pressing [End] will turn the cursor off and erase any markers that may have been active. If the cursor is on any point other than the rightmost point of the display, pressing [End] will move the cursor to the rightmost point of the display, pressing [End] will move the cursor to the rightmost point of the display.

Pressing [F5] causes a marker to appear on the screen at the data cursor location. Two markers may be placed on the graph. The markers control data deletion and restoration, zooming, fitting, baseline drawing, and area calculations. If no cursor is displayed when [F5] is pressed, the cursor is turned on, placed at the leftmost data point displayed, and the marker placed at this position also.

The value of X at each marker is printed on the bottom line of the screen between square brackets, e.g.,

[45.5] X = 45.5 Y = -124.8 or, if two markers are active,

[45.5, 72.3] X = 36.7 Y = 77.66

where X and Y are the values at the cursor.

Markers are erased by placing the data cursor on the marker and pressing [F5]. Repeated pressings of [F5] alternately turn a marker on and off. If two markers are displayed and the cursor is not at either marker, pressing [F5] will cause the marker nearest the cursor to shift to the cursor.

Markers are either active or inactive depending on whether they are displayed or not. Pressing [Ctrl-F5] will deactivate active markers and they will disappear from the screen. If a marker is deactivated, its position is remembered until a new marker is chosen. If no marker is active, pressing [Ctrl-F5] reactivates both markers on the screen at their previous positions. The starting, default positions for inactive markers are the first and last points of the display.

If one marker is active, pressing **[Home]** will cause the cursor to jump between the first data point and the marker. If two markers are active, pressing **[Home]** will cause the cursor to move to the next lowest marker or to the first data point. If the cursor is already at the first data point, pressing **[Home]** will move the cursor to the highest marker.

In summary:

[F5]	•	places a marker at the data cursor position or erases a marker at the data cursor position or moves the nearest marker to the data cursor.
[Ctrl-F5]	•	erases any active markers from the screen or restores both markers at their previous positions
[Home]	•	moves the data cursor from one marker to the next to the first position on the display or turns the cursor off
[End]	٠	moves the data cursor to the last position on the display, and if already there, erases any markers and turns the cursor off.

If a marker is on, the data cursor is always on.

Two markers cannot be coincident unless only one data point is displayed, whereupon the marker behavior is unpredictable.

The data cursor and the markers are associated with the data *indices*. Switching the display between [F1], [F2], [Shift-F2], and [F9] or transforming the data ([Ctrl-F1]) leaves the markers at their equivalent positions in the new display. Loading a new data file leaves the markers at their former indices. If higher than the number of data in the new file, the highest is placed at the highest index in the new data. If a plot is zoomed (see below) the markers are placed at the leftmost and rightmost positions on the zoomed screen (their locations before zooming), but in their inactive state.

Full-Screen Cursor: [Ctrl-Home]

The data cursor can be replaced with a full-screen cursor with [Ctrl-Home]. The data cursor is replaced with a cross-hair that can be moved to any position within the displayed axes with cursor keys. A help message appears on the top line:

$\uparrow \downarrow \longleftrightarrow$ PgUp PgDn Home End Ins]: Move [Enter]: Value X,Y (Twice to e	gDn Home End Ins]: Move [Enter]: Value X,Y (Twice to end)	
--	---	--

In contrast with the data cursor, the full screen cursor moves pixel by pixel in response to the cursor keys. A coarse scale cursor move is available with $[Ctrl \rightarrow]$ and $[Ctrl \leftarrow]$ or with the shift keys in combination with the cursor keys on the **numerical keypad**. Although it would be desireable to maintain the convention of Ctrl keys in combination with cursor keys for all coarse cursor moves, this is not possible with the up and down cursor keys which do not register in combination with the Ctrl keys on many keyboards. Hence the shift-cursor combination is offered. The coarse cursor moves one cursor width at a time. [Page-Up] moves the full screen cursor to the top margin of the display, [Page-Down] to the bottom, [Home] to the left margin, and [End] to the right margin. Pressing [Enter] causes the values of X,Y (or X,t or Y,t) to be printed on the bottom line, usually with more significant figures than printed with the data cursor. Since the default initial position for the full screen cursor is the X,Y position of the data cursor, the coordinates at the data cursor can be displayed with greater precision by first pressing [Ctrl-Home] followed by [Enter].

The full screen cursor can be moved directly to any X,Y value on the display by pressing **[Insert]** and providing the values of X and Y (or X and t or Y and t) in response to prompts. The current position of the full screen cursor is the default for each of the prompts.

A straight line can be drawn between any two positions on the screen by moving the full screen cursor to the first position and pressing [Ctrl-Home]. The top line prompt changes to:

Move to second position. Draw line with [Ctrl-Home] or [Ctrl-End]

After moving the cursor to the second position, pressing [Ctrl-Home] draws the straight line and moves the full screen cursor back to the first position. [Ctrl-End] draws the line and leaves the full screen cursor at the second position. If [Esc] is pressed before [Ctrl-End] or the second [Ctrl-Home] is pressed, the procedure is aborted.

Pressing [Esc] or pressing [Enter] twice in succession returns to the data cursor and the Analysis Options Menu.

The full screen cursor routine is automatically used for editing or adding data points ([Shift-F1]) and for vertical placement of reference graphs ([F4]), other file graphs ([F3]), residual graphs ([Shift-F9]), and graphs of fitting or reference functions.

Zooming and Unzooming: [Shift-F5] [Alt-F5]

If two markers are active (i.e., displayed), pressing [Shift-F5] will zoom the display to the region including and between the two markers. Note the following:

Following zooming, the Y axis is always rescaled.

Zooming always leaves the markers in their inactive state at the leftmost and rightmost positions on the zoomed screen.

Zooming can be performed on any display ([F1], [F2], [Shift-F2] or [F9]). Once zoomed, all subsequent graphs will be presented in the zoomed scale.

A zoomed graph can always be zoomed further.

Restoration of the original scale (unzooming) is accomplished by pressing [Shift-F5] with no markers active. When the original scale is restored, markers will be placed on the screen at the former leftmost and rightmost positions of the zoomed graph. One can thus return to the zoomed graph by pressing [Shift-F5] a second time.

The final state of a repetitively zoomed graph can be considered as a nested set of zoomed graphs. [Alt-F5] allows for a stepwise unzooming of a nested set of zoomed graphs. [Alt-F5] will take a zoomed graph to the state it was in before the most recent zoom. [Alt-F5] performs this function regardless of the status of the markers.

Data Deactivation, Selection, and Editing: [Del] [Ins] [Alt-F1] [Shift-F1]

Data can be deactivated with the delete key, [Del], and can be restored with the insert key, [Ins]. If no marker or only one marker is active, pressing [Del] or [Ins] deactivates or restores the point at the cursor and moves the cursor to the next data point in the direction of the most recent cursor move. If both markers are active when the [Del] key is depressed, all of the data between and at the markers are deactivated and the markers are turned off. Similarly, the [Ins] key will restore all of the data between two markers, but in this case, the markers will remain active. Note the following:

Deactivated data appear on the screen but with a different color or lowered intensity. Active data are displayed with higher intensity. Depending on the contrast and intensity settings of the monitor, the distinction between active and inactive data may be more or less obvious. CGA displays do not have high and low intensities in graphics modes. Consequently, for CGA displays, deactivated data are not displayed on the screen. Their presence can be detected with the cursor which still moves from point to point whether those points are displayed or not.

If the data between two markers are accidentally deactivated, they can be restored by first restoring the markers with [Ctrl-F5] followed by the data with [Ins].

In a least squares fit of the data, inactive data are *not* included in the fit. This is the purpose of data deactivation.

As an example of the use of data deactivation, the area under a DSC profile can be determined by

- 1. Deactivating the points throughout the transition interval
- 2. Fitting the remaining data to a power series in X with [F7], thus establishing a baseline
- 3. Restoring the deactivated data with [Ctrl-F5] to restore the markers followed with [Ins] to restore the data
- 4. Calculating the area between the data and the baseline with [Shift-F6].

As another example, two DSC profiles can be analyzed by first deleting all of the points of one profile and then analyzing the remaining profile. This works even if the profiles are overlapping.

Data can be replaced **permanently** with the displayed, active data by pressing [Alt-F1]. Inactive points and points excluded in a zoomed range are permanently deleted. (The original data file is unaltered.) For example, the data could be replaced with a small region of the integrated display by selecting [Shift-F2] to display the integral, zooming to the region of interest with [Shift-F5], and selecting [Alt-F1]. Y would then be replaced with the displayed integral. All points outside of the zoomed region would be permanently deleted. If X vs t were to be subsequently displayed, only those points in the originally zoomed range would remain. Once the original data have been replaced, they can only be retrieved by rereading the original data file. In order to protect against accidental replacement, the prompt:

Replace X, Y with currently displayed graph? Y/N [N]

is offered with corresponding prompts when X vs t or Y vs t is being analyzed. If, for example, the integral of X vs t is displayed, X will be replaced by the integral. The values of t and (undisplayed)

Y will not be changed. The values of the undisplayed variable, t, Y, or X, corresponding to points that are deleted are, of course, also deleted.

Individual data points can be inserted, deleted, or changed with [Shift-F1]. However this can only be performed when the primary data, i.e., [F1], are displayed. Attempts to edit data with [Shift-F1] from displays other than [F1] results in the message:

```
Data can only be edited from normal ([F1]) display
```

[Shift-F1] displays the Data Editing Options menu on the top line:

```
Data Editing Options: [1] Delete [2] Change [3] Insert [none]
```

If [1] is selected, the datum at the cursor is removed, the cursor is moved to the next position, and the Data Editing Options menu remains displayed. This contrasts with deactivated data which are, for the most part, merely ignored until they are restored. Deleting data from the Data Editing Options menu has the same effect as first deactivating data and then permanently deleting them with [Alt-F1].

If [2] is selected, the Data Editing Options menu is replaced with the help message:

DATA EDITING. Select datum with cursor. [Enter] to change. [Esc] to end.

Move the cursor to the point to be changed and press enter to select it. The data cursor is replaced with the full screen cursor in the shape of a cross hair (see [Ctrl-Home]) and the help message is changed to:

 $[\uparrow\downarrow\longleftrightarrow$ or Ins] to select new X,Y value. [Enter] end [Esc] cancel

The full screen cursor can be moved to any point on the screen with the cursor arrow keys. Pressing an arrow key with the shift key simultaneously depressed moves the cursor in coarser steps. [Home] moves the cursor to the left axis. [End] to the right edge of the display, [Page Up] to the top, and [Page Down] to the bottom. [Ins] prompts for keyboard entry of the value of X (or t) and the value of Y (or X). Values beyond the range of the display can be entered using [Ins]. [Esc] leaves the data editing mode. [Enter] replaces the values of X (or t) and Y (or X) of the selected data point with their values at the cross hair. After the values of a selected data point are replaced, selection of another point is prompted and so on until [Esc] is pressed. Note that changing the value of X and/or Y has no effect on the undisplayed coordinate. That is, if X and Y of a particular data point are changed, the t value remains unchanged. To change t also, the graph must first be replaced with X vs t or Y vs t.

Option [3], insertion of data, is offered only if Y vs X is displayed. When [3] is selected, the full screen cursor appears on the screen and the X,Y coordinates of the added point are selected either by moving the cross hair to a selected position or by entering the values from the keyboard after pressing [Ins]. If new values of X and Y are inserted, a corresponding value of t is calculated by interpolation or

extrapolation.

Data Transformations [Ctrl-F1]

[Ctrl-F1] enters the Data Transformation Menu directly without returning to the Main Menu. On exit from the Data Transformation Menu, the program returns to the display. Otherwise, [Ctrl-F1] is identical to Option 4 selected from the Main Menu. Details of the data transformations are described in the section dealing with the Main Menu.

Baseline Drawing: [F6] [F8]

Pressing [F6] when both markers are active calculates and draws a straight baseline between the two data at the markers or between the average values of a range of data at each of the markers. If one marker is active, a straight line is drawn between the marker and the cursor. If no marker is active, a straight line is drawn through the cursor (or the average of a range of points surrounding the cursor) parallel to the X axis. When [F6] is pressed, the prompt

Number of points to be averaged [1] $_$

appears on the top line. The default value is 1 or the most recent value entered. The averaging range is placed symmetrically about the cursor or marker unless the cursor or marker is near the minimum or maximum value of X. If X is near its maximum or minimum, as much of the range is selected as possible.

If a baseline or a calculated function line is displayed before [F6] is pressed, it is first erased before the new baseline is drawn. When a new baseline is drawn, a new set of residuals (Y(obs) - baseline) can be calculated and displayed by pressing [F9]. If a residuals graph itself is resident when the baseline is drawn, pressing [F9] will calculate and display new residuals, i.e., old residuals - new baseline, and the original residuals values will be lost.

The baseline may be erased by pressing [F8], but the values of the coefficients defining the baseline remain in memory and the baseline may be redrawn at any time by pressing [F8] again. The parameters defining the slope and intercept of the baseline may be displayed by pressing [Ctrl-F7].

The drawing of a baseline is equivalent to a least squares fit of a straight line to the data surrounding each of the markers (or marker and cursor).

Curve Fitting: [F7] [Shift-F7]

General Information, selecting data ranges, bottom line information

[F7] and [Shift-F7] initiate least squares fits of selected functions to the data. These functions include:

Polynomial in X (or t) Exponential decay or growth in X (or t) Two-state model for DSC type data Two-state model for spectroscopic type data **[Shift-F7]** has the same effect as **[F7]** except the origin for the X axis is taken to be X_0 , that is, the fit is for the equation: $Y = f(X-X_0)$. X_0 is entered in response to a prompt with the position of the cursor serving as the default value for X_0 . If the cursor is not active, the lowest displayed value of X is the default value of X_0 .

Not all of the data are necessarily fit. The data to be fit can be limited through (1) use of zoomed displays, (2) range definition with the cursor and/or the markers, and/or (3) deactivation of data. If no marker is active, pressing [F7] fits all of the active, displayed data, ignoring any data falling outside the range of a zoomed graph. If one marker is active, pressing [F7] will result in fitting of active data between and including the cursor and the marker. Note:

For power series and exponential functions, only active data between and including the markers are fit.

For the two-state analyses, the displayed data outside of the markers represent asymptotic baseline limits and also contribute to the analyses.

Pressing [F7] or [Shift-F7] produces the following prompt at the top of the screen:

 $[1] \cdot \Sigma Cn X^{n} \quad [2] \cdot Exp(X/\tau) \quad [3] \cdot DSC \quad [4] \cdot SPEC \quad [0] \cdot Resume \quad [F7] \cdot Repeat \quad [1]$

The default function to be fit is a power series in X, $Y = \sum CnX^n$. If, in response to this prompt, **[F7]** or **[Shift-F7]** is pressed a second time, the previous least squares fit is repeated for the current display, with current marker and cursor positions, but with all the options as selected in the previous fit. The option [0], resume, initiates the previous fit with the values from the previous fit as initial values. It is offered only if the previous analysis was a two state (DSC or SPEC) function.

Following any iterative, non-linear least squares fit, the bottom line contains information on the convergence conditions. Pressing any key restores the usual cursor and marker data on the bottom line but with the following difference: The value of Y printed on the bottom line of the display corresponding to the position of the cursor is followed in parentheses by the value of Y calculated from the parameters of the least squares fit. The value of Y(calc) is displayed only if the graph of Y(calc) is displayed. If the resulting residuals are displayed, the number in parentheses following the value of Y (the residual) is the value of the residual divided by the standard deviation of the fit. That is, it is the number of standard deviations by which Y(obs) differs from its expected value of Y(calc).

Power Series Fit:

If a power series function is selected, the order of the power series (from 0 to 8) is selected in response to the prompt:

 $Y = \Sigma CnX^n$. Enter value (0 -> 8) for n [1]

or, if [Shift-F7] is pressed,

 $Y = \Sigma Cn(X - X0)^n$ Enter value (0 -> 8) for n

in which case the value of X_o is entered in response to a prompt of the form:

```
Enter value for X0 [56.301] _
```

Following completion of the fit, the calculated value of Y is drawn, the values of the parameters and their standard deviations are printed in a box on the display, and the title line is restored and appended with an identification of the fit and the standard deviation of the fit, e.g.,

[Title]: $\Sigma Cn(X - 56.301)^n$ for $n = 4 \rightarrow \sigma = .0345$ (.0012)

The quantity in parentheses is the estimate of the precision of the data from third differences. (See the section on *The Least Squares Fit* for the description of third differences.) If the standard deviation of the fit is significantly greater than the estimate of the standard deviation from third differences, the presence of systematic errors is indicated and the standard deviations of the parameters should be viewed with skepticism.

In power series fits, especially at high orders, the coefficients can be highly correlated and significant round-off problems can occur. Round-off errors will contribute to unusual graphs of residuals or confidence intervals. Under such circumstances, fitting Y to $\Sigma \operatorname{Cn}(X-X_{o})^{n}$ with a value of X_{o} near the center of the range of the displayed X values may provide a more satisfactory analysis.

The power series fit is useful for establishing baselines for DSC profiles, particularly when the baselines are obviously nonlinear. Using the data deactivation feature described in the section above on *Data Deactivation, Selection, and Editing*, a baseline can be drawn by deactivating the data in the transition region, estimating a polynomial through the remaining data, restoring the transition data, subtracting the calculated baseline with **[F9]** and replacing the original data with the baseline-subtracted data with **[Alt-F1]**.

Exponential Fit:

An exponential fit is included with the program in order to allow calculation of instrumental time constants. The exponential function is $Y = B + (A-B)e^{(X-X_0)/\tau}$. A is the value of Y at the point where the exponential decay is switched on or where the exponential growth is switched off. If $(X-X_0)/\tau > 0$, the value of the exponent is set equal to 0. The value of X_0 is therefore the cutoff value below which Y = A for exponential decay or above which Y = A for exponential growth. B is the value of Y approached asymptotically for exponential decay or departed from for exponential growth. τ is the time constant and is negative for exponential decay and positive for exponential growth. If a value of X_0 cannot be determined, it is set equal to the first (or last) value of X in the range of data being fit, the value of A is set equal to the value of X_0 , and the values of τ and B are determined from the least squares fit. If [Shift-F7] rather than [F7] has been pressed, the value of X_0 is the value of X at the cursor position even if the cursor position is beyond the range of the markers.

Since the fit is nonlinear, it is conducted iteratively and the number of iterations is printed on the top line while the fit is progressing. The iterative procedure is based on a Taylor's Series expansion of the function around the current values of the parameters, cutting the expansion off at the first term, and calculating the corrections to the coefficients from a linear fit of Y(obs) - Y(calculated). With each iteration, the current values of the parameters are printed in a box on the display with the values of

those parameters being varied displayed with emphasized intensity and indicated with arrows. (Some displays do not allow emphasized intensity, hence the arrows to indicate which parameters are being varied.) A status message of the form:

ITERATION NO. 3 took 11.25 secs $\sigma/\sigma^{\circ} = 1.35/0.687$

is printed on the top line. The value of σ is the value of the standard deviation calculated directly from Y(obs) - Y(calc) using the value of the parameters from the previous iteration (the second iteration in the above example). The value of σ° is the value of the linearized standard deviation resulting from the current iteration. The difference between the two reflects the effects of nonlinearity and the approach to convergence.

Several different tests are applied to determine whether the iterative procedure has converged. If the correction δ_i^n to each and every parameter i at the end of the nth iteration is less than 10% of the standard deviation for that parameter returned by the least squares fit then the procedure is deemed to have converged. A second test is based on the size of the corrections compared to the values of the parameters themselves. If the absolute value of δ_i^n/C_i^n for each and every parameter being varied is less than one part in 10^6 , then the procedure is deemed to have converged. Finally, the standard deviation σ^0 returned by the linearized least squares fit is compared to the standard deviation σ calculated directly from Y(obs) - Y(calc) before the linearized least squares fit is performed. If the value of $(\sigma - \sigma^0)/\sigma < .005$, that is, if the standard deviation has not decreased significantly, the procedure is assumed to have converged. If any of the three convergence tests is successful, the iterative procedure is ended. On the other hand, if the absolute value of the correction to and the standard deviation of each of two or more parameters increases from one iteration to the next, that is, if $|\delta_i^n| > |\delta_i^{n-1}|$ and $\sigma_i^n > \sigma_i^{n-1}$ for two or more parameters, the procedure may be diverging. The iterative process is stopped and re-initiated holding X₀ fixed. If X₀ is already being held fixed, the iterative proces is re-initiated holding both X₀ and A fixed. If both are being held fixed, the user is prompted to intervene.

Analysis may be diverging. Continue? Y/N [Y]

In such circumstances it is generally advised to proceed, judging from the rationality of the values of the parameters themselves whether the procedure is truly diverging or not. Often it is not.

Pressing any key during the iterative process causes the calculation to pause at the end of the iteration. On fast computers, the iterations can be performed so rapidly that the information following each iteration cannot be assimilated. If desired, the program can be instructed to pause automatically after each iteration and to prompt whether another iteration should be performed or not. This automatic pause feature is activated by selecting [8] in the main menu and making the appropriate change when prompted.

A final note: the initial estimates of B and τ are taken from a linear fit of $dY/dX = -B/\tau + Y/\tau$, where dY/dX is calculated from differences between adjacent points. The initial value of A is taken from the average of $(Y-B)exp(X0-X)/\tau$, and the value of X0 from the datum with the largest change between adjacent points. If the least squares fit is unstable, the analysis returns these initial estimates with all parameters indicated as held fixed.

DSC and SPEC: The Two-State Analysis of Calorimetric and Spectroscopic Data.

The first step in the two-state analysis of calorimetric or spectroscopic data is the determination of asymptotic baselines and initial parameter estimates. Therefore when DSC or SPEC is selected, the following message appears on the top line of the screen:

Calculating baselines and initial estimates...

If this is a repeat (**[F7]** selected in response to the least squares fit menu), values set by the user in the previous fit will be used as initial estimates in the present fit, while the remaining parameters are estimated from the data. If this is the resumption of a previous fit (**[0]**), the values from the final iteration of the previous fit are used as initial estimates of the present fit.

For parameters not set by the user in previous fits, the initial estimates depend on the placement of the markers to delineate the transition region and the asymptotic baseline regions. In the most common mode for performing a two-state analysis of calorimetric or spectroscopic data, markers are placed above and below the transition interval to define the pre-transition, transition, and post-transition regions. The initial estimates of B_a and B'_a are determined by a linear fit of the data in the pre-transition region, of B_b and B'_b by a linear fit of the data in the post-transition region, and of the transition parameters from the area and from the peak height and position of the transition profile for calorimetric data and from the transition midpoint and slope for spectroscopic data. The details of how the initial estimates are obtained are described in the section on *The Least Squares Fit*.

If either marker is at the edge of the display, the baseline corresponding to that marker will be horizontal to the X axis passing through the point nearest the edge of the display. If no marker is active when the DSC function is selected, the leftmost and rightmost data define horizontal baselines. If the transition appears to be incomplete within the display range of a DSC profile, a single baseline is defined by the baseline point at either end of the display. Note however that the program can mistake a nearly complete transition for a complete one. Using [Shift-F7] with the cursor at the peak of the displayed profile will force the initial value of T_0 to be held fixed at that value or at a value entered from the keyboard.

If only one marker is active when the DSC or SPEC function is selected, the cursor and the marker behave like two markers. That is, they define the transition interval and the asymptotic baseline regions.

Following the calculation of the initial estimates of the parameters, a menu of the two-state parameters is presented to allow adjustment of any value that may have been incorrectly estimated and to define which parameters are to be varied, which are to be held constant, and which are to be constrained to be functions of each other. A typical Parameter Menu will look something like:

$ \begin{array}{c} \mathbf{n} \\ \rightarrow \mathbf{T0} \\ \rightarrow \Delta \mathbf{H} \\ \rightarrow \mathbf{N} \\ \Delta \mathbf{Cp} \\ \Delta \mathbf{Cp}' \\ \mathbf{Ba} \\ \mathbf{Ba}' \\ \mathbf{Bb} \\ \mathbf{Bb}' \end{array} $	= 1

Note that for spectroscopic data, the value of N is not offered as a variable parameter.

In the Parameter Menu, parameters displayed with high intensity will be varied in the least squares fit. Parameters displayed with low intensity will be held fixed at constant values or will be constrained in the least squares fit. In addition, varied parameters are indicated with an arrow to accomodate those displays that do not have high and low intensities in graphics modes. The up and down cursor keys move the pointer ««« to select a parameter. [Enter] changes the status of the selected parameter between varied, not varied, or, for the baseline and heat capacity parameters, between varied, not varied, and constrained. Any parameter can be held fixed at any value by entering the numerical value from the keyboard. Similarly, a parameter can be varied beginning with a value entered from the keyboard. Certain parameters can be constrained to be functions of the others. In the above example, ΔCp is constrained to be $\Delta B/N = (B_b-B_a)/N$. After each iteration of the least squares fit, ΔC_p is reset to $\Delta B/N$. Similarly, B_a can be constrained to be $B_b-N\Delta C_p$ and so on. However, only one of ΔC_p , B_a , and B_b can be constrained to be a function of the other two. Similarly only one of $\Delta C'_p$, B'_a , and B'_b can be constrained to be a function of the other two. Also, only two of ΔC_p , B_a , and B_b and two of $\Delta C'_p$, B'_a , and B'_b can be varied simultaneously.

If the baseline parameters are held fixed and ΔC_p is varied, ΔC_p is determined solely by the shape of the transition profile. The change in the baseline intercepts at T_0 has no direct effect on its value. If ΔC_p is constrained to be $(B_b-B_a)/N$, then ΔC_p is calculated from $\Delta B/N$ following each iteration and held fixed at this value on the following iteration. It changes from one iteration to the next because N and T_0 change. If, on the other hand, B_b is constrained to equal $B_a + N\Delta C_p$, then $B_a + N\Delta C_p$ is substituted for B_b in eq. (10) in the section describing the two-state model and ΔC_p is determined both by the transition profile and the data in the baseline regions.

The stoichiometry parameter n cannot be varied in the least squares fit but can be set to any positive (for $A \rightarrow nB$) or negative (for $nA \rightarrow B$) integer.

In addition to the printing of the values of the parameters, the pre- and post-transition baselines are drawn on the screen and a vertical line is drawn at T_o to the value of the peak height at $T=T_o$ for an analysis of calorimetric data. For spectroscopic data, the midpoint of the sigmoidal profile is marked with a vertical line and a line tangent to the profile at the midpoint is drawn. This gives a graphical indication of how the least squares fit is going to look at the beginning of the first iteration.

The meanings of each of the parameters (see eqs. (10) and (11) in the section on *The Two-State Model*) are as follows:

n:	Stoichiometry $A \rightarrow nB$ ($nA \rightarrow B$ for $n < 0$)
T0:	Midpoint of transition (where α , the fraction of transition = 1/2)
ΔH:	Enthalpy change from State A to State B at TO
N:	Number of moles of material present (for calorimetric data only)
∆Cp:	Transition change in C _p
ACp':	Transition change in dC_{μ}/dT
Ba:	Pre transition Baseline
Ba':	Pre transition Baseline Slope
Bb:	Post transition Baseline
Bb':	Post transition Baseline Slope

If the four baseline parameters B_a , B'_a , B_b , and B'_b are held fixed, the transition parameters T0, ΔH , N, ΔC_p , and $\Delta C'_p$ are determined only by data **between the markers**. If one or more of the baseline parameters is varied then the data from the entire displayed range are used to determine transition and baseline parameters.

When the conditions for conducting the least squares fit have been selected, pressing [F7] again will initiate the iterative fit. Since the fit is a nonlinear one, it is conducted iteratively and the number of iterations is printed on the top line while the fit is progressing. As described in the section on *The Least Squares Fit*, the iterative procedure is based on a Taylor's Series expansion of the function around the current values of the parameters, cutting the expansion off at the first term, and fitting the corrections to the coefficients to Y(obs) - Y(calc). With each iteration, the current values of the parameter box on the display:

-		_	
\rightarrow	то	=	50.0001
\rightarrow	$\Delta \mathbf{H}$	=	600.544
\rightarrow	N	=	0.998088 E-3
	∆Cp	=	3.00522 (AB/N)
	∆Cp′	=	0.0
	Ba	=	-0.001 E-3
	Ba'	=	0.15 E-3
	Bb	=	0.002998
	Bh'	=	0.058254 E-6

The values of those parameters being varied are displayed with emphasized intensity and/or indicated with an arrow. Constrained parameters such as $\Delta Cp = \Delta B/N = (B_b - B_a)/N$ are also so indicated following their values. Parameters that cannot be expressed conveniently in a fixed decimal format are expressed in floating point form as multiples of $10^{\pm 3n}$ by $E \pm 3n$. In addition, a status message of the form:

ITERATION NO. 3 took 11.25 secs $\sigma/\sigma^{\circ} = 1.35/0.687$

is printed on the top line.[†] The value of σ is the value of the standard deviation calculated directly from Y(obs) - Y(calc) using the value of the parameters from the previous iteration (the second iteration in the above example). The value of σ° is the value of the linearized standard deviation resulting from the iteration just completed. Until convergence, σ is always greater than σ° . The difference between the two reflects the effects of nonlinearity and the approach to convergence.

Several different tests are applied to determine whether the iterative procedure has converged. If the correction δ_i^n to each and every parameter i at the end of the *nth* iteration is less than 10% of the standard deviation for that parameter returned by the least squares fit then the procedure is deemed to have converged. A second test is based on the size of the corrections compared to the values of the parameters themselves. If the absolute value of δ_i^n/C_i^n for each and every parameter being varied is less than one part in 10⁶, then the procedure is deemed to have converged. Finally, if the value of $(\sigma - \sigma^0)/\sigma < .005$, that is, if the standard deviation has not decreased significantly, the procedure is assumed to have converged. If any of the three convergence tests is successful, the iterative procedure is ended. On the other hand, if the absolute value of the correction to and the standard deviation of each of two or more parameters increases from one iteration to the next, that is, if $|\delta_i^n| > |\delta_i^{n-1}|$ and $\sigma_i^n > \sigma_i^{n-1}$ for two or more parameters, the procedure may be diverging and a prompt to intervene is issued.

Analysis may be diverging. Continue? Y/N [Y]

In such circumstances it is generally advised to proceed, judging from the rationality of the values of the parameters themselves whether the procedure is truly diverging or not. Often it is not.

Upon convergence, the calculated profile is drawn as are the asymptotic baselines. For calorimetric data, the baseline is drawn as is a vertical line at the midpoint of the transition (where $\alpha = 1/2$). For spectroscopic data, a vertical line is drawn from the X axis to the midpoint of the transition as is a line tangent to the profile at the midpoint. On the top line is printed information about the least squares fit similar to the following:

 $\texttt{A} \rightarrow \texttt{3B}$ DSC, $\sigma\texttt{=0.0479}$ (0.0382)

The final standard deviation reported by EXAM is σ and the number in parentheses is the estimate of precision from third differences as described in the section on *The Least Squares Fit*. The final value of σ°/σ is printed on the bottom line of the display along with an indication of which test(s) determined that the procedure had converged. This information is replaced with the cursor information when any key is pressed.

If, following a least squares fit of the data, another fit is desired using the same configuration as the previous fit, selecting [F7] in response to the least squares fit prompt will repeat the preceding fit. Those parameters that were held fixed or constrained in the previous fit will, as a default, be held fixed or constrained again. Those parameters given initial or permanent values in the previous fit will

[†] On some computers, the iterations are conducted so rapidly that the top line messages may be difficult to assimilate. The program can be instructed to pause automatically at the end of each iteration by selecting [8] from the main menu and following prompts. Also, pressing any key during the iterative process causes the calculation to pause at the end of the iteration.

be given the same initial values in this fit. However, all of these settings can be changed from the Parameter Menu prior to initiating the iterative fit. Selecting [0] as the least squares fit option starts the next least squares fit with the values of the parameters from the previous least squares fit and also allows additional setting of parameters.

Because systematic errors typically overwhelm the random errors, the standard deviations of the parameters are not printed with the values of the parameters. If the values of the standard deviations of the parameters are desired, the residuals must first be displayed by pressing [F9]. Pressing [Ctrl-F7] then displays the values of the parameters and their standard deviations:

\rightarrow	то	=	50.0001 ± 0.0001
\rightarrow	$\Delta \mathbf{H}$	=	600.544 ± 0.031
\rightarrow	N	=	0.998088 ± 0.000085 E-3
	ΔCp	=	3.00522 (∆B/N)
	∆Cp′	=	0.0
	Ba	=	-0.001 E-3
	Ba′	=	0.15 E-3
	Bb	=	0.002998
	Bb′	=	0.058254 E-6

The values of the standard deviations of the parameters can thus be reviewed only within the context of the systematic errors displayed by pressing **[F9]**. If no systematic errors are apparent, the standard deviations of the parameters may have some significance in assessing the true values of the parameters and their confidence intervals. This will seldom happen however and the standard deviations of the parameters will be of little use. Note that for parameters presented in floating point representations the uncertainty is scaled by the same exponential factor as the parameter value. That is, in the above example, N = 0.998088×10^{-3} and $\sigma(N) = 0.000085 \times 10^{-3}$.

It is important to remember that the dimensional units of N are determined by the units of Y as discussed in the earlier section on *Data and Units*. Provisions have been made in EXAM for scaling Y through a menu of data transformations. For example, a data file from a DSC measurement frequently consists of power and temperature values. The power is the power required to maintain the sample cell at the same temperature as the reference cell. To determine the heat needed to raise the temperature of the cell by a unit amount, the power measurements must be divided by the temperature scan rate. The resulting ratio will have the proper units of heat capacity, namely, energy per degree centigrade. This transformation from power to excess heat capacity is performed by transformation #2, Y->Y/dX/dt, in the menu of available data transformations. (Selected from the Main Menu or by [Ctrl-F1] from the graphics display. See the section on Data Transformations for more detail.) Since it must be performed before conducting the two-state analysis of DSC data in order for the units of N to be correct, attempts to perform a two-state DSC analysis when the transformation has not been performed results in the warning message and prompt:

Y has not been divided by the scan rate. Continue? Y/N [Y]

If no, the least squares fit is aborted in order to allow the transformation to be performed. If the units of Y are power rather than heat capacity, the value of N resulting from the least squares fit will be scaled by the scan rate, and the values of ΔH , T_o , and ΔC_p will be slightly shifted due to the temperature dependence of the scan rate which is usually small.

If the values of Y are in units of voltage, they must first be converted to power values through an instrument calibration function before dividing by the scan rate. One such calibration function that provides a temperature dependence for the calibration constant,

$$Y_{power} = Y_{volts} \sum_{n=0}^{4} C_n X^n$$

where X is the temperature, is included in the data transformation menu. The values of C_n are entered by the user and stored in the *(PAR).EXM* file which is read each time the program is run. Each time the transformation is performed, the values of the coefficients will have their former values unless changed by the user.

Least Squares Parameter Reports: [Ctrl-F7], [Alt-F7]

If the display is changed by, for example, zooming the graph or plotting the residuals, the values of the parameters from a least squares fit are no longer displayed. The values can be recalled by pressing [Ctrl-F7]. [Ctrl-F7] behaves as a toggle. If the parameter table is displayed on the screen, pressing [Ctrl-F7] will erase it.

The results of a least squares analysis may be saved as a disk file with [Alt-F7]. The file name is entered in response to the prompt:

If the file already exists, the results report is appended to the end of the file. The results report is an ASCII file that can be read by any word processor or printed directly. The default file name is RESULTS.EXM. If another file name is entered, it becomes the default. The results report contains a description of the data that are fit including the file name, the title line descriptor (which might also be the file name) and the data type, e.g., Y vs X, dX/dt, Y-f(t), etc.; the fitting function; the marker positions at the time of the fit (or marker and cursor positions if only one marker is active, or the display limits if no marker is active); the values of the parameters; and the standard deviations of those parameters varied in the fit. Parameters not varied appear without standard deviations. Also in the report are the standard deviation of the fit, the estimate of the standard deviation from third differences, and the number of degrees of freedom (= the number of data fit minus the number of parameters varied). For DSC type functions, the report also includes the numerically integrated area above the baseline between the marker positions and the profile height above the baseline at T_0 . The latter is linearly interpolated between the two data above and below T_0 and not calculated from the DSC profile function (beyond using the baseline of this function). Note that the theoretical area under the profile is approximately $N\Delta H(T_0)$. If any data transformations have been performed, these are also listed in the results report. Results reports can also be generated for reference functions (see [Ctrl-F3] below) but without statistical information.

Saving and Recalling Fitting Functions: [Alt-F3], [Ctrl-F3]

[Alt-F3] saves in memory the current fitting function for future reference. When EXAM is exited, all functions saved as references are lost. The associated statistics are not saved and when the function

is recalled, all uncertainties are set to zero and every parameter is indicated as held constant. The principal purpose of [Alt-F3] and its companion [Ctrl-F3] is to aid in the analysis of overlapping profiles. Note how [Alt-F3] differs from [Shift-F3], [Alt-F4], and [Alt-F7]. The graph calculated from the fitting function can be saved as a file of X,Y values by selecting the appropriate option with [Shift-F3]. With [Alt-F4], the graph of the calculated function is saved as a reference graph. [Alt-F7] saves a report of the least squares analysis to a file. [Alt-F3] merely remembers the parameters of a least squares fit and the function type (polynomial, two-state, etc.).

When [Alt-F3] is pressed, a prompt may be issued depending on the number of functions that have been saved. If none, the function is saved with a title that is printed on the top line:

Saved as $F(1) = A \rightarrow B$ DSC, $\sigma = 0.247$ [Ctrl-F3 to recall]

If one has been saved, the prompt is:

```
Ref. Function: [1]»Replace [2]»Save [0]»Delete
```

in which case the existing function in memory can be replaced or the current fitting function can be saved as a second reference function or the existing function can be deleted leaving no function in memory. If more than one function exists in memory, the prompt is:

```
Ref. Function: [1-n]»Replace selection [n+1]»Save [0]»Purge all [none]
```

where n is an integer up to 9. If 9 is selected, all reference graphs in memory are shifted by one and the first one saved is deleted. A title is assigned to each reference to aid in future recall.

If no fitting function parameter table is displayed when [Alt-F3] is pressed and reference functions exist, the option of purging all existing reference functions is offered:



[Ctrl-F3] recalls the reference function if only one function is saved in memory. If more than one, a prompt is issued to select the function:

```
Recall function [1-n] (None prompts for each) [none]
```

If **[Enter]** alone is pressed, a prompt is issued for each function in turn consisting of the title that was printed on the top line when the function was saved. When a function is recalled, it becomes the resident function replacing any fitting function that may have existed when it was recalled. No statistics are associated with the function and it is displayed as if a least squares fit had been performed with the parameters fixed at their displayed values. When a reference function is to be drawn on the display (**[F8]**), a prompt is always issued to allow the graph to be positioned on the display. This is the same prompt offered when a reference graph (**[F4]**), the residuals (**[Shift-F9]**), or a file graph (**[F3]**) is to be drawn on the current display. After a reference function has been

recalled, residuals can be calculated for this function by pressing **[F9]**. Aside from the lack of statistics and the prompt for placement of the graph of the function, a recalled function behaves like any fitting function.

As noted, the principal reason for including the provision of saving fitting functions is to allow analysis of overlapping profiles, saving the functions corresponding to each component as individual references. If there are n profiles in all, residues can be formed by sequentially subtracting n-1 of the functions from the residues themselves until only one profile remains which can then be fit and the procedure repeated for one of the other profiles. This will be described in detail in one of the examples included with this documentation.

Erasing and Drawing Graphs of Fitting Functions: [F8], [Alt-F8]

[F8] acts as a toggle switch, either drawing or erasing the graph of Y(calc) and, if appropriate, its corresponding baseline. Y(calc) is calculated from the parameters of the most recent least squares fit or recalled reference function and is always drawn on the screen following a fit. For two-state models, either spectroscopic or calorimetric, [Alt-F8] draws a dotted line for each of the asymptotic baselines, extending them from the limits of the graph to the midpoint of the transition interval. A vertical line is drawn from the X-axis to the midpoint of the transition. For spectroscopic data, a tangent line is drawn through the midpoint of the transition. If both asymptotes and calculated lines are displayed, pressing [F8] will erase both.

The least squares fitting function is linked to the graph from which it was derived. With the exception of the two-state calorimetric fit, attempting to draw the calculated fit derived from primary data (i.e., from an [F1] graph) on a first-derivative display (i.e., on an [F2] graph) will draw the first derivative of the fitting function on the display. For all other situations where the displayed graph is not the same as the graph from which the most recent fitting function was derived, pressing **[F8]** will result in either of the following prompts:

Calculated line is not derived from current display. Draw it anyhow? Y/N

Draw line generated from previous data set? Y/N [Y]

If selected, the fitting function will be drawn. Depending on the relative ranges and scales of the fitting function graph and the display, various prompts for vertical scaling and/or placement of the fitting function graph may be offered:

```
Range of graph exceeds display range. Rescale? Y/N [Y] __
Vertical position of graph is off scale. Shift for comparison? Y/N [Y] _
Select vertical position for display of graph? Y/N [N] __
```

If the scaling/placement option is declined, the fitting function graph is drawn or not according to the scale of the display. If accepted, the full screen cursor (see [Ctrl-Home]) in the form of a cross hair appears on the screen at the current cursor position and the top line prompt reads:

Select vertical position with: [Ins 1 PgUp PgDn] [Enter] when done

The cross hair can be moved with the arrow keys or it can be positioned by pressing **[Ins]** and following the prompt for the entry of Y. The vertical position of the fitting function graph will be shifted to pass through the center of the cross hair. The horizontal position of the cross hair does not shift the horizontal position of the graph. It only determines the value of X through which the fitting function graph must pass.

The drawing and erasing of fitting functions operates differently for monochromatic and color displays. For color displays, the individual pixels of the fitting function are exclusive ORed with the current status of the corresponding display pixels. This will result in the color of the fitting function curve complementing the data display. If the data graph is white and the fitting function graph is yellow then wherever those two graphs overlap, the difference is blue. Although this gives the display a somewhat jagged appearance, it is easy to see where the fitting function more directly fits the data. Similarly, the asymptote graph is exclusive ORed with the current display and, for DSC two-state functions, the baseline is exclusive ORed with the display. In this latter case, if the baseline is drawn in low intensity yellow (brown) then where it and the fitting function overlap, the resulting color is low intensity white and where the sum of the two overlap with the data, the resulting color is high intensity black. The exclusive ORing also allows the graphs to be easily erased by repetition.

For monochrome displays, the fitting function, baselines, and/or asymptotes are merely drawn on the screen and, instead of erasure, the data graph is redrawn.

Residuals [F9] [Shift-F9] [Ctrl-F9] [Alt-F9]

[F9] calculates and/or displays the residuals graph formed by subtracting a fitting function or a reference function from a selected display: *[F1]*, *[F2]*, *[Shift-F2]*, or *[F9]*. The display need not be the one from which the fitting function was derived. **[Shift-F9]** is similar to **[F9]** except the residuals are drawn on the current display. This does not have to be the display for which the residuals were calculated nor the display from which the fitting function was derived. The fitting function is defined by the most recent least squares fit or baseline fit performed or reference function recalled (**[Ctrl-F3]**). A table of function parameters and a title of the current function, if not displayed, can be displayed by pressing **[Ctrl-F7]**. When EXAM is first run and before any fitting function has been defined, the fitting function is the nul function, i.e., a linear fit with all coefficients equal to zero. When **[F9]** or **[Shift-F9]** is first pressed following a least squares fit or a baseline fit or recall of a reference function, the residuals are calculated by subtracting from the displayed values of Y the values of Y calculated from the fitting, baseline, or reference function. Thereafter, **[F9]** and **[Shift-F9]** will display these same residuals until

- another least squares fit is performed
- a reference fitting function is recalled with [Ctrl-F3]
- a new baseline is calculated with [F6]
- [Ctrl-F9] or [Alt-F9] is selected
- the data that were fit are replaced

If a least squares fit of a residual graph itself is performed, pressing **[F9]** will display the residuals of the fit of the residuals, **replacing** the residuals from the initial least squares fit. Thereafter these new residuals of residuals will remain the current residuals until one of the above actions mandates recalculation of the residuals. Also, if a least squares fit of a residual graph is performed and **[F9]** is first pressed while any *other* graph is displayed, a prompt will be issued to display the old residuals

rather than calculating the residuals for the current display. Still, once residuals are calculated, they are not recalculated until one of the above conditions is met.

If a new data file is entered, reference data are recalled with [Ctrl-F4], or the data are transformed, and a residuals graph exists, pressing [F9] will issue a prompt:

Recalculate residuals for the (new) (reference) (transformed) data? Y/N [N]

If yes, residuals are calculated for the new, transformed, or reference data using the current fitting or reference function. If no, the current residuals remain in effect and are drawn.

Depending on the relative ranges and scales of the residuals graph and the display, various prompts for vertical scaling and/or placement of the residuals graph are offered following [Shift-F9]:

Range of graph exceeds display range. Rescale? Y/N [Y] __ Vertical position of graph is off scale. Shift for comparison? Y/N [Y] __ Select vertical position for display of graph? Y/N [N] __

If the scaling/placement option is declined, the residual graph is drawn or not according to the scale of the display. If accepted, the full screen cursor (see [Ctrl-Home]) in the form of a cross hair appears on the screen at the current cursor position and the top line prompt reads:

Select vertical position with: [Ins 1 PgUp PgDn] [Enter] when done

The cross hair can be moved with the arrow keys or it can be positioned by pressing **[Ins]** and entering Y in response to a prompt. The vertical position of the residual graph will be shifted to pass through the center of the cross hair at the indicated value of X. The horizontal position of the cross hair does not shift the horizontal position of the graph. It only determines the value of X through which the residual graph must pass.

While **[F9]** calculates and then displays the residuals formed by subtracting the fitting function from the *current* display, **[Ctrl-F9]** calculates and then displays the residuals formed by subtracting the fitting function from the *[F1]* display, i.e., from the primary data. And whereas **[F9]** thereafter merely displays the already calculated residuals graph, **[Ctrl-F9]** always recalculates the residuals, even if these are unchanged from the current residuals. Note that the residuals displayed by **[F9]** or **[Shift-F9]** following **[Ctrl-F9]** will be the residuals calculated by **[Ctrl-F9]**. The primary use of **[Ctrl-F9]** and **[Alt-F9]** is in the analysis of overlapping profiles as will become apparent in the examples described in this documentation.

[Alt-F9] is identical to [Ctrl-F9] except when the fitting function is the two-state DSC function. For the two-state DSC function, [Alt-F9] deactivates all data for which the difference between Y_{calc} and the calculated baseline is greater than two standard deviations from the most recent two-state, DSC fit. This aids in analyzing overlapping profiles by ignoring, for the fitting of the second profile, those data strongly influenced by the first profile.
Extrapolation and Interpolation: [Shift-F8]

Following a least squares fit, pressing [Shift-F8] will prompt for the entry of a value of X. From X the corresponding value of Y and its standard deviation will be calculated using the function most recently fit and the variance-covariance matrix from the fit. The default value of X is the value of X at the cursor. If the function is a DSC function, the value of Y minus the baseline is calculated and displayed. If, immediately following the calculation of Y, the full screen cursor is selected with [Ctrl-Home], the cross hair is placed at the selected value of X and the calculated value of Y.

Confidence Intervals: [Ctrl-F8]

[Ctrl-F8] allows curves corresponding to $\pm t\sigma$ to be drawn on the graph. The value of t is entered from the keyboard in response to the prompt:

```
For 77 degrees of freedom enter value of t for to plots: [2] _
```

Note that these are confidence intervals of the fitting function, not the data. The number of degrees of freedom is the number of data fit minus the number of parameters varied. As many of these confidence intervals can be drawn on the display as desired. The confidence intervals are based on the standard deviation of the least squares fit and the correlation of errors among the parameters. If the residuals graph [F9] is displayed when [Ctrl-F8] is selected, two horizontal lines at $\pm t\sigma$ are drawn, where σ is the standard deviation of the fit. These two horizontal lines are thus confidence intervals for the measured values of the data. Confidence interval calculations are based on an assumption of random, uncorrelated errors in the data. They are mostly useful for applications other than two-state analyses because the model errors inherent in the two-state model frequently dominate the random measurement error. The confidence intervals are not drawn if a recalled reference function is the current fitting function.

Area Calculation: [Shift-F6]

If a fitting function or a baseline is currently displayed, the area bounded by the data, the baseline, and two markers (not necessarily the markers used for the fit or the baseline drawing) is calculated when [Shift-F6] is pressed. If the fitting function is a DSC two-state function, the area is calculated between the data and the sigmoidal baseline. If any other fitting function is displayed, the area between the marker and the cursor is calculated. If no marker is displayed, the area from the leftmost to rightmost displayed points is calculated. The area is calculated from the sum of the areas of the trapezoids between adjacent points. Portions of the graph below the baseline or fitted line, subtract area from the total. This is the same as the algorithm used for calculated is that relative to the X axis (i.e., relative to Y=0). The area is printed on the top line in a message of the form:



Again, if a DSC profile has been analyzed with the two-state model, both a calculated transition line

and a baseline appear on the display. The area calculated by pressing [Shift-F6] is the area between the baseline and the *measured* values of Y. The area between the baseline and the *fitting function* can be calculated approximately from Area = $N\Delta H$ (but not exactly because of the temperature dependence of ΔH).

Peak Height Estimation: [Alt-F6]

The height of a peak (or depth of a valley) can be calculated by pressing **[Alt-F6]**. The peak height is based on a least squares fit of a parabola to a segment of data specified by the user. The operation of **[Alt-F6]** depends on the status of the markers. If a cursor but no marker is active when **[Alt-F6]** is pressed, the user is prompted to enter the number of points to be used for the calculation of the parabola:

Number of points to be used in calculation: [9] _

After entering N, the number of points, a parabola is fit to N data values *centered about the cursor position*. The parabola is drawn, a vertical line is drawn from the Y axis to the calculated maximum or minimum of the parabola, and the X,Y values of the peak or valley are printed on the top line. If the maximum or minimum of the parabola fails to fall within the range of the N points used in the fit, an error message is printed:

No peak or valley within the range of calculation

If two markers are active when **[Alt-F6]** is pressed, the range between the markers is searched for a potential peak or valley based on the height of the data above the baseline. A \land mark for a peak or a \lor mark for a valley is placed at the located position and a slightly different prompt for the number of data points to be used in calculating the parabola appears on the top line:

Number of points to be used in calculation ([Esc] for all) [9] _

The purpose for the \land or \lor mark is to indicate how the calculation is progressing. Whether the point is identified as a peak or a valley will not affect the calculation of the parabola. Pressing **[Esc]** results in a prompt to confirm using all the data in the range specified by the markers. If no (or if **[Esc]** is pressed a second time) the calculation is aborted. Otherwise the calculation proceeds as in the case of a cursor alone.

If a single marker is active and at a different location than the cursor, the cursor and the marker behave like two markers. If no cursor is active, **[Alt-F6]** functions as if there were markers at the leftmost and rightmost points on the graph.

If a baseline is drawn on the screen from a least squares fit [F7] or a baseline calculation [F6], the peak height is calculated relative to this baseline when [Alt-F6] is pressed.

64

Reference Graphs: [Ctrl-F4] [Alt-F4] [F4] [Shift-F4]

Any displayed graph, fitting function, or baseline may be saved as a reference graph. At any time the reference graph can be drawn on the current display. A displayed graph can be added to, subtracted from, or ratioed with the reference graph. A calculated graph can be added to or subtracted from the reference graph. The reference graph can replace the displayed graph or the reference and displayed graphs can be exchanged. The reference graph is used in much the same way as a memory register on a calculator.

If no reference graph exists, pressing [Ctrl-F4] saves the current display as a reference graph with the message:

Current display saved as reference graph

If a reference graph exists, **[Ctrl-F4]** allows saving, recalling, exchanging, adding to, subtracting from, or ratioing with the reference graph in response to the prompt:

[1]»Replace [2]»Exchange [3]»Recall [4]»Add [5]»Subtr [6]»Ratio

If a reference graph is recalled, it replaces the primary data which will be lost. If exchanged, the data in the current display are saved as the reference graph and the reference data become the primary data. If the displayed graph is a zoomed graph, the option is offered of saving only the data in the zoomed range or the full range of data. Checks are made and warnings issued if the reference graph and display graph are of different type, for example, Y vs X or Y vs t. If a Y vs X graph is saved, the corresponding t scale is not. If a Y vs t graph is saved, the corresponding X scale is not. When a Y vs X reference graph is recalled, a new t scale is generated from the index of each datum and the default value of Δt . (See the section on reading and interpreting data for a description of Δt .) When a Y vs t reference graph is recalled, a new X scale is generated using the same value of Δt . In general therefore the scale of the second independent variable t is of limited significance after a reference graph has been recalled.

Reference and display graphs are added, subtracted or ratioed in the following way. The overlap range of the two graphs is determined and only data in this overlap range are included. Data from either graph outside the range are discarded. The number of points of the two graphs in this range is compared. If N_{min} is the lesser of these two numbers, then N_{min} is the number of points in the sum, difference or ratio graph. The range is divided into N_{min} segments and the averages of all the Y values and X values (possibly only one value) falling in each segment for the display graph and for the reference graph are calculated. The average of the average values of X from the two graphs is taken as the value of X to be saved in the sum, difference, or ratio. The sum, difference, or ratio of the average value of Y from each graph is taken as the value of Y to be saved. If the X scales of the two graphs overlap exactly, then the sum, difference, or ratio is a simple sum, difference, or ratio of corresponding values of Y. The more general approach allows any two graphs to be added, subtracted, or ratioed so long as their X scales overlap for some region. The sum, difference, or ratio graph replaces the reference graph and can be recalled with [Ctrl-F4] or drawn with [F4].

[Alt-F4] allows saving a fitting function or baseline function as the reference graph, adding a fitting function or baseline function to the reference graph, or subtracting a fitting function or baseline function from the reference graph. If the fitting function to be saved is a DSC two-state function, the

option of saving the function or the baseline is offered with the prompt:

```
Calculated values: [1]»Function [2]»Baseline WHICH?
```

The fitting function can be saved, added to an existing reference graph or subtracted from an existing graph in response to the prompt:

 \rightarrow Reference: [1]»Calculated values [2]»Ref+Calc [3]»Ref-Calc

If the fitting function is to be added to or subtracted from an existing reference graph, the value of Y is calculated for each value of X in the reference graph and added to or subtracted from the corresponding value of Y in the reference graph. If the fitting function is a calorimetric two-state function, either the function itself or the corresponding sigmoidal baseline can be saved, added, or subtracted.

If no reference graph has been saved or if the option of replacing the reference graph with the fitting function is selected, a further option is offered to save calculated values corresponding to all the displayed data or corresponding to 500 points evenly spaced throughout the display range:

Save Y(calc) corresponding to [1]»Data [2]»Displayed Range WHICH? [1] _

[Alt-F4] differs from [Alt-F3] in that for the latter, the parameters defining the function (from which the graph of the calculated values can be generated) are saved whereas for [Alt-F4] it is the graph itself of the calculated values that is saved.

[F4] draws the reference graph on the current display. Depending on the relative scales of the reference graph and the current display, various prompt messages are offered for vertical scaling and/or placement of the reference graph:

Range of graph exceeds display range. Rescale? Y/N [Y] __ Vertical position of graph is off scale. Shift for comparison? Y/N [Y] _ Select vertical position for display of graph? Y/N [N] __

If the scaling/placement option is declined, the reference graph is drawn or not according to the scale of the display. If accepted, the full screen cursor (see [Ctrl-Home]) in the form of a cross hair appears on the screen at the current cursor position and the top line prompt reads:

Select vertical position with: [Ins $\uparrow \downarrow$ PgUp PgDn] [Enter] when done

The cross hair can be moved with the arrow keys or it can be positioned by pressing **[Ins]** and entering the value for Y when prompted. The vertical position of the reference graph will be shifted to pass through the center of the cross hair at the value of X indicated. The horizontal position of the cross hair does not shift the horizontal position of the graph. It only determines the value of X through which the reference graph must pass.

If the X scale of the reference is off the X scale of the current display, an error message is printed:

X scale of reference is off scale of current display

The graph type (Y vs X, X vs t, or Y vs t) of the reference graph is remembered and an attempt to draw a reference graph of one type on a display of a different type will result in a warning message and prompt of the following form:

Caution! Reference is Y vs X and display is X vs t. Continue Y/N? [N]

[Shift-F4] subtracts the reference graph from the *current display*, saves the difference as residuals, Y-f(X), and displays them. This is similar to [F9] only instead of subtracting the fitting function from each value of Y, the corresponding value of the reference graph is subtracted. When [F9] or [Shift-F9] is selected after this, these are the residuals that are displayed until a new set is calculated by performing a new least squares fit, reading new data, or pressing [Ctrl-F9] or [Alt-F9] which always recalculates the residuals. [Shift-F4] only functions when the X scales of the reference and display are identical. They are tested before performing the subtraction and if they do not match, an error message is printed and the calculation is aborted. If subtraction of two graphs with unequal X scales is desired, then the display must be subtracted from the reference graph using [Ctrl-F4] and the reference graph recalled with [Ctrl-F4]. This operation replaces [F1] whereas [Shift-F4] replaces [F9], the residuals, leaving the primary [F1] data intact.

Drawing data file graphs on the current display: [F3]

A data file can be read and the data graphed on an existing display by pressing [F3] and entering the file name in response to the prompt:



The default directory for the file is the directory that is resident when EXAM is first run. If a data file from another directory is sought, the full path must be provided with the file name. Following entry of the data file name, various status messages will be printed on the top line while the file is read and the graph of the data is drawn. This is a display feature only and the principal data remain intact. If the display is replaced the graph is lost. **[F3]** allows multiple graphs to be drawn on the current display whereas the reference graph facility (**[F4]**) allows only one. The reference graph facility however allows exchanges, additions, subtractions, and ratioing of graphs.

Depending on the relative ranges and scales of the data file graph and the display, various prompts for vertical scaling and/or placement of the data files graph are offered:

Range of graph exceeds display range. Rescale? Y/N [Y] ______ Vertical position of graph is off scale. Shift for comparison? Y/N [Y] _____ Select vertical position for display of graph? Y/N [N] ______

If the scaling/placement option is declined, the data file graph is drawn or not according to the scale

of the display. If accepted, the full screen cursor (see [Ctrl-Home]) in the form of a cross hair appears on the screen at the current cursor position and the top line prompt reads:

Select vertical position with: [Ins $\uparrow \downarrow$ PgUp PgDn] [Enter] when done

The cross hair can be moved with the arrow keys or it can be positioned by pressing **[Ins]** and entering the value of Y when prompted. The vertical position of the data file graph will be shifted to pass through the center of the cross hair at the indicated value of X. The horizontal position of the cross hair does not shift the horizontal position of the graph. It only determines the value of X through which the data file graph must pass.

Saving Displays as Files: [Shift-F3]

A displayed graph or a fitting function or both can be saved as an ASCII file of X,Y values by pressing [Shift-F3]. However, a fitting function can be saved as an ASCII data file only if it is displayed. Even if a fitting function exceeds the axis range of the display, all values between the minimum and maximum values of X will be saved. If no fitting function is displayed, only the displayed graph can be saved for future reference. Assuming the most general situation in which a fitting function is displayed, pressing [Shift-F3] results in the prompt:

Save in file: [1]»data [2]»calculated values [3]»both [1] _

If option 2 or 3 is selected and if the fitting function is a two-state calorimetric function, a further option is offered:

Calculated values: [1]»function [2]»baseline [3]»both [3]

Once the selection is made, entry of a file name is prompted:

Save data in file named [Default File Name] _

The default file name is the name of the current file being analyzed. If the file name selected already exists (as will be the case if the default file is accepted) a warning prompt is issued:

REPLACE filename? Y/N [N]

If yes, the old file is overwritten. If no, the file name prompt is repeated. Up to this point, the procedure can be aborted by pressing [Esc] in response to any of the prompts.

Following entry of a file name, the option of saving a title as the first record in the file is offered:

68

TITLE ([Esc] for none) [Default Title]: _

A title of up to 30 characters in length can be entered and will title the data when the file is next analyzed. The default title is the current title appearing in top line messages. The title can be used, for example, for noting whether or not data transformations have been performed.

If the display consists of a zoomed plot, the option of saving only those data in the zoomed range is offered:

Save only the data in the zoomed range? Y/N [Y]

One reason for saving the calculated and observed values in a single file is to use commercial software for preparing graphs of publication quality as was done to produce the figure below. Note that the data density is not changed with [Shift-F3]. The X values of the data remain the same. To save a fitting or reference function graph with a data density of 500 points, the fitting function must first be saved as a reference graph with [Alt-F4], recalled or exchanged with [Ctrl-F4] and then saved as a file with [Shift-F3].



Printing the Display [F10]

If the default printer is capable of printing graphics images using the escape sequence $\text{Esc L} m_1 m_2$ to print a line of 8 bit images at 120 dots per inch density (such as Epson, IBM Proprinter, IBM Laser Printer in PPDS mode, Hewlett-Packard Think Jet in Epson emulation, etc.), pressing [F10] will cause the screen to be printed. This is a slow operation taking anywhere from half a minute to four or five minutes depending on the computer. Each pixel on the display is tested and the pixels are assembled

into 8 bit bytes with each bit corresponding to a pixel. Approximately 80% of the time is spent reading the pixels. Since the printed graph is a direct representation of the display, it is generally not adequate for publication. It is most useful for a notebook of analyses or for overhead projection in a presentation. If a figure of publication quality is required, the best approach is to use [Shift-F3] to save the calculated and observed values to an ASCII text file that can be read and interpreted by commercial presentation graphics software.

If the printer is not on, connected or compatible, an appropriate message appears at the top of the screen. If more than one printer is available, it may be necessary to specify which printer, LPT1, LPT2, etc. This can be set from option 8 under the Main Menu.

Printing Labels and Messages on the Display [Ctrl-F10]

Pressing [Ctrl-F10] places on the screen a text cursor, , which can be moved about with the number pad cursor arrows and which will not erase any text or graphics from the screen. Any character typed from the keyboard (including the alternate characters entered by holding the [Alt] key depressed and entering the decimal equivalent from the number pad, e.g., Alt-179 for the line drawing character "|") will be printed on the screen, erasing all data and text in the character cell. The cursor is then shifted right one position. As with the full screen cursor, [Page Up] moves the text cursor to the top of the screen, [Page Down] to the bottom, [Home] to the left edge and [End] to the right edge. Typing past the right edge causes the line to wrap to the left edge. Character strings may be erased by typing over them with the space bar. The backspace key and the delete key also erase the character at the cursor.

Warning beeps

All prompts and some warning messages are accompanied by a warning beep as a program default. If the display is capable of enhanced and normal intensity in its graphics mode, all prompts and warning messages are displayed with enhanced intensity. The warning beep can be turned off by selecting Option 8 from the main menu.

Returning to the Main Menu and Exiting Program: [Esc]

Pressing [Esc] while in the graphics display will terminate the analysis and return to the Main Menu leaving markers, fitting or reference functions, and reference and residuals graphs in their current state. If, from the Main Menu, [Esc] is pressed, the program offers one final prompt:

Exit Program? Y/1	N [N]_
-------------------	--------

and, if [Y], [y], or [Esc], ends.

70

EXAMPLES

Every attempt has been made to make *EXAM* an intuitive program. However, as in all programs, some experience with its operation is required and reading through the documentation on program options can be tedious. The following examples have been selected to serve as a learning aid for the *EXAM* program and to demonstrate several points about two-state analyses in general. They use a number of test data files supplied with the program.

1. A simple example of model errors

The data file *TEST1.DAT* contains a simulation of a differential scanning calorimetric analysis of a simple protein. The "true" values of the parameters are:

 $A \rightarrow B$ transition, n=1 °C $T_0 = 50$ kJ mol⁻¹ $\Delta H = 600$ N = 1.0umoles of material $\Delta C_p = 3$ kJ K⁻¹ mol⁻¹ kJ K⁻² mol⁻¹ $\Delta C'_{\rm p} = -0.15$ kJ K⁻¹ $B_{a} = 10$ kJ K⁻² $B'_{a} = 0.15$ kJ K⁻¹ $B_{\rm b} = 13$ kJ K⁻² $B_{\rm b}' = 0$

The data file consists of 641 records. Each record consists of a temperature value and the instrument response in microwatts. Normally distributed errors have been added to the temperature and the cell response.

Begin the analysis by typing EXAM. If a parameter file *{PAR}.EXM* cannot be found in the default directory, the directory that is active when *EXAM* is run, a message to that effect will be printed on the screen. Select Option 3 to create one in the current directory. If the parameter file does exist, the program will begin with the title page. Note that pressing [?/] at this point prints all the analysis options on the screen. None of these is effective however until a data graph appears on the screen. Pressing any other key leads to a prompt for entry of the name of a data file:

```
Enter name of data file
   [Enter] alone prompts for data entry from the keyboard
   <Filename>/P allows previewing of data file
   RUN <command> executes DOS command
   [Esc] to cancel
Entry: __
```

Enter TEST1.DAT. The program will respond with a message that the file is being read. If the file cannot be found in the default directory, the directory that is active when *EXAM* is started, a message to that effect is printed on the screen.

After the file has been read successfully, the data definition menu appears on the screen:

```
The reading of 641 data records took 3 secs.
TEST1.DAT contains 2 data fields. Program variables are defined as:
    X (temperature) = Field 1
    Y (response) = Field 2
    t (time or index) = index
Select: [Enter] to continue
    [X,x], [Y,Y], or [T,t] to redefine variable
    [R,r] to review beginning lines of data
    [Esc] to abandon this data file
```

In this example, the definitions of X, the temperature, as Field 1 and Y, the response, as Field 2 are correct as can be seen by reviewing the beginning lines of the file by pressing the R (for review) key.

The data were simulated with 10 s time intervals and an instrument response time constant of 100 s. The time scale will be needed to calculate the scan rate and to correct for the response time of the instrument. Therefore, if the t scale is not defined as 10 s per point, it must now be so defined. Pressing T (or t) will result in the prompt:

Which field (1-2, 0 if none) contains the values of t [0]

Select 0, which is the default for data files with only two fields per record. This signifies that none of the data fields is to be interpreted as the time scale. A second prompt is offered to select the time increment to generate a time scale proportional to the data index:

Define t = $\Sigma \Delta t$ where Δt = time increment (in seconds). Enter value of Δt [1]: _

Enter the value of 10. The data definition menu is repeated, this time with the [t (time or index) = $\Sigma \Delta t$, $\Delta t = 10.0$] definition for t.

Once the time scale has been defined, pressing [Enter] leads to the Main Menu:

M	AIN MENU				
Data file: TEST1.DAT	Title line: A \rightarrow B \triangle H=600 T0=50 τ =100				
The following options are available:					
1: Analyze Y vs X 5 2: Analyze X vs t 6 3: Analyze Y vs t 7 4: Transform Data 8	: Analyze a new file : Change title line : Execute DOS command : Program Environment				
ENTER NUMBER or [Esc] to exit program	1 [1] _				

Before beginning the analysis, transform the data from units of microwatts to milliJoules/° by dividing all the values of Y by the scan rate and then multiplying by 0.001. Select [4] from the Main Menu to reach the Data Transformation Menu:

[Enter] alone to end and resume analysis Select Transformation: 1: Tian Transformation 7: Y -> $s(Y + \Delta Y)$ 13: X -> $s(X + \Delta X)$ 8: $Y \rightarrow Y^n$ 14: $X \rightarrow X^n$ 2: $Y \rightarrow Y/(dX/dt)$ (t=N Δ t) 9: Y -> 1/Y15: $X \to 1/X$ 3: Interchange X and Y 4: Reverse order of X, Y 10: $Y \rightarrow Ln(Y)$ 16: $X \to Ln(X)$ 11: $Y \rightarrow Exp(Y)$ 17: $X \rightarrow Exp(X)$ 5: Compress (average) Data 6: Y -> $(2Y)^{n}/2(1-Y)$ 12: Y -> Y· Σ CnXⁿ 18: X -> Σ Cntⁿ [Esc] to cancel all transformations performed in this session Enter Selection ([Enter] alone to end, [Esc] to cancel):

Divide the values of Y by the scan rate by pressing [2][Enter]. An explanatory message about the transformation is printed:

Transform power values to heat capacity values by dividing each value of the power by the scan rate: Y = Y/(dX/dt), where X is the temperature and t is time. The scan rate, dX/dt, is calculated as a function of t by fitting X to a cubic power series in t. The scan rate at each time t is calculated from the analytical derivative of the cubic polynomial. If a time scale is not included with the data, it is generated from $t=N\Delta t$, where Δt is the time increment per datum. ([Esc] to cancel)

Pressing any key other than [Esc] proceeds with the transformation. The scan rate is calculated in accord with the above message. Since on some computers this may take several seconds, a message is printed to the screen:

Calculating $T = \Sigma Cnt^n$ for $n = 0 \rightarrow 3$ Please wait....

After the scan rate has been calculated as a function of temperature, the option of replacing the

temperature scale with a smoothed temperature scale based on the scan rate function is offered:

Replace X values with their smoothed values? Y/N [N]

In general, the option is declined unless compelling reasons should otherwise dictate. Following this, the transformation is performed, a message appears on the top line reporting the beginning and ending scan rate in degrees centigrade per minute, and on the bottom line a message of the transformations performed to date on this data file. Pressing [Enter] should result in exit from the transformation menu and return to the main menu except that the magnitude of the values of Y is now sufficiently large to suggest rescaling by a factor of 0.001. The units of Y are μ J K⁻¹ as a result of dividing by the scan rate. Multiplying the values by 0.001 will result in units of mJ K⁻¹. If the values of Y are in mJ K⁻¹ the value of N returned by the two state analysis will be in µmole. (This can be easily deduced by noting that the integral of Y, which will be in mJ, is approximately N Δ H and Δ H is in kJ/mole. Thus, mJ/(kJ/mole) = µmole.) Accept the option by entering Y and return to the main menu.

Accept all program defaults by pressing [Enter] and [Enter] again when prompted to autoscale the displays. The transformed data are then displayed on the screen as in Figure 1.



Proceed immediately to perform a two-state analysis of the data by first differentiating the asymptotic regions from the transition interval. Place markers at X=26.24 and X=73.74 as in Figure 2. This is done most easily by moving the cursor from its home position to X = 26.24 with 5 presses of [Ctrl \rightarrow]. [F5] places a marker at this position. Move the cursor to the far right with [End] and then left to 73.74 with 5 presses of [Ctrl \leftarrow]. [F5] places a marker at this position. Remember that [\rightarrow] and [\leftarrow] move the cursor from one datum to the next whereas [Ctrl \rightarrow] and [Ctrl \leftarrow] move the cursor $\sqrt{N_{pts}}$ where N_{pts} is the number of data points displayed on the screen. The markers define the pretransition asymptote, the transition region, and the posttransition asymptote. The analysis is begun by pressing the [F7] key and selecting [3], the DSC analysis, in response to the prompt.

The program calculates initial estimates of the baseline and transition parameters and, as shown in Figure 2, draws dotted asymptotic baselines and a dotted vertical line at the estimated value of T_0 . A table of the initial estimates of all the parameters is printed on the display. The units of ΔH are



kJ mol⁻¹, of ΔC_p kJ K⁻¹ mol⁻¹, $\Delta C'_p$ kJ K⁻² mol⁻¹, and the units of T₀ degrees Centigrade. The units of B_a and B_b are the same as the units of Y. The units of B'_a and B'_b are in units of Y K⁻¹. The units of N depend on the units of Y. In the present case, with Y in units of mJ K⁻¹, the units of N are in µmoles. The top line provides brief help information. Parameters that will be varied in the least squares fit are indicated with an arrow and, if allowed by the monitor and display, in enhanced intensity. Initial estimates can be changed by moving the ««« marker to the parameter of choice using the up and down cursor keys and entering the value from the keyboard. Pressing [Enter] alone selects or deselects parameters for varying. The value of n is the stoichiometry of the transition $A \rightarrow nB$. $nA \rightarrow B$ transitions are indicated by negative values of n. n is not varied in the least squares fit but is set at its appropriate value at this time. For the case at hand, we accept the program default configuration and proceed directly to the analysis by pressing [F7].

The parameter table, asymptotes, and vertical line at T_0 are erased. As the analysis progresses, a message on the top line announces the number of the iteration and the values of the standard deviation from the linearized fit, σ° , and the standard deviation calculated directly from Y_{ebs} - Y_{cale} , σ . (See the section on the least squares fit for a more detailed description of σ° and σ .) With each iteration the values of the parameters are updated and printed on the screen. When convergence is reached the asymptotes and vertical line at T_0 are again drawn as well as the calculated fitting function and the baseline function, as shown in Figure 3.

On the top line is a description of the fit including the standard deviation and, in parentheses, an estimate of the standard deviations from third differences in the data. The fitting function appears to agree well with the data. However, this is misleading. The first clue that the agreement is not as good as it seems is the difference between the standard deviation of the fit, $\sigma = 0.676$, and the value obtained from third differences, 0.047. The standard deviation is an order of magnitude greater than expected. This is usually an indication of model errors. Another indication of model errors is the value of N, which, as mentioned above is in units of µmoles. If the molecular weight of the material under investigation is known, the number of moles in the calorimeter is known. In this case it is known to be 1 µmole. The value returned by the two-state analysis however is 1.1215 µmoles, a significant difference from the "known" value. This difference suggests model errors. In the absence of knowledge of the magnitude of N, the presence of model errors is most clearly demonstrated by pressing **[F9]** and inspecting the residuals, Y_{obs} - Y_{calc} , as shown in Figure 4.





If model errors did not dominate the random errors, the residuals would appear as a scatter of points more or less filling the screen. Instead, one sees the pronounced pattern of Figure 4. The most important lesson to be drawn from this is the unreliability of the standard deviations of the parameters returned by the least squares fit when systematic errors arising from an inaccurate model dominate the random measurement errors. That is why the standard deviations of the parameters are displayed for the two-state model only when the residuals are displayed and when trends in the residuals can be visually assessed. Press [Ctrl-F7] to display the values of the parameters and their standard deviations as in Figure 5.

Taking ΔH as an example, the true value of ΔH as already noted is 600 kJ/mole, but the value returned by the least squares fit is 536.5 kJ/mole, an error of 63.5 kJ/mole which is 63.5 times the standard deviation of 1 returned by the least squares fit! In the presence of the systematic, non-statistical error apparent in Figure 5, the statistical interpretation of the standard deviations of the parameters from the variance-covariance matrix of the least squares fit is meaningless.



The nature of the least squares fit is to spread the error as uniformly as possible throughout the data and among the parameters. For the variation of three parameters, T_0 , ΔH , and N, the residuals curve crosses zero three times. The residuals returned by the least squares fit can be compared with the differences between the data and a calculation based on the true values of the parameters given at the beginning of this example. To do this first save the residuals as a reference graph with [Ctrl-F4]. Return to the primary data display with [F1]. Refit the data to the two state model with two presses of [F7], the first to initiate a least squares fit, the second to repeat the previous fit. This time set each of the values of the parameters to their "true" values by moving the ««« marker to each parameter and entering the following values in turn:

n = 1 T_0 = 50 $\Delta H = 600$ N = 1.0 $\Delta C_{p} = 3$ $\Delta C'_{p} = -0.15$ Ba = 10 B'_a = 0.15 $B_{\rm b}$ = 13 $B_{\rm h}'$ = 0

When the parameters have been set, start the least squares fit with [F7]. Since all parameters have been set and none are being varied, no fit is actually performed. Instead, the profile for the entered values of the parameters is calculated and drawn. A standard deviation is still calculated and is 9.63, more than an order of magnitude greater than the standard deviation that resulted from the least squares fit to the two state model! A large fraction of the model error is absorbed by distributing the errors among the parameters in the least squares fit.

To see this graphically, display the residuals with [F9]. Move the cursor to the lowest point on the display with $[\downarrow]$ and redisplay the primary data with [F1], noting that the cursor still marks the position where the model error is the most negative. Now draw the model error, i.e., the residuals

from the true values of the parameters, on the display with [Shift-F9]. A prompt will appear to select the vertical position the residuals graph:

```
Select vertical position for display of graph? Y/N [N] \_
```

Select [Y] and the data cursor is replaced by a cross hair and the top line prompt becomes:

Select vertical position with [Ins], \uparrow , or \downarrow . [Enter] when done

[Page Down] moves the cross hair to the bottom of the display. Since the cross hair marks the X position of the minimum of the residuals plot, we know that the residuals plot will be drawn on the display with the minimum point passing through the center of the cross hair touching the bottom axis. Press [Enter] to draw the residuals. To compare these model errors with the residuals from the two-state analysis, recall that the residuals from the least squares fit were saved as the reference graph. Draw the reference graph with [F4]. When offered the option of positioning the reference graph, enter [Y] and with the cursor keys, move the cross hair to coincide with the horizontal asymptotes at either end of the previously drawn residual graph. Press [Enter] and the reference graph is drawn passing through the center of the cross hair. The resulting display should resemble Figure 6 which shows graphically how the full extent of the model errors can be masked by the least squares fitting process.



The source of the model error in this example is the neglect of the response time of the calorimeter. The profile can be adjusted to accomodate the effects of the response time using the Tian equation. For the simulated data in this example, the time constant is $\tau = 100$ s. For real data, the time constant must be measured independently for the instrument. Depending on the time constant and the scan rate, the effect of instrument response time may be important or not. To test this, the data can be analyzed both with and without performing the Tian transformation and the results compared. If the Tian transformation does not alter the values of the parameters or if the difference between the two-state model fits of the data before and after the Tian transformation is small compared to the residuals of the fit (we will examine how to show this later in this discussion), then the instrument response time can be ignored. In this example, it cannot be ignored as will now be demonstrated.

Pressing [Ctrl-F1] replaces the graphical display with the Transformation Menu. Selecting 1, the Tian transformation, provides an explanatory message:

Transform Y according to the Tian Equation: $Y \rightarrow Y + \tau(dY/dt)$ where t is the time and τ is the instrument time constant. If a time scale is not included in the data file, it is calculated from $t = I\Delta t$ where Δt is the time increment per datum and I is the index of each datum in the data file. In calculating dY/dt, Y is expressed as a power series in t, fit to M points placed symmetrically about each value of Y(I),t(I) and the derivative dY/dt is evaluated from the coefficients of the fit.

Pressing any key except [Esc] replaces the message with the prompt:

Enter value for time constant τ (in secs) [1]: _

Enter the value of 100. The next two prompts determine how the derivative dY/dt is to be calculated as a function of t. A range of data is fit to a polynomial and dY/dt is calculated for the datum at the midpoint of the range by taking the analytical derivative of the polynomial and evaluating it at the value of t at the midpoint of the range. Since the least squares fit is repeated for each of the data, the transformation can be a very time consuming operation. A three point linear fit gives a very rapid response but typically magnifies the noise in the data by an order or magnitude or more. A 25 point cubic fit can take as long as several minutes but does not alter the random noise in the data. However, taking too large a range can distort the data if the polynomial is not a very good approximation to the data in that range. With real data, several combinations should be tried before choosing the most suitable for the data and the scan rate. For the example at hand, a cubic polynomial of seven points will be used to calculate the derivative dY/dt for the Tian equation. Therefore, in response to the prompt:

Enter polynomial order (1-3) for calculating derivative [1]: _

enter 3 and in response to the prompt:

Enter number of points to be used in calculating the derivative [7]: _

accept the default value of 7. As a warning for those situations when the transformation is expected to take several minutes, a rough estimate of the required time is made and a message similar to the following is printed:

Transformation will take approximately 0:31 minutes

After the Tian transformation is performed, it is added to the list of transformations printed on the bottom of the display. It appears in higher intensity to indicate that it was performed in this most recent entry into the transformation menu. It also indicates that this transformation, and only this one of the three performed, can be undone by pressing [Esc]. Pressing [Enter] returns to the graphical display.

Repeat the two-state DSC fit by pressing [F7] twice, the second time to indicate a repeat of the previous fit. Note that all the parameters are held fixed at their true values. When repeating a least squares fit, all of the conditions of the previous fit are invoked and, in this case, the previous fit was a calculation of the profile with the true values of the parameters. Since this is not the fit we wish to perform, press [Esc] to abort the analysis, press [F7] to initiate a new least squares fit and select [3]. Now the program's default settings are displayed. Press [F7] to start the least squares fit.

The value of the standard deviation from the fit is now 0.263 whereas the estimate from third differences is 0.158. The standard deviation has increased because of the relative magnitudes of the cell response time and the time intervals in the Tian transformation. With τ =100, Δ t=10, and a 7 point cubic fit with 4 parameters (and thus three degrees of freedom), we would expect the noise in Y=Y+ τ (dY/dt) to increase by a factor on the order of the square root of 1+(1/3)(τ/Δ t)=5.86. When the Tian transformation using a seven point cubic fit is performed on random noise, the standard deviation increases by a factor of 5.22 in good agreement with the expected value. However the standard deviation estimated from third differences increases by only a factor of 3.5. This is because the errors in the data are now correlated. That is, any two adjacent points following the Tian transformation were calculated using six common data points. This correlation causes the third differences to underestimate the standard deviation and demonstrates why the third differences can serve only as a qualitative guide to the possibility of model errors.



As already mentioned, the best indication of model errors is the appearance of trends in the residuals. Pressing [F9] displays the residuals as in Figure 7. In contrast to Figure 4, the residuals now appear as a random scatter of data filling the screen but clustering about Y - f(X)=0.

Pressing [Ctrl-F7] prints the values of the parameters on the residuals plot as shown in Figure 8. It can be seen that the calculated values of T_0 , ΔH , and N now fall within three standard deviations of their true values. However, they are still more than two standard deviations away, a point that will be discussed further in Example 3 below.

Note that when the parameters are printed on the display, a number in parentheses follows the value of Y printed below the graph ((-1.0291) in Figure 8). For the residuals plot, this is the number of standard deviations by which the value of Y differs from its calculated value. When the parameters are not displayed as in Figure 7, only the value of the residual, Y - $f(X) = Y_{obs} - Y_{calc}$, at the cursor is



displayed.

2. The order of performing the Tian transformation and dividing by the scan rate doesn't matter.

To demonstrate further uses of the reference graph, we now ask whether performing the Tian transformation before or after dividing by the scan rate makes a difference in the analysis (it should not since dividing by the scan rate is merely a scale change). To answer this, two analyses will be compared: one for data for which the Tian transformation is performed *before* division by the scan rate with one for which the Tian transformation is performed *after* division by the scan rate. The latter analysis has just been performed, so begin by saving the results of this calculation as a reference.

First replace the residuals graph of Figure 8 with the graph of the primary data by pressing [F1]. Attempting to save the fitting function as a reference graph by pressing [Alt-F4] will result in the top line warning message:

Calculated graph can be saved only if displayed

Draw the fitting function on the display by pressing [F8]. Now [Alt-F4] results in the top line prompt:

Calculated values: [1]»Function [2]»Baseline WHICH? [1]

Select [1] to save the function rather than the baseline. Because a reference graph already exists (Remember that the residuals were saved in Example 1.) the next prompt asks if the calculated function is to replace the reference graph, to be added to the reference graph, or to be subtracted from the reference graph:

 \rightarrow Reference: [1]»Calculated values [2]»Ref+Calc [3]»Ref-Calc [1]

Select [1]. A final prompt is offered:

Save Y(calc) corresponding to [1]»Data [2]»Displayed Range WHICH? [1]

If [1] is chosen, Y is calculated for each value of X and these calculated values along with the X values of the data are saved as the reference graph. If [2] is chosen, the displayed range of data is divided into 500 equally spaced values of X and for each value of X, the value of Y is calculated. These X,Y values are then saved as the reference graph. When only a few data exist, [2] may be preferable because the reference graph will present a smoother graph when displayed. In this example with 641 values select [1].

Now verify that a reference graph exists by drawing it on the current display with [F4]. Before the reference graph is drawn, the option selecting its vertical position is offered:

```
Select vertical position for display of graph? Y/N [N] _
```

Selecting [N] draws the reference graph on the current display. It should appear to be identical to the fitting function.

Next verify that the reference graph is indeed the fitting function by replacing the primary data with the reference graph. [Ctrl-F4] offers six options prompted on the top line:

```
[1]»Replace [2]»Exchange [3]»Recall [4]»Add [5]»Subtr [6]»Ratio [1] _
```

Select [3]. The screen is cleared and a new graph is drawn. To verify that this is indeed the fitting function calculate the residuals with [F9]. Because the data have been replaced, the residuals are recalculated as indicated by the message on the top line. The residual graph is a straight line at Y - f(X)=0, thus demonstrating that the reference graph is the fitting function.

Next generate the fitting function for the same data, but performing the Tian transformation before dividing by the scan rate. To do this, return to the main menu with **[Esc]** and select **[5]**, "Analyze a new file." When prompted for the file name, again enter *TEST1.DAT*. Note that when the file has been read and the Data Definition Menu appears, the default definitions are the same definitions as those previously selected. The program remembers the definitions entered and uses these as defaults until a new data file with a different number of fields per record (entries per line) is read or until the program is exited.

Press [Enter] to return to the Main Menu. Select [4] to display the Transformation Menu. Notice that the bottom of the screen, which previously contained a list of the transformations performed, is now blank because a new data file has been read. Perform the Tian transformation first. As before, select a cubic fit with 7 points. Note that the default value for the time constant τ is the value 100 entered in the previous analysis. This value will be saved in the *{PAR}.EXM* file and will remain the default value the next time *EXAM* is run. Following the Tian transformation, divide by the scan rate with [2] $Y \rightarrow Y/(dX/dt)$, and exit from the Transformation Menu again accepting the option to multiply the values of Y by 0.001. From the Main Menu proceed to the analysis of Y vs X. Note that the markers have retained their positions. Again perform the two-state DSC analysis with [F7] accepting all program defaults as before. Following the analysis, press [Alt-F4] to subtract the calculated values of the fitting function from the reference graph. Select the [1]»Function rather than [2]»Baseline when prompted and when further prompted with:

```
→ Reference: [1]»Calculated values [2]»Ref+Calc [3]»Ref-Calc [1]
```

Select [3]. The reference graph now contains the difference between the fitting function for the data with the Tian transformation performed *after* dividing by the scan rate and the fitting function for the data with the Tian transformation performed *before* dividing by the scan rate. Compare this difference to the residuals by first displaying the residuals with [F9] resulting in a scatter plot similar to Figure 7. Press [F4] to draw the reference graph this time declining to position the graph on the display. This will force the reference graph to be drawn on the scale of the residuals graph as in Figure 9.



Even on the scale of the residuals, the reference graph appears to be a flat line at Y - f(X)=0. In fact it should not be exactly 0 since interchanging the two transformations should lead to slightly different round off errors. This can be seen by exchanging the reference graph with the primary data by pressing [F1] followed by [Ctrl-F4] and in response to the prompt:

[1]»Replace [2]»Exchange [3]»Recall [4]»Add [5]»Subtr [6]»Ratio [1] ____

select [2] to exchange the primary data with the reference graph. After the reference graph has been drawn press [Tab] or [F1] to connect the points with straight lines. The resulting graph should be similar to that shown in Figure 10. Interestingly enough, the shape of the original profile is reproduced. The large jumps from one point to the next represent the six digit precision limitation inherent in representing floating point numbers with five bytes. The Y scale is more than five orders of magnitude smaller than the original data and of a magnitude expected for round-off errors.



3. Even with an apparent random scatter in the residuals, the model errors are greater than suggested by the standard deviation of the parameters.

In the first example, it was shown that after the Tian transformation had been performed, the parameters from the least squares analysis agreed with their true values to within three standard deviations. However, for the analysis being performed, one would expect the parameters to differ from the true values by more than two standard deviations less than 5% of the time. That the differences from the true values are still a little larger than expected suggests remaining model errors. In this example the source of those errors will be investigated to give additional practice with the reference graph.

Begin by recalling the primary data which had been exchanged with the reference graph by pressing **[F4]** and selecting **[3]** for recall. The following should be noted:

The reference graph still contains the primary data as can be seen by drawing it with [F4].

The t scale has been replaced with $i\Delta t$ where i is the index of each datum and Δt is the value most recently entered by the user in the Data Definition Menu. The default value of Δt is saved in the *EXAM* parameter file *(PAR).EXM*. In the present example, this is precisely the t scale and so the t scale is still valid for the reference graph displayed.

The data title is now *Reference Graph*. If the top line happens to be a help prompt, pressing **[Enter]** will display the data title on the top line.

The X values at the marker positions are unchanged demonstrating that the X scale has been retained in the reference graph.

To continue the discussion of model errors, the most recent least squares fit will be compared with a more accurate model. The model errors that still remain arise from two sources:

The temperature dependence of the heat capacity which has been ignored by setting $\Delta C'_p = 0$

The separation of the calculation of the baselines from the calculation of the transition

parameters. There may be a slight contribution to the baselines from the tails of the transition giving rise to slightly erroneous baseline parameters.

The most recent least squares fit is remembered by the program as can be seen by pressing [Ctrl-F7] and comparing the values of the parameters with those in Figure 8. Save the fitting function as a reference graph with [Alt-F4]. As a precaution to avoid possible confusion, the graph of the fitting function must be displayed ([F8]) before it can be saved. The graph of the fitting function is not drawn without further prompting however. The program recognizes that a reference graph is now displayed whereas the most recent least squares fit was performed with the previous data. Both happen to be the same, but the program does not keep track of this. Hence, when [F8] is pressed, a top line prompt is issued:

```
Draw line generated from previous data set? Y/N [N] _
```

Select **[Y]**. Save the fitting function as the reference graph with **[Alt-F4]**, following prompts to save (rather than add to or subtract from) the calculated values of the function (rather than the baseline) corresponding to the data (rather than the display range).

Repeat the least squares fit by pressing [F7] followed by [0] which will use the values of the parameters from the most recent analysis as initial values. From the inspection of the residuals from the most recent analysis, the values of the parameters are likely to be good approximations to the true values. When the parameter menu appears, move the ««« indicator to $\Delta C'_p$ and press [Enter]. An arrow appears to the left of $\Delta C'_p$ indicating that $\Delta C'_p$ will be varied in the fit. Since the baseline parameters are to be varied, $\Delta C'_p$ cannot also be varied independently but must remain fixed or be calculated from $\Delta B'/N = (B'_b-B'_a)/N$. Press [Enter] again. Now $\Delta C'_p = \Delta B'/N$ indicates that $\Delta C'_p$ will be constrained to this value. Next set all the baseline parameters to vary by pressing [Enter] as the ««« indicator is moved to each one. When completed, pressing [F7] will initiate the least squares fit.

Not only are all the parameters being varied in the least squares fit, but all the data points are being fit as well. When only transition parameters, i.e., T_0 , ΔH , and/or N, are varied, only the data between the markers are used for the least squares fit. Therefore, each iteration takes longer than those of the previous fit.[†]

When the least squares fit converges and the fitting function and baseline have been drawn on the display, subtract the fitting function from the reference graph with [Alt-F4]. Now the reference graph contains the difference between two models: One in which the transition and baseline parameters are determined separately and $\Delta C'_p$ is held fixed at 0 and one in which all the parameters are varied simultaneously subject to the constraints $\Delta C_p = \Delta B/N$ and $\Delta C'_p = \Delta B'/N$. Examine the model difference within the context of the residuals. Press [F9] to display the residuals and [F4] to display the reference graph which contains the difference between the two models. Do not alter the vertical position of the reference graph. The resulting display should resemble Figure 11:

Two points should be noted. First the standard deviation of the fit for the more accurate model is almost unchanged, having decreased from 0.263 to 0.260. Secondly, the graph of the difference between the two models is very small compared to the random scatter in the data. Specifically, the root mean square of the model error is only 0.0273. (This can be calculated by exchanging the primary

^TThe time required for the least squares fit varies in proportion to the number of data being fit and in proportion to the square of the number of parameters being varied.



data and the reference graph with [F4], deactivating the markers with [Ctrl-F5], and fitting all the data to a horizontal line with [F7] [1] [0]. However, do not do this yet.) Thus, the model errors are an order of magnitude smaller than the random errors.



Pressing [Ctrl-F7] displays the values of the parameters with their standard deviations as shown in Figure 12. The standard deviations of the baseline parameters are also given because they were varied in the fit. The following points should be noted:

The values of the standard deviations of the parameters have increased slightly in spite of the fact that the standard deviation of the fit decreased slightly. This is a result of the correlation of errors among the parameters. The larger the number of parameters varied, the greater the opportunity for correlation. The effect of correlation is always to increase the standard deviations of the parameters compared to the standard deviations of a set of uncorrelated parameters.

The values of all the parameters now agree with their true values to within one standard deviation.

We have shown that even slight model errors result in parameter values that differ from their true values by more than would be statistically inferred. This is the reason why the standard deviations of the parameters are displayed only within the context of a display of the residuals. Even then, some caution must be exercised and differences in values of the parameters assuming different reasonable models might give a more realistic estimate of the uncertainties in the parameters.

4. $\triangle H$ from the two-state analysis differs slightly from $\triangle H_{vH}$ calculated from a van't Hoff analysis.

In a van't Hoff analysis the logarithm of the equilibrium constant K is plotted against 1/T where T is the absolute temperature. The slope of the plot is $-\Delta H_{vH}/R$ where R is the gas constant. The intercept is $\Delta H_{vH}/RT_0$ where T_0 is the midpoint of the transition defined by $\alpha = 1/2$. ΔH_{vH} is the so-called van't Hoff enthalpy and should be the same as ΔH obtained from the two-state analysis of the DSC profile. The equilibrium constant is calculated from the extent of the transition α by eq. (6a) on p.4: $K/K_0 = (2\alpha)^n/2(1-\alpha)$ or, for $nA \rightarrow B$ transitions, by eq. (6b), $K/K_0 = (2\alpha)/2^n(1-\alpha)^n$. The extent of the transition α is taken to be the fractional completion of the transition, i.e., the normalized integral of the DSC profile. A van't Hoff analysis can be performed using transformations provided in the Data Transformation Menu. In this example, we will perform such an analysis and compare ΔH_{vH} with ΔH obtained from the two state analysis.

The data to be analyzed are found in the file *TEST2A.DAT* and were simulated using the following parameters:

 $\begin{array}{lll} A \to 2B \mbox{ transition, n=2} \\ T_0 &= 50^{\circ} C \\ \Delta H &= 600 \mbox{ kJ mol}^{-1} \mbox{ (of A)} \\ N &= 1.0 \mbox{ \mumol of A} \mbox{ (2 } \mbox{ \mumoles of B when fully denatured)} \\ \Delta C_p &= 3 \mbox{ kJ } K^{-1} \mbox{ mol}^{-1} \mbox{ (of A)} \\ \Delta C'_p &= -0.15 \mbox{ kJ } K^{-2} \mbox{ mol}^{-1} \mbox{ (of A)} \\ B_a &= 10 \mbox{ kJ } K^{-1} \\ B'_a &= 0.15 \mbox{ kJ } K^{-2} \\ B_b &= 13 \mbox{ kJ } K^{-1} \\ B'_b &= 0 \mbox{ kJ } K^{-2} \end{array}$

The time constant is 100 s, and random errors have been added to the calculated values.

After the data file has been read and the Data Definition Menu displayed, ensure that the time scale is defined as $t=\Sigma\Delta t \Delta t=10.0$. From the Main Menu, select [4] transform data and, from the Data Transformation Menu, perform the Tian transformation using a time constant of 100 and a 7 point cubic fit for calculating the derivatives. Divide by the scan rate. Return to the Main Menu and when prompted with:

> Maximum value of Y is greater than 10,000. Do you wish to rescale Y values by .001? Y/N [N]

select Y. The values of Y will then be in mJ/° so that the value of N returned by the two state analysis will be in μ mole. (This can be easily deduced by noting that the integral of Y, which will be in mJ, is equal to N Δ H and Δ H is in kJ/mole. Thus, mJ/(kJ/mole) = μ mole.)

Perform a two-state analysis of the data with markers at X=26.24 (5 [Ctrl \rightarrow] moves from the left margin) and X=73.74 (5 [Ctrl \leftarrow] moves from the right margin), setting n=2, varying all parameters simultaneously, and forcing $\Delta C_p = \Delta B/N$ and $\Delta C'_p = \Delta B/N$. Notice that the iterations are somewhat slower than for n=1 because of the iterative calculation of α as a function of T.

The values of the parameters returned by the least squares fit appear in Figure 13. Notice that T_0 , indicated by the dotted vertical line in Figure 13, is not at the peak of the transition. T_0 is defined as the midpoint of the transition, the temperature at which $\alpha=1/2$. Try to verify this by measuring the area between the baseline and the data on the low and high temperature sides of T_0 .



Measure the area between X=10.0, the lowest temperature displayed and X=T₀ by first erasing the markers with [Ctrl-F5], placing a marker at X=T₀=50.0, the cursor at X=10.0 (notice that hitting [Home] moves the cursor from X=50.0 to the first point on the display which happens to be X=10.0), and [Shift-F6] to calculate the area. As shown in Figure 13, the value obtained is 293.93 kJ, but the value from N Δ H/2 is 299.99 kJ. Why the difference?

Referring to eq. (1) on p. 1, notice that the integral of $\Delta H(\partial N_a/\partial T)_p dT = J\Delta H dN_a = N\Delta H$ only if ΔH is not a function of temperature. But, because ΔC_p and $\Delta C'_p$ are non-vanishing, ΔH is a function of temperature. Because $\Delta C_p > 0$ and because ΔC_p is the most significant contribution to the temperature dependence of ΔH , we would expect from $\Delta H = \Delta H_0 + \Delta C_{p0}(T-T_0) + \frac{1}{2}\Delta C'_{p0}(T-T_0)^2$ that the half area below T_0 would be less than $N\Delta H/2$ which is what is observed. Moreover, because the contribution from ΔC_p above T_0 is opposite in sign, we would expect the difference between the area above T_0 and $N\Delta H/2$ to compensate for the error below T_0 . If we make the measurement above T_0 we obtain 303.53 kJ and the total area is thus 597.46. This is close to the value of $N\Delta H = 599.98$ kJ but not exact because the contribution from $\Delta C'_p$ is everywhere negative. We anticipate from these arguments that the van't Hoff analysis will present us with similarly small but significant errors.

The van't Hoff analysis is performed as follows:

Clear the markers with [Ctrl-F5] and move cursor to $X_0 = 50.0$. This position will be needed as a reference later on. Integrate the data numerically above the baseline by pressing [Shift-F2] which results in Figure 14. Replace the primary data with the integral with [Alt-F1] which first displays the warning prompt:



[Y] clears the display and redraws the integral with the Y axis legend /YdX replaced with Y vs X. [Ctrl-F1] displays the Transformation Menu. Since Y is now the integral of dQ/dT, dividing it by the total area will give α the fraction of transition. This is done by selecting 7: Y \rightarrow s(Y+ Δ Y) and, in response to the prompts:



accept the defaults. The list of transformations performed on the bottom of the display should now read: Tian Y/(dX/dt) Y/1000 $s(Y + \Delta Y)$ and on the top line the message **NOTE:** Y has been **normalized** appears. Y is now α . The equilibrium constant is calculated from α by selecting 6: $Y \rightarrow (2Y)^n/2(1-Y)$, and in response to the prompt:

```
Replace Y, the extent of reaction, with Keq, the equilibrium constant,

For A \rightarrow nB, Keq = (2Y)^n/2(1-Y), enter a positive value for n.

For nA \rightarrow B, Keq = 2Y/(2(1-Y))^n, enter a negative value for n.

Enter value for n [1]: _
```

enter [2]. The entry is confirmed by the message on the top line: NOTE: Y replaced with $(2Y)^n/2(1-Y)$ with n = 2. Y is now the equilibrium constant K. Next select 10: $Y \rightarrow Ln(Y)$ which

replaces the equilibrium constant with its logarithm. This completes the transformations required for the Y coordinate. Note that if a mistake is made while performing the transformations, pressing **[Esc]** will undo those transformation performed since entering the Transformation Menu.

X is transformed from °C to K by adding 273.15 with transformation 13: $X \rightarrow s(X+\Delta X)$ by entering 273.15 in response to the prompt for ΔX and accepting the default value of 1.0 for s. Selecting 15: $X \rightarrow 1/X$ then gives us the reciprocal of the temperature which is the required quantity for a van't Hoff analysis. Note that when 15: $X \rightarrow 1/X$ is selected, the prompt

Maximum value of X is less than .01 Do you wish to rescale X values by 1000? Y/N [N] _

is offered. This allows the values of X to be less cumbersome, but confuses the dimensionality. Decline the option. A warning message **NOTE:** Ordering of points with respect to N has been reversed! is printed on the top line. Since the data had been in order of increasing T, the ordering of 1/T must be reversed in order that X be increasing from left to right in the display. Hence the left side of the graph now corresponds to high temperature and the right side to low temperature. The list of transformations performed appearing on the bottom line should now read: Tian Y/(dX/dt) $Y/1000 \ s(Y + \Delta Y) \ (2Y)^n/2(1-Y) \ Ln(Y) \ s(X + \Delta X) \ 1/X$

[Enter] completes the transformations and displays the transformed data as shown in Figure 15.



The van't Hoff analysis is performed by measuring the slope and intercept of $\ln(K)$ vs 1/T. We note in Figure 15 that a straight line is not obtained. This is because at the extremes of the transition, random error dominates the value of $\ln(K)$. Even if the data were noise free, the extrema would be poorly defined because of round-off errors. Thus, the slope and intercept must be determined by a linear least squares fit of the data in the vicinity of $1/T_0$ where $\ln(K)$ is well determined. The cursor position still corresponds to $X = T_0 = 50^\circ$. Place markers as shown in Figure 15 (with one [Ctrl \leftarrow] to position the lower marker followed by three [Ctrl \rightarrow]s to place the upper marker). To perform the linear fit, press [F7] followed by [1] when prompted for the function and [1] again when prompted for the order of the polynomial. The results of the fit are shown in Figure 15. From C0 and C1 we calculate $T_0 = (-C1/C0)-273.15 = 50.06$ which agrees reasonably well with the value of 50.00 from the

90

two-state analysis. The value of ΔH_{vH} = -RC1 (where R=.000831451 kJ/mol^o) = 601.75 ± 0.31 kJ/mol. We note that the van't Hoff enthalpy differs from the two-state analysis value of ΔH by six standard deviations suggesting a model error. The model error arises, as we have stated, from the temperature dependence of ΔH .

This can be shown by repeating the calculation for the data in file *TEST2B.DAT* which is a simulation identical to *TEST2A.DAT* except that ΔC_p and $\Delta C'_p$ are both identically 0 so that ΔH has no temperature dependence. On repeating the preceding analysis and measuring the area of the transition profile below T_0 the value 299.93 kJ is obtained and is exactly N $\Delta H/2$ as shown in Figure 16.





Repeating the van't Hoff analysis outlined above and as shown in Figure 17, gives the value for ΔH_{vH} = -RC1 = 599.42 ± 0.0167 kJ/mol which is in closer agreement with ΔH than when ΔH was a function of temperature. However, now ΔH_{vH} differs from its true value of 600.0 kJ/mol by 35 standard deviations! What's going on here? Again we have a model error but now the error is not in the thermodynamic model but in the *assumptions underlying the distribution of errors in the van't Hoff analysis*. The linear least squares fit in *EXAM* assumes the same distribution of errors for each datum and therefore assigns a unit weight to each value. In fact, the errors are far from uniform in ln(K) as well as in 1/T. As noted earlier in this example and as can be see in Figures 15 and 17, the errors become much more pronounced far from the midpoint of the transition interval. This is true even for the narrow range over which the linear least squares fit was performed. Therefore, perform the linear least squares fit again, this time taking only the range of five points above 1/T₀ to five points below, i.e., from 1/T = 0.003087 to 1/T = 0.003099, where the assumption of the uniformity of errors over the data range is not quite so incorrect.

This time we obtain $C1 = -72156.1 \pm 41$ corresponding to $\Delta H_{vH} = 599.94 \pm 0.34$ kJ/mol which is in excellent agreement with the true value of 600. One might well ask if this was not also the source of error in the van't Hoff analysis of the data in *TEST2A.DAT*. When the same narrow range 1/T = 0.003087 to 0.003099 is chosen for the linear least squares fit of the van't Hoff plot, a value of $\Delta H_{vH} = 606.03 \pm 0.43$ kJ/mol is obtained, which is in error by 14 standard deviations due, as before, to the temperature dependence of ΔH . Thus, ΔH_{vH} is not quite equal to ΔH because of the erroneous assumption of the temperature independence of ΔH which underlies the van't Hoff analysis.



5. The analysis of overlapping profiles.

Data consisting of overlapping profiles can be analyzed in an iterative manner by analyzing portions of one component, subtracting the calculated value from the data leaving the remaining component. The second component is then analyzed and *its* function is subtracted from the data leaving the first component. The first component is then analyzed a second time, but this time free from the influence of the second component. This should give a better estimate of the first function. The process is repeated until there is no further change in the values of the parameters. The procedure is tedious and judgement must be exercized on how to treat the baseline and the heat capacity change.

In the example to be discussed, three overlapping profiles will be analyzed using [Alt-F3] to save fitting functions and [Ctrl-F3] to recall them. The data to be analyzed are found in the file *TEST3.DAT* and were simulated using the following parameters:

\mathbf{T}_0	= 40	50	60	°C
ΔH	= 600	400	200	kJ mol ⁻¹
N	= 1.0	1.0	1.0	µmol
$\Delta C_{\rm p}$	= 3.0	2.0	1.0	$kJ \ K^{\text{-1}} \text{mol}^{\text{-1}}$
ΔC_{p}^{\prime}	= 0.0	0.0	0.0	kJ K ⁻² mol ⁻¹
B_a	= 10 kJ K	-1		
B'_a	= 0.15 kJ	K^{-2}		
\mathbf{B}_{b}	= 13 kJ K	-1		
$\rm B_b'$	= 0.15 kJ	K^{-2}		

For all three profiles the stoichiometry number n = 1, i.e., $A \rightarrow B$ transitions. As in the previous examples, the time constant is 100 seconds, and random errors have been added to the calculated data.

After the data file has been read and the Data Definition Menu is displayed, ensure that the time scale is defined as $t=\Sigma\Delta t \Delta t=10.0$. Proceed as in the previous examples. From the Main Menu, select [4] transform data and, from the Data Transformation Menu, perform the Tian transformation using a time constant of 100 and a 7 point cubic fit for calculating the derivatives. Divide by the scan rate

with $Y \rightarrow Y/(dX/dt)$. Return to the Main Menu and, when prompted with:

Maximum value of Y is greater than 10,000. Do you wish to rescale Y values by .001? Y/N [N]

select Y. The values of Y will be in mJ/° and the value of N returned by the two state analysis will be in µmole. (This can be determined by noting that the integral of Y, which will be in mJ, is equal to N Δ H. Δ H is always in kJ/mole. Thus, mJ/(kJ/mole) = µmole.)

Display the data by pressing [Enter] twice from the main menu. The transformed data may be saved as a new file by pressing [Shift-F3] and following the prompts, taking care to give the new file a new name. Saving the transformed data as a new file allows the resumption of work on the transformed data at some future time.

Begin by analyzing the data on the low temperature half of the low temperature profile. Pressing the [\uparrow] cursor key moves the data cursor to the peak maximum. Place a marker at this point with [F5]. Press [End] to move the cursor to the last point and [Ctrl←] to move it to the left to X=86.75°. Place a marker at this point with [F5]. Deactivate all the points between the two markers with [Del]. Reactivate the marker at X=86.75° with [F5]. Move the cursor to the first data point with [Home] and then to X=19.75 with three presses of [Ctrl→]. Place a marker at this point with [F5].

By deactivating all points above the midpoint of the first profile the contributions of the higher temperature profiles to the analysis of the first profile will be minimized. Start the least squares analysis with [F7] and [3] in response to the prompt for a least squares fitting function. Do not accept the default analysis options. Instead, move the ««« marker to the entry for B_a and press [Enter]. An arrow appears to the left of B_a (and, if allowed by the computer's graphics display, the parameter and its value are highlighted). This indicates that B_a is to be varied in the least squares analysis along with the transition parameters. In a similar fashion allow B'_a and B_b to vary. B'_b on the other hand will be constrained to be $B'_a + N\Delta C'_p$ (= B'_a because $\Delta C'_p = 0$) by hitting the [Enter] twice for B'_b . By this arrangement, a baseline is established throughout the region and the pre- and post-transition slopes are the same. When the time comes to analyze the remaining profiles, their corresponding baselines will be assumed to be zero.

Begin the least squares fit with [F7]. Following the fitting procedure, the display should resemble Figure 18.

Save the results of this first analysis of the low temperature profile with [Alt-F3]. The top line message:

Saved as $F(1) = A \rightarrow B$ DSC, $\sigma = 0.271$ [Ctrl-F3] to Recall

indicates that the function has been saved as reference function (1).

Reactivate all the data between the markers by pressing [Insert]. [Alt-F9] subtracts the function from the data and displays the residuals. In addition, all data whose *calculated* values fall more than two standard deviations away from the baseline are *deactivated*. The deactivated points are therefore those for which the low temperature profile contributes significantly to the overall profile. Notice that most of the points in the middle profile are deactivated indicating that it will be difficult



to obtain a good estimate of the middle profile because it is influenced by the the two flanking profiles.

In order to have enough data to analyze, reactivate the data between X=47.25° and 50.49° by placing markers at these values and pressing [Insert]. Although data will be analyzed in a region where the low temperature profile contributes, this low profile has been subtracted from the data in forming the residuals. The contributions from the low temperature profile are therefore minimized but not negligible. Next deactivate all the data between 51.37° and 86.75° by placing markers at these values and pressing [Delete]. Finally, place markers at 19.75° and 86.75° and initiate the least squares fit of the few active data between these markers to a two-state DSC model with [F7] [3].

Before initiating the least squares fit, set all baseline parameters B_a , B'_a , B_b , and B'_b to 0 since the analysis just performed of the low temperature profile should have fit the pre- and post-transition baselines exactly. Of course this will introduce a model error into the calculation because we have assumed that the cumulative heat capacity changes of all the component profiles with their effects on the baseline have been represented by a single heat capacity change in the low temperature profile. Since this profile is the largest, this should not be a bad approximation. Later, an alternative approach will be examined. As with the fit of the low temperature profile, constrain $\Delta C_p = \Delta B/N$ (=0). Begin the least squares fit with [F7]. Following the fit, the display should resemble Figure 19.

Save the fitting function as Function 2, F(2), with [Alt-F3] followed by [2] in response to the prompt.

Ref. Function: [1]»Replace [2]»Save [0]»Delete

Note that the prompt reflects the fact that a reference function already exists and a choice is offered of saving the function as F(2) or replacing F(1). Select [2] and the fitting function will be saved as $F(2) \land A \rightarrow B DSC, \sigma = 0.277$

Press [Insert] to reactivate all the data between the markers. Recalling that the display represents the data minus the contribution from F(1), press [F9] to subtract F(2) also. The resulting residuals, which are really residuals of residuals, will be the result of subtracting both the low and middle profiles from the data and should resemble the data in Figure 20.



Figure 19





Deactivate data between X=19.75° and X=60.99° (the latter being the peak of the profile and reached with $[\uparrow]$) thus minimizing residual contributions from the two low temperature profiles. With markers at X=19.75° and X=86.75°, perform the same analysis of the high temperature profile as was performed with middle profile by pressing [F7] three times in succession, the second to repeat the previous analysis and the third to initiate the least squares fit. Save the fitting function as Function 3, F(3), with [Alt-F3] followed by [3] in response to the prompt:

Note that either Function 1 or Function 2 could have been replaced or all functions deleted. The top line message should read Saved as $F(3) \land A \rightarrow B DSC$, $\sigma=0.247$ [Ctrl-F3] to Recall. Next inspect the residuals with [F9] which should generate a display resembling Figure 21.

Save this display as a reference graph with [Ctrl-F4]. As can be seen from the residuals, pronounced



systematic errors remain. These are most likely due to the contributions of profiles to the analysis of their neighbors although the residuals could have arisen from the assumption that all the heat capacity change in the baseline belonged to the low temperature component of the profile. Under the assumption that the pronounced systematic errors arise from the former, proceed to minimize the contributions of profiles to their neighbors by redoing the analysis of all profiles, this time subtracting the contributions of all others.

The currently active function is the function for the high temperature profile. Pressing [Ctrl-F9] subtracts this function from the primary data which consists of the three overlapping profiles. The resulting display is a residuals display in which the currently active function, F(3), has been subtracted from the data. The display therefore represents only the two lower temperature profiles. Recall F(2), the middle profile, with [Ctrl-F3] and in response to the prompt:

select [2]. The values of the parameters of F(2) are printed on the screen. Pressing [F8] will draw this profile on the display to confirm that the correct function has been recalled. Pressing [F9] subtracts the function from the display and the resulting residuals of residuals display, which should resemble the data in Figure 22, will represent the low temperature profile alone. The slight dip on the high temperature side of the profile suggests that the F(2) and F(3) were not accurate and is a reflection of the residuals in Figure 21. This can be verified by drawing the reference graph on the display with [F4] which shows a dip at the same temperature of similar magnitude.

Next perform the same analysis as was originally performed on the low temperature profile including deactivation of all points between the peak at 40.12° and 83.50°. The points are again deactivated because there remains some influence from the lower two profiles due to the errors in their calculation. Allow the three baseline parameters B_a , B'_a , and B_b to be varied in the fit and constrain B'_b to be $B'_a + N\Delta C'_p$ (= B'_a because $\Delta C'_p$ = 0). The resulting analysis should resemble Figure 22.

The current fitting function should be a better representation of the low temperature profile than F(1) because the contributions to this profile from the higher temperature profiles have been reduced. Therefore, replace F(1) with the current fitting function with [Alt-F3] followed by [1] in response to



the prompt. A second prompt will be issued for confirmation to prevent accidental replacement.

Next analyze the middle component after subtracting the contributions of the low and high temperature components. Press [Insert] to restore all deactivated data to their active state. Press [Ctrl-F9] to subtract the current fitting function, improved F(1), the low temperature profile, from the primary data. Recall the high temperature function, F(3), with [Ctrl-F3] followed by [3] in response to the prompt. Subtract this function from the display with [F9]. The resulting display, residuals of residuals, will represent the middle profile and should resemble Figure 23. Deactivate the data between 19.75° and 47.25° and between 53.74° and 80.24° to minimize the contributions of *residuals* of the low and high temperature components to the middle component. Place markers at 19.75° and 80.24°. As with the previous analysis of the middle component, initiate the least squares fit with [F7] [3] and set all baseline parameters B_a , B'_a , B_b , and B'_b to 0 since the fit of the low temperature profile should have fit the pre- and post-transition baselines exactly. The current fitting function should be a better representation of the middle component than F(2) because the contributions to this component from the other two components have been reduced. This can be seen by comparing Figure 23 with Figure 19. Therefore, replace F(2) with the current fitting function with [Alt-F3] followed by [2] in response to the prompt. Again, a confirmatory prompt will be issued.

With improved fitting functions for the lower and middle components, finish this round with a reanalysis of the high temperature component. Press [Insert] to restore all deactivated data to their active state. Press [Ctrl-F9] to subtract the current fitting function, improved F(2), the middle component, from the primary data. Recall the low temperature function, F(1), with [Ctrl-F3] followed by [1] in response to the prompt. Subtract this function from the display with [F9]. The resulting display, residuals of residuals, is the high temperature component of Figure 24. Comparison with Figure 20 shows that the low temperature side of the profile appears less perturbed by the residuals of the two low temperature components.

Deactivate data between X=19.75° and X=60.99° thus minimizing contributions from the residuals of the two low temperature components. With markers at X=19.75° and X=86.75°, perform the same analysis of the high temperature profile as was performed with the middle profile by pressing [F7] three times in succession, the second to repeat the previous analysis and the third to initiate the least squares fit. The result of the analysis should resemble the parameter table in Figure 20. The current fitting function should be a better representation of the high temperature component than F(3) because



the contributions to this component from the other two components have been reduced. Therefore, replace F(3) with the current fitting function with [Alt-F3] followed by [3] in response to the prompt.

Now compare the residuals at the end of round two with those from the initial round of analyses. **[F9]** will display the residuals (actually residuals of residuals of residuals) in which the contributions from all three fitting functions have been removed. The residuals from the first round were saved as a reference graph. To compare the current residuals with the reference graph return to the display of the full data with **[F1]**. Now *exchange* these data with the data in the reference graph with **[Ctrl-F4]** and, in response to the prompt:

[1]»Replace [2]»Exchange [3]»Recall [4]»Add [5]»Subtr [6]»Ratio [1]

select [2]. Note, if the default [1] is selected, the reference graph will be replaced by the current [F1] display. If [3] is selected, the display will be replaced by the reference graph and the principal data
can then only be retrieved by rereading the data file. After [2] is selected, the reference graph will contain the original data and the residuals from the first round of analyses will be displayed. Superimpose on this display the graph of the residuals from the second round of analyses with [Shift-F9]. Because the data have been replaced, the residuals may not be germane to the display and a prompt is issued:

```
Recalculate residuals for the transformed data? Y/N [N]
```

Select [N] since we want to compare the existing residuals with the ones we stored in the reference graph and which are now displayed. A prompt is then issued to position the residuals plot:

```
Select vertical position for display of graph? Y/N [N]
```

Accept the default [N].





The resulting graph should resemble Figure 25. Several points are worth noting. First of all, the residuals are reduced from those of the first round by perhaps a factor of 5. The scatter in the wings is the same. A noticeable trend is still visible in the residuals from 42° to 58°. The center component is the most poorly determined judging from the standard deviations of the fits of the three components in the second round of analyses. The standard deviation of the fit is not much worse than the standard deviation obtained in the first *example* when nearly all sources of model error had been removed (Figure 8). Consequently, the likelihood of obtaining independent values of ΔC_p for the three components is not high. Indeed when an attempt is made to vary ΔC_p for each component, the procedure diverges.

One could continue along these lines, refitting each of the three components after subtracting improved values for the remaining components. However, little improvement is noted after the second round. Instead, start from the beginning with a different assumption about the heat capacity change associated with each profile. Keep in mind that the change in heat capacity affects not only the transition profile, but the baseline as well.

Assume instead that the value of ΔC_p associated with each component is proportional to the value of ΔH for that component. There is little basis for this assumption other than the bonding in different proteins or oligonucleotides and the number of available states will be roughly proportional to the number of residues or units in the macromolecule. Taking the total heat capacity change of 6 kJ/mol, the proportional values for ΔC_p for the three components will be roughly 3, 2, and 1 which are indeed just the values used in the simulation. Thus, this "fortunate assumption" should give a best case example of how well the overlapping profiles can be analyzed.

Return to the primary data with [F1]. Deactivate all the data between the peak at 40.12° and the end at 89.99°. Place the low temperature marker at 19.75° and the high temperature marker at the uppermost point, 89.99°. Initiate the least squares fit with [F7] [3]. Set ΔC_p to 3.0 by moving the parameter marker to ΔC_p and entering the number 3 from the keyboard. Allow B_a and B'_a to vary and constrain B_b to be $B_a + N\Delta C_p$ and B'_b to be $B'_a + N\Delta C'_p$ (= B'_a because $\Delta C'_p$ = 0). Note that no high temperature points will be used in this analysis because on this first round the high temperature baseline contains contributions from ΔC_p for all three components. The high temperature baseline will be generated from $B_a + N\Delta C_p$. Perform the fit with [F7] and, following the fit, save the parameters as F(1) with [Alt-F3] [1]. The display should resemble Figure 26.



Form the residuals with **[F9]**. Press **[Insert]** to reactivate all deactivated data. Deactivate data from 19.75° to 47.25° and from 53.74° to the end. Place markers at 19.75° and the uppermost point, 89.99°. Perform the two-state analysis setting ΔC_p to 2, B_a and B'_a to 0, and constraining B_b to be $B_a + N\Delta C_p$ and B'_b to be $B'_a + N\Delta C'_p$. The result of the fit should resemble Figure 27. With **[Alt-F3] [2]**, save the fitting function as F(2).

Form the residuals with **[F9]**, reactivate all data with **[Insert]**, deactivate all data from 19.75° to 61.62°, the latter being the temperature of the maximum. Place markers at 19.75° and at the uppermost point, 89.99°. Initiate the analysis pressing **[F7]** twice to use the same conditions as invoked with the middle component. Set $\Delta C_p = 1$ and continue with **[F7]**. Save the fitting function as F(3) with **[Alt-F3] [3]**. The result should resemble Figure 28.

This completes the first round of analyses and the reference functions are the first approximations. Each is an approximation because each of the components distorts the shapes of its neighbors. They can be improved by repeating the analyses, each time subtracting the calculated profiles of the two





components not being analyzed. Display the residuals with **[F9]**. As can be seen in Figure 29, the residuals show a pronounced systematic error, perhaps larger than that encountered when the analysis of these overlapping profiles was first performed with all of the baseline change in the low temperature profile. Save the residuals for future comparison as a reference graph with **[Ctrl-F4]** [1].

Repeat the analyses of the three profiles subtracting from each the functions of the other two before performing the analysis. Begin with [Ctrl-F9] which subtracts the current fitting function, F(3), from the primary data. Follow with [Ctrl-F3] [2] [F9] to recall F(2) and subtract it from the display. The resulting residuals of residuals will be the low temperature component with the middle and upper components subtracted (to within the accuracy of their corresponding fitting functions). Repeat the analysis of the low temperature profile holding $\Delta C_p = 3$, allowing B_a and B'_a to vary and constraining B_b to be $B_a + N\Delta C_p$ and B'_b to be $B'_a + N\Delta C'_p$ (= B'_a because $\Delta C'_p = 0$). Similarly, repeat for the middle and high temperature profiles using the same conditions as in round 1. At the end of round 2, display the residuals with [F9].



In comparison to the first analysis in which all the heat capacity change in the baseline was subsumed into the low temperature profile, the improvement in the systematic error in the baseline is not as dramatic, even though the assumed values for the change in heat capacity for each component are correct. Partly, this is because fewer data are being used in the analysis of each component. As in the first analysis, the residuals from round 2 can be compared with those from round 1 by displaying the primary data with [F1], exchanging them with the reference data with [Ctrl-F4] [2], and drawing the residuals from round 2 with [Shift-F9]. The display should resemble Figure 30.

Exchange the reference graph with the display to restore the primary data with **[Ctrl-F4] [2]**. Perform round 3, repeating the analysis but with more points included in the least squares fits than in rounds 1 and 2. For F(1), the low temperature component, deactivate data from 40.12 to 86.75 (rather than to the end as in rounds 1 and 2). For F(2), the middle component, deactivate data from 19.75 to 47.25 and from 53.74 to 86.75. For F(3), the high temperature component, deactivate data from 19.75 to 60.99 (the peak maximum for this component). Do round 4, repeating the conditions of round 3. Finally, do 2 more rounds but fitting all the data for all the components. Now inspect the



Figure 30

residuals. Compare these residuals with those from round 1 with [F1] [Ctrl-F4] [2] [Shift-F9]. The display should resemble Figure 31.



Comparison of Figure 31 with Figure 25 shows that trend observed in the residuals of the first model in which all the heat capacity change was contained in the low temperature component is now missing in the model in which the heat capacity change is distributed among all three components. This is not surprising since we know a priori that the latter model is indeed correct. However, it is somewhat surprising that the correct model converges so slowly. The slow convergence may be due to the following. The contribution from overlapping profiles cause a significant shift in the determined value of N. N from the fit of the first, that is the low temperature profile, gives an inaccurate, extrapolated high temperature baseline from the constraint $B_b = B_a + N\Delta C_p$. This baseline proves to be not as accurate as assuming all of the baseline change is in the first profile, but forcing it through the high temperature data. This, plus the residual contribution of F(1) to the middle profile gives an even more inaccurate value of N than that obtained during the first round of the first analysis (1.244 in Figure 27 compared with 1.057 in Figure 19). This is magnified in the error in the baseline for the high temperature profile and an even more inaccurate value for N (0.5697 in Figure 28 compared with 0.8342 in Figure 20). Eventually the procedure converges, but the poor first round causes the convergence to be rather slow.

If we had attempted to determine the true values of ΔC_p for each of the components by invoking the constraint $\Delta C_p = \Delta B/N$ and fitting only the lower half of the transition profile, the iterative process would begin to **diverge**!

Perhaps a global fit of several components to many parameters would converge more definitely, but such an analysis is currently beyond the memory capacities of personal computers. In the meantime care should be taken not to overinterpret the values of the parameters obtained from the analysis of overlapping profiles. With this caveat in mind, the analysis just described can be used to obtain useful information on the thermodynamic functions and stoichiometries of complex profiles.

NIST_{Technical} Publications

Periodical

Journal of Research of the National Institute of Standards and Technology – Reports NIST research and development in those disciplines of the physical and engineering sciences in which the Institute is active. These include physics, chemistry, engineering, mathematics, and computer sciences. Papers cover a broad range of subjects, with major emphasis on measurement methodology and the basic technology underlying standardization. Also included from time to time are survey articles on topics closely related to the Institute's technical and scientific programs. Issued six times a year.

Nonperiodicals

Monographs – Major contributions to the technical literature on various subjects related to the Institute's scientific and technical activities.

Handbooks – Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Special Publications – Include proceedings of conferences sponsored by NIST, NIST annual reports, and other special publications appropriate to this grouping such as wall charts, pocket cards, and bibliographies.

Applied Mathematics Series – Mathematical tables, manuals, and studies of special interest to physicists, engineers, chemists, biologists, mathematicians, computer programmers, and others engaged in scientific and technical work.

National Standard Reference Data Series – Provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated. Developed under a worldwide program coordinated by NIST under the authority of the National Standard Data Act (Public Law 90-396). NOTE: The Journal of Physical and Chemical Reference Data (JPCRD) is published bi-monthly for NIST by the American Chemical Society (ACS) and the American Institute of Physics (AIP). Subscriptions, reprints, and supplements are available from ACS, 1155 Sixteenth St., NW., Washington, DC 20056.

Building Science Series—Disseminates technical information developed at the Institute on building materials, components, systems, and whole structures. The series presents research results, test methods, and performance criteria related to the structural and environmental functions and the durability and safety characteristics of building elements and systems.

Technical Notes – Studies or reports which are complete in themselves but restrictive in their treatment of a subject. Analogous to monographs but not so comprehensive in scope or definitive in treatment of the subject area. Often serve as a vehicle for final reports of work performed at NIST under the sponsorship of other government agencies.

Voluntary Product Standards – Developed under procedures published by the Department of Commerce in Part 10, Title 15, of the Code of Federal Regulations. The standards establish nationally recognized requirements for products, and provide all concerned interests with a basis for common understanding of the characteristics of the products. NIST administers this program in support of the efforts of private-sector standardizing organizations.

Consumer Information Series – Practical information, based on NIST research and experience, covering areas of interest to the consumer. Easily understandable language and illustrations provide useful background knowledge for shopping in today's technological marketplace. *Order the above NIST publications from: Superintendent of Documents, Government Printing Office, Washington, DC 20402.*

Order the following NIST publications—FIPS and NISTIRs—from the National Technical Information Service, Springfield, VA 22161.

Federal Information Processing Standards Publications (FIPS PUB) – Publications in this series collectively constitute the Federal Information Processing Standards Register. The Register serves as the official source of information in the Federal Government regarding standards issued by NIST pursuant to the Federal Property and Administrative Services Act of 1949 as amended, Public Law 89-306 (79 Stat. 1127), and as implemented by Executive Order 11717 (38 FR 12315, dated May 11, 1973) and Part 6 of Title 15 CFR (Code of Federal Regulations). **NIST Interagency Reports (NISTIR)** – A special series of interim or final reports on work performed by NIST for outside sponsors (both government and non-government). In general, initial distribution is handled by the sponsor; public distribution is by the National Technical Information Service, Springfield, VA 22161, in paper copy or microfiche form.

U.S. Department of Commerce National Institute of Standards and Technology Gaithersburg, MD 20899

Official Business Penalty for Private Use \$300 .