*Journal of Thermal Analysis, Vol. 39 (1993) 2779-2790*

# THESEUS: A NEW SOFTWARE PACKAGE FOR THE HANDLING AND ANALYSIS OF THERMAL DENATURATION DATA OF BIOLOGICAL MACROMOLECULES

*G. Barone, P. Del Vecchio, D. Fessas, C. Giancola and G. Graziano*

DEPARTMENT OF CHEMISTRY, UNIVERSITY 'FEDERICO II' VIA MEZZOCANNONE 4, 80134 NAPLES, ITALY

A new software package (THESEUS) has been assembled for the analysis of the DSC data, concerning the thermal denaturation of biological macromolecules. The system is useful to obtain accurate physico-chemical information, bypassing the casual and systematic errors, very common in these experiments. It can also be used for handling data from other instruments and methodologies giving thermodynamic, spectroscopic or other kind of data as a function of temperature. Because many of the researches in this field are of exploratory nature and continuously new unfolding mechanisms are described or hypothesized in the current literature, we have written and assembled this powerful and flexible program of general applicability, in order to put the operator in a position to control each step of the calculation procedure and use his own experience for choosing the better way to solve unexpected problems.

**Keywords:** denaturation, new software package

## Introduction

Denaturation studies on proteins and other natural macromolecules are of wide interest, both for basic molecular biology (as they allow to get an insight into the complex problem of the folding) [1–9] and for modern biotechnologies [10, 11]. In particular, the detailed knowledge of the limits of resistance of enzymes to the turnover in extreme non-physiological conditions (high temperatures, extreme ranges of *pH* and/or ionic strength, high concentration of organic solvents or presence of immobilizing matrixes) is an important objective [12].

Differential scanning microcalorimetry is a powerful methodology to investigate thermal denaturation, as it allows to measure directly the changes of enthalpy associated with the conformational transitions. In fact, all the thermodynamic quantities related to the denaturation process can be, in principle, deduced from

*John Wiley & Sons, Limited, Chichester*
*Akadémiai Kiadó, Budapest*

the 'excess average enthalpy' $<\Delta H(T)>$, referred to a particular thermodynamic state (the 'native' state f. i. ), and averaged on all the accessible macroscopic states at a given temperature $T$. This definition do not imply any p[articular model or mechanism for the specific process considered. Using the general relationships of the statistical thermodynamics it results:

$$<\Delta H\,(T)> = RT^2\,[\mathrm{d}\ln Q\,(T) / \mathrm{d}T] \tag{1}$$

where $Q = \Sigma_i F_i$ is the partition function over all the states $i$, each characterized by the probability $F_i$. This function can allow in turn to calculate the number of the discrete and stable macroscopic states and the changes of all the thermodynamic parameters, for each transition from one state to another, and finally the relative populations as a function of temperature. Methods for carrying on these analyses are reported in the recent literature [13, 14]. The quantity $<\Delta H(T)>$ is experimentally determined by means of the relation:

$$<\Delta H\,(T)> = \int_{T_o}^{T} [C_p(T) - C_p^N(T)]\,\mathrm{d}T = \int_{T_o}^{T} C_p^E\,(T)\,\mathrm{d}T \tag{2}$$

where $C_p^E(T)$, the excess molar heat capacity of the protein, relative to the 'native' state, is the difference between the apparent heat capacity of the solution per mole of protein, (the observable quantity $C_p(T)$) and the apparent molar heat capacity of the solution containing the protein in its native state, $C_p^N(T)$ at each considered temperature $T$. In conclusion, the detailed and accurate knowledge of this function is the starting basis for evaluating the thermodynamic parameters or testing statistical mechanical models.

For this kind of analysis specialized and upper class microcalorimeters are required, to obtain very reproducible and reliable data on small amounts of biological material. Besides few very fine home-made instruments [15, 16] up to now only three kinds of commercial apparatus (in the experience of the authors) are actually qualified to be used for these studies: the series of DASM instruments of the Academy of Science of the ex-U.R.S.S. (few examples are diffused in U.S. and Western Europe), the more recent equipment of the Microcal (U.S.) and the MICRO DSC and MICRO DSC-II of the Setaram (France). Our laboratory is equipped with the last two instruments.

It is well known to all the users of DSC instruments, working in the field of biological macromolecules in solution, that the scanning rate must be chosen carefully. Practical reasons (low stability of biological materials at high temperatures in presence of secondary reactions or of traces of hydrolytic enzymes in the case of repeated heating and cooling cycles, time consume, etc.) are in contrast with the necessity that the system attains the equilibrium at each programmed temperature.

Irreversibility of the denaturation process, frequently observed, can be due to:

– secondary reactions

– post denaturation macromolecular incoherent aggregation

– very slow kinetics for the rebuilding of ordered structures or very low probability for restoring all the correct interactions (DNA f.i.). In all these cases, however, the lack of equilibrium during the heating does not invalidate the measure of the enthalpies and heat capacities but only the application of equilibrium models.

The determination of the denaturation calorimetric data, however, do not require only very good instruments and accurate experimental procedures, but a very delicate refinement of the signals and handling of the data to extract all the information disposable. Small errors in the correction of the base lines, for instance, can be responsible for wide uncertainties that can prevent until the classification of the transition as a one-step N <=> D equilibrium process or not. The temperature dependence of thermodynamic parameters can be affected by dramatic errors. For a more complex unfolding, the possibility to test whatever statistical mechanical approach could be compromised.

## THESEUS program

### *Generalities*

THESEUS system handles the instrument signals giving finally a numerical detailed description of the apparent specific heat of the sample solution as a function of temperature. From this description the complexity of the unfolding process is immediately focused. In the case of a one step reversible denaturation mechanism, the program gives very easily a set of thermodynamic parameters characterizing the transition and a preliminary estimate of the 'cold denaturation' temperature.

For more complex processes THESEUS gives a set of graphical and numerical description of the data useful for testing different statistical mechanical treatments as well as for the mathematical deconvolution of the apparent specific heat curves.

We discuss the principal features in the next sections starting from the general tools regarding the handling of experimental data. In the last section we discuss a complete thermodynamical analysis performed by the program in a case of one step transition experiment.

The program are compiled in the microsoft Quick Basic 4.5 environment. It runs on PC's IBM compatible with VGA graphic Card. Mathematical coprocessor is desired. RAM memory request depends on the dimension of the input data files.

*Data input*

'THESEUS Input format' consists of two files of data in ASCII format, one containing the experimental values of the observable quantity (see the calibration tool described later) and the other contains in a specific order all the experimental conditions and information indispensable for the analysis. In the case of our MICRO-DSC instrument, as for other apparatus of our laboratory these files are created automatically by the MIDAS hardware-software system which accumulates the calorimeter output electrical signals by means of an ADC card, provided of other auxiliary digital outputs devoted to the regulation of the operating modes of the instrument. For our calorimeter as well as for other instruments the Theseus package contains a short program which converts all data files from the ASCII or IEEE format in 'THESEUS Input format' (for our Micro-DSC-II/ the IEEE format is produced directly by the DG11 Setaram Controller device).

The package offers the possibility to create simulated calorimetric data as an input, using as a model the sum of a set of single step independent transition [17]. The operator can choose to apply a random noise too in order to test and evaluate the methods of handling described in the next section.

*Handling*

The data, when loaded, are represented in a graphic window, the dimensions of which are automatically set. The ZOOM utility permits to change the dimensions of the graphic representation (optical ZOOM) and also permits to select from the data file the range of interest, excluding initial and eventually final ranges of instrumental instability, reducing the elaboration time (effective ZOOM).

All electrical signals present an undesirable noise. The SMOOTH utility offers three complementary possibilities to avoid this problem:

– a three point moving average, using a binomial distribution [18]; this method is not powerful but quick and it is used to gain time, when the noise is very weak. The results from the other two methods are practically identical.

– a second or a third degree polynominal least square 'movable strip' smooth. The choice of the dimension of the 'movable strip' determines the effectiveness of the smooth. The process can be iterated.

– in special cases, the operator can choose different parameters of smoothing in different regions, avoiding on this way the loss of information in delicate range; this method is very powerful and benefit from advantages and disadvantages of the least square methods. In the case of data difficult to handle, which do not permit a satisfactory movable strip smooth (as they present matrix problems, or other difficulties), we implemented a Fourier smooth. After many tests we chose a method [19] which does not use the artifice of the Fast Fourier Transforms in order to obtain a good smooth even in extreme conditions of data sets. Of course, this is payed in calculation time.

In case of counterfeit electrical spikes, the operator can use the utility 'PERS' which permits to make corrections 'by hand', point by point, using the cursors, or vertical shifts through all the selected ranges.

As shown in Fig. 1, the program permits also to carry out algebraical additions of a constant factor, multiplication by a constant or polynomial function, to derive or integrate in selected ranges, to research maxima or minima, to compare different files etc.

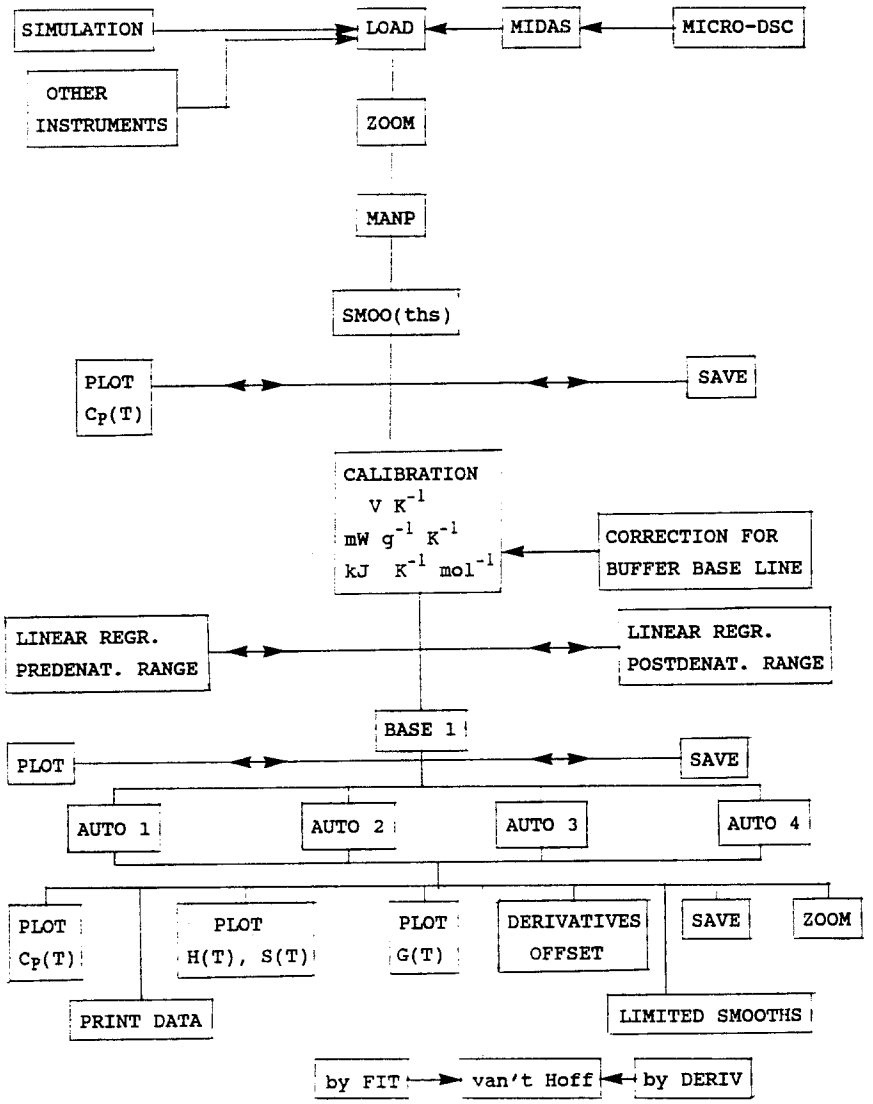


**Fig. 1** Typical iter for a thermodynamic analysis in the case of one-step denaturation process

*Calibration*

The output signal of the calorimeter is usually measured in μV or, if amplified, in V. The instrument calibration curve (as a function of temperature) obtained by using the Joule effect from an electrical resistor and represented by a polynomial expansion, allows to evaluate for each electrical signal a corresponding power, expressed in mV.

The program, using the coefficients of this calibration curve and the information file, transforms the power terms in specific or molar apparent heat terms, expressed as $kJ \cdot K^{-1}$ (g solution)$^{-1}$ or $kJ \cdot K^{-1}$ (mol protein)$^{-1}$.

The information file contains the dimension of the numerical values, so the data input file is handled only for the required number of preceding steps.

*Base line*

For a reference against reference DSC measurement one could expect a corrected $C_p(T)$ line, constant in the whole temperature range. It does not happen in the reality, several balancing problems arise from small asymmetries of the sensors of the cell holders, from the different wear of the cells, etc.

To correct this effect the program operates a fit of the data from the reference-reference $C_p(i)$ file with the polynomial expansion RIF($i$) and finds the minimum (MINRIF) of this function in the range of the experimental temperatures. The quantities |RIF($i$–th) – (MINRIF| are then subtracted, point by point, from the reference solution measurement file $C_p(i$–th). In such a way the effect of the non-constant trend of the reference-reference measurements is removed.

As it was mentioned in the introduction, the quantity to obtain is the excess heat capacity $C_p^E(T) = [C_p(T) - C_p^N(T)]$. At temperatures low enough when no thermal effects occur, $C_p(T)$ and $C_p^N(T)$ have the same value. The temperature dependences of these parameters are commonly assumed to be almost linear [4, 7]. A linear regression can be applied to the set of data of this region, using the utility BASE 1 to build up a reference line extended in all the experimental range of explored temperatures. The $C_p(T)$ data are referred now to this line to obtain the $C_p^E(T)$ values. This point is very delicate. Incorrect choices for the predenaturation region may have strong influence on the slope of the reference line, making the elaboration unreliable. To avoid this problem, the operator, choosing the range of interest containing $n$ points, selects a number of points $m$ (with $m \leq n$, usually $m = 0.7\ n$): then the utility BASE1 operates an automatic research of the regression line on the $n$ consecutive points (inside the selected range) with the best correlation coefficients, and takes this line as the reference line. This kind of file is useful in several cases of deconvolution analysis [20]. In other types of deconvolution approaches the difference $C_p^E(T_i) - C_p^E(T_f)$ experimentally observed is not taken in account. In this case the procedure BASE2, applied after the pro-

cedure BASE1, refers the file in the denaturation range to a curve starting from the point $C_p^E(T_i)$ and distributes the difference $C_p^E(T_i) - C_p^E(T_f)$ values in all the range of the denaturation, affecting them with a weight given by the ratio (incremental area)/(total area) and iterating the process until convergence. In the first step of the iteration the total area is measured against a line that connects $C_p^E(T_i)$ point to $C_p^E(T_f)$ point.

## Thermodynamic analysis of one step reversible denaturation data

Figure 1 shows a typical iter for a thermodynamic analysis in the case of one step denaturation process. After the handling, calibration, and eventually BASE1 procedure, the operator is now ready to process the data.

Four procedures can be chosen for evaluating the thermodynamic parameters characterizing the denaturation (AUTO1, 2, 3, 4).

– In the first one the slope of the function $C_p^D(T)$ defined as the $C_p(T)$ function in the post-denaturation range, is assumed equal to that of the pre-denaturation region, (a linear regression can be utilized if BASE1 utility was not applied before). This is useful when the high temperature region is scarcely reproducible for secondary phenomena.

– Using AUTO2 on the other hand, the slope of $C_p^D(T)$ is evaluated by means of a linear regression in the post-denaturation region in the same way as mentioned in the utility BASE1.

– In both these cases the $\Delta_d C_p^E(T_d) = C_p^D(T_d) - C_p^N(T_d)$ and the integrated $\Delta_d H$ are obtained by using the two regression lines as a reference: $C_p^N(T)$ up to $T_d$ and $C_p^D(T)$ beyond $T_d$.

– The third and fourth procedures are analogous to the first and second respectively, but the reference line, delimiting the area to be integrated to obtain the total $\Delta_d H$ is represented by a sigmoid which distributes the difference $C_p^E(T_f) - C_p^E(T_i)$ as we mentioned in the procedure BASE2. On this basis the other quantities can be evaluated, according to the given option.

First of all the initial assumption that we are dealing with a one-step process can be immediately tested, by using the calorimetric data as an analytical tool for measuring the degree of progress of the transition $\theta(T)$, where $\theta(T)$ is given by the ratio between the incremental area and the total area. For a two-state equilibrium transition can be defined at each temperature an equilibrium constant:

$$K(T) = \theta(T) / [1 - \theta(T)] \tag{3}$$

and from the van't Hoff relation:

$$\frac{\mathrm{d}\ln K(T)}{\mathrm{d}T} = \frac{\Delta_\mathrm{d} H^{\mathrm{v.H.}}}{RT^2} \tag{4}$$

it is possible to calculate $\Delta_\mathrm{d}H^{\mathrm{v.H.}}(T_\mathrm{d})$. The coincidence (within ±10%) of the values of this quantity and enthalpy $\Delta_\mathrm{d}H$ is a necessary condition that the denaturation model is correct [7, 8]. THESEUS program automatically performs not only the $\Delta_\mathrm{d}H^{\mathrm{v.H.}}$ evaluation but also gives graphical representation of the function $\mathrm{d}\ln K(T)/\mathrm{d}T$. These data set can be used to validate, by means of other tests, the one-step transition hypothesis [20].

The determined value of $\Delta_\mathrm{d}C_\mathrm{p}^\mathrm{E}(T_\mathrm{d})$ or, if the case, its functional dependence on temperature, $\Delta_\mathrm{d}C_\mathrm{p}^\mathrm{E}(T)$, permits to calculate the state functions $\Delta_\mathrm{d}H(T)$, $T{\cdot}\Delta_\mathrm{d}S(T)$ and $\Delta_\mathrm{d}G(T)$ for each temperature $T$, according to the relationships:

$$\Delta_\mathrm{d}H(T) = \Delta_\mathrm{d}H(T_\mathrm{d}) + \int_{T_\mathrm{d}}^{T} \Delta_\mathrm{d}C_\mathrm{p}^\mathrm{E}\,\mathrm{d}T \tag{5}$$

$$\Delta_\mathrm{d}S(T) = \frac{\Delta_\mathrm{d}H(T_\mathrm{d})}{T_\mathrm{d}} + \int_{T_\mathrm{d}}^{T} \frac{\Delta_\mathrm{d}C_\mathrm{p}^\mathrm{E}\,\mathrm{d}T}{T_\mathrm{d}} \tag{6}$$

$$\Delta_\mathrm{d}G(T) = \Delta_\mathrm{d}H(T) - T{\cdot}\Delta_\mathrm{d}S(T) \tag{7}$$

Equations (5–7) are of general validity, so that the defined functions are significant even out of the denaturation range. It must be outlined that for a true equilibrium process N <=> D the $\Delta_\mathrm{d}G(T)$ is identically zero in the whole range $T_\mathrm{i} \rightarrow T_\mathrm{f}$, but it occurs for the standard Gibbs energy $\Delta_\mathrm{d}G^\circ(T)$ only for $T = T_\mathrm{d}$, when $K(T) = 1$. Then it is more rigorous to introduce into Eqs (5–7) only the standard molar quantities, each defined for the particular considered temperature $T$. In this manner the process will be considered to be a chemical equilibrium more than a phase equilibrium (fusion).

Equation (7) gives a description of the thermal stability of the protein: the $\Delta_\mathrm{d}G^\circ(T)$ passes through a maximum for $T_\mathrm{s} \ll T_\mathrm{d}$, where $\Delta_\mathrm{d}S^\circ = 0$, and, in principle, another denaturation temperature, $T_\mathrm{d}'$, the cold denaturation temperature, much lower than the normal freezing point of water, can be forecast.

For more complex behaviours, one option allows to analyze the $C_\mathrm{p}^\mathrm{E}(T)$ curve by means of the derivative method, focusing each maximum, minimum and flex points. The slopes at these and in all points and the corresponding offsets are also evaluated. In any case the data can be stored in files for subsequent analyses.

*Examples*

Figure 2a shows the registered calorimetric curves for protein solution and buffer solution in the case of pancreatic Ribonuclease A (RNAase A) at *pH* 5.0, and protein concentration $1.8 \cdot 10^{-4}$ *M*, measured at 25°C. Acetate buffer, at total ionic strength $I = 0.1$ *M*, was used.

In Fig. 2b the same curve is presented after buffer-buffer correction, calibration and smoothing. The thermodynamic parameter evaluated by THESEUS are $T_d = 336.0$ K, $\Delta_d H = 492$ kJ·mol$^{-1}$, $\Delta_d C_p^E = 7.2$ kJ·mol$^{-1}$K$^{-1}$, $\Delta_d H^{v.H.} = 483$ kJ·mol$^{-1}$. Since $\Delta_d H/\Delta_d H^{v.H.} = 1.01$, and thedenaturation process is reversible, the $\Delta_d G^\circ(T)$ can be calculated and it is reported in Fig. 3: parameters at maximum: $T_s = 274$ K, $\Delta_d G^\circ(T_s) = 46.8$ kJ·mol$^{-1}$ and $\Delta_d S^\circ(T_s) = 0$, by definition. Cold denaturation temperature: $T_d' = 216.5$ K (hypothetical).

In Fig. 4 the corrected curve for the dimeric bovine seminal Ribonuclease (RNAase BS) [21–23] is reported at $pH = 5.0$; acetate buffer at total ionic strength $I = 0.1$ *M* and protein concentration $3.7 \cdot 10^{-5}$ *M*. The thermodynamic parameter values calculated were: $T_d = 334.2$ K, $\Delta_d H = 863$ kJ·mol$^{-1}$, $\Delta_d C_p^E = 10.1$ kJ·mol$^{-1}$K$^{-1}$, $\Delta_d H^{v.H.} = 367.6$ kJ·mol$^{-1}$, $\Delta_d H/\Delta_d H^{v.H.} = 2.29$. The denaturation

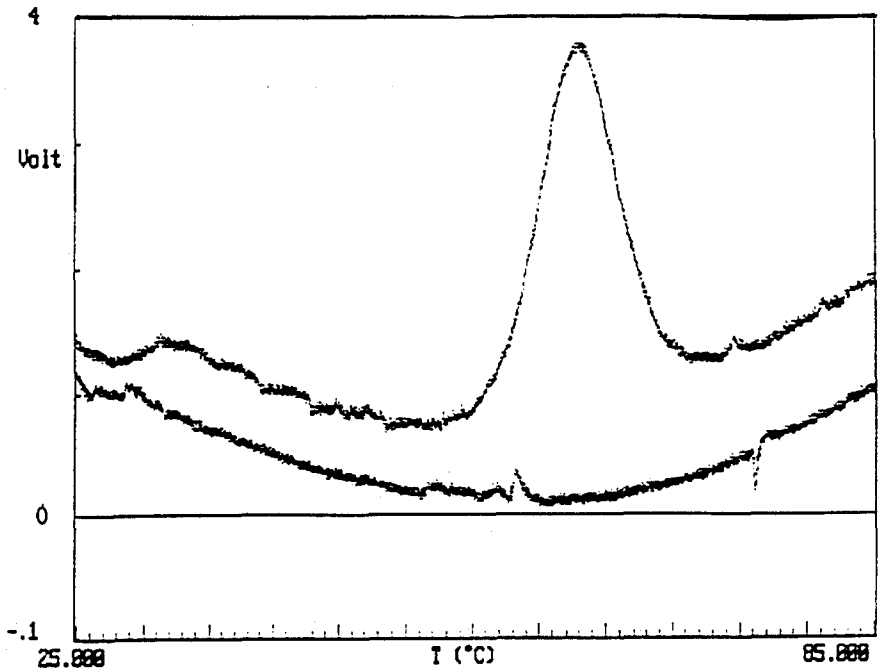


**Fig. 2a** Thermal denaturation of RNAase A at $pH = 5.0$; protein concentration $1.8 \cdot 10^{-4}$ *M*, acetate buffer, ionic strength $I = 0.1$ *M*; direct calorimetric registrations of protein solution and of buffer-buffer measurement

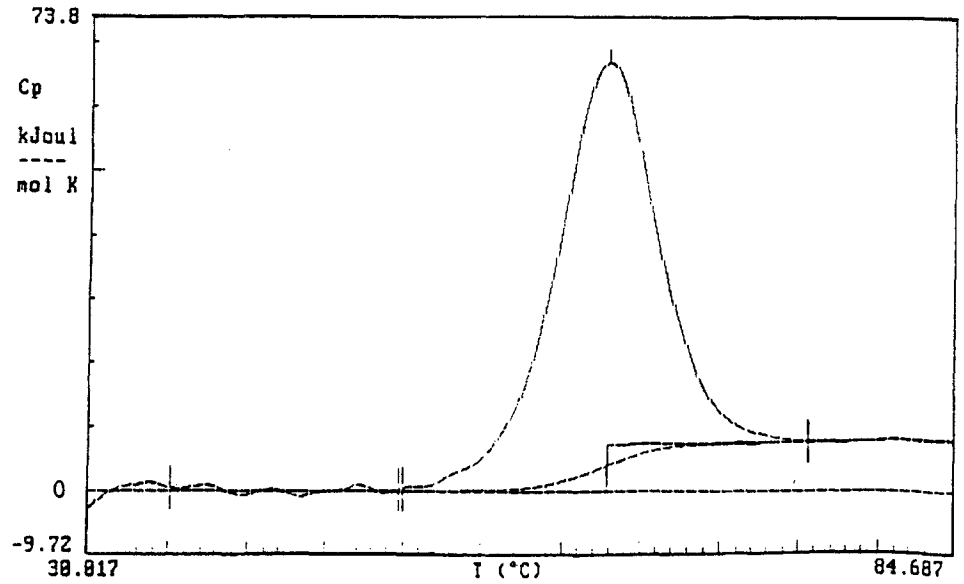


**Fig. 2b** $C_p^E(T)$ profile after all corrections carried out by THESEUS as described in the text

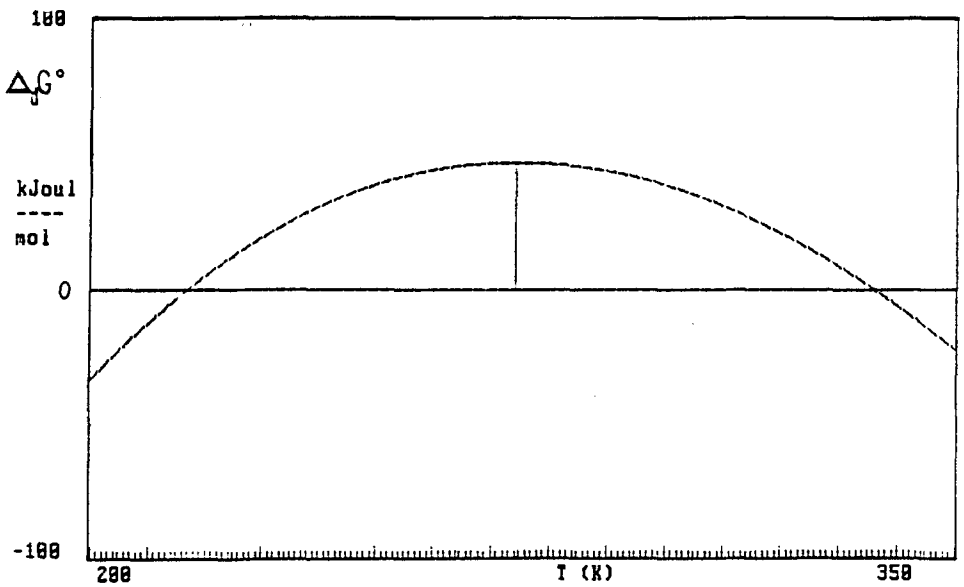


**Fig. 3** Thermal stability of RNAase A as a function of temperature

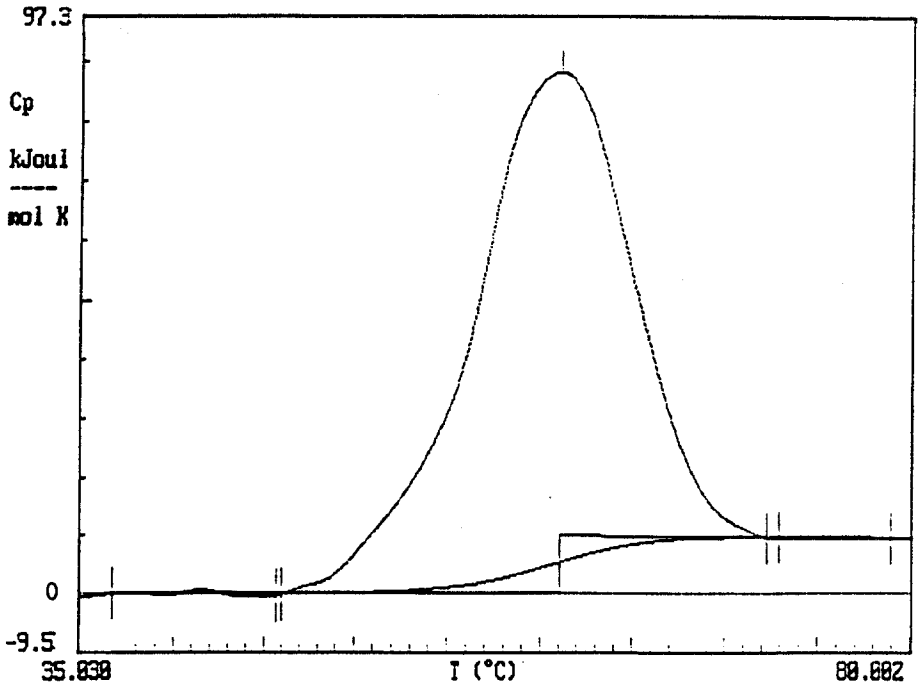


**Fig. 4** Thermal denaturation of RNAase BS after correction for calibration and reference line; $pH = 5.0$; acetate buffer at total ionic strength $I = 0.1\ M$ and protein concentration $3.7 \cdot 10^{-5}\ M$

cannot be assumed as a two-state process. For the deconvolution of the peaks the programs DEDALUS and MINOS have been written and they will be the subject of further publications.

* * *

This work was carried out with the financial support of Italian CNR (Target Program 'Chimica Fine')(Rome) and of Ministry of University and Scientific and Technological Research.

## References

1 C. B. Anfinsen, J. Biol. Chem., 221 (1956) 405.
2 W. Kauzmann, Adv. Protein Chem., 14 (1959) 1.
3 J. F. Brandts, in 'Structure and Stability of Biological Macromolecules', S. N. Timascheff and G. D. Fasman Eds., Dekker-New York 1969, p. 213.
4 P. L. Privalov and N. N. Khechinashvili, J. Mol. Biol., 86 (1974) 665.
5 R. R. Baldwin, Ann. Rev. Biochem., 44 (1975) 453.

6 J. M. Sturtevant, Proc. Nat. Acad. Sci. USA, 74 (1977) 2236.
7 P. L. Privalov, Adv. Protein Chem., 33 (1979) 179.
8 P. L. Privalov, Adv. Protein Chem., 35 (1982) 1.
9 W. J. Becktel and J. A. Schellman, Biopolymers, 26 (1987) 1859.
10 V. V. Mozhaev, I. V. Berezin and K. Martinek, Crit. Rev. Biochem., 23 (1988) 235.
11 V. V. Mozhaev and K. Martinek, Enz. Microb. Technol., 4 (1982) 299.
12 B. Zentgraf, M. Ringpfeil, K. Martinek, V. V. Mozhaev and T. J. Ahern, Pure Appl. Chem., 63 (1991) 1527.
13 E. Freire and R. Biltonen, Biopolymers, 17 (1978) 463.
14 E. Freire and R. Biltonen, Biopolymers, 17 (1978) 481.
15 O. L. Mayorga, W. W. van Osdol, J. L. Lacomba and E. Freire, Proc. Natl. Acad. Sci. USA, 85 (1988) 9514.
16 J. M. Sturtevant, Ann. Rev. Phys. Chem., 38 (1987) 463.
17 L. H. Chang, S. J. Li, T. L. Ricca and A. G. Marshall, Anal. Chem., 56 (1984) 1502.
18 P. R. Bevington, 'Data Reduction and Error Analysis for the Physical Sciences', McGraw-Hill Book Company, 1969.
19 E. E. Aubanel and K. B. Oldham, Byte, (February 1985) 207.
20 G. Barone, P. Del Vecchio, D. Fessas, C. Giancola and G. Graziano, in Proc. 2nd Joint Greek-Italian Meeting on Chemistry of Biological Systems and Molecular Chemical Engineering',
J. Anastassopoulou, G. Barone, N. Russo Eds., Kluwer Ac. Pub., Dordrecht 1993 (in press).
21 G. D'Alessio, A. Floridi, R. De Prisco, A. Pignaro and E. Leone, Eur. J. Biochem., 26 (1972) 153.
22 L. Mazzarella, C. A. Mattia, S. Capasso and G. Di Lorenzo, Gazetta Chimica Italiana, 117 (1987) 91.
23 R. Ambrosino, G. Barone, G. Castronuovo, O. Cultrera, A. Di Donato and V. Elia, Biopolymers, 28 (1989) 1403.

**Zusammenfassung** — Bei der Analyse von DSC-Daten der thermischen Zersetzung von biologischen Makromolekülen wurde ein neues Softwarepaket (THESEUS) angewendet. Unter Umgehung der in derartigen Experimenten oft vorkommenden gelegentlichen und systematischen Fehler eignet sich dieses System, um genaue Physikalisch-chemische Informationen zu erhalten. Außerdem kann es zur Verarbeitung von Daten verwendet werden, welche von anderen Instrumenten und Methoden stammen, die thermodynamische, spektroskopische oder andere Daten als Funktion der Temperatur liefern. Da viele der Untersuchungen auf diesem Gebiet Forschungs-Charakter tragen und in der gegen-wärtigen Literatur ständig neue Dekonvolutionsmechanismen beschrieben oder angenommen werden, haben wir dieses leistungsstarke, flexible und allgemein anwendbare Programm geschrieben und umgesetzt, um den Operator in die Lage zu versetzen, jeden Phase des Rechenvorganges kontrollieren und seine eigene Erfahrung benutzen zu können, um den besten Weg zur Lösung unvorhergesehener Probleme zu finden.