Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2683207 A1 2010/04/17

(21) 2 683 207

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(22) Date de dépôt/Filing Date: 2009/10/16

(41) Mise à la disp. pub./Open to Public Insp.: 2010/04/17

(30) Priorité/Priority: 2008/10/17 (US12/253,676)

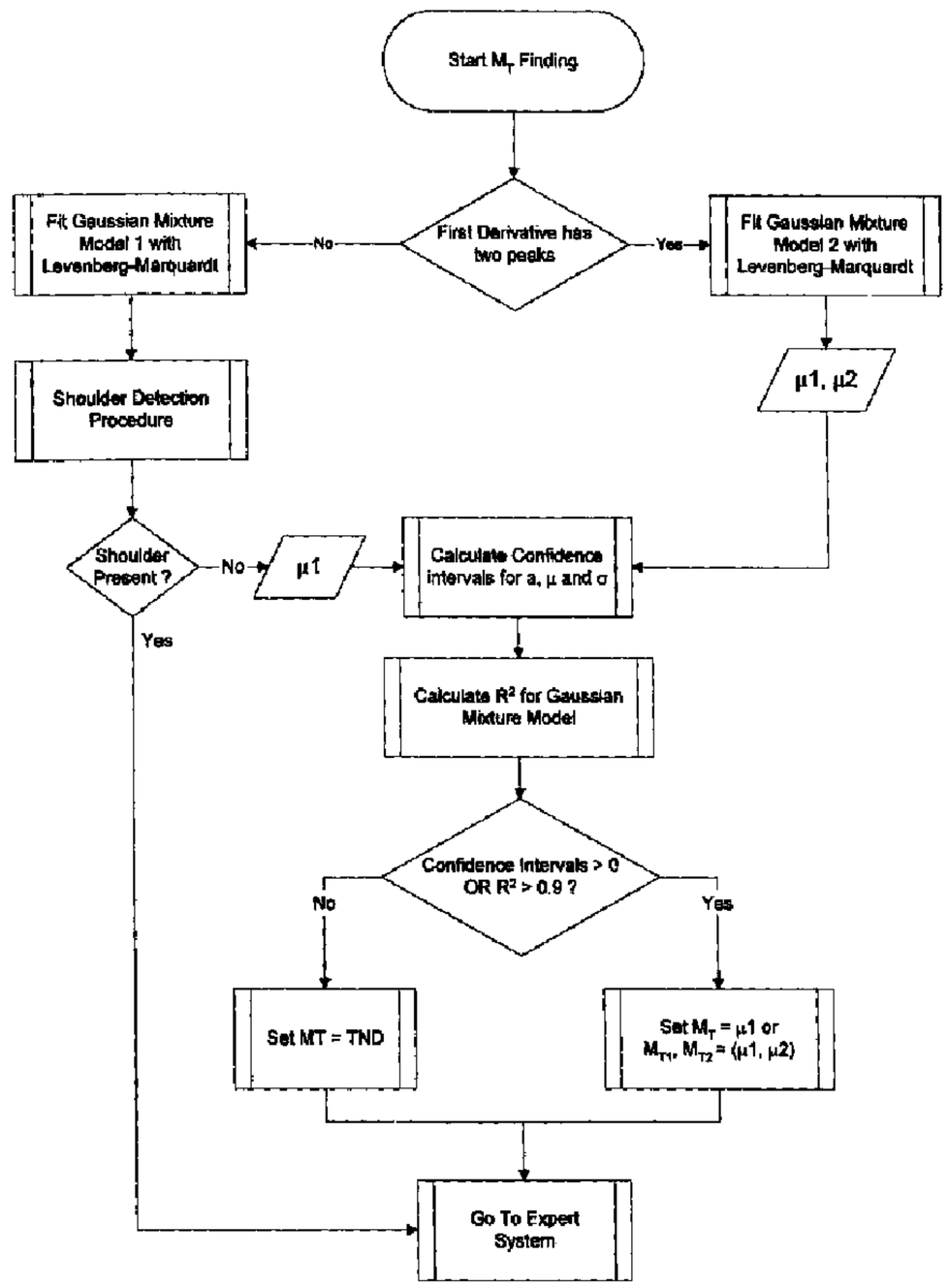
(51) Cl.Int./Int.Cl. *C12Q 1/68* (2006.01),
G06F 19/00 (2006.01)

(71) Demandeur/Applicant:
F. HOFFMANN-LA ROCHE AG, CH

(72) Inventeur/Inventor:
KURNIK, RONALD T., US

(74) Agent: BORDEN LADNER GERVAIS LLP

(54) Titre : DETERMINATION DE TEMPERATURES DE FUSION PAR METHODES SANS EQUATION
(54) Title: DETERMINATION OF MELTING TEMPERATURES BY EQUATION-LESS METHODS



Major Processing Steps of the Equation-Less Melting Algorithm

(57) Abrégé/Abstract:
Numerical determinations of the first derivatives of a melt curve data set are made. A baseline is determined for the first derivative values and the baseline is subtracted from the first derivative values to produce modified first derivative values. A first maximum

(21) 2 683 207

(13) A1

(57) Abrégé(suite)/Abstract(continued):
value of the modified first derivative values is determined and said first maximum value represents a melting temperature Tm of a DNA sample. A model function, such as a Gaussian Mixture Model (GMM) function, with parameters determined using a Levenberg-Marquardt (LM) regression process can also be used to find an approximation to the first derivative curve. The maximum values of the numerically determined first derivative values are used as initial conditions for parameters of the model function. The determined parameters provide one or more fractional melting temperature values, which can be returned, for example, displayed or otherwise used for further processing.

ABSTRACT

Numerical determinations of the first derivatives of a melt curve data set are made. A baseline is determined for the first derivative values and the baseline is subtracted from the first derivative values to produce modified first derivative values. A first maximum value of the modified first derivative values is determined and said first maximum value represents a melting temperature Tm of a DNA sample. A model function, such as a Gaussian Mixture Model (GMM) function, with parameters determined using a Levenberg-Marquardt (LM) regression process can also be used to find an approximation to the first derivative curve. The maximum values of the numerically determined first derivative values are used as initial conditions for parameters of the model function. The determined parameters provide one or more fractional melting temperature values, which can be returned, for example, displayed or otherwise used for further processing.

DETERMINATION OF MELTING TEMPERATURES BY EQUATION-LESS METHODS

BACKGROUND OF THE INVENTION

The present invention relates generally to processing data representing melt characteristics of oligonucleotides, and more particularly to systems and methods for determining one or more melting temperatures for an oligonucleotide sample based on melt curve data.

Determination of DNA melting temperatures, usually performed directly after a PCR experiment, is an important method to distinguish genotypes. For example, recently discussed in the literature is the use of examining the KRAS gene to determine which patients might be candidates for a treatment for non-small cell lung cancer. Patients whose KRAS gene is of a wild-type would benefit from the treatment, whereas if the patient had a mutant variation of this gene, the treatment would be of no benefit. As these treatments often have major side effects, it is very important to determine the patient's correct genotype. The use of a melting KRAS assay can help distinguish the genotype of the patient. Therefore, it is desirable to provide systems and methods that accurately and efficiently determine the melting temperature of a DNA sample.

BRIEF SUMMARY OF THE INVENTION

The present invention provides systems and method for determining melting temperatures, Tm, for oligonucleotides based on melt curve data. In various embodiments, numerical determinations of the first derivatives of a melt curve data set are made. A model function, such as a Gaussian Mixture Model (GMM)

function, with parameters determined using a Levenberg-Marquardt (LM) regression process is used to find an approximation to the first derivative curve. The maximum values of the numerically determined first derivative values are used as initial conditions for parameters of the model function. The determined parameters provide fractional melting temperature values, which can be returned, for example, displayed or otherwise used for further processing.

According to one aspect of the present invention, a computer-implemented method of determining a melting temperature, Tm, of DNA is provided. The method typically includes receiving a dataset representing a melt curve for a DNA sample, the dataset including a plurality of data points each having a pair of coordinate values. The method also typically includes numerically determining first derivative values for data points of the melt curve, determining a baseline for the first derivative values, subtracting out the baseline from the first derivative values to produce modified first derivative values, and determining a first maximum value of the modified first derivative values. The method also typically includes outputting said maximum value, wherein said first maximum value represents a melting temperature Tm of the DNA sample. The method may further comprise the steps of calculating an approximation of a curve that fits the modified first derivative values by applying a Levenberg-Marquardt (LM) regression process to a Gaussian Mixture Model function to determine one or more parameters of the function, wherein said parameters include initial conditions, and wherein the first maximum value is used as an initial condition for a first parameter, and outputting the first parameter, wherein the determined first parameter represents a melting temperature, Tm, of the DNA sample. In certain aspects, the method further includes determining whether the first derivative values include a shoulder value proximal to the first maximum value. In certain embodiments the Gaussian Mixture Model includes an expression of the form:

$$GMM_1 = Exp(-a_1)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right),$$

wherein μ_1 is the first parameter, and wherein a_1 and σ_1 are additional parameters. In another aspect the method further includes displaying the first parameter value. In yet another aspect the regression process includes a Levenberg-Marquardt (LM) regression process. In certain aspects the method further comprises the steps of determining a second, a third or a fourth maximum value of the modified first derivative values, wherein the second, third or a fourth maximum value is used as an initial condition for a second, third or fourth parameter; and outputting the second, third or fourth parameter, wherein the determined second, third or fourth parameter represents a second, third or fourth melting temperature, Tm2, Tm3 or Tm4 of the DNA sample. In another aspect of the method the Gaussian Mixture Model includes an expression of the form:

$$GMM_2 = Exp(-a_1)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right) + Exp(-a_2)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right),$$

wherein μ_1 is the first parameter, μ_2 is the second parameter and wherein a_1, σ_1, a_2 and σ_2 are additional parameters. In yet another aspect of the method the Gaussian Mixture Model includes an expression of the form:

$$GMM_3 = Exp(-a_1)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right) + Exp(-a_2)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right) +$$
$$Exp(-a_3)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_3}{\sigma_3}\right)^2\right)$$

wherein μ_1 is the first parameter, μ_2 is the second parameter, μ_3 is the third parameter and wherein a_1, σ_1, a_2, σ_2, a_3 and σ_3 are additional parameters. In yet another aspect of the method the Gaussian Mixture Model includes an expression of the form:

$$GMM_4 = Exp(-a_1)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right) + Exp(-a_2)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right) +$$

$$Exp(-a_3)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_3}{\sigma_3}\right)^2\right) + Exp(-a_4)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_4}{\sigma_4}\right)^2\right)$$

wherein μ_1 is the first parameter, μ_2 is the second parameter, μ_3 is the third parameter μ_4 is the fourth parameter and wherein a_1, σ_1, a_2, σ_2, a_3, σ_3, a_4 and σ_4 are additional parameters. In another aspect of the method determining a first maximum value includes applying a windowing process to the modified first derivative values, wherein the windowing process also determines one or more additional initial conditions for the Gaussian Mixture Model function parameters.

According to another aspect of the present invention, a computer readable medium is provided that stores code for controlling a processor to determine a melting temperature, Tm, of DNA. The stored code typically includes instructions to receive a dataset representing a melt curve for a DNA sample, the dataset including a plurality of data points each having a pair of coordinate values, to numerically determine first derivative values for data points of the melt curve, to determine a baseline for the first derivative values, to subtract out the baseline from the first derivative values to produce modified first derivative values, and to determine a first maximum value of the modified first derivative values. The code also typically includes instructions to output the first maximum value, wherein said first maximum value represents a melting temperature Tm of the DNA sample. The code also typically includes instructions to calculate an approximation of a curve that fits the modified first derivative values by applying a Levenberg-Marquardt (LM) regression process to a Gaussian Mixture Model function to determine one or more parameters of the function, wherein said parameters include initial conditions, and wherein the first maximum value is used as an initial condition for a first parameter, and to output the first parameter, wherein the determined first parameter represents a melting temperature, Tm, of

the DNA sample. Herein, in certain aspects the Gaussian Mixture Model includes an expression of the form:

$$GMM_1 = Exp(-a_1)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right),$$

wherein μ_1 is the first parameter, and wherein a_1 and σ_1 are additional parameters. In certain aspects the code further includes instructions to determine whether the first derivative values include a shoulder value proximal to the first maximum value. In another aspect the code further includes instructions to display the first parameter value. In yet another aspect the code further includes instructions to determine a second maximum value of the modified first derivative values, wherein the second maximum value is used as an initial condition for a second parameter; and to output the second parameter, wherein the determined second parameter represents a second melting temperature, Tm2, of the DNA sample. In certain aspects the Gaussian Mixture Model includes an expression of the form:

$$GMM_2 = Exp(-a_1)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right)+Exp(-a_2)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right),$$

wherein μ_1 is the first parameter, μ_2 is the second parameter and wherein a_1, σ_1, a_2 and σ_2 are additional parameters. In yet another aspect the Gaussian Mixture Model includes an expression of the form:

$$GMM_3 = Exp(-a_1)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right)+Exp(-a_2)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right)+$$
$$Exp(-a_3)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_3}{\sigma_3}\right)^2\right)$$

wherein μ_1 is the first parameter, μ_2 is the second parameter, μ_3 is the third parameter and wherein a_1, σ_1, a_2, σ_2, a_3 and σ_3 are additional parameters. In yet another aspect the Gaussian Mixture Model includes an expression of the form:

$$GMM_4 = Exp(-a_1)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right) + Exp(-a_2)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right) +$$

$$Exp(-a_3)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_3}{\sigma_3}\right)^2\right) + Exp(-a_4)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_4}{\sigma_4}\right)^2\right)$$

wherein μ_1 is the first parameter, μ_2 is the second parameter, μ_3 is the third parameter μ_4 is the fourth parameter and wherein a_1, σ_1, a_2, σ_2, a_3, σ_3, a_4 and σ_4 are additional parameters. In another aspect of the computer readable medium the instructions to determine a first maximum value include instructions to apply a windowing process to the modified first derivative values, wherein the windowing process also determines one or more additional initial conditions for the Gaussian Mixture Model function parameters.

According to yet another aspect of the present invention, a kinetic Polymerase Chain Reaction (PCR) system is provided that typically includes a kinetic PCR analysis module that generates a melt curve dataset representing a DNA melt curve, the dataset including a plurality of data points, each having a pair of coordinate values, and an intelligence module adapted to process the melt curve dataset to determine a Tm value. The intelligence module typically determines a Tm value by numerically determining first derivative values for data points of the melt curve, by determining a baseline for the first derivative values, by subtracting out the baseline from the first derivative values to produce modified first derivative values, and by determining a first maximum value of the modified first derivative values. The intelligence module also typically determines a Tm value by using the determined first maximum value of the modified first derivative as the Tm value and outputting said Tm value. In certain aspects the intelligence module determines a Tm value by calculating an approximation of a curve that fits the modified first derivative values by applying a Levenberg-Marquardt (LM) regression process to a Gaussian Mixture Model function to determine one or more parameters of the function, wherein said parameters include initial conditions, and wherein the first maximum value is used as an

initial condition for a first parameter, and by outputting the first parameter, wherein the determined first parameter represents a melting temperature, Tm, of the DNA sample. In certain aspects, a Gaussian Mixture Model includes an expression of the form:

$$GMM_1 = Exp(-a_1)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right),$$

wherein μ_1 is the first parameter, and wherein a_1 and σ_1 are additional parameters. In another aspect the intelligence module is further adapted to determine whether the first derivative values include a shoulder value proximal to the first maximum value. In yet another aspect the intelligence module is further adapted to determine a second maximum value of the modified first derivative values, wherein the second maximum value is used as an initial condition for a second parameter; and to output the second parameter, wherein the determined second parameter represents a second melting temperature, Tm2, of the DNA sample. In certain aspects, a second maximum value of the modified first derivative values is determined, wherein the second maximum value is used as an initial condition for a second parameter, and wherein a Gaussian Mixture Model includes an expression of the form:

$$GMM_2 = Exp(-a_1)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right) + Exp(-a_2)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right),$$

wherein μ_1 is the first parameter, μ_2 is the second parameter and wherein a_1, σ_1, a_2 and σ_2 are additional parameters. In certain aspects, a third maximum value of the modified first derivative values is determined, wherein the third maximum value is used as an initial condition for a third parameter, and wherein a Gaussian Mixture Model includes an expression of the form:

$$GMM_3 = Exp(-a_1)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right) + Exp(-a_2)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right) + Exp(-a_3)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_3}{\sigma_3}\right)^2\right)$$

wherein μ_1 is the first parameter, μ_2 is the second parameter, μ_3 is the third parameter and wherein a_1, σ_1, a_2, σ_2, a_3 and σ_3 are additional parameters. In certain aspects, a fourth maximum value of the modified first derivative values is determined, wherein the fourth maximum value is used as an initial condition for a fourth parameter, and wherein a Gaussian Mixture Model includes an expression of the form:

$$GMM_4 = Exp(-a_1)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right) + Exp(-a_2)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right) + Exp(-a_3)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_3}{\sigma_3}\right)^2\right) + Exp(-a_4)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_4}{\sigma_4}\right)^2\right)$$

wherein μ_1 is the first parameter, μ_2 is the second parameter, μ_3 is the third parameter, μ_4 is the fourth parameter and wherein a_1, σ_1, a_2, σ_2, a_3, σ_3, a_4 and σ_4 are additional parameters. In another aspect of the system the intelligence module determines a first maximum value by applying a windowing process to the modified first derivative values, wherein the windowing process also determines one or more additional initial conditions for the Gaussian Mixture Model function parameters.

Reference to the remaining portions of the specification, including the drawings and claims, will realize other features and advantages of the present invention. Further features and advantages of the present invention, as well as the structure and operation of various embodiments of the present invention, are described in detail below with respect to the accompanying drawings. In the drawings, like reference numbers indicate identical or functionally similar elements.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 illustrates an example of a melt curve.

FIG. 2 illustrates a first derivative curve for the melt curve of FIG. 1.

FIG. 3 shows the derivative curve of FIG. 2 after baseline subtraction.

FIG. 4 illustrates a process for determining melting temperature(s) according to one embodiment.

FIG. 5 illustrates a Melting Temperature computation process according to one embodiment.

FIG. 6 illustrates a shoulder detection process according to one embodiment.

FIGS. 7a and 7b show an example of a raw data curve and a melting curve, respectively, in the case of a single melting peak.

FIGS. 8a and 8b show an example of a raw data curve and a melting curve, respectively, in the case of a two melting peaks.

FIGS. 9a and 9b show an example of a raw data curve and a melting curve, respectively, in the case of a two melting peaks.

FIGS. 10a and 10b show an example of a raw data curve and a melting curve, respectively, in the case of a single melting peak plus a shoulder.

FIGS. 11a and 11b show an example of a raw data curve and a melting curve, respectively, in the case of a single melting peak plus a shoulder.

FIG. 12 shows a general block diagram explaining the relation between software and hardware resources that can be used to implement the processes and system of the invention.

FIG. 13 illustrates the interaction between a thermocycler device and a computer system.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides systems and methods for determining melting temperatures, Tm, for DNA.

The present invention provides systems and methods for determining a melting temperature by analyzing data representing a melt curve. In certain aspects, numerical determinations of the first derivatives of a melt curve data set are made. A Gaussian Mixture Model (GMM) function with parameters determined using a Levenberg-Marquardt (LM) regression process is used to find an approximation to the first derivative curve. The maximum values of the numerically determined first derivative values are used as initial conditions for parameters of the GMM function. The determined parameters provide fractional melting temperature, Tm, values. The Tm value(s) are then returned and may be displayed or otherwise used for further processing.

One example of a melt curve in the context of a PCR process is shown in FIG. 1. As shown in FIG. 1, data for a typical melt curve can be represented in a two-dimensional coordinate system, for example, with temperature defining the x-axis and an indicator of accumulated polynucleotide defining the y-axis. Typically, the indicator of accumulated polynucleotide is a fluorescence intensity value as the use of fluorescent markers is perhaps the most widely used labeling scheme. However, it should be understood that other indicators may be used depending on the particular labeling and/or detection scheme used. Examples of other useful indicators of accumulated signal include luminescence intensity,

chemiluminescence intensity, bioluminescence intensity, phosphorescence intensity, charge transfer, voltage, current, power, energy, temperature, viscosity, light scatter, radioactive intensity, reflectivity, transmittance and absorbance.

General Process Overview

Consider a typical melt curve as shown in FIG. 1. It is desired to obtain one or more melting temperatures from the data represented in FIG 1. According to one embodiment, a process 100 for determining a melting temperature can be described briefly with reference to FIG. 4. In step 110, an experimental data set representing the melt curve is received or otherwise acquired. An example of a plotted melt curve data set is shown in FIG. 1, where the y-axis and x-axis represent fluorescence intensity and temperature, respectively, for a melt curve. In certain aspects, the data set should include data that is continuous and equally spaced along an axis.

In the case where process 100 is implemented in an intelligence module (e.g., processor executing instructions) resident in a PCR device such as a thermocycler, the data set may be provided to the intelligence module in real time as the data is being collected, or it may be stored in a memory unit or buffer and provided to the intelligence module after the experiment has been completed. Similarly, the data set may be provided to a separate system such as a desktop computer system or other computer system, via a network connection (e.g., LAN, VPN, intranet, Internet, etc.) or direct connection (e.g., USB or other direct wired or wireless connection) to the acquiring device, or provided on a portable medium such as a CD, DVD, floppy disk or the like. In certain aspects, the data set includes data points having a pair of coordinate values (or a 2-dimensional vector). For melt data, the pair of coordinate values typically represents the temperature and the fluorescence intensity value. After the data set has been received or acquired in step 110, the data set may be analyzed to determine the melting temperature(s).

In step 120, the data is numerically processed to determine derivative values. The melting temperatures for these curves is obtained by finding the (fractional) temperature value (x-axis) corresponding to the maximum of the first derivative (y-axis) of the fluorescence intensity with respect to temperature. Using the data shown in FIG. 1, corresponding graphs for the first derivative is shown in FIG. 2. In one embodiment, baseline subtraction is performed on the derivative curve data to produce modified derivative data. In one embodiment, baseline subtraction is performed by first defining "MedianLeft" as the median of the fluorescent values of the first five points in FIG 2, then defining "MedianRight" as the median of the fluorescent values of the last five points in FIG. 2. A straight line is then defined connecting the "MedianLeft" point (x, y) in FIG. 2 with the "MedianRight" point (x, y). The slope and intercept of this straight line is then subtracted from all coordinate pairs. FIG. 3 shows the derivative curve after baseline subtraction.

In one embodiment, derivatives are obtained by use of the Savitzky-Golay (SG) method. [See A. Savitzky and Marcel J.E. Golay (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Analytical Chemistry, 36: 1627-1639 and Press, W.H., et al. "Numerical Recipes in C, 2nd Ed.," Savitzky-Golay Smoothing Filters, Section 14.8, 650-655.]. A SG-2-2-2 configuration (meaning 2 points to the left, 2 points to the right, and 2nd degree polynomial) is used, in one embodiment, to calculate the first derivative of the raw melting data curve. In general, other configurations of the SG method may be used, for example, SG-1-1-2 to SG-50-50-2. More generally, SG-x-y-z may be used where x and y are numbers from 1 to 50 and z is a number from 1 to 5.

Scale Invariant Forms of derivatives

In certain embodiments, alternative methods are used to calculate derivatives in order to allow the melting temperature to be scale-invariant. Scale invariant means that if fluorescence values are multiplied by a constant, the resulting Tm value(s) are unchanged.

According to one method, the fluorescence value is divided by the mean fluorescence value before calculation of the derivative melting curve, e.g., y is replaced by:

$$y / y_{mean}, \text{ where } y_{mean} = \frac{1}{n}\sum_{i=1}^{n} y_i .$$

(1)

According to another method, the fluorescence value is divided by the (maximum fluorescence - minimum fluorescence) before calculation of the derivative melting temperatures.

According to another method, the derivative of the fluorescence value is divided by the fluorescence value before calculation of the melting temperatures.

According to yet another method, the derivative of the fluorescence value is divided by the mean derivative of the fluorescence value before calculation of the melting temperatures.

According to still another method, the derivative of the fluorescence value is divided by the (maximum-minimum) derivative of the fluorescence value before calculation of the melting temperatures.

Returning to FIG. 4, in step 130, a number of peaks in the first derivatives data is determined. In one embodiment, a local maximization process is used to determine zero, one, two three or four peaks in the first derivatives data as will be discussed in more detail below. The raw derivatives data may be used, or the baseline subtracted derivatives data may be used.

In one embodiment, fractional Tm values are determined in step 140. To find the maximum values of the curves, for example as shown in FIGS. 2 or 3, in one embodiment, a Gaussian Mixture Model is fit to the data. The mean of the Gaussian Mixture Model corresponds to the maximum, and hence the Tm value.

In one embodiment, a curve fit is done by calculating an approximation of a curve that fits the determined first derivative values, or modified (baseline subtracted) derivative values, by applying a regression process to a Gaussian Mixture Model function to determine one or more parameters of the function. In certain aspects, a Levenberg-Marquardt regression process is used. In one embodiment, for the case of a single peak, a Gaussian Mixture Model for one peak is used as shown in equation (2). If two peaks are present, a Gaussian Mixture Model for two peaks is used as shown in equation (3). If three peaks are present, a Gaussian Mixture Model for three peaks can be used as shown in equation (4). If four peaks are present, a Gaussian Mixture Model for four peaks can be used as shown in equation (5). The regressed values for the coefficients μ_1 or (μ_1, μ_2) correspond to the Tm values for one and two peaks respectively. Gaussian Mixture Models are used in one embodiment, rather than taking additional derivatives to find the maximum, as higher order derivatives (3^{rd}, and 4^{th}) can become unstable.

$$GMM_1 = Exp(-a_1)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right)$$

(2)

$$GMM_2 = Exp(-a_1)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right)+Exp(-a_2)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right)$$

(3)

$$GMM_3 = Exp(-a_1)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right)+Exp(-a_2)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right)+$$

$$Exp(-a_3)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_3}{\sigma_3}\right)^2\right)$$

(4)

$$GMM_4 = Exp(-a_1)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right)+Exp(-a_2)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right)+$$

$$Exp(-a_3)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_3}{\sigma_3}\right)^2\right)+Exp(-a_4)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_4}{\sigma_4}\right)^2\right)$$

(5)

It should be appreciated that other models/functions could be used instead of a Gaussian Mixture Model as would be apparent to one skilled in the art. Examples of other models include Beta, Binomial, Cauchy, Chi, ChiSquare, Exponential, Extreme Value, FRatio, Gamma, Gumbel, Laplace, Logistic, Maxwell, Pareto, Rayleigh, StudentT, and Weibull models.

It is to be understood that the above embodiment can be applied with first derivative data comprising more than two peaks. In this case, the initial estimates (regressed values) for the coefficients $\mu_1, \mu_2, \mu_3, \mu_4$ by the local maximum corresponds to the final melting temperature.

In one embodiment, the Levenberg-Marquardt (LM) method is used to curve fit equation (2) (or equation (3), equation (4), equation (5)). Details of this method can be found in the reference [Moré, J.J., "The Levenberg-Marquardt Algorithm, Implementation and Theory," *Numerical Analysis*, ed. Watson, G.A. Lecture Notes in Mathematics 630, Springer-Verlag, 1977]. It should be appreciated that other regression methods as are well known may be used. In general, the LM regression method includes an algorithm that requires various inputs and provides output. In one aspect, the inputs include a data set to be processed, a function (*e.g.*, Gaussian Mixture Model) that is used to fit the data, and an initial guess for the parameters or variables of the function. The output includes a set of one or more parameters for the function that minimizes the sum of the squares of the distance between the function and the data set. It should be appreciated that other regression processes as would be apparent to one skilled in the art may be used.

One feature of the Levenberg-Marquardt method is that it requires good estimates of the parameter values prior to performing the regression. For the parameters a_1 (or a_1, a_2, a_3, a_4) and σ_1 (or σ_1, σ_2, σ_3, σ_4) the initial conditions can be set equal to a constant (e.g., 1 or 2) in all cases. These parameters are generally not sensitive and will generally converge regardless of the initial conditions used. The parameters μ_1 (or μ_1, μ_2, μ_3, μ_4) may require more accurate initial conditions that should be determined for each curve. In one embodiment, a windowing method is used to calculate initial conditions for the parameters μ_1 (or μ_1, μ_2, μ_3, μ_4) as described in more detail below.

In optional step 150, one or more expert system checks are performed to assess whether the results are valid as will be discussed in more detail below. For example, if implemented, an expert system check may determine that the determined result is invalid.

In step 160, the Tm value(s) are returned, e.g., for display or further processing. Graphical displays may be rendered with a display device, such as a monitor

screen or printer, coupled with the system that performed the analysis of FIG. 4, or data may be provided to a separate system for rendering on a display device.

In some embodiments, the R^2 statistic and/or confidence (e.g., 95% confidence) intervals are calculated for the GMM_1, GMM_2, GMM_3 GMM_4 parameters. These values assess the quality of the curve fit and may be used in an Expert System (described below) to help determine whether the calculated Tm values are valid, invalid, or zero (no sample present). These values may also be displayed in step 160.

Determination of Maximum in Curves

In one embodiment, a windowing process is used on the data set to determine initial conditions for the parameters μ_1, μ_2, μ_3, μ_4. The following windowing process is described for a first derivative comprising up to two peaks. Nevertheless, as mentioned above, it is to be understood that this windowing process can be applied to first derivative data presenting up to four peaks. In one embodiment, the windowing process searches for potential local maxima by using the following procedure:

1. Starting at the first point, examine the first several (e.g., five) points (points 1-5) of the data set.

2. If the middle y-Point is not the maximum in these five points, then there is no potential maximum in these five points. If the middle y-point is the maximum of these five points and it has a value greater than 0 (to avoid adding the middle points of a longer sequence of points with an exact value of 0 into the set of potential maxima), then there is a potential maximum. Add this point to the set of potential maxima **S**.

3. Advance the sliding window by one point (e.g., now points 2-6), and repeat the process described in item 2, again accepting only maxima at

index 3 of these five points. Continue this process throughout the entire data set.

4. Examine the result set **S** of possible maxima, representing the set of possible maxima at index 3, and find the maximum data point in this set **S** of possible maxima (**Smax**).

5. If **Smax** is equal or less than a MaxNoise Input Parameter (a noise parameter that may be input by a user, or automatically determined), there is no peak in the curvature data.

6. Keep the remaining possible maximum data points from this set **S**, providing that they are greater than **Smax** x a RelativeMin Input Parameter and greater than an AbsoluteMin Input Parameter.

7. If there is only one data point left, there is only one peak, and the curve has only one maximum. Define this single peak as pk_1. If there are two data points left, then this represents a curve with two maxima. If there are more than two peaks, take the two peaks with the highest values of the data set **S** and return the peak with the lower cycle number of these two as pk_1 and the peak with the higher cycle number as pk_2.

8. The initial condition for μ_1 is then pk_1 and the initial condition for (μ_1, μ_2) is (pk_1, pk_2).

Melting Temperature Computation

A Melting Temperature computation according to one embodiment is illustrated in FIG. 5. In step 210, a determination is made as to whether the first derivative data set includes one or two (or three, four) peaks. If the first derivative of the melting data set is identified as having a single peak, in step 220, the maximum of the melting curve is found by a nonlinear regression (e.g., using the Levenberg-

Marquardt method or other regression method) of the single component Gaussian function as described in equation (2). The initial conditions are given by the local maxima search. If the first derivative of the melting data set is identified as having two peaks (or three, four), in step 230, both local maxima of the melting curve are found by a nonlinear regression (e.g., using the Levenberg-Marquardt method) of the double component Gaussian function as described in equation (3), and the parameters μ_1, μ_2 (or μ_3 *and* μ_4) (corresponding to Tm1, Tm2, or Tm3 and Tm4) are returned. The initial conditions are given by the local maxima search. It should be appreciated that even if FIG. 5 shows a melting temperature computation for one or two peaks, the same melting temperature computation used for two peaks can also be used for three and four peaks.

In step 240, a shoulder detection process determines whether a shoulder is present for the single derivative peak situation according to one embodiment. If two (or three, four) peaks are found in the MELT algorithm, no further action is taken. If, instead, one peak is found, it might be possible that a shoulder is present on this peak. Details of a shoulder detection process according to one embodiment are discussed below. If no shoulder is detected, μ_1 (corresponding to Tm1) is returned. In step 250 confidence intervals are determined for the various parameters. In step 260, an R^2 value is calculated for the Gaussian Mixture Model used. In one embodiment, if the R^2 value of the GMM1 (or GMM2, GMM3, GMM4) fit is > 0.9, or the confidence intervals for the parameters (μ_1, σ_1), or (μ_1, σ_1, μ_2, σ_2, μ_3, σ_3, μ_4, σ_4) do not include zero, then the values of Tm1 (or Tm1, Tm2, Tm3, Tm4) are accepted in step 270. Otherwise, Tm1, (or Tm1, Tm2, Tm3, Tm4) are set as target not detected (TND) in step 280, and the optional Expert System in step 290 makes the final call of TND or Invalid.

Shoulder Detection

A shoulder detection process according to one embodiment is shown in FIG. 6. In step 310, a double component Gaussian model is fit to the same first derivative

data using $a_{1,2} = 2$, $\mu_{1,2} = m \pm 2$, $\sigma_{1,2} = 2$ as initial conditions, where m is the melting temperature of the single peak. In step 320, the potential main peak and shoulder (lower) peak are determined as above. In step 330, a determination is made as to whether the height of the lower peak is at least a certain percentage (e.g., 0.05 or more times the higher peak) of the higher peak. If not, a shoulder not present call is returned and Tm determination processing continues as per step 250 in FIG. 5. If so, in step 340, confidence intervals are determined for the various parameters and, in step 350, an R^2 value is determined for the GMM2 equation. In step 360, a determination is made as to whether the confidence intervals of the means are at least a certain number of degrees Centigrade apart from each other (e.g., 3 degrees or more). If not, a shoulder not present call is returned and Tm determination processing continues as per step 250 in FIG. 5. If so, in step 370, a determination is made as to whether the confidence intervals do not contain zero or whether the R^2 value is greater than a threshold (e.g., 0.9). If not, a shoulder not present call is returned and Tm determination processing continues as per step 250 in FIG. 5. If so, in step 380, a determination of whether the second (2d) derivative of the double component Gaussian model (GMM2) evaluated at both μ_1 and μ_2 is negative. If so, a shoulder detected call is made.

Thus, assuming the LM-method converged, in one embodiment, a shoulder is detected if ALL of the following conditions hold true:

- The first derivative data was identified to have exactly one local maximum.
- The lower peak is at least 0.05 times the higher peak in height (e.g., using "exp(-a)").
- The **confidence intervals of the means** are at least a 3 degrees Centigrade apart from each other.

- The confidence intervals of μ_1, μ_2, σ_1 and σ_2 do not contain zero or $R^2 >$ 0.9.

- The **second derivative** of the double component Gaussian model (GMM2) evaluated at **both** μ_1 and μ_2 is negative.

In the case in which these conditions are fulfilled, a Shoulder Detection flag is raised. The above shoulder detection process is described in connection with FIG. 6 for a first derivative comprising one or two peaks. Nevertheless, as mentioned above, it should be understood that this shoulder detection process can be applied to first derivative data presenting up to four peaks.

Expert System Checks

In one embodiment, one or more expert system checks are implemented in step 290 (FIG. 5). In one check, it is determined whether the median of all the entire melting curve data is > 0. If this is false, then the result is reported as Invalid. In another check, it is determined whether the absolute value of the maximum of the Melt Peak fluorescent data is greater than the absolute value of the minimum of the Melt Peak fluorescent data. If this is false, then the result is reported as Invalid. In another check, it is determined whether the median of the raw fluorescence vs. temperature values from the first several (e.g., five) cycles is greater than the median of the raw fluorescence vs. temperature values from the last several (e.g., five) cycles. If this is false, then result is reported as Invalid. In yet another check, a second order polynomial is fit to the raw fluorescence vs. temperature data set. If the R^2 on this fit is greater than a threshold (e.g., 0.99), then no peaks are present and the result is reported as Target Not Detected (TND).

Examples

FIGS. 7a and 7b show a raw data curve and a melting curve, respectively, in the case of a single melting peak. The calculated melting temperature is Tm = 67.8.

FIGS. 8a and 8b show a raw data curve and a melting curve, respectively, in the case of a two melting peaks. The calculated melting temperatures are Tm1 = 58.8 and Tm2 = 67.8.

FIGS. 9a and 9b show a raw data curve and a melting curve, respectively, in the case of a two melting peaks. The calculated melting temperatures are Tm1 = 57.9 and Tm2 = 68.6.

FIGS. 10a and 10b show a raw data curve and a melting curve, respectively, in the case of a single melting peak plus a shoulder. The calculated melting temperatures are Tm1 = 62.4 and Tm2 = 68.8.

FIGS. 11a and 11b show a raw data curve and a melting curve, respectively, in the case of a single melting peak plus a shoulder. The calculated melting temperatures are Tm1 = 61.3 and Tm2 = 68.1.

In a certain embodiment the Tm determination processes according to the invention can be implemented by using conventional personal computer systems including, but not limited to, an input device to input a data set, such as a keyboard, mouse, and the like; a display device to represent a specific point of interest in a region of a curve, such as a monitor; a processing device necessary to carry out each step in the method, such as a CPU; a network interface such as a modem, a data storage device to store the data set, a computer code running on the processor and the like. Furthermore, the processes may also be implemented in a PCR process according to the invention or in a PCR system according to the invention.

It should be appreciated that the Tm determination processes, including the derivative and shoulder determination processes, may be implemented in computer code running on a processor of a computer system. The code includes instructions for controlling a processor to implement various aspects and steps of the Tm determination processes. The code is typically stored on a hard disk, RAM or portable medium such as a CD, DVD, etc. Similarly, the processes may be implemented in a PCR device such as a thermocycler, or other specialized device, including a processor executing instructions stored in a memory unit coupled to the processor. Code including such instructions may be downloaded to the device memory unit over a network connection or direct connection to a code source or using a portable medium as is well known.

One skilled in the art should appreciate that the Tm determination processes of the present invention can be coded using a variety of programming languages such as C, C++, C#, Fortran, VisualBasic, etc., as well as applications such as Mathematica® which provide pre-packaged routines, functions and procedures useful for data visualization and analysis. Another example of the latter is MATLAB®.

An example of system according to the invention is displayed in FIGS. 12 and 13. FIG. 12 shows a general block diagram explaining the relation between software and hardware resources that may be used to implement the processes and system of the invention. The system depicted on FIG. 13 comprises a kinetic PCR analysis module which may be located in a thermocycler device and an intelligence module which is part of the computer system. The data sets (PCR data sets) are transferred from the analysis module to the intelligence module or vice versa via a network connection or a direct connection. The data sets may for example be processed according to the flowcharts as depicted on FIGS. 4, 5 and 6. These flowcharts may conveniently be implemented by a software stored on the hardware of a computer system for example according to the flowchart as depicted on FIG. 12. Referring to FIG. 12, computer system (400) may comprise

receiving means (410) for example for receiving fluorescence data obtained during PCR reactions, calculating means (420) for processing said data according to the processes of the invention, applying means (430) for applying the result obtained by the calculation means, and displaying means (440) for displaying the results on a computer screen. FIG. 13 illustrates the interaction between a thermocycler device and a computer system. The system comprises a kinetic PCR analysis module which may be located in a thermocycler device and an intelligence module which is part of the computer system. The data sets (PCR data sets) are transferred from the analysis module to the intelligence module or vice versa via a network connection or a direct connection. The data sets may be processed according to FIG. 12 by computer code running on the processor and being stored on the storage device of the intelligence module and after processing transferred back to the storage device of the analysis module, where the modified data may be displayed on a displaying device.

While the invention has been described by way of example and in terms of the specific embodiments, it is to be understood that the invention is not limited to the disclosed embodiments. To the contrary, it is intended to cover various modifications and similar arrangements as would be apparent to those skilled in the art.

CLAIMS:

1. A computer-implemented method of determining a melting temperature, Tm, of DNA, the method comprising:

- receiving a dataset representing a melt curve for a DNA sample, the dataset including a plurality of data points each having a pair of coordinate values;

- numerically determining first derivative values for data points of the melt curve;

- determining a baseline for the first derivative values;

- subtracting out the baseline from the first derivative values to produce modified first derivative values;

- determining a first maximum value of the modified first derivative values; and

- outputting said first maximum value, wherein said first maximum value represents a melting temperature Tm of the DNA sample.

2. The method of claim 1, further comprising the steps of:

- calculating an approximation of a curve that fits the modified first derivative values by applying a Levenberg-Marquardt (LM) regression process to a Gaussian Mixture Model function to determine one or more parameters of the function, wherein said parameters include initial conditions, and wherein the first maximum value is used as an initial condition for a first parameter; and

- outputting the first parameter, wherein the determined first parameter represents a melting temperature, Tm, of the DNA sample.

3. The method of claim 2, wherein the Gaussian Mixture Model includes an expression of the form:

$$GMM_1 = Exp(-a_1)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right),$$

wherein μ_1 is the first parameter, and wherein a_1 and σ_1 are additional parameters.

4. The method of claim 1, further including determining whether the first derivative values include a shoulder value proximal to the first maximum value.

5. The method of claim 2, further including displaying the first parameter value.

6. The method of claim 1, further comprising the step of:

- determining a second, a third or a fourth maximum value of the modified first derivative values, wherein the second, third or a fourth maximum value is used as an initial condition for a second, third or fourth parameter; and

- outputting the second, third or fourth parameter, wherein the determined second, third or fourth parameter represents a second, third or fourth melting temperature, Tm2, Tm3 or Tm4 of the DNA sample.

7. The method of claim 6, wherein the Gaussian Mixture Model includes an expression of the form:

$$GMM_2 = Exp(-a_1)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right) + Exp(-a_2)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right)$$

,

wherein μ_1 is the first parameter, μ_2 is the second parameter and wherein a_1, σ_1, a_2 and σ_2 are additional parameters.

8. The method of claim 6, wherein the Gaussian Mixture Model includes an expression of the form:

$$GMM_3 = Exp(-a_1)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right) + Exp(-a_2)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right) + Exp(-a_3)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_3}{\sigma_3}\right)^2\right)$$

wherein μ_1 is the first parameter, μ_2 is the second parameter, μ_3 is the third parameter and wherein a_1, σ_1, a_2, σ_2, a_3 and σ_3 are additional parameters.

9. The method of claim 6, wherein the Gaussian Mixture Model includes an expression of the form:

$$GMM_4 = Exp(-a_1)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right) + Exp(-a_2)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right) + Exp(-a_3)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_3}{\sigma_3}\right)^2\right) + Exp(-a_4)\cdot Exp\left(-\frac{1}{2}\left(\frac{x-\mu_4}{\sigma_4}\right)^2\right)$$

wherein μ_1 is the first parameter, μ_2 is the second parameter, μ_3 is the third parameter μ_4 is the fourth parameter and wherein a_1, σ_1, a_2, σ_2, a_3, σ_3, a_4 and σ_4 are additional parameters.

10. A computer readable medium including code for controlling a processor to determine a melting temperature, Tm, of DNA, the code including instructions to:

- receive a dataset representing a melt curve for a DNA sample, the dataset including a plurality of data points each having a pair of coordinate values;

- numerically determine first derivative values for data points of the melt curve;

- determine a baseline for the first derivative values;

- subtract out the baseline from the first derivative values to produce modified first derivative values;

- determine a first maximum value of the modified first derivative values; and

- output said first maximum value, wherein said first maximum value represents a melting temperature Tm of the DNA sample.

11. The computer readable medium of claim 10, wherein the code is adapted to perform the steps of any one of claims 2 to 9.

12. A kinetic Polymerase Chain Reaction (PCR) system, comprising:

- a kinetic PCR analysis module that generates a melt curve dataset representing a DNA melt curve, said dataset including a plurality of data points, each having a pair of coordinate values; and

- an intelligence module adapted to process the melt curve dataset to determine a Tm value, by

- receiving a dataset representing a melt curve for a DNA sample, the dataset including a plurality of data points each having a pair of coordinate values;

- numerically determining first derivative values for data points of the melt curve;

- determining a baseline for the first derivative values;

- subtracting out the baseline from the first derivative values to produce modified first derivative values;

- determining a first maximum value of the modified first derivative values; and

- outputting said first maximum value, wherein said first maximum value represents a melting temperature Tm of the DNA sample.

13. The system of claim 12, wherein the kinetic PCR analysis module is resident in a kinetic thermocycler device, and wherein the intelligence module includes a processor communicably coupled to the analysis module or wherein the intelligence module includes a processor resident in a computer system coupled to the analysis module by one of a network connection or a direct connection.

14. The system of claim 12, further including a display module, wherein outputting includes displaying the Tm value on the display module.

15. The system of claim 12, wherein the system is adapted to perform the steps of any one of claims 2 to 9.

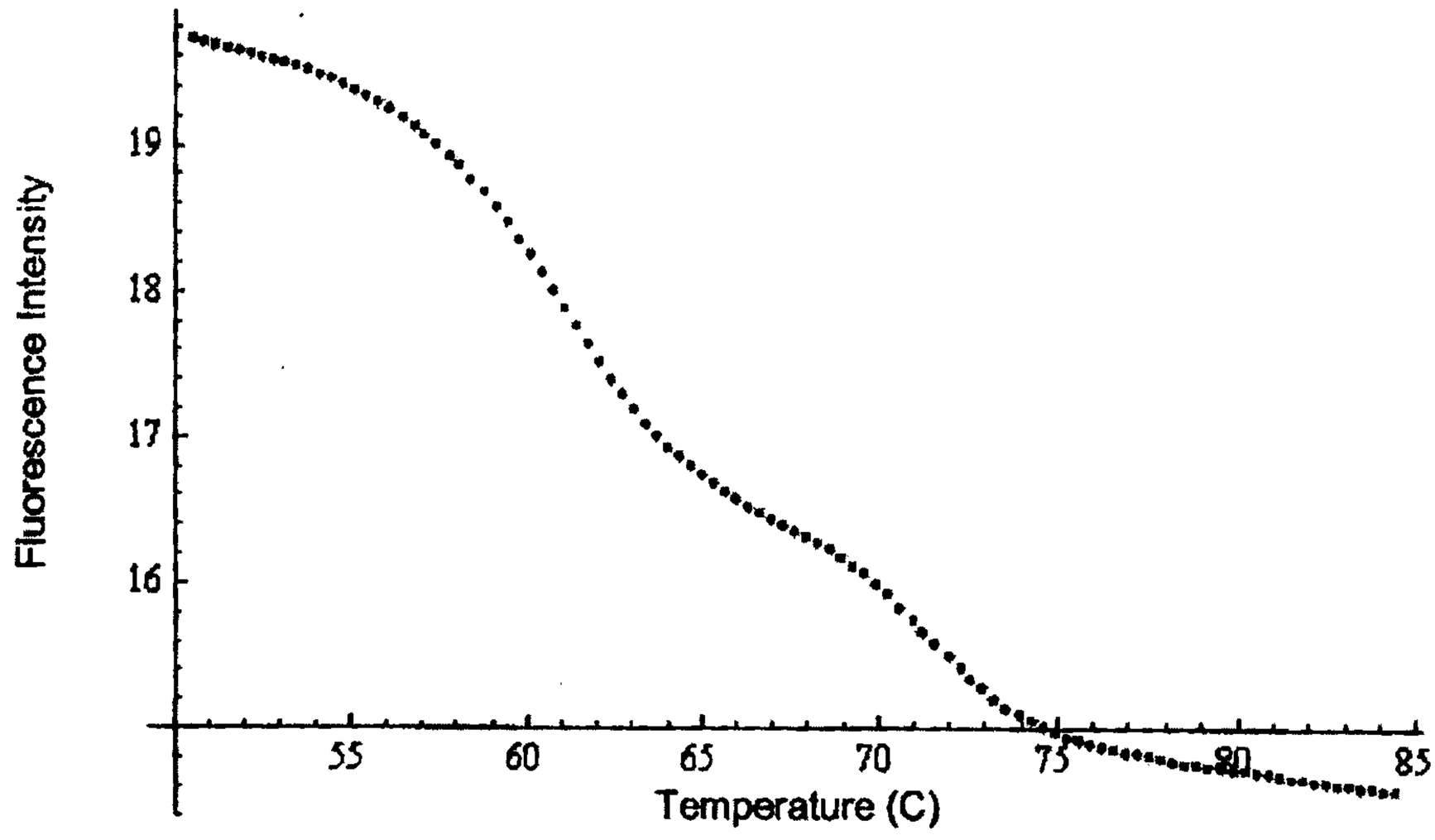


Figure 1: Fluorescence Intensity vs. Temperature

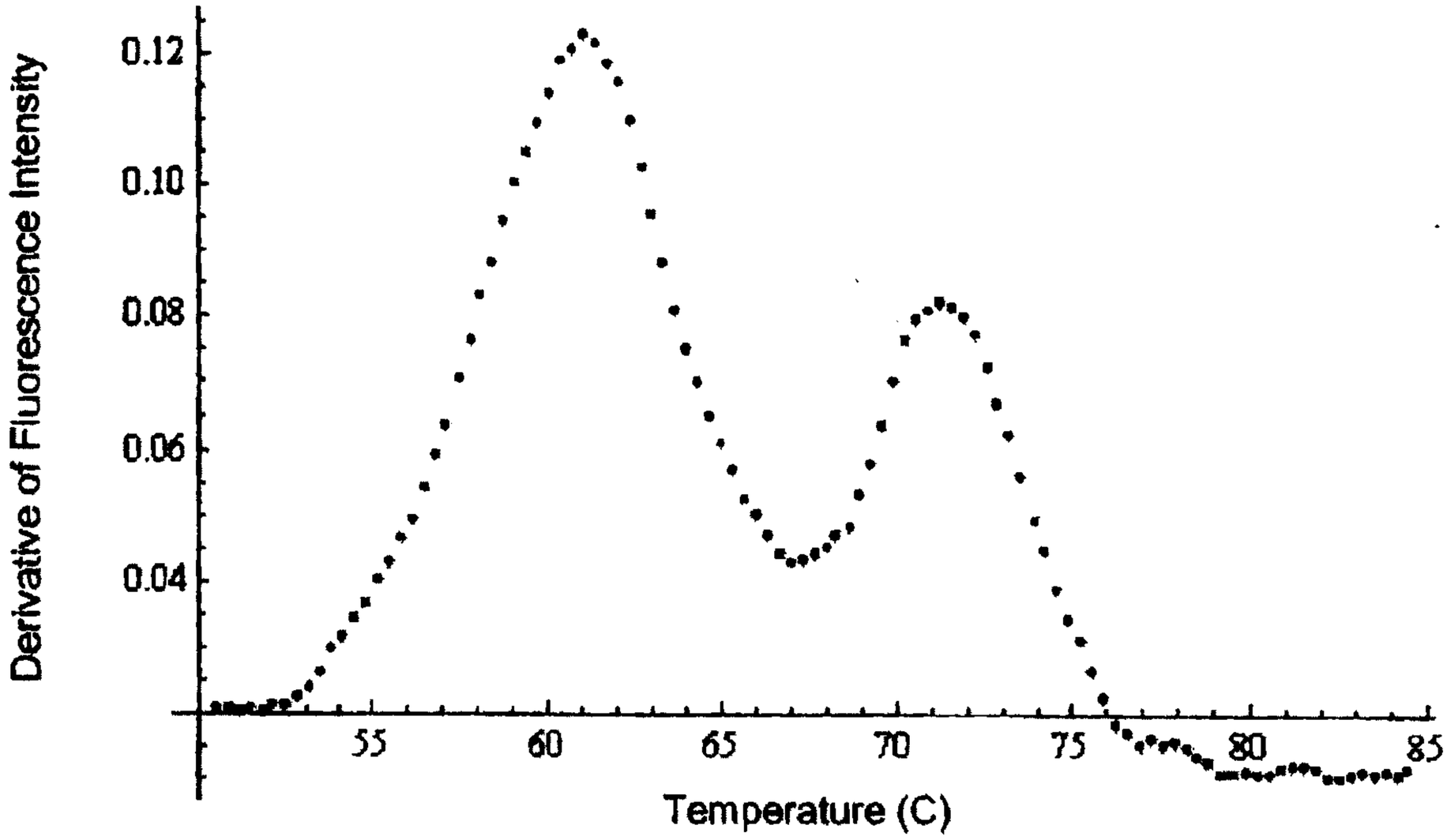


Figure 2: First Derivative of Fluorescence Intensity vs. Temperature

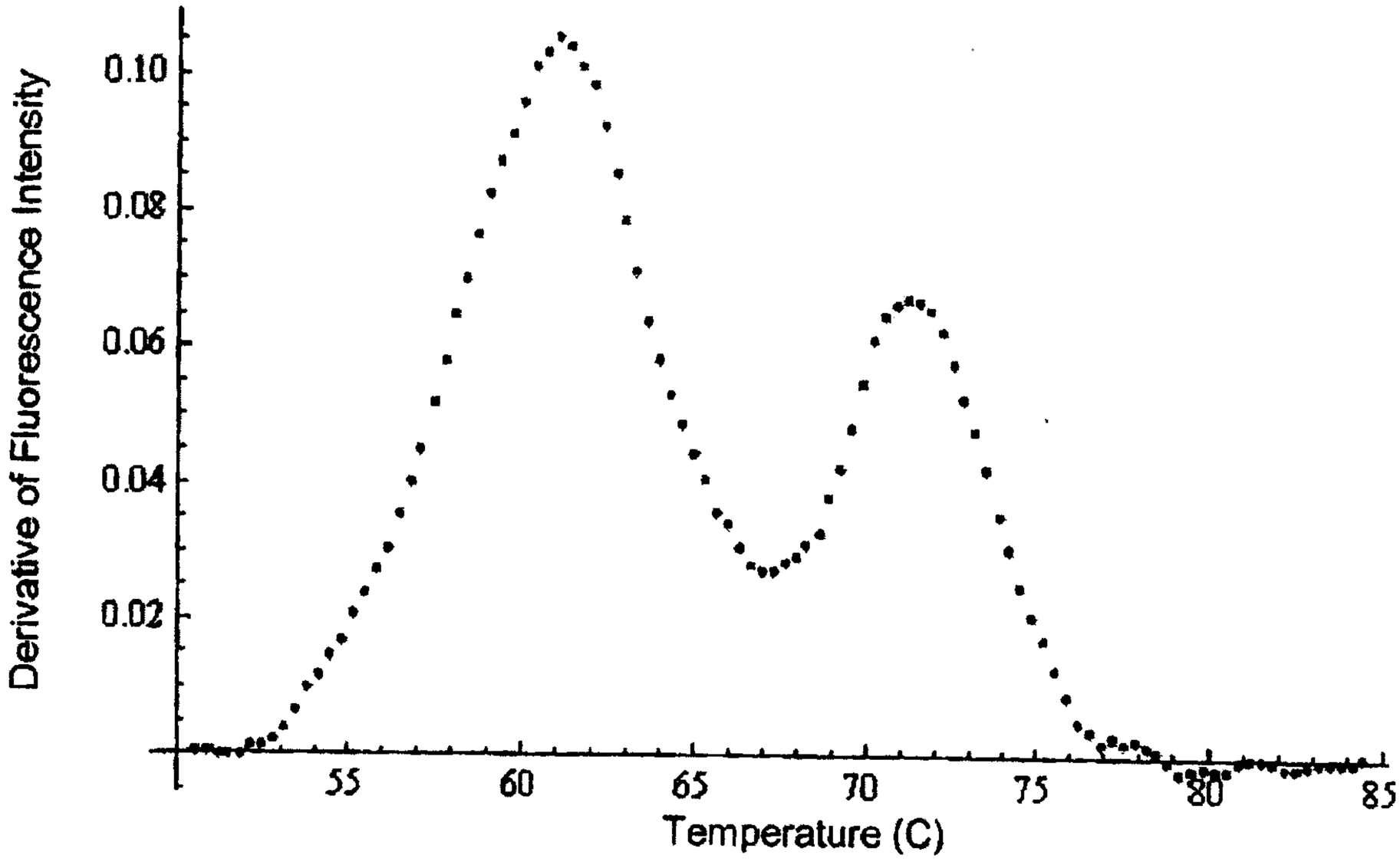


Figure 3: Baseline Subtracted First Derivative of Fluorescence Intensity vs. Temperature

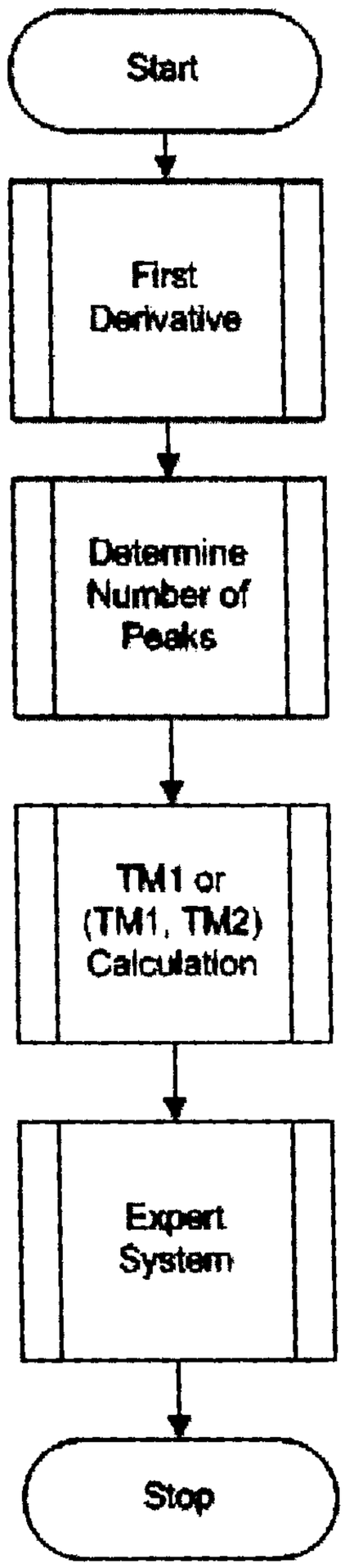


Figure 4: Major Processing Steps of the Equation-Less Melting Algorithm

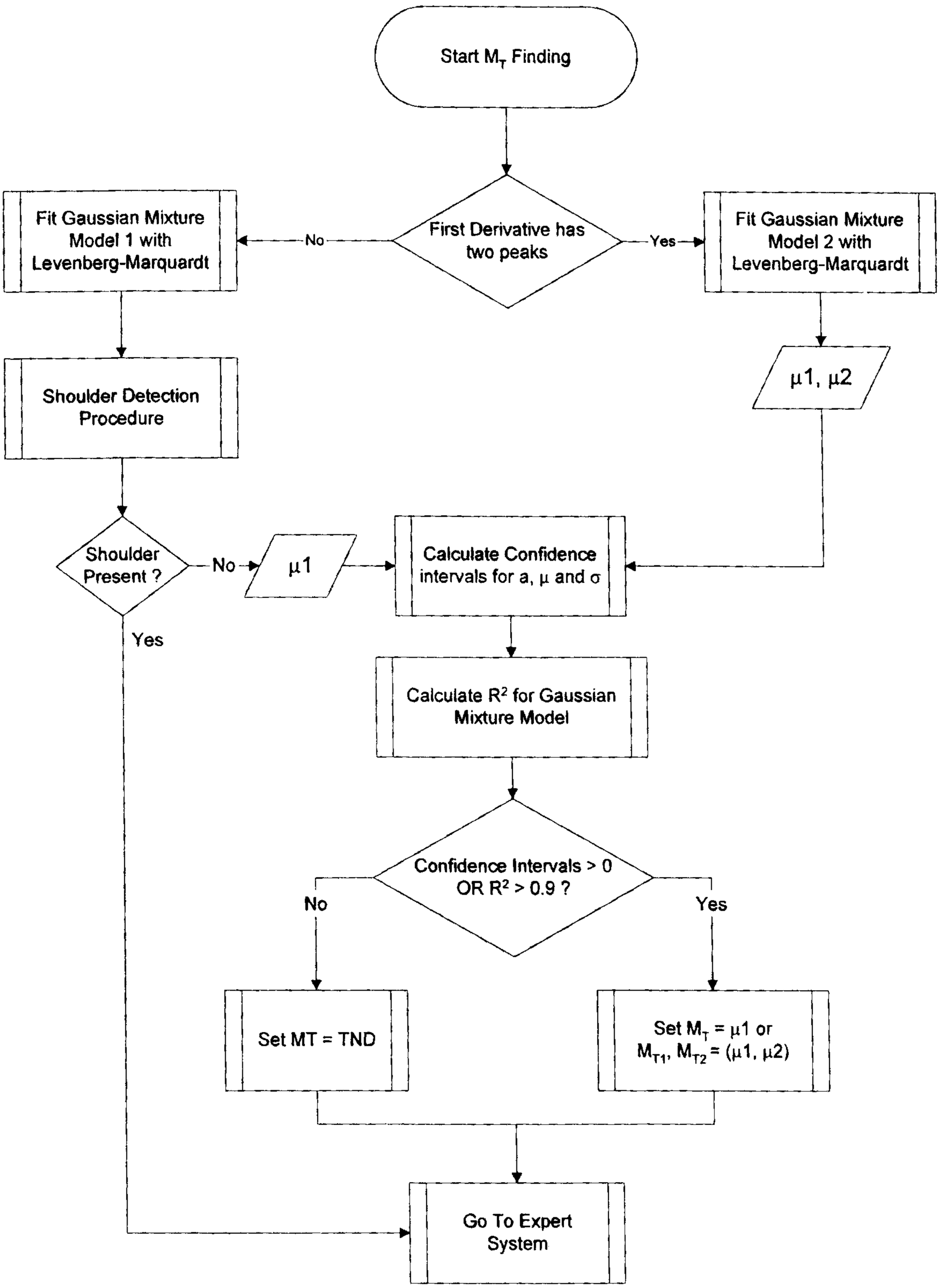


Figure 5: Major Processing Steps of the Equation-Less Melting Algorithm

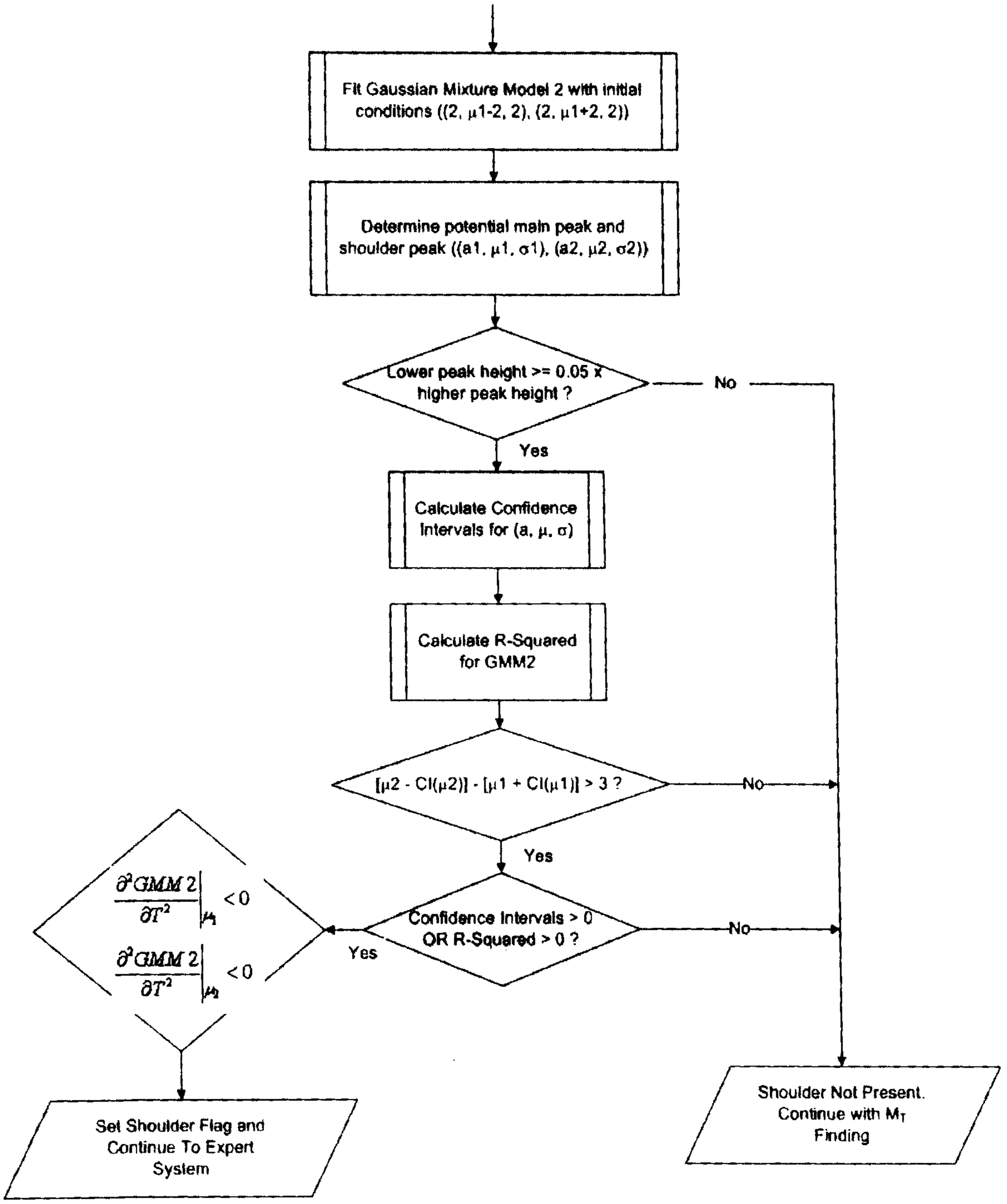


Figure 6: Major Processing Steps of the Equation-Less Melting Algorithm

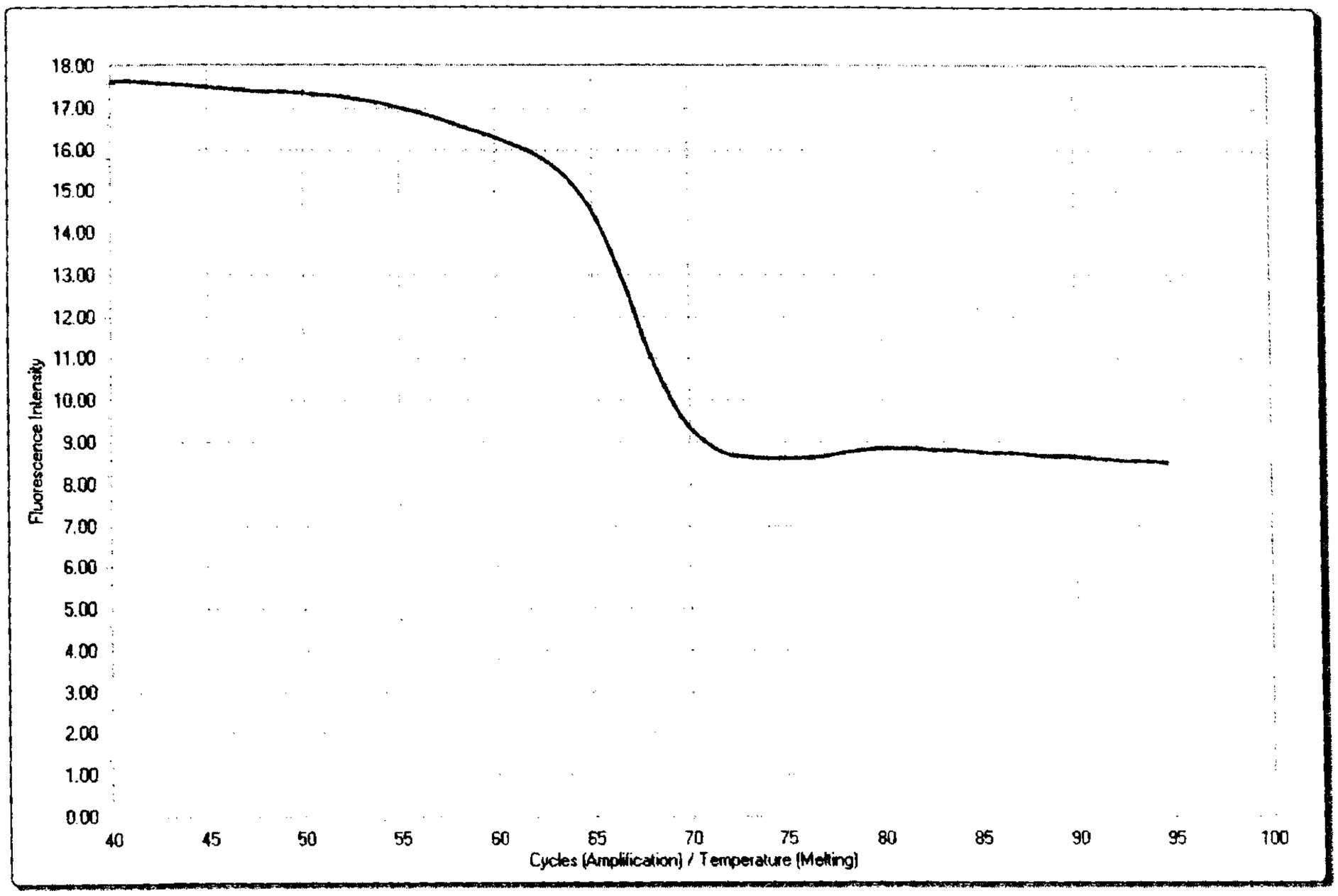


Figure 7a

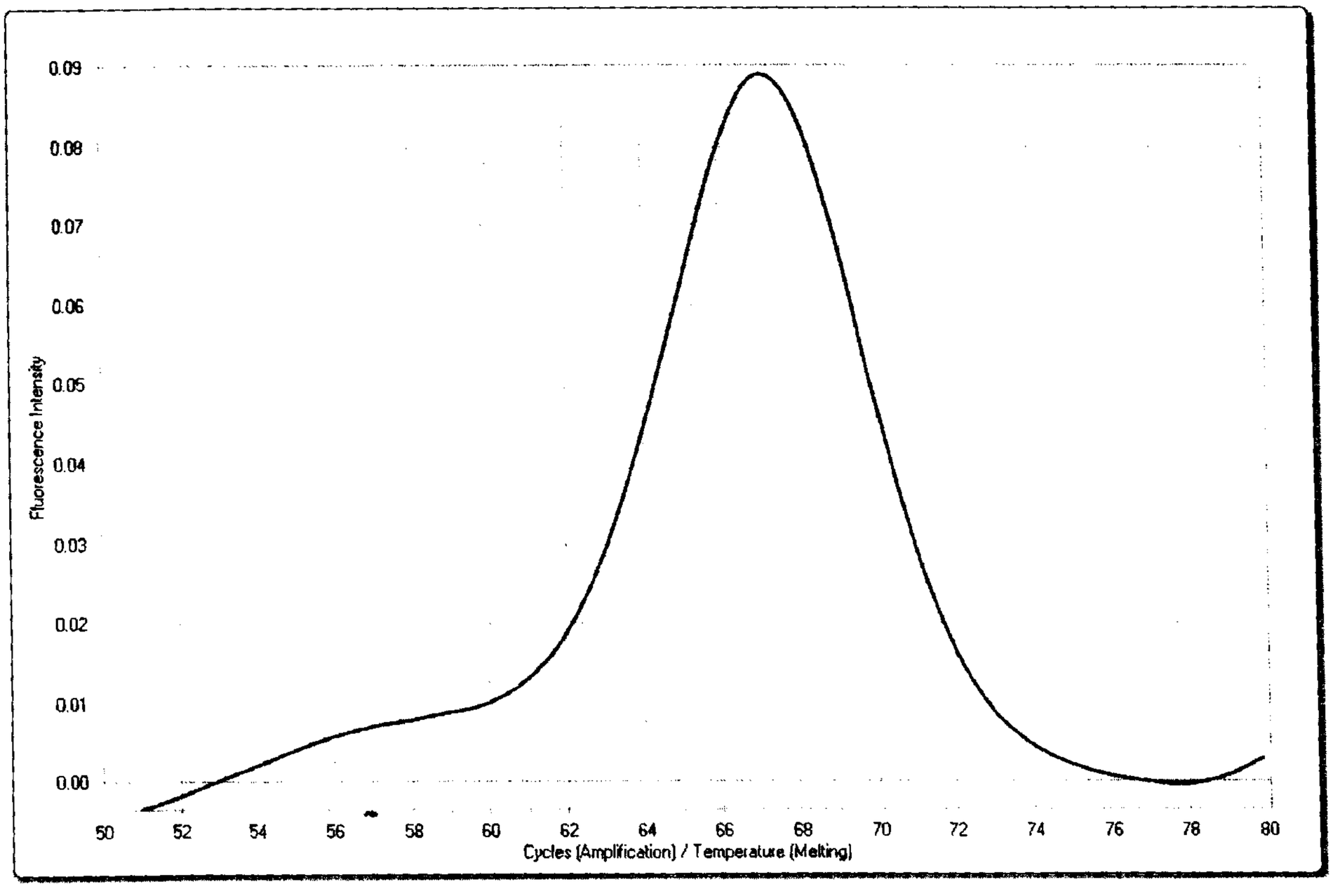


Figure 7b

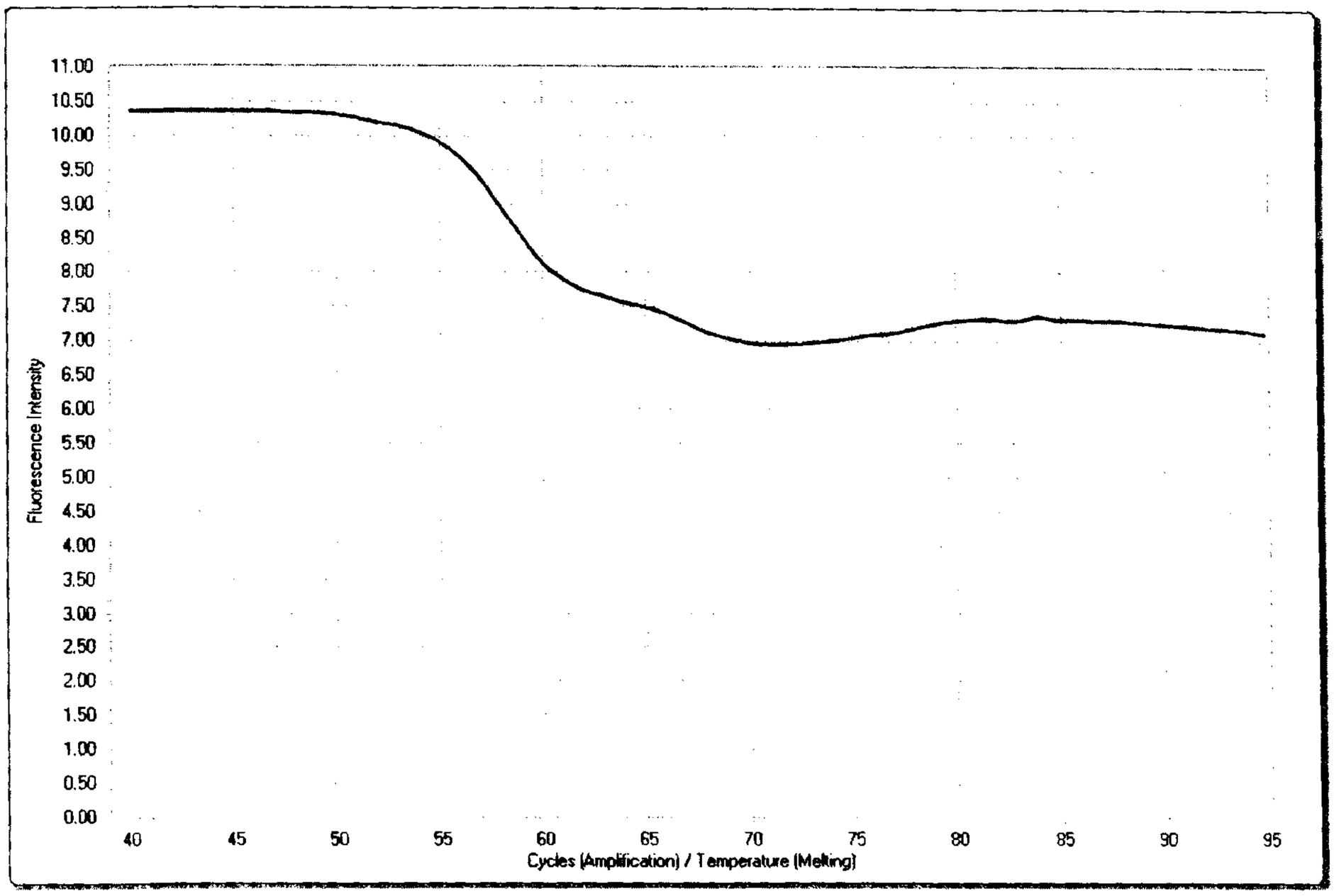


Figure 8a

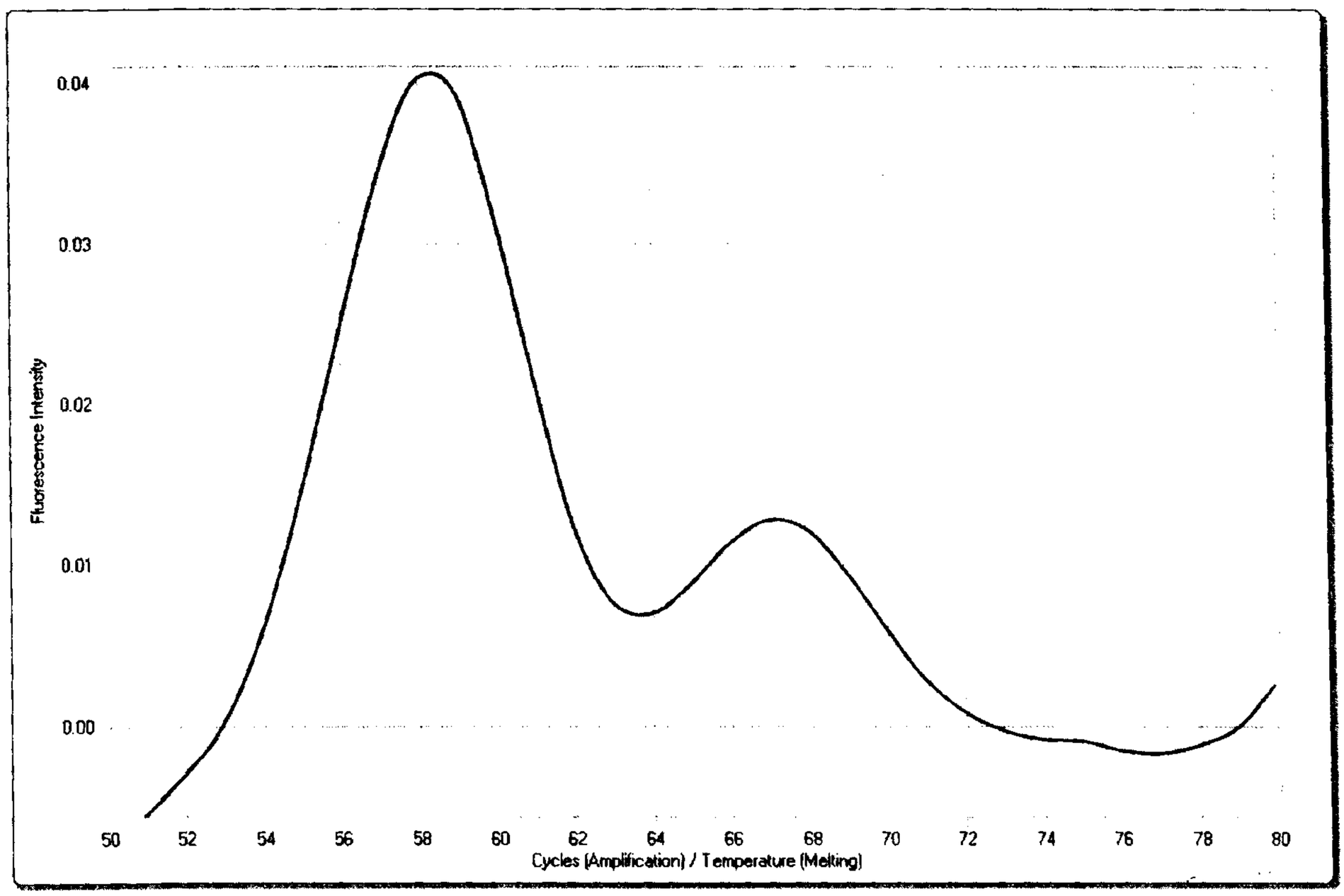


Figure 8b

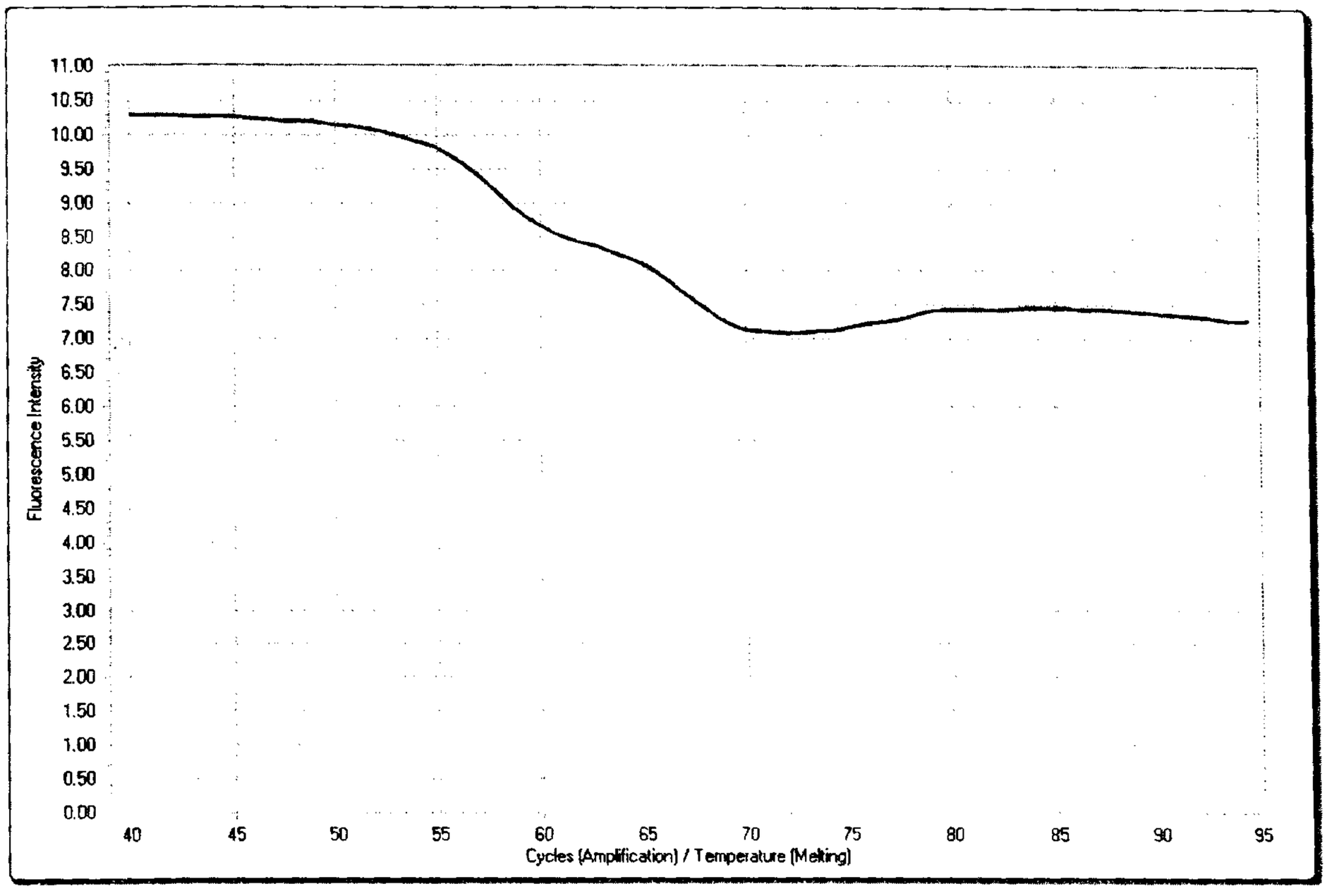


Figure 9a

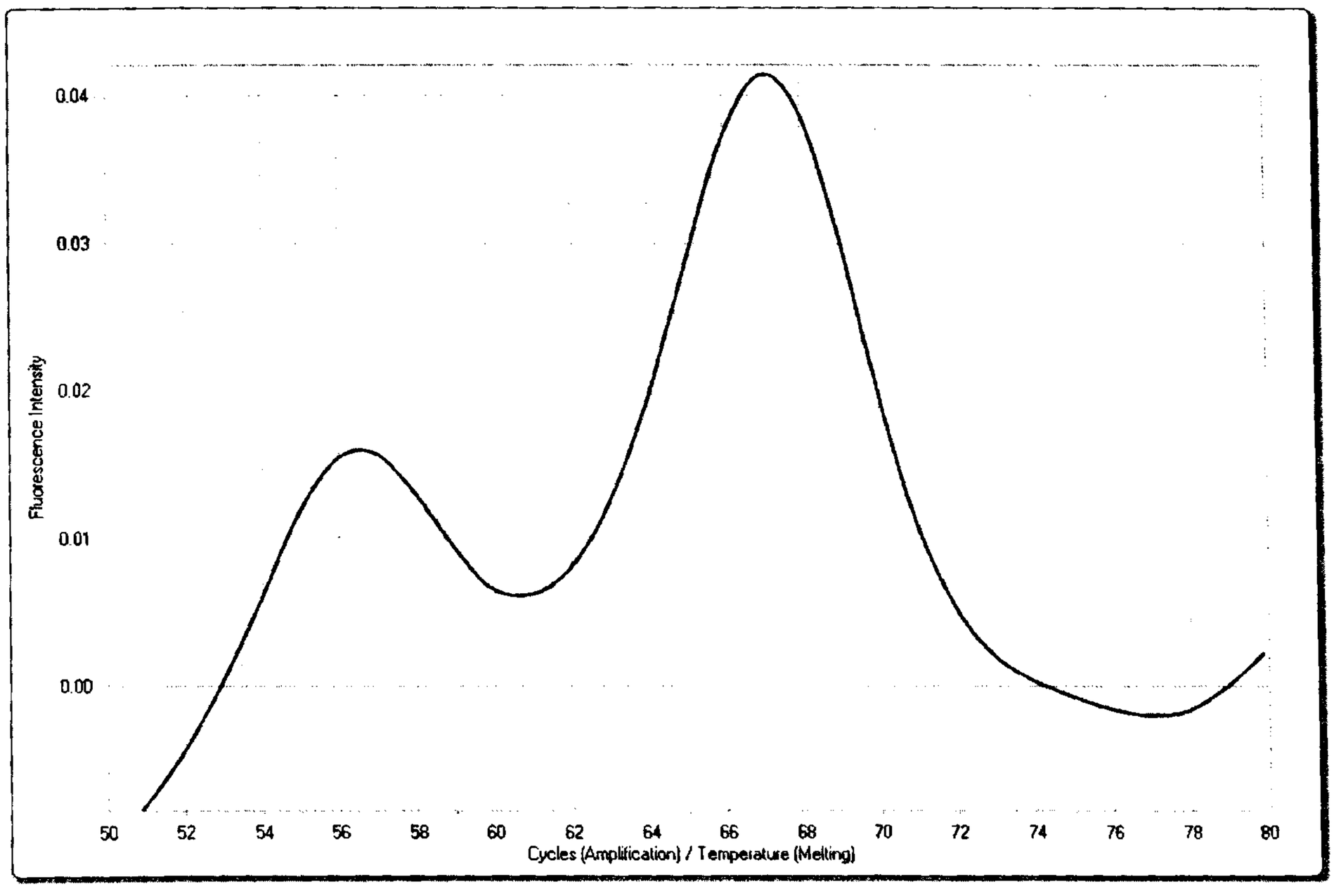


Figure 9b

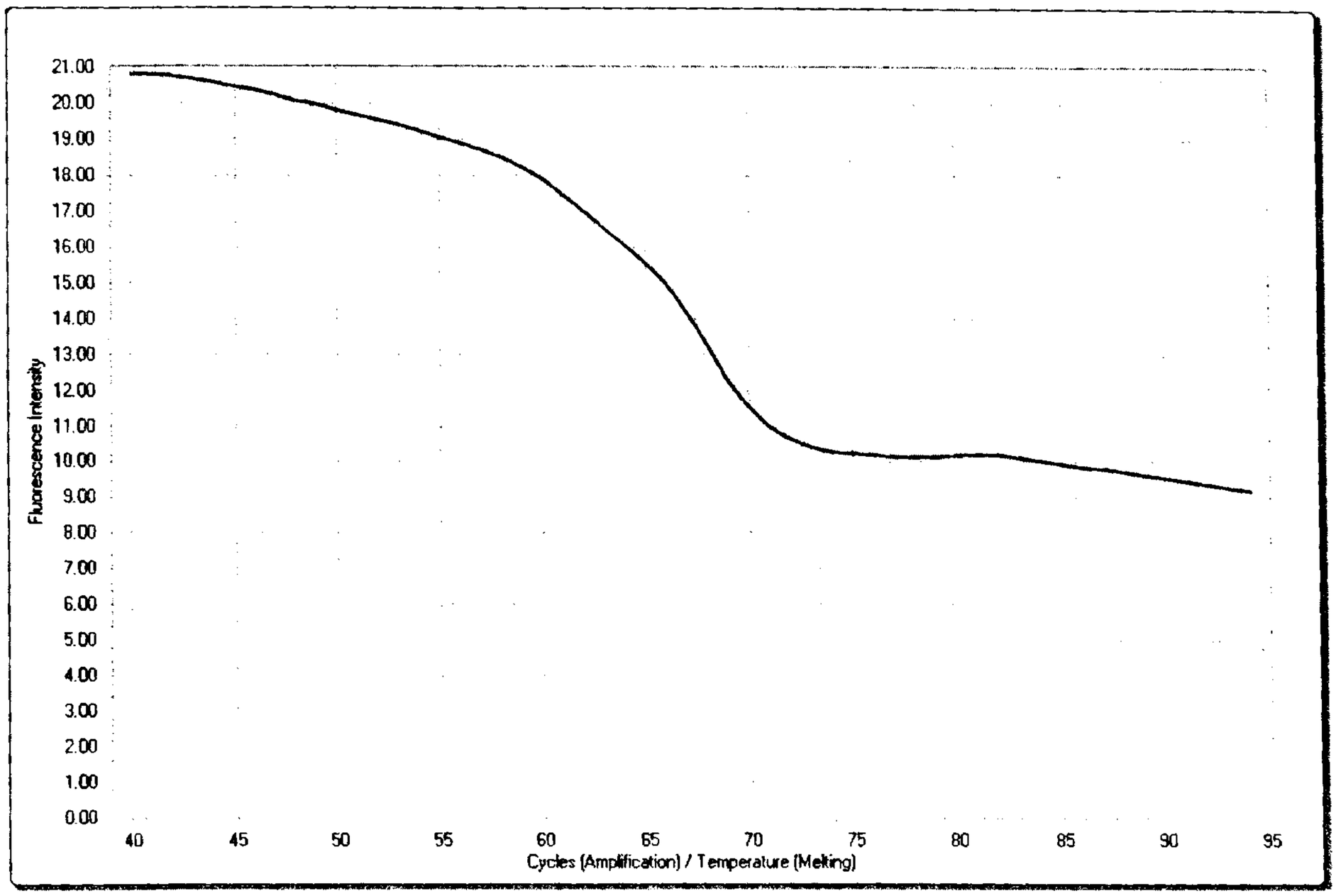


Figure 10a

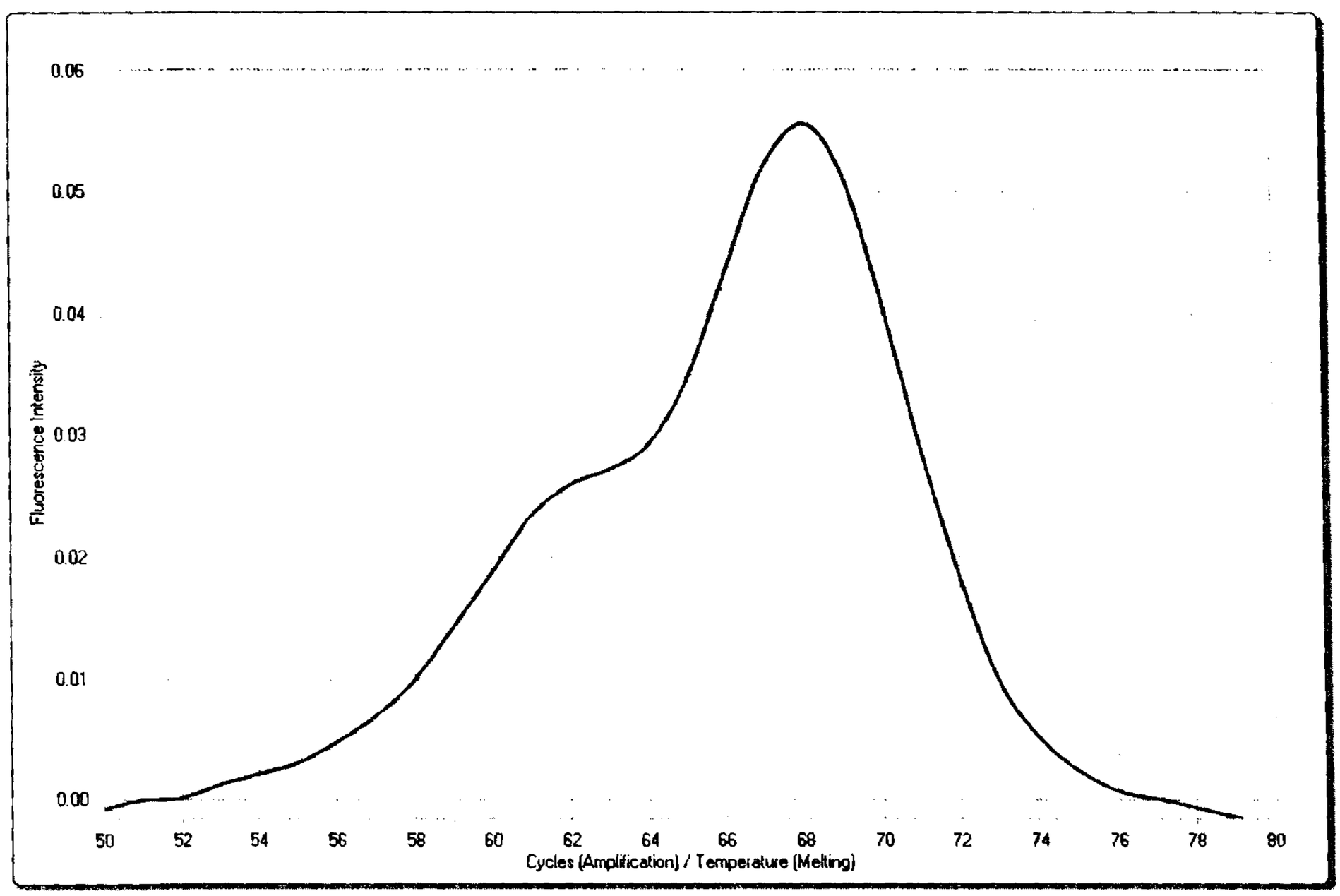


Figure 10b

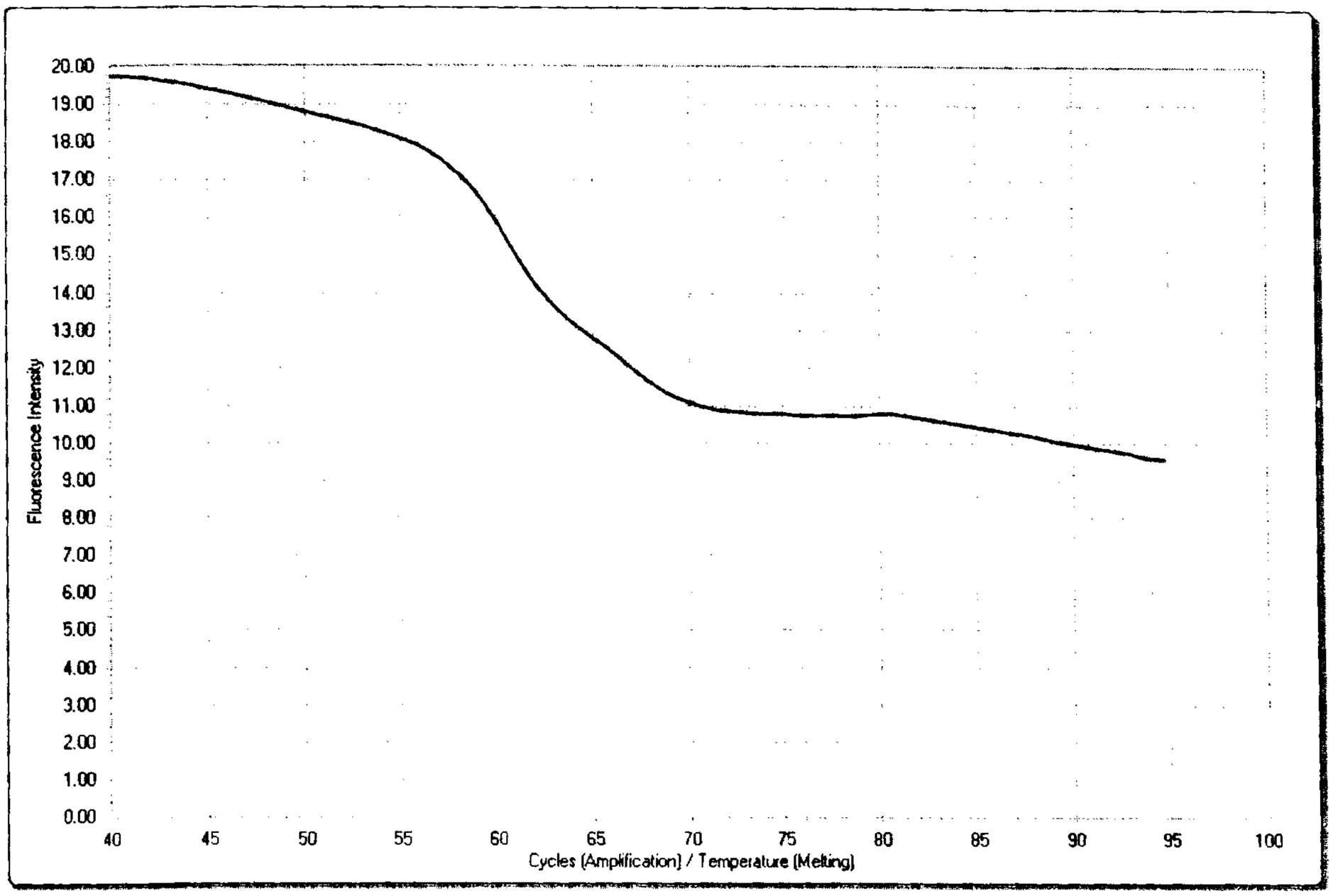


Figure 11a

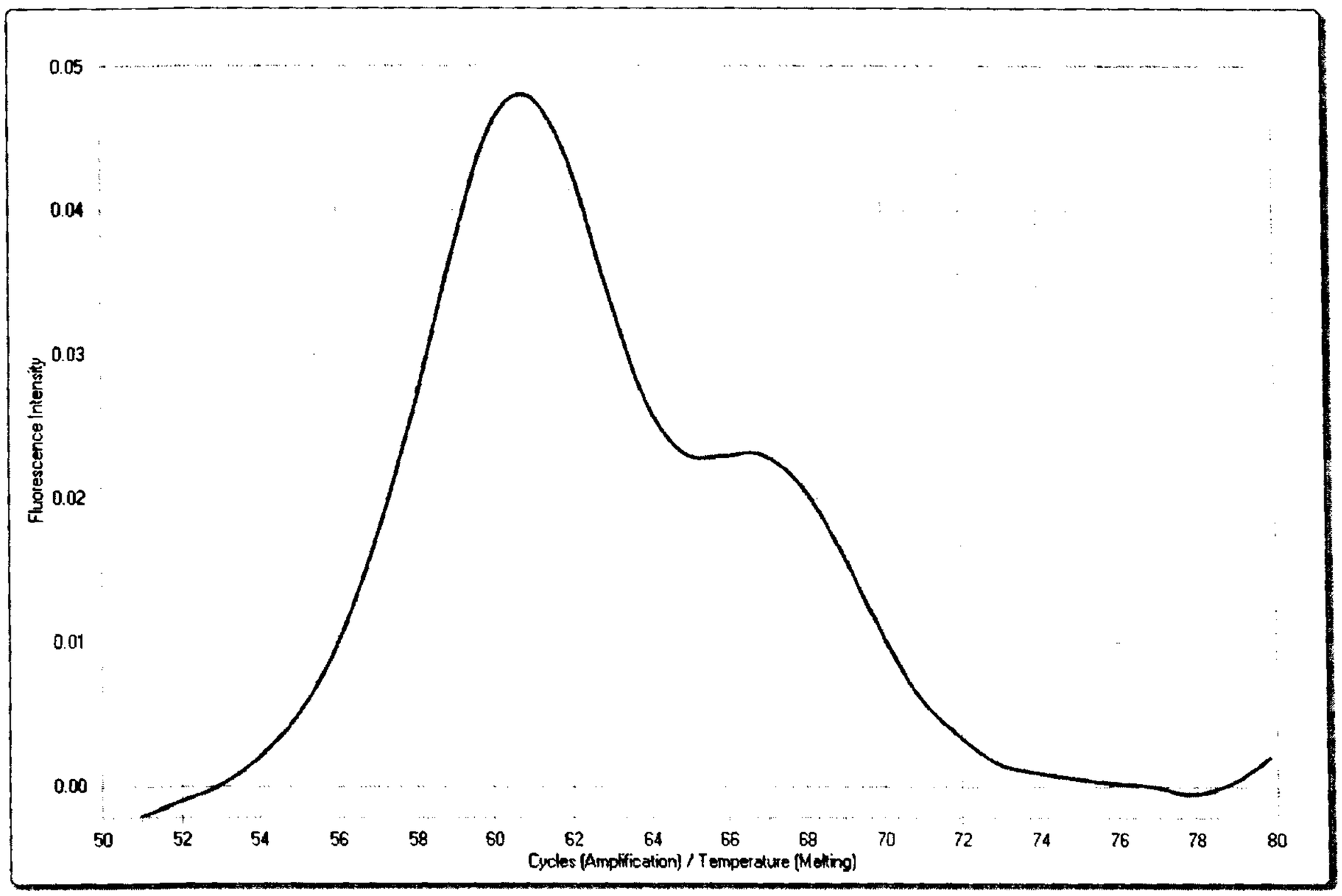


Figure 11b

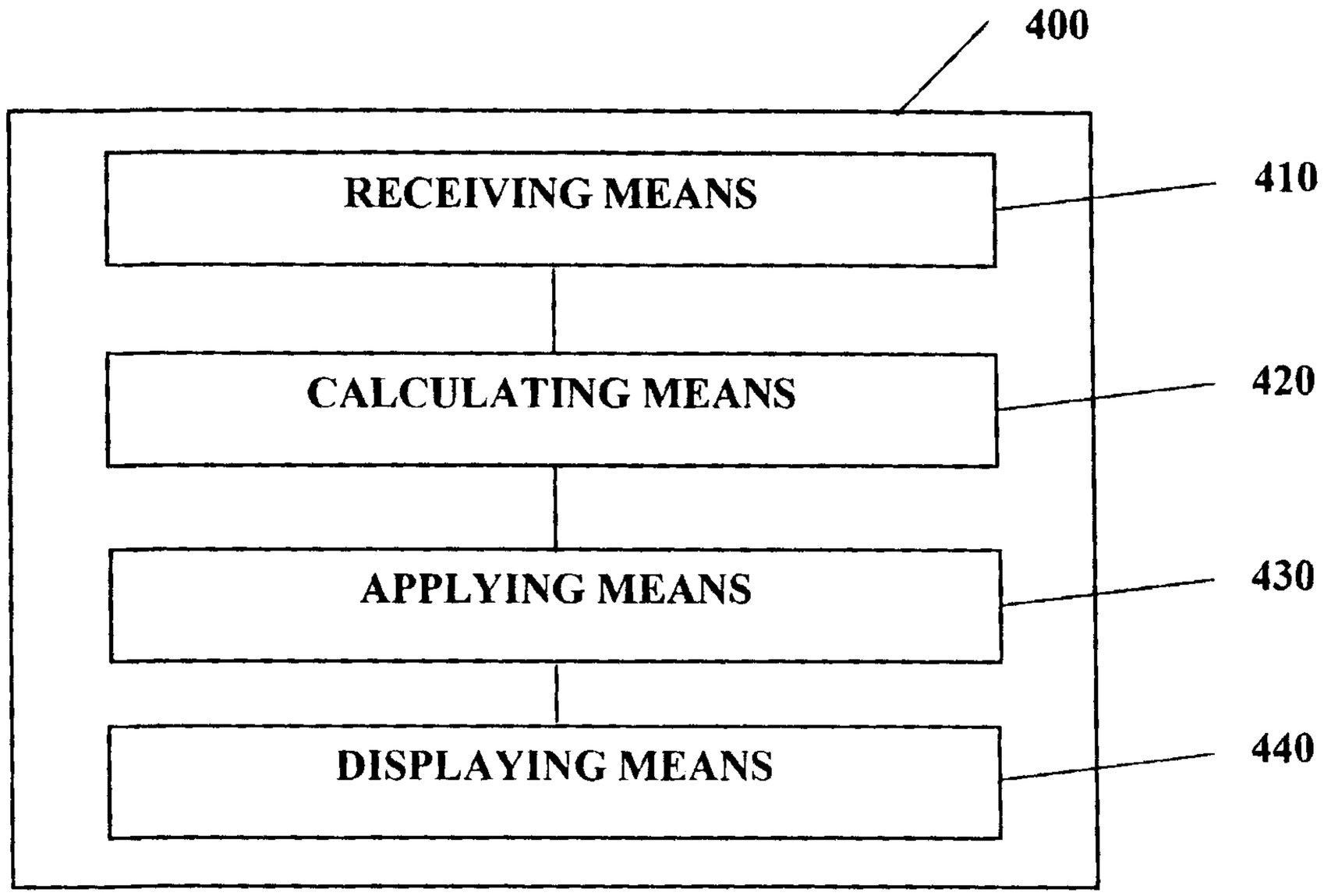

400
RECEIVING MEANS
410
CALCULATING MEANS
420
APPLYING MEANS
430
DISPLAYING MEANS
440

Figure 12

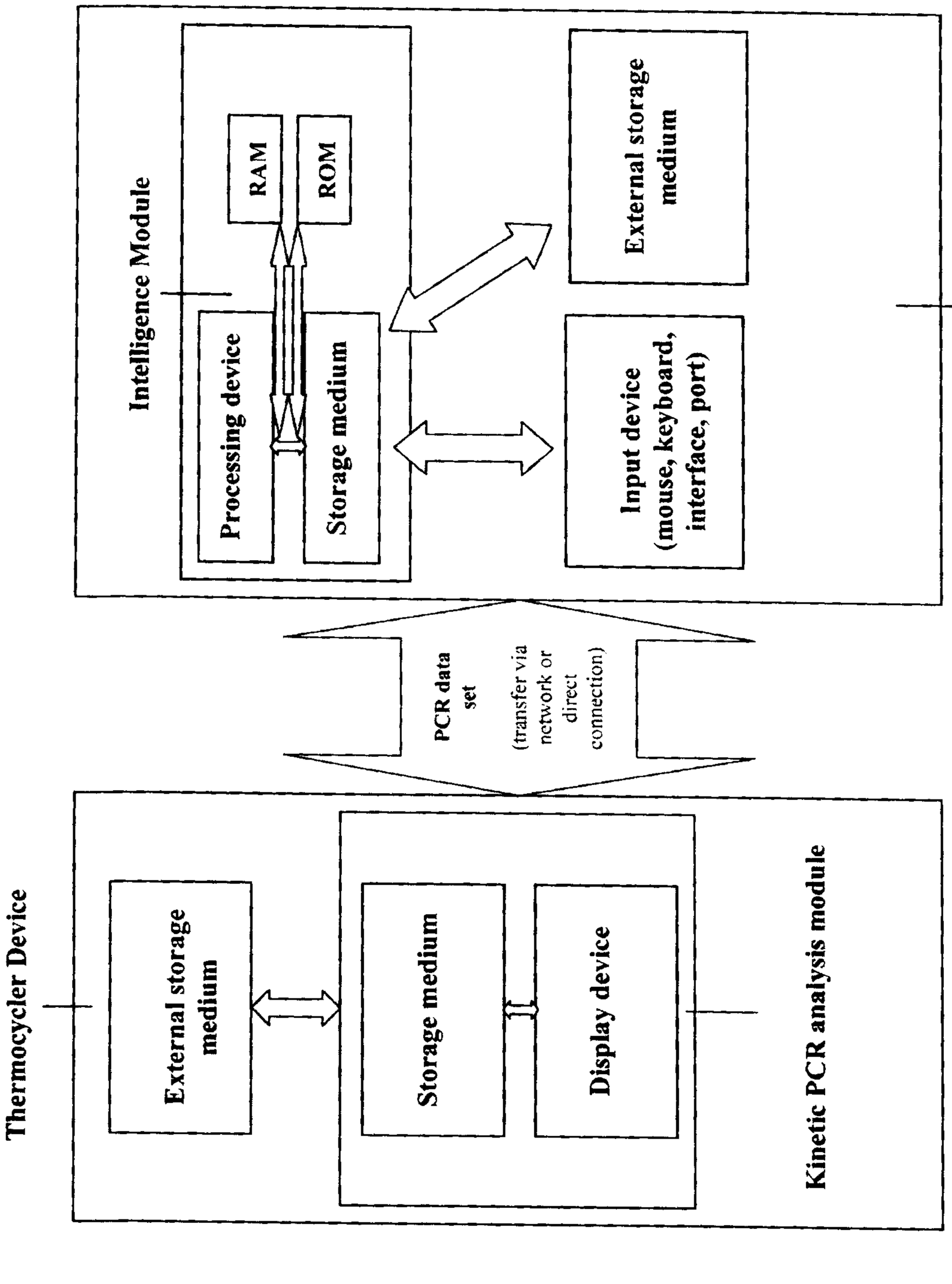
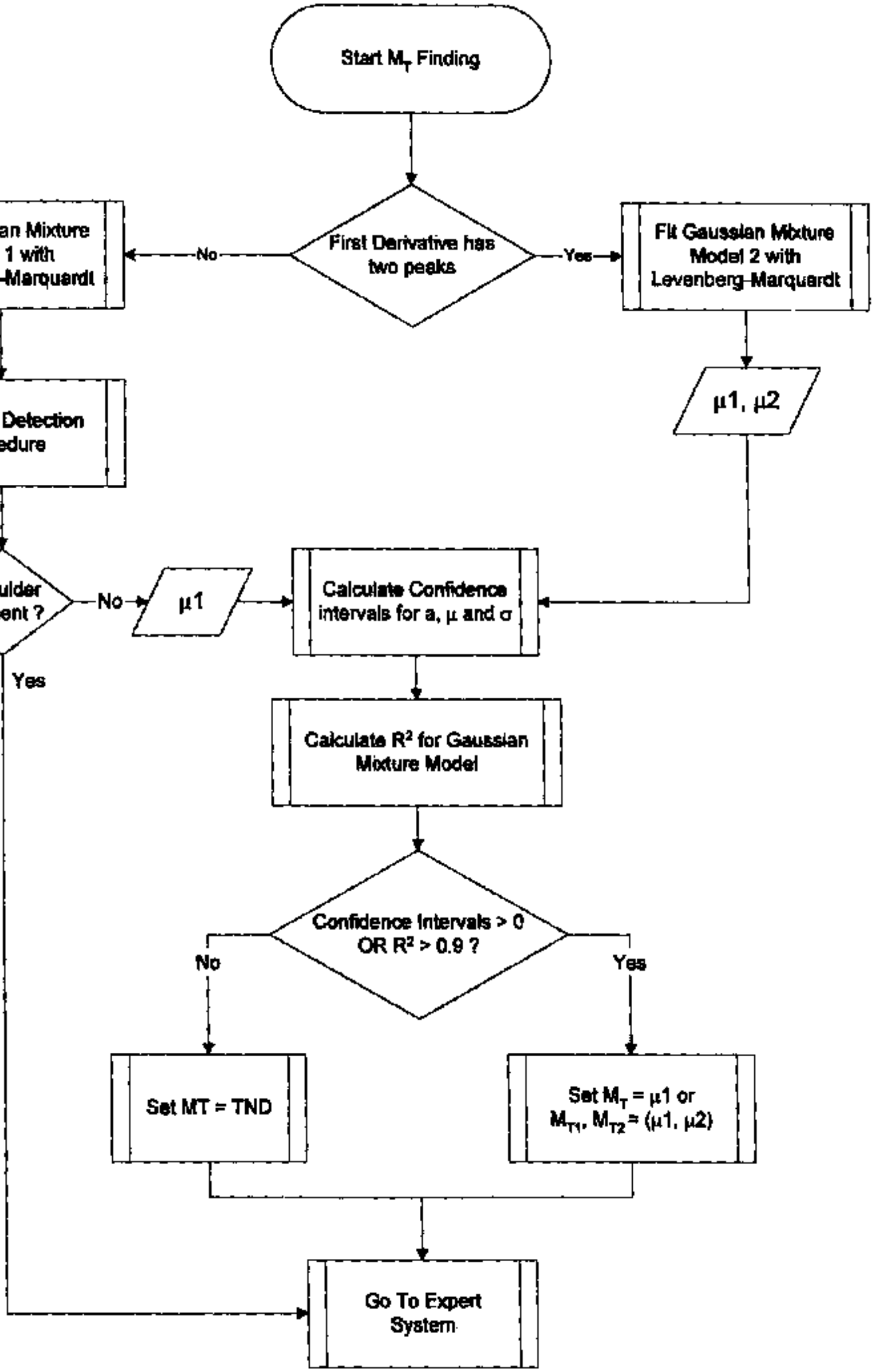

Intelligence Module
RAM
ROM
Processing device
Storage medium
External storage medium
Input device (mouse, keyboard, interface, port)
Computer System
PCR data set
(transfer via network or direct connection)
Thermocycler Device
External storage medium
Storage medium
Display device
Kinetic PCR analysis module

Figure 13



Major Processing Steps of the Equation-Less Melting Algorithm