
**МАТЕМАТИЧЕСКИЕ МЕТОДЫ
И МОДЕЛИРОВАНИЕ В ПРИБОРОСТРОЕНИИ**

УДК 543.426; 543.9

© Д. А. Белов, Ю. В. Белов, В. В. Манойлов

**РАЗРАБОТКА МЕТОДИК ОБРАБОТКИ
СИГНАЛОВ ПЛАВЛЕНИЯ ДНК**

Выбраны условия и параметры реализации методик обработки сигналов плавления ДНК. В разработанных методиках использованы усовершенствованные модели сигнала на базе нелинейной сигмоидальной функции и ее производной, модель на основе функции Гаусса, модель на основе степенных полиномов 3-й и 7-й степеней. Выполнено сравнение результатов применения методик для обработки реальных данных 8 проб, полученных на приборе АНК-32. Разработан алгоритм полной автоматизации вычислений в программной среде MATLAB. Показана возможность разделения систематической и случайной погрешностей. Разработанные методики использованы для определения неравномерности температуры пробирок анализатора нуклеиновых кислот.

Кл. сл.: ДНК, анализатор нуклеиновых кислот, методики плавления, неравномерность температуры пробирок

ВВЕДЕНИЕ

Метод плавления ДНК основан на эффекте перехода двойной спирали молекулы ДНК в клубкообразное состояние при повышении температуры. При этом наблюдается резкое изменение интенсивности сигналов флуоресценции красителей, присоединенных к молекуле ДНК [1, 2].

В качестве температуры плавления ДНК (T_m) обычно принимается положение на температурной оси максимума первой производной сигналов плавления. Значение T_m зависит от длины фрагмента ДНК и процентного содержания суммы нуклеотидов С и G [2–6].

Для выполнения операции плавления ДНК необходимо выбрать пределы изменения температуры, шаг изменения температуры (от 0.1 до 1 °С) и время наблюдения флуоресценции при каждом шаге (20 с и более). Эти величины определяют трудоемкость (длительность) эксперимента, а также отношение сигнала к шуму. Все эти параметры влияют на погрешность измерения величины T_m .

В настоящей статье для формулировки адекватной задачи методики обработки сигналов плавления ДНК (СП) использованы и сравниваются аппроксимационные усовершенствованные модели СП на базе нелинейной сигмоидальной функции (СФ) и ее производной [7, 8], модель на основе функции Гаусса (ФГ) и модели на основе степенных полиномов. Определены особенности этих моделей и выполнено сравнение результатов их применения. В работе использованы вычислительные средства пакета MATLAB.

**УСЛОВИЯ И ПАРАМЕТРЫ РЕАЛИЗАЦИИ
МЕТОДА ПЛАВЛЕНИЯ ДНК**

Влияние шумов, дрейфа нулевой линии и дискретизации сигналов может быть в значительной степени уменьшено путем обработки сигналов с помощью аппроксимирующей функции.

Можно принять в качестве критериев успешности методики обработки СП величину погрешности измерения T_m , длительность эксперимента, количество параметров и критичность выбора параметров аппроксимирующей функции.

Исходные (сырые) данные кривой плавления ДНК получены на анализаторе нуклеиновых кислот АНК-32, который серийно выпускается в Институте аналитического приборостроения РАН (ИАП РАН). В качестве образца был использован комплекс зонд—ДНК—мишень с флуорофором (ROX) в комбинации с гасителем производства ЗАО "СИНТОЛ" [9].

Для обеспечения погрешности T_m порядка 0.1 °С выбраны следующие условия: диапазон температуры 82–90 °С, шаг изменения температуры 0.2 °С, время наблюдения при каждом шаге 60 с (общее время анализа примерно 50 мин).

**РЕЗУЛЬТАТЫ РЕАЛИЗАЦИИ МЕТОДИКИ
ОБРАБОТКИ СП НА ОСНОВЕ
УСОВЕРШЕНСТВОВАННОЙ СФ-МОДЕЛИ**

Усовершенствованную модель кривой плавления на базе СФ предлагается выразить для использования в программе MATLAB в виде формулы

Характеристики СП, вычисленные по разным методикам для 8 последовательно расположенных проб

№ пробы	Точки T_m , вычисленные по методикам обработки СП						Уточнение данных для методики Полином-7	
	СФ	Произв. СФ	Гаусс	Гаусс М	Полином-3	Полином-7	T_{mt}	T_{md}
	Д. 1	Д. 2	Д. 3	Д. 4	Д. 5	Д. 6		
B1	85.54	85.52	85.70	85.5	84.99	85.52	85.53	-0.01
B2	85.50	85.52	85.66	85.46	84.95	85.51	85.49	0.02
B3	85.49	85.48	85.67	85.46	84.93	85.46	85.47	-0.01
B4	85.49	85.49	85.65	85.45	84.96	85.48	85.49	-0.01
B5	85.56	85.57	85.73	85.53	85.06	85.59	85.54	0.06
B6	85.58	85.61	85.75	85.55	85.08	85.58	85.61	-0.03
B7	85.68	85.75	85.842	85.65	85.21	85.68	85.72	-0.04
B8	85.84	85.93	86.00	85.81	85.45	85.89	85.86	0.03
Стандартные отклонения								
dT_{mco}	0.12	0.16	0.12	0.12	0.18	0.14	—	—
dT_{ms}	0.03	0.02	0.02	0.02	0.03	0.03	—	0.03

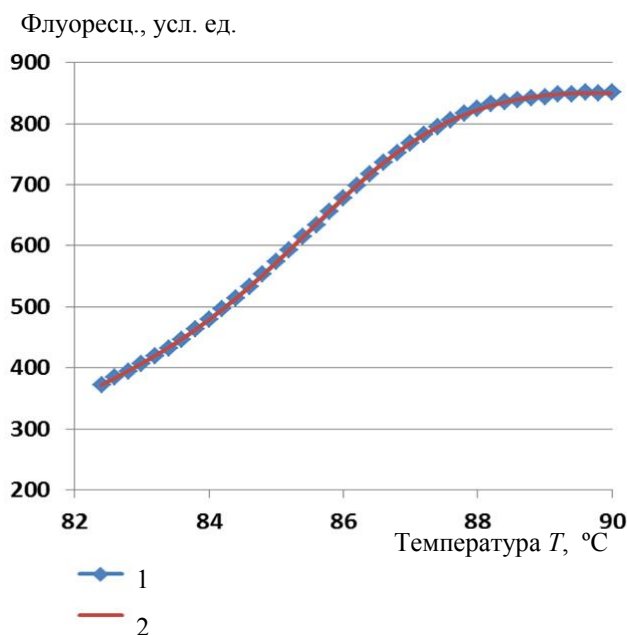


Рис. 1. График кривой плавления ДНК одной из проб (1) и график модельной СФ (2)

$$f = @ (x, u) \quad (x(2) ./ (1 + \exp ((x(1) - u) ./ x(3)))) + x(6) * u.^2 + x(5) * u + x(4), \quad (1)$$

где

f — относительная величина интенсивности флуоресценции при температуре образца u (°C);

$x(1)$ — значение температуры плавления T_m (°C) с шагом 0.1 (°C);

$x(2)$ — относительная интенсивность (размах) участка плавления;

$x(3)$ — коэффициент, который определяется наклоном кривой;

$x(4)$, $x(5)$, $x(6)$ — коэффициенты, учитывающие смещение, а также линейную и квадратичную зависимости нулевой линии от температуры.

Результаты обработки СП по различным рассматриваемым в статье аппроксимационным методикам для 8 проб сведены в таблице. Столбцам данных для краткого их обозначения присвоены номера Д. 1–6.

Значения вычисленных T_m по методике "Аппроксимация СФ" приведены в столбце СФ (Д. 1).

На рис. 1 приведен график кривой плавления ДНК одной из проб. На него наложен график модельной СФ с оптимизированными параметрами.

Можно отметить хорошее совпадение этих графиков.

РЕЗУЛЬТАТЫ РЕАЛИЗАЦИЯ МЕТОДИКИ ОБРАБОТКИ СП НА ОСНОВЕ УСОВЕРШЕНСТВОВАННОЙ МОДЕЛИ ПРОИЗВОДНОЙ СФ

Можно отметить некоторые особенности дифференцирования СП.

Дифференцирование предложено выполнять по формуле (2):

$$V(j) = (s(j+2) - s(j)) / 2, \quad (2)$$

где $V(j)$ — результат дифференцирования на j -м шаге; $s(j+2)$ и $s(j)$ — исходные данные для $(j+2)$ и (j) шагов соответственно.

В результате дифференцирования значительно ухудшается отношение сигнала к шуму, что можно объяснить следующими причинами. В результате вычитания двух сигналов, обладающих шумовыми составляющими в виде случайных величин, шум увеличивается в 1.41 раза.

Максимальное значение дифференцированного сигнала пропорционально шагу изменения температуры. В качестве сигнала сырых данных можно принять размах СФ. Этот сигнал содержит N шагов изменения температуры. При дифференцировании выполняется вычитание сигналов сырых данных, отличающихся на 1 шаг, поэтому сигнал ориентировочно уменьшается в N раз.

Улучшить отношение сигнала к шуму можно путем увеличения времени наблюдения при каждом шаге, но это приводит к значительному увеличению длительности эксперимента.

Результаты обработки СП на основе производной СФ приведены в столбце Д. 2 таблицы. На рис. 2 приведены график производной кривой плавления ДНК одной из проб и график производной СФ. Можно отметить эффективное сглаживание шума и компенсацию дрейфа нулевой линии за счет применения аппроксимирующей кривой.

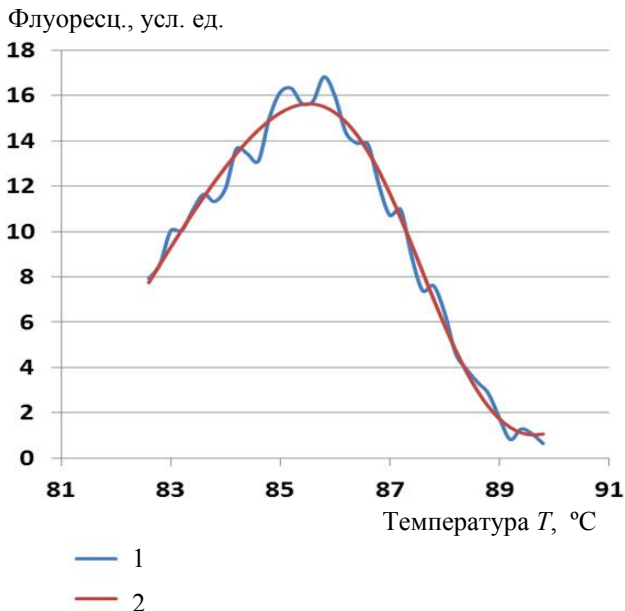


Рис. 2. График производной кривой плавления ДНК одной из проб (1) и график производной СФ (2)

МЕТОДИКА ПОСТРОЕНИЯ УСОВЕРШЕНСТВОВАННОЙ МОДЕЛИ И ОПТИМИЗАЦИИ ПАРАМЕТРОВ ПРОИЗВОДНОЙ КРИВОЙ ПЛАВЛЕНИЯ НА ОСНОВЕ ФУНКЦИИ ГАУССА

Усовершенствованная модель производной кривой плавления на базе функции Гаусса использована в программе MATLAB в виде формулы (3):

$$f = \frac{x(2) \cdot \exp(-(u - x(1))^2 / (2 \cdot x(3)^2)) + x(4) + x(5) \cdot u}{1.77245 \cdot x(3)} \quad (3)$$

где f — относительная величина интенсивности флуоресценции при температуре образца u (°C);

$x(1)$ — значение температуры (математическое ожидание), соответствующее максимальному значению функции Гаусса (°C);

$x(2)$ — максимальное значение функции Гаусса;

$x(3)$ — полуширина функции приблизительно для уровня 0.6 от максимального значения функции.

Для учета смещения и линейной зависимости нулевой линии от температуры в усовершенствованную модель предложено ввести слагаемые $x(4) + x(5) \cdot u$.

Начальные значения параметров можно вычислить путем обработки данных производной кривой

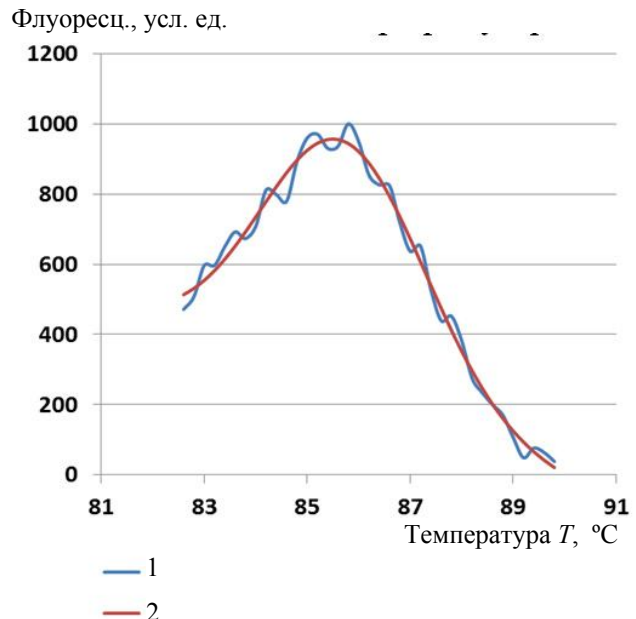


Рис. 3. График производной кривой плавления ДНК одной из проб (1) и график усовершенствованной модели на базе ФГ (2)

плавления. По графику "на глаз" находится максимальное значение $x(2)$ и соответствующее этому значению величина $x(1)$.

Уточнить начальные параметры можно путем аппроксимации данных производной кривой плавления функцией Гаусса без слагаемых $x(4)$ и $x(5) * u$.

Команды программы MATLAB сохраняются в виде М-файла, при этом расчет температуры плавления может выполняться в автоматическом режиме.

Можно отметить некоторые особенности приведенной формулы. Значение $x(1)$ относится к симметричной функции Гаусса. Можно использовать значение $x(1)$ в качестве температуры плавления T_m . Эти результаты обработки СП приведены в столбце Гаусс (Д. 3) таблицы.

Однако если использовать определение T_m как максимальное значение кривой плавления, то можно добавить операцию нахождения этого максимального значения по формулам (4)–(7):

$$s = (t(1):0.01:t(m)), \quad (4)$$

$$Vnn = f(xx, s), \quad (5)$$

$$[pks, locs] =$$

$$= \text{findpeaks}(Vnn, 'minpeakdistance', 15), \quad (6)$$

$$an = s(locs), \quad (7)$$

где

Vnn — значение модельной функции в диапазоне s с шагом 0.01 (°C);

an — значение T_m как результат нахождения максимального значения этой функции.

Результаты обработки производной СП приведены в столбце Гаусс М (Д. 4) таблицы.

График усовершенствованной модели на базе функции Гаусса одной из проб (рис. 3) практически аналогичен графику производной СФ на рис. 2.

МЕТОДИКА ПОСТРОЕНИЯ УСОВЕРШЕНСТВОВАННОЙ МОДЕЛИ И ОПТИМИЗАЦИИ ПАРАМЕТРОВ ПРОИЗВОДНОЙ КРИВОЙ ПЛАВЛЕНИЯ НА ОСНОВЕ СТЕПЕННЫХ ПОЛИНОМОВ

Модель производной кривой плавления на базе полиномов предлагается выразить для использования в программе MATLAB в виде формулы (8):

$$koaf = \text{polyfit}(t, V, n), \quad (8)$$

где V — значения производной СП при температуре t ;

n — значения степени полинома (в приведенных примерах $n = 3$ или 7).

Флуоресц., усл. ед.

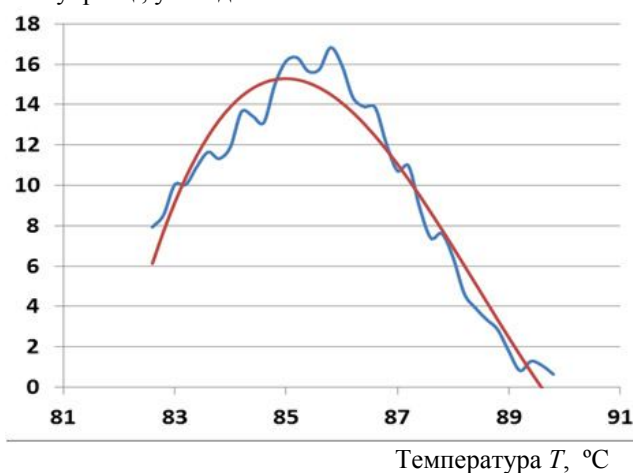


Рис. 4. График производной СП (1) и график аппроксимации полиномом 3-й степени (2)

Флуоресц., усл. ед.

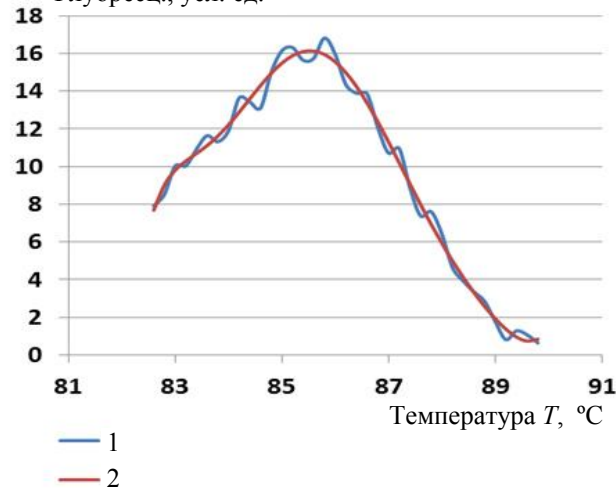


Рис. 5. График производной СП (1) и график аппроксимации полиномом 7-й степени (2)

Результаты обработки производной СП на основе усовершенствованной модели на базе полиномов приведены в столбцах данных Д. 5 и Д. 6 таблицы.

Аппроксимация полиномом 3-й степени имеет существенный недостаток: ее график (рис. 4) значительно смещен относительно графика производной СП.

График аппроксимации полиномом 7-й степени (рис. 5) практически аналогичен графику производной СП.

водной СФ (рис. 2) и модели на базе функции Гаусса (рис. 3).

ОЦЕНКА ПОГРЕШНОСТИ ОПРЕДЕЛЕНИЯ ТЕМПЕРАТУРЫ ПЛАВЛЕНИЯ ДНК ПРИ ИСПОЛЬЗОВАНИИ ПРЕДЛОЖЕННЫХ МЕТОДИК ОБРАБОТКИ СИГНАЛОВ ПЛАВЛЕНИЯ И ИХ ПРОИЗВОДНЫХ

Результаты определения температуры плавления T_m приведены в таблице. Эти же результаты приведены в виде графиков зависимостей от номера пробирки с пробой на рис. 6 (номера пробирок и проб одинаковы). Для наглядности точки графиков соединены прямыми линиями. На графиках величины T_m (вертикальная ось, °C) зависят от номера пробирки (горизонтальная ось, числа 1–8 определяют место последовательного расположения в держателе пробирок) и от методик обработки СП.

При сравнении графиков на рис. 6 можно отметить следующие особенности: кривые Гаусс М (Д. 4) и Полином-7 (Д. 6) почти полностью совпадают.

Зависимость величины T_m от номера пробы можно объяснить неравномерностью температурного поля прибора (систематическая погрешность измерения). Отклонение от этой зависимости можно считать случайной погрешностью.

Для количественной оценки полной погрешности определения температуры плавления ДНК вычислены в среде Excel значения стандартного отклонения в каждом столбце Д. 1–6 таблицы

по формуле (9):

$$dT_{mco} = \text{СТАНДОТКЛОН}(X1:X8), \quad (9)$$

где $X1:X8$ — значения в 8 ячейках для В1–8 соответствующих столбцов.

СПОСОБ ОПРЕДЕЛЕНИЯ НЕРАВНОМЕРНОСТИ ТЕМПЕРАТУРЫ ПРОБИРОК

Неравномерность температуры пробирок анализатора нуклеиновых кислот можно считать систематической погрешностью определения значения T_m .

Можно отметить нелинейность зависимости значения T_m от номера пробирки. Пробирки с номерами 1 и 8 расположены вблизи противоположных краев держателя пробирок. Их значения T_m больше, чем у пробирок, расположенных ближе к центру держателя пробирок (так называемый "краевой эффект").

Поэтому для выделения систематической погрешности для каждой модели построена полиномиальная (со степенью 2) линия тренда и определены 3 коэффициента уравнения аппроксимации. Согласно этому уравнению, найдены уточненные табличные значения T_{mt} . В качестве примера графики для методики Полином-7 приведены на рис. 7.

Для примера в таблице для методики Полином-7 приведены значения T_{mt} и $T_{md} = T_m - T_{mt}$. Затем по формуле (9) вычислены значения стандартного отклонения dT_{ms} для методики Полином-7 и др.

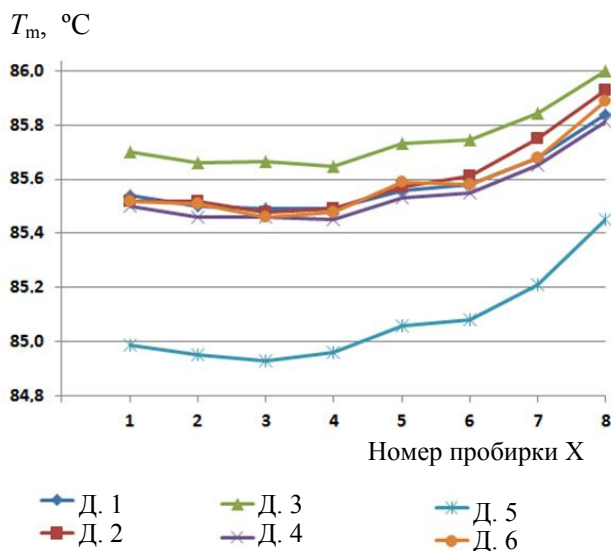


Рис. 6. Зависимости T_m от номера пробирки $X = 1-8$ по данным таблицы для различных методов

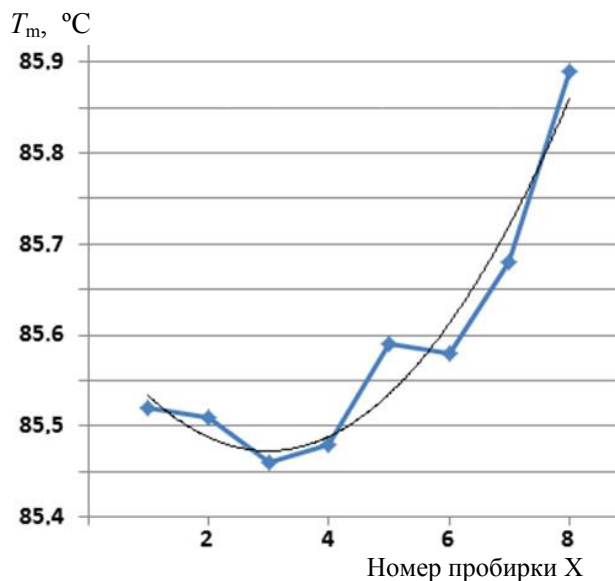


Рис. 7. Зависимость значения T_m от номера пробирки для методики Полином-7 (ломаная линия) и полиномиальная (со степенью 2) линия тренда
 $y = 0.0154x^2 - 0.0922x + 85.611$

Поскольку величины dT_{ms} для всех методик находятся в пределах 0.02–0.03 °C, то можно сделать вывод о том, что все методики обеспечивают приблизительно равные случайные погрешности. Однако модель на основе полиномов 7-й степени позволяет стандартными средствами MATLAB полностью автоматизировать вычисления без предварительного выбора параметров.

ЗАКЛЮЧЕНИЕ

1. Выбраны условия и параметры реализации методики плавления ДНК.

2. Разработаны методики обработки сигналов плавления ДНК, в которых использованы следующие усовершенствованные модели: модель на базе нелинейной сигмоидальной функции; модель на основе производной от сигмоидальной функции; модель на основе функции Гаусса и модель на основе степенных полиномов.

3. Показана возможность автоматизации вычислений при использовании программы MATLAB. Отмечено преимущество методики с применением усовершенствованной модели на основе степенных полиномов, полностью автоматизирующей вычисления стандартными средствами MATLAB без предварительного выбора параметров.

4. Показана возможность разделения систематической и случайной погрешностей определения температуры плавления ДНК.

5. Предложен способ определения неравномерности температуры пробирок.

Авторы выражают благодарность сотрудникам ЗАО «СИНТОЛ» (Москва) за предоставленные образцы, необходимые для получения исходных данных плавления ДНК.

СПИСОК ЛИТЕРАТУРЫ

1. Веденов А.А., Дыхне А.М., Франк-Каменецкий М.Д. Переход спираль—клубок в ДНК // Успехи физических наук. 1971. Т. 105, № 11. С. 479–519. Doi: 10.3367/UFNr.0105.197111d.0479.
2. ДНК плавление. URL: <http://humbio.ru/humbio/dnastructure/0002a247.htm>.
3. Календарь Р.Н., Сиволап Ю.М. Полимеразная цепная реакция с произвольными праймерами // Биополимеры и клетка. 1996. Т. 11, № 3-4. С. 55–65. URL: <http://www.biopolymers.org.ua/pdf/ru/11/3/055/biopolym.cell-1995-11-3-055-ru.pdf>.

4. Сиволап Ю.М., Календарь Р.Н., Чеботарь С.В. Генетический полиморфизм злаковых растений при помощи ПЦР с произвольными праймерами // Цитология и генетика. 1994. Т. 28, № 6. С. 54–61. URL: <http://www.biocenter.helsinki.fi/bi/genomedynamics/Pdfs/zyt.pdf>.
5. Гибридизация ДНК. Вычисление температуры плавления. URL: https://ru.wikipedia.org/wiki/Гибридизация_ДНК.
6. Ребриков Д.В., Саматов Г.А., Трофимов Д.Ю., Семенов П.А., Савилова А.М., Кофиади И.А., Абрамов Д.Д. ПЦР "в реальном времени". М.: БИНОМ. Лаборатория знаний, 2009. 224 с. URL: <http://nashol.com/2014072579193/pcr-v-realnom-vremeni-rebrikov-d-vsamatov-g-a-trofimov-d-u-2009.html>.
7. Белов Д.А., Корнева Н.А., Альдекеева А.С., Белов Ю.В., Киселев И.Г. Повышение разрешающей способности генетических анализаторов при определении температуры плавления ДНК // Научное приборостроение. 2016. Т. 26, № 2. С. 17–22. URL: <http://213.170.69.26/mag/2016/abst2.php#abst2>.
8. Белов Д.А., Белов Ю.В., Манойлов В.В. Методика обработки данных при плавлении продуктов полимеразной цепной реакции в реальном времени // Научное приборостроение. 2016. Т. 26, № 3. С. 10–14. URL: <http://213.170.69.26/mag/2016/full3/Art2.pdf>.
9. Эйдельштейн М.В., Алексеев Я.И., Никулин А.А., Романов А.В., Козлов Р.С. Способ детекции специфических нуклеотидных последовательностей и нуклеотидных замен с помощью ПЦР в режиме реального времени с эффектом гашения флуоресценции зонда праймером. Патент на изобретение РФ № 2451086. МПК C12Q 1/68. Оpubл. 20.05.2012. Бюл. № 14.

**Институт аналитического приборостроения РАН,
г. Санкт-Петербург (Белов Д.А., Белов Ю.В.,
Манойлов В.В.)**

**Университет ИТМО, г. Санкт-Петербург
(Манойлов В.В.)**

Контакты: Белов Дмитрий Анатольевич,
onoff_10@mail.ru

Материал поступил в редакцию: 12.01.2017

DNA MELTING DATA PROCESSING TECHNIQUES DEVELOPMENT

D. A. Belov¹, Yu. V. Belov¹, V. V. Manoylov^{1,2}¹Institute for Analytical Instrumentation of RAS, Saint-Petersburg, Russia²ITMO University, Saint-Petersburg, Russia

The conditions and parameters of the implementation of the DNA melting techniques were chosen. There were developed the DNA melting data processing techniques with advanced models based on sigmoidal nonlinear function and its derivative, a model based on a Gaussian function, a model based on power polynomials of degree 3 and 7. The comparison of the results of the application of techniques for processing real data of 8 samples received on the device ANK-32 was carried out. An algorithm for computing the full automation in MATLAB software environment was developed. The possibility of the separation of systematic and random error are shown. Developed techniques used to determine the temperature unevenness in nucleic acid analyzer tubes.

Keywords: DNA, the nucleic acid analyzer, melting techniques, uneven temperature tubes

REFERENCES

1. Vedenov A.A., Dychne A.M., Frank-Kamenezkiy M.D. [The helix-coil transition in DNA]. *Physics-Uspekhi* [Advances in Physical Sciences], 1972, vol. 14, no. 6, pp. 715–736. Doi: 10.1070/PU1972v014n06ABEH004752.
2. *DNK plavlenie* [DNA melting]. URL: <http://humbio.ru/humbio/dnastructure/0002a247.htm>. (In Russ.).
3. Kalendar' R.N., Sivolap Yu.M. [Polymerase chain reaction with the arbitrariness primers]. *Biopolimery i kletka* [Biopolymers and cell], 1996, vol. 11, no. 3-4, pp. 55–65. URL: <http://www.biopolymers.org.ua/pdf/ru/11/3/055/biopolym.cell-1995-11-3-055-ru.pdf>. (In Russ.).
4. Sivolap Yu.M., Kalendar' R.N., Chebotar' S.V. [Genetic polymorphism of cereals by means of PCR with the arbitrariness primers]. *Citologiya i genetika* [Cytology and genetics], 1994, vol. 28, no. 6, pp. 54–61. URL: <http://www.biocenter.helsinki.fi/bi/genomedynamics/Pdfs/zyt.pdf>. (In Russ.).
5. *Gibridizatsiya DNK. Vychislenie temperatury plavleniya* [DNA hybridization. Melting point calculation]. URL: https://ru.wikipedia.org/wiki/Гибридизация_ДНК. (In Russ.).
6. Rebrikov D.V., Samatov G.A., Trofimov D.Yu., Semenov P.A., Savilova A.M., Kofiadi I.A., Abramov D.D. *PCR "v real'nom vremeni"* [PCR "in real time"]. Moscow, BINOM. Laboratoriya znaniy Publ., 2009. 224 p. URL: <http://nashol.com/2014072579193/pcr-v-realnom-vremeni-rebrikov-d-vsamatov-g-a-trofimov-d-u-2009.html>. (In Russ.).
7. Belov D.A., Korneva N.A., Aldekeeva A.S., Belov Yu.V., Kiselev I.G. [Genetic analyzer resolution increasing at DNA melting temperature determination]. *Nauchnoe Priborostroenie* [Scientific Instrumentation], 2016, vol. 26, no. 2, pp. 17–22. Doi: 10.18358/np-26-2-i1722. (In Russ.).
8. Belov D.A., Belov Yu.V., Manoylov V.V. [Method of processing data in melting of real-time polymerase chain reactions]. *Nauchnoe Priborostroenie* [Scientific Instrumentation], 2016, vol. 26, no. 3, pp. 10–14. Doi: 10.18358/np-26-3-i1014. (In Russ.).
9. Eydel'shteyn M.V., Alekseev Ya.I., Nikulin A.A., Romanov A.V., Kozlov R.S. *Sposob detekzii spetsificheskikh nukleotidnykh posledovatel'nostey i nukleotidnykh zamen s pomosh'yu PCR v rezhime real'nogo vremeni s efektom gasheniya fluoressenzii zonda praymerom* [Way of detection of the specific nucleotide sequences and nucleotide replacements by means of PTsR in real time with effect of clearing of a bloom of the probe with a primer]. Patent RF No. 2451086. 20.05.2012. (In Russ.).

Contacts: Belov Dmitriy Anatol'evich,
onoff_10@mail.ru

Article received in edition: 12.01.2017