

Ю.И. Пивоваров, А.С. Сергеева, Т.Е. Курильская

СПОСОБ МАТЕМАТИЧЕСКОЙ ОБРАБОТКИ НАБОРА БЕЛКОВЫХ ПОЛОС,
ПОЛУЧЕННЫХ С ПОМОЩЬЮ ЭЛЕКТРОФОРЕЗА

ФГБУ «Научный центр реконструктивной и восстановительной хирургии» СО РАМН (Иркутск)

Предлагается программа математической обработки электрофореграммы белковых полос, которая предназначена для определения массы белков в исследуемом образце и имеет цель автоматически установить координаты максимальной концентрации белков в изучаемых полосах и рассчитать площадь под их пиком путем использования алгоритма интегрирования с адаптивным выбором шага, после считывания отсканированной полосы электрофореза с расширением (.pcx) и обработки полученной цифровой матрицы.

Ключевые слова: мембранные белки, электрофорез

MATHEMATICAL PROCESSING OF THE SET OF PROTEIN STRIPS
TAKEN WITH THE HELP OF ELECTROPHORESIS

Yu.I. Pivovarov, A.S. Sergeyeva, T.E. Kurilskaya

Scientific Center of Reconstructive and Restorative Surgery SB RAMS, Irkutsk

The program of mathematical processing of the electrophoregram of protein strips is suggested. It is intended for definition of weight of proteins in a researched sample and has the purpose to automatically establish coordinates of the maximal concentration of proteins in investigated strips and to calculate the area under their peak using the algorithm of integration with an adaptive choice of a step, after reading the scanned strip of electrophoresis with expansion (.pcx) and processing of the received digital matrix.

Key words: proteins of the membranes, electrophoresis

Многочисленные исследования показали, что, несмотря на некоторые отличия в количестве и типах белков, связанных с мембранным цитоскелетом, общая структура мембраны одинакова для всех клеток, и она включает многие белки, обнаруженные в эритроцитах. В связи с этим вполне закономерен интерес к изучению особенностей структурно-функциональной организации мембраны эритроцитов периферической крови при патологии, как высокочувствительный тест состояния клеточных мембран организма [1].

Исследования мембранных белков эритроцитов в норме и при патологии предусматривают изучение структурно-функциональных свойств и количества различных белков, входящих в состав мембраны этих клеток. Однако, как в норме, так и при различных заболеваниях, состояние мембранных белков с уже известными свойствами и молекулярной массой (kD) может быть оценено, прежде всего, по их количественному представительству в мембране клеток.

В исследовательской практике для определения концентрации белков, разделенных, в частности, на полосы путем электрофореза, обычно используют денситометрирование электрофореграмм с помощью лазерных денситометров различной модификации, которые рассчитывают площадь под пиком содержания белков различной молекулярной массы. После чего производят пересчет содержания отдельных белков в исследуемом образце на 1 мг общего белка по известному количеству (в мкг) маркерного белка.

Однако в случае отсутствия денситометра концентрация различных белков, полученных с помощью

электрофореза, может быть рассчитана программно с применением системы компьютерной математики – «Mathcad 2001 Professional», т.к. обработка электрофореграммы в среде Mathcad имеет одно важное преимущество – она математически прозрачна, что незаменимо при создании новых алгоритмов обработки изображений. В связи с этим нами предлагается собственный алгоритм математической обработки электрофореграммы, который позволяет рассчитать количество каждого из изучаемых мембранных белков эритроцитов.

На рис. 1 представлен пример исходных, отсканированных электрофореграмм белковых полос мембраны эритроцитов, взятых у донора и два вида маркерных белков: фирмы «BIO-RAD» с известным количеством белка различной молекулярной массы – 50 kD (0,750 мкг) и 100 kD, 20 kD (по 0,120 мкг) и фирмы «Termoscientific». Наличие последних маркеров желательно (но необязательно) для более точного определения молекулярной массы белковых фракций в исследуемом образце. Обычно последовательность белковых полос клеточных мембран по международному соглашению обозначается номером, начиная с самой тяжелой из них.

Отмеченные электрофореграммы отделяются друг от друга с помощью графического редактора Photoshop и запоминаются в виде трех самостоятельных файлов с расширением .pcx. Здесь важно, чтобы ширина полосы электрофореграммы исследуемого образца не была меньше, чем ширина полосы маркерных белков «BIO-RAD». Подготовленные таким образом файлы могут быть использованы в дальнейшем для обработки на ПК в среде Mathcad согласно предлагаемому алгоритму, листинг которо-

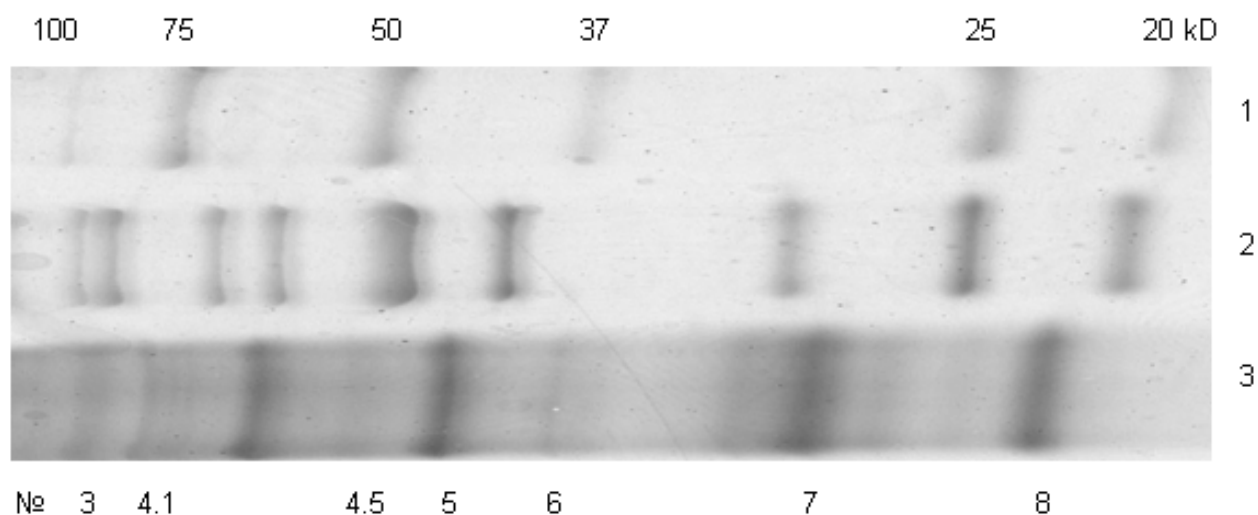


Рис. 1. Электрофорез белков мембран эритроцитов донора (3) совместно с маркерами фирмы «BIO-RAD» (1) и «Termoscientific» (2): № – номера набора белковых полос. Концентрация полиакриламидного геля – 15 %.

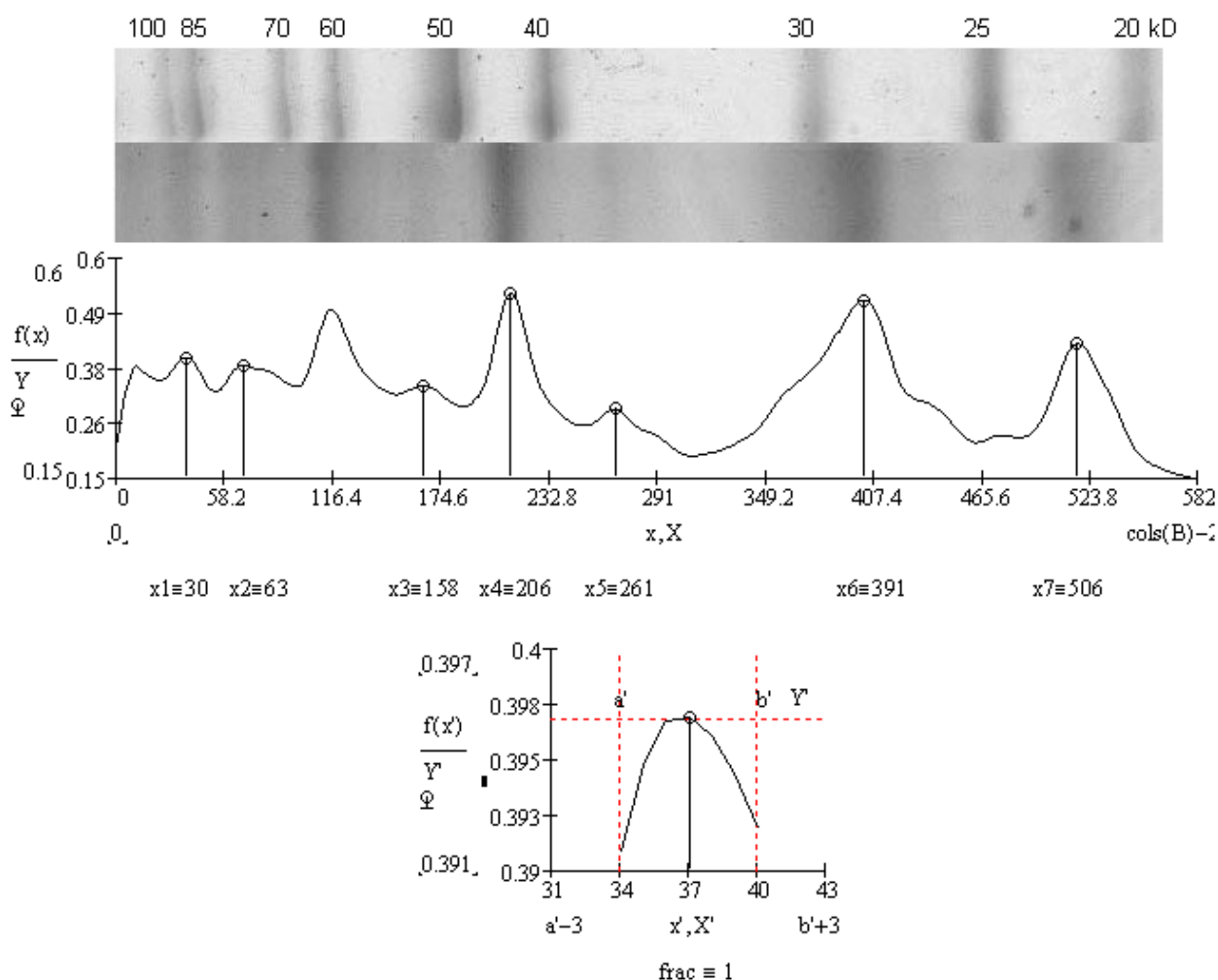


Рис. 2. Слева – график функции $f(x)$ с маркировкой пиков максимальной концентрации изучаемых белковых полос мембран эритроцитов. Маркировка полосы в 60 kD пропущена, поскольку структурно-функциональные свойства этих белков пока неизвестны. Над графиком представлены совмещенные матрицы электрофореграмм маркеров «Termoscientific» и белков мембран эритроцитов. Справа – график дополнительного визуального контроля над точностью определения начального интервала максимального пика белковой полосы (в данном случае – полосы 3).

го представлен ниже, а его графические результаты показаны на рис. 2.

Листинг математической обработки электрофореграммы белков мембраны эритроцитов:

(1) $f := \text{concat}("D:/EFOREZ(3-8)")$ $ff := \text{concat}("/Termo")$
 $ff1 := \text{concat}("/Donor")$ $fff := ".pcx"$

(2) $A := \text{READ_BLU}\text{ff}(\text{concat}(f, ff, fff))$
 $A1 := \text{READ_BLU}\text{ff}(\text{concat}(f, ff1, fff))$

(3) $n := \text{rows}(A1)$ $m := \text{cols}(A1)$ $n = 160$ $m = 585$

(4) $h := n \div 2$ $s := 71.5$
 $s1 := h - s$ $s2 := h + s$ $s1 = 8.5$ $s2 = 151.5$

(5) $B := \text{submatrix}(A1, s1, s2 - 1, 0, m - 2)$
 $\text{rows}(B) = 143$ $\text{cols}(B) = 584$

(6) $j := 0.. \text{cols}(B) - 1$ $i := 0.. \text{rows}(B) - 1$ $x_1 := i$

(7) $Q_j := \frac{1}{\text{rows}(B)} \sum_{i=0}^{\text{last}(x)} (B_{i,j})$ $vy := 1 - \frac{Q}{\max(Q)}$ $vx_j := j$
 $x_j := j$ $k := 6$

(8) $\text{span} := 0.05$ $z := \text{loess}(vx, vy, \text{span})$
 $f(x) := \text{interp}(z, vx, vy, x)$ $\tau := 40$

(9) $t1 := (x1 \ x2 \ x3 \ x4 \ x5 \ x6 \ x7)^T$ $t2 := t1 + \tau$
 $i := 0.. k$ $c := 3$

(10) $C_i := \begin{cases} \text{for } x \in t1_i..t2_i \\ p_x \leftarrow f(x) \\ Y \leftarrow \max(p) \\ X \leftarrow \sum_{x=t1_i}^{t2_i} x \cdot (f(x) = Y) \\ \begin{pmatrix} Y \\ X \end{pmatrix} \end{cases}$

(11) $Y_i := (C_i)_0$ $X_i := (C_i)_1$ $a_i := X_i - c$ $b_i := a_i + 2c$

(12) $S_i := \int_{a_i}^{b_i} f(x) dx$ $Y' := Y_{\text{frac}-1}$ $X' := X_{\text{frac}-1}$

(13) $a' := X' - c$ $b' := a' + 2c$ $x' := a'.. b'$
 $u1 := \text{concat}("polosa")$ $u2 := \text{concat}("S")$
 $u3 := \text{concat}("Y")$ $u4 := \text{concat}("X")$
 $u := \text{stack}(u1, u2, u3, u4)$
 $r := \text{augment}(3, "4.1", "4.5", 5, 6, 7, 8)$
 $d := \text{stack}(r, S^T, Y'^T, X'^T)$ $D := \text{augment}(u, d)$
 $A' := \text{submatrix}(A, 0, \text{rows}(A) - 1, 0, \text{cols}(A) - 2)$

$E' := \text{stack}(A', B)$

$E := \text{submatrix}(E', 80, \text{rows}(E') - 90, 0, \text{cols}(E') - 2)$

(14) $j := 1.. 7$ $K := 0.750 \div 1.453$

(15) $p_{j-1} := \left[\left(K \cdot D_{1,j} \right) \div 10 \right] \cdot 1000$

$\text{Prot}_{\mu\text{g_on_mg.prot}} := \text{stack}(r, p^T)$

Смысл предлагаемого алгоритма заключается в том, что после считывания электрофореграмм маркерных белков фирмы «Termoscientific» (A) и белков мембраны эритроцитов ($A1$) [(1), (2) строки листинга], путем использования встроенной функции *submatrix*, выделяется центральная область электрофореграммы $A1$ [(3), (4) и (5) строки] с числом строк матрицы (B) равным количеству строк матрицы электрофореграммы маркерных белков «BIO-RED».

Затем производится усреднение переменных матрицы B (Q), приводятся к нулю их максимальные значения (vy) [(7) строка] и осуществляется сглаживание данных с помощью регрессии фрагментами полиномов (z) и $f(x)$ [(8) строка]. Формирование векторов начальных ($t1$) и конечных ($t2$) интервалов местоположения максимальных пиков изучаемых белковых полос [(9) строка] производится вручную под визуальным контролем с помощью глобальных переменных $x1... x7$. После чего автоматически (C) [(10) строка] определяются точные координаты максимальных пиков ($X; Y$), промежутки ($a; b$) кривой $f(x)$ [(11) строка], на отрезке которых рассчитывается площадь под пиком (S) путем численного интегрирования с адаптивным выбором шага [(12) строка]. Кроме того, программа предусматривает дополнительный визуальный контроль над точностью определения интервала местоположения максимального пика любой из белковых полос (Y', X', a', b' и x') [(12), (13) строки].

Такая обработка данных позволяет в последующем рассчитать массу каждого изучаемого белка по формуле: $m(x) = S(q) \cdot 0,750 \text{ мкг} / S(qm)$, где $m(x)$ – масса изучаемой белковой фракции, $S(q)$ – площадь изучаемой полосы под пиком, $S(qm)$ – площадь маркерного белка в 50 kD «BIO-RED» под пиком (в данном случае – 1,453), которая определяется на электрофореграмме данных маркеров аналогичным образом, что и в исследуемом образце. Расчет массы изучаемых белков каждой полосы с учетом постоянного множителя K [(14) строка] проводится на 1 мг общего белка [(15) строка]. Введенный в уравнение (p) делитель «10» отражает количество белка в исследуемом образце (10 мкг).

Использование встроенных функций *stack* и *augment* позволяют сгруппировать полученные результаты в сводные таблицы данных (D и Prot).

$D := \begin{pmatrix} \text{"polosa"} & 3 & \text{"4.1"} & \text{"4.5"} & 5 & 6 & 7 & 8 \\ \text{"S"} & 2.41 & 2.313 & 2.068 & 3.193 & 1.755 & 3.111 & 2.582 \\ \text{"Y"} & 0.402 & 0.386 & 0.345 & 0.532 & 0.293 & 0.518 & 0.43 \\ \text{"X"} & 37 & 68 & 165 & 211 & 268 & 400 & 517 \end{pmatrix}$

$\text{Prot}_{\mu\text{g_on_mg.prot}} := \begin{pmatrix} 3 & \text{"4.1"} & \text{"4.5"} & 5 & 6 & 7 & 8 \\ 124 & 119 & 107 & 165 & 91 & 161 & 133 \end{pmatrix}$

Таким образом, предлагаемый алгоритм математической обработки электрофореграмм может быть довольно успешно применен для определения количественного содержания белков различной молекулярной массы в биологических образцах, как в клинических, так и в экспериментальных исследованиях.

ЛИТЕРАТУРА REFERENCES

1. Фаллер Д.М., Шилдс Д. Молекулярная биология клетки. Руководство для врачей. – М.: Бином-Пресс, 2006. – 256 с.
Faller D.M., Shields D. Molecular biology of the cell. The Guidelines for physicians. – Moscow: Binom-Press, 2006. – 256 p.

Сведения об авторах

Пивоваров Юрий Иванович – доктор медицинских наук, профессор, ведущий научный сотрудник научного отдела коронарного атеросклероза ФГБУ «Научный центр реконструктивной и восстановительной хирургии» СО РАМН (664079, г. Иркутск, мкр. Юбилейный, 100; тел.: 8 (3952) 46-55-56)

Сергеева Анна Сергеевна – старший научный сотрудник научного отдела коронарного атеросклероза ФГБУ «Научный центр реконструктивной и восстановительной хирургии» СО РАМН. (664079, г. Иркутск, мкр. Юбилейный, 100; тел.: 8 (3952) 46-55-56, sergeeva1111@yandex.ru)

Курильская Татьяна Ефимовна – доктор медицинских наук, заведующая научным отделом коронарного атеросклероза ФГБУ «Научный центр реконструктивной и восстановительной хирургии» СО РАМН. (664079, г. Иркутск, мкр. Юбилейный, 100; тел.: 8 (3952) 46-55-56)

Information about the authors

Pivovarov Yuriy Ivanovich – M.D., professor, leading scientific worker of the scientific department of coronary atherosclerosis at Scientific Center of Reconstructive and Restorative Surgery SB RAMS (664079, Irkutsk, Yubileyniy, 100; Tel.: 8 (3952) 46-55-56)

Sergeyeva Anna Sergeyevna – senior scientific worker of the scientific department of coronary atherosclerosis at Scientific Center of Reconstructive and Restorative Surgery SB RAMS (664079, Irkutsk, Yubileyniy, 100; Tel.: 8 (3952) 46-55-56; e-mail: sergeeva1111@yandex.ru)

Kurilskaya Tatyana Yefimovna – M.D., the head of the scientific department of coronary atherosclerosis at Scientific Center of Reconstructive and Restorative Surgery SB RAMS (664079, Irkutsk, Yubileyniy, 100; Tel.: 8 (3952) 46-55-56)