

МНОЖЕСТВЕННОЕ ВЫРАВНИВАНИЕ СОВОКУПНОСТИ АМИНОКИСЛОТНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ НА ОСНОВЕ ВЕРОЯТНОСТНОЙ МОДЕЛИ ЭВОЛЮЦИИ

© Сулимова В.В.¹, Разин Н.А.², Моттль В.В.³, Мучник И.Б.⁴

¹ Тульский государственный университет, Россия 300600, г.Тула, пр.Ленина, 92

² Московский физико-технический институт, Россия 141700, г.Долгопрудный, Институтский пер. 9

³ Вычислительный центр РАН, Россия 117967, г.Москва, ГСП-1, ул.Вавилова, 40

⁴ УНИВЕРСИТЕТ РАТГЕРС, HILL CENTER, BUSCH CAMPUS, NEW BRUNSWICK, USA, NJ 080901

E-MAIL: vmottl@yandex.ru, vsulimova@yandex.ru

Abstract. The problem of measuring group similarity of amino acid sequences is one of fundamental issues of the modern bioinformatics. Existing algorithms for decision this problem (so called multiple alignment procedures) are not based on any formal problem definition and any model of evolution of proteins. In this paper we propose a new approach for measuring group similarity of proteins, which is founded on probabilistic evolutionary model of transformation of amino acid sequences.

ВВЕДЕНИЕ

Последовательности нефиксированного числа элементов из некоторого заданного множества примитивов являются типичным видом представления информации об объекте во многих задачах интеллектуального анализа данных. В частности, массовым источником последовательностей разной длины является молекулярная биология, в которой один из важнейших видов задач анализа данных заключается, в конечном счете, в прогнозировании биологических свойств предъявленной молекулы белка, заданной своей так называемой первичной структурой в виде последовательности аминокислот, образующих полимерную цепь. Различают 20 существующих в природе аминокислот, поэтому первичная структура полимерной молекулы каждого конкретного белка образует символьную последовательность индивидуальной длины над 20-буквенным алфавитом.

В настоящее время накоплено большое число баз данных, в которых собраны аминокислотные последовательности белков, биологические свойства которых удалось установить экспериментально. При этом в них выделены группы похожих последовательностей, в которых белки обладают одинаковыми или похожими биологическими свойствами.

Именно эта группировка позволяет практикам-биологам переносить систематизированные знания о белках, накопленные в базах данных, на новые неизученные белки, решая такие широко известные задачи, как задача локализации консервативных регионов [1], прогнозирования вторичной [2] и третичной (пространственной) структуры белков [3], построения филогенетических деревьев [4]. Одной из фундаментальных проблем молекулярной биологии, связанной с использованием группированных баз данных, является проблема измерения общего сходства аминокислотных последовательностей по всей группе, а также сходства отдельных последовательностей со всей группой.

В качестве инструментов измерения группового сходства в настоящее время используются алгоритмы множественного выравнивания [5, 6, 7, 8, 9, 10]. Многие из них сопровождают результат так называемым профилем анализируемой совокупности последовательностей, под которым понимается некоторый самостоятельный "обобщенный" белок в виде последовательности дискретных распределений вероятностей над множеством всех аминокислот. Однако проблема заключается в том, что существующие алгоритмы, во-первых, не основаны на какой-либо формальной постановке задачи и, во-вторых, не базируются на какой-либо единой модели эволюционной модификации белков.

В данной работе предлагается принципиально новый подход к измерению группового сходства белков, в основе которого лежит вероятностная эволюционная модель преобразования аминокислотных последовательностей.

1. ИЗМЕРЕНИЕ СХОДСТВА АМИНОКИСЛОТ НА ОСНОВЕ МОДЕЛИ ЭВОЛЮЦИИ М. ДЭЙХОФФ

Очевидно, что измерение эволюционного сходства белков неизбежно должно базироваться на измерении сходства составляющих их аминокислот. Пусть A - конечное множество аминокислот $A = \{\alpha^{(1)}, \dots, \alpha^{(20)}\}$.

Под сходством двух аминокислот $\alpha^i, \alpha^j \in A$ принято понимать их склонность к взаимному мутационному превращению. Одной из основных теоретических концепций сравнения аминокислот является вероятностная модель эволюции Маргарет Дэйхофф [11], получившая название РАМ. Данная модель предполагает, что склонность аминокислот к взаимному мутационному превращению количественно выражается квадратной матрицей условных вероятностей

$$\Psi = (\psi^{ij}, i, j = 1, \dots, n), \psi^{ij} = \psi(\alpha^j | \alpha^i), \alpha^i, \alpha^j \in A, n = 20, \quad (1.1)$$

интерпретируемых как вероятности того, что на очередном шаге эволюции аминокислота α^i превратится в аминокислоту α^j . Таким образом, $\sum_{\alpha^j \in A} \psi(\alpha^j | \alpha^i) = 1$ для всех $\alpha^i \in A$.

Основным математическим понятием модели Дэйхофф является понятие марковской цепи эволюции аминокислоты в отдельно взятой позиции цепи $h_s, s = 1, 2, 3, \dots$, определяемой матрицей условных вероятностей $\psi(\alpha^j | \alpha^i)$. Предполагается, что данная Марковская цепь представляет собой эргодический и обратимый случайный процесс, т.е.: 1) он характеризуется финальным распределением вероятностей $\xi(\alpha^j)$: $\sum_{\alpha^i \in A} \xi(\alpha^i) \psi(\alpha^j | \alpha^i) = \xi(\alpha^j)$ и 2) выполняется условие обратимости:

$$\xi(\alpha^i) \psi(\alpha^j | \alpha^i) = \xi(\alpha^j) \psi(\alpha^i | \alpha^j). \quad (1.2)$$

Сходство пары аминокислот α^i и α^j естественно оценивать, вычисляя вероятность их происхождения в результате двух независимых ветвей эволюции от одной и той же неизвестной аминокислоты

$$\mu(\alpha^i, \alpha^j) = \sum_{k=1}^n \xi(\alpha^k) \psi(\alpha^i | \alpha^k) \psi(\alpha^j | \alpha^k). \quad (1.3)$$

В биоинформатике принято выражать эти величины в логарифмической форме в виде семейства так называемых матриц РАМ при разном понимании величины эволюционного шага:

$$\mathbf{D} = (D_{ij}, i, j = 1, \dots, n), D_{ij} = 10 \log_{10} [K(\alpha^i, \alpha^j) / \xi(\alpha^i) \xi(\alpha^j)].$$

Однако следует отметить, что мера сходства по Дэйхофф (1.3) представляет собой скалярное произведение

$$\mu(\alpha^i, \alpha^j) = \sum_{k=1}^n \xi(\alpha^k) \psi(\alpha^i | \alpha^k) \psi(\alpha^j | \alpha^k) = \sum_{k=1}^n \left(\sqrt{\xi(\alpha^k)} \psi(\alpha^i | \alpha^k) \right) \left(\sqrt{\xi(\alpha^k)} \psi(\alpha^j | \alpha^k) \right) = \sum_{k=1}^n x_{ik} x_{jk} = \mathbf{x}_i^T \mathbf{x}_j$$

Таким образом, функция (1.3) является потенциальной функцией [12], погружающей конечное множество аминокислот в линейное пространство $A = \{\alpha^i, i = 1, \dots, n\} \subset \tilde{A}$, в котором она играет роль скалярного произведения [13]. Это свойство позволяет ввести понятие "обобщенной аминокислоты" как линейной комбинации реальных аминокислот с коэффициентами, определяемыми вектором $\vec{\beta} \in \tilde{A}$, $\vec{\beta} = (\beta_k, k = 1, \dots, 20)$, таким, что $\sum_{k=1}^{20} \beta_k = 1$. Компоненты β_k , $k = 1, \dots, 20$ могут быть интерпретированы как вероятности того, что соответствующий элемент $\vec{\beta} \in \tilde{A}$ является реальной аминокислотой $\alpha^k \in A$. В частности, если все компоненты вектора $\vec{\beta}$, кроме одной, k -й компоненты, равны нулю, то это означает, что данный вектор соответствует реальной аминокислоте α^k .

2. ВЕРОЯТНОСТНАЯ МОДЕЛЬ ЭВОЛЮЦИОННЫХ ПРЕОБРАЗОВАНИЙ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Пусть Ω - множество всех возможных аминокислотных последовательностей конечной длины $\omega = (\alpha_t, t = 1, \dots, N_\omega)$, $\omega \in \Omega$, образованных реальными аминокислотами $\alpha_t \in A$, и $\tilde{\Omega} \supset \Omega$ - множество всех возможных обобщенных аминокислотных последовательностей $\vartheta = (\vec{\beta}_s, s = 1, \dots, n_\vartheta)$, $\vartheta \in \tilde{\Omega}$, составленных из обобщенных аминокислот $\vec{\beta} \in \tilde{A}$ и $\tilde{\Omega}^n \supset \tilde{\Omega}$ - множество обобщенных последовательностей длины n . При этом очевидно, что множество $\tilde{\Omega}$ включает в себя все последовательности ω из множества Ω .

Далее мы будем использовать для обозначения последовательности символ ϑ в случае, если она может состоять как из обобщенных, так и из реальных аминокислот, и символ ω в случае, если необходимо подчеркнуть, что имеется ввиду именно реальная последовательность, состоящая только из реальных аминокислот.

Идея предлагаемой в данной работе вероятностной модели эволюционных преобразований белков основана на понятии семейства условных распределений вероятностей на множестве всех конечных последовательностей аминокислот. Предлагается связать с каждой последовательностью $\vartheta = (\vec{\beta}_s, s = 1, \dots, n)$ некоторое специфичное по отношению к ней распределение вероятностей $f(\omega | \vartheta)$ на множестве всех реальных последовательностей $\omega \in \Omega$, интерпретируемое как совокупность вероятностей ее преобразования в другую последовательность $\omega = (\alpha_t, t = 1, \dots, N)$ некоторой другой длины N за один шаг процесса эволюции.

Модель случайного преобразования последовательности $\vartheta = (\vec{\beta}_s, s = 1, \dots, n)$ в другую последовательность $\omega = (\alpha_t, t = 1, \dots, N)$, предлагается представить в виде двухкомпонентного случайного процесса (X, Y) .

Первая компонента $X = (x_i, i = 1, \dots, n)$ со значениями $x_i = k, k \in \{1, \dots, N\}$, представляет собой скрытый марковский случайный процесс с переходными вероятностями $q_i^{lk} = p(x_i = k | x_{i-1} = l)$, имеющими смысл вероятностей того, что i -й элемент первой последовательности преобразуется в k -й элемент второй последовательности, при условии, что $(i - 1)$ -й элемент первой последовательности преобразовался в l -й элемент второй последовательности, $l \in \{1, \dots, N\}$. Начальные вероятности преобразования 1-го элемента последовательности ϑ , являющиеся безусловными, будем обозначать $q_1^{0k} = q_1^k = p(x_1 = k)$.

Распределения вероятностей q_1^k и q_1^{lk} могут быть заданы различными способами, но так, чтобы выполнялись следующие условия

1) условие отсутствия пропусков элементов исходной последовательности:

$$q_i^{lk} = 0, \text{ если } k \leq l \text{ или } k > N - (n - i). \tag{2.1}$$

Это условие обеспечивает, чтобы каждый элемент исходной последовательности $\vartheta_i, i = 1, \dots, n$ с ненулевой вероятностью преобразовался в некоторый k -й элемент второй последовательности длины N , причем такой, чтобы его номер был больше номера предыдущего l -го элемента.

2) условие симметричности:

$$p(x_n = k) = p(x_1 = N - k + 1), k = 1, \dots, N, \tag{2.2}$$

где $p(x_i = k) = \sum_{s=1}^N q_i^{ks} p(x_{i-1} = s)$ - безусловная вероятность того, что i -й элемент первой последовательности преобразуется в k -й элемент второй последовательности.

3) условные вероятности преобразования i -го элемента исходной последовательности q_i^{lk} и начальные вероятности q_1^k , должны составлять полную группу:

$$\sum_{k=1}^N q_i^{lk} = 1 \quad \text{и} \quad \sum_{k=1}^N q_i^k = \sum_{k=1}^N p(x_1 = k) = 1 \tag{2.3}$$

Данный случайный процесс определяет структуру преобразования последовательности $\vartheta \in \tilde{\Omega}^n$ длины n в последовательность $\omega \in \Omega$ длины N , т.е. выбирает позиции в последовательности ω , в которые будут преобразовываться обобщенные аминокислоты, составляющие ϑ . Следует обратить внимание, что, в соответствии с условием 2.1, данная вероятностная схема запрещает пропуски элементов первой последовательности. В результате, количество производимых замен всегда строго равно n , а длина второй последовательности оказывается всегда большей или равной длине исходной последовательности. В связи с этим мы называем такое преобразование "несжимающим".

Вторая компонента представляет собой условный случайный процесс $(Y|X)$, $Y = (y_i, i = 1, \dots, n)$ замены аминокислот в позициях, выбранных процессом X , с множеством состояний $y_i \in \{\alpha_k, k = 1, \dots, N\}$. Это наблюдаемый случайный процесс, определяющийся условными распределениями $\eta^k(y_i) = \eta(\alpha_k | \vec{\beta}_i, x_i = k)$, которые

представляют собой совокупности вероятностей преобразования ϑ_i в α_k :

$$\eta(\alpha_k|\vec{\beta}_i, x_i = k) = \frac{\psi(\alpha_k|\vec{\beta}_i)}{\sum_{t=1}^N \psi(\alpha_t|\vec{\beta}_i)}, k = 1, \dots, N, \quad (2.4)$$

где $\psi(\alpha_k|\vec{\beta}_i) = \sum_{l=1}^{20} \beta_i^l \psi(\alpha_k|\alpha_l)$ - вероятность замены обобщенной аминокислоты $\vec{\beta}_i$ на аминокислоту α_k .

Структура преобразования последовательности ϑ в последовательность ω , определяемая первым случайным механизмом, может быть представлена в виде случайного пути на графе (рис.1).

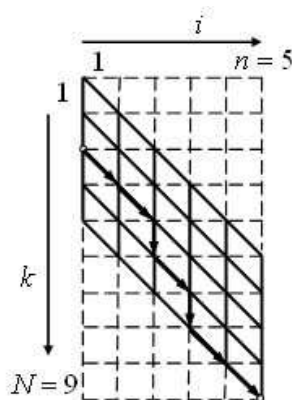


Рис. 1. Структура преобразования последовательности ϑ длины n в последовательность ω длины N

Возможно множество различных вариантов преобразования последовательности ϑ в последовательность ω . Однако не все из них являются допустимыми (имеющими ненулевые вероятности) в соответствии с предложенной схемой. Ребра, принадлежащие области допустимых преобразований, показаны на данном графе сплошными линиями. Стрелками показан один из допустимых вариантов преобразования.

Здесь продвижение по диагонали соответствует преобразованию i -го элемента последовательности ϑ в k -й элемент последовательности ω , а продвижение по вертикали – вставке нового элемента в последовательность ω . Следует обратить внимание, что допустимыми в соответствии с принятой схемой являются только те преобразования, которые соответствуют путям на графе, проходящим ровно через n диагональных ребер, в связи с чем на данном графе запрещены любые горизонтальные продвижения.

Каждый вариант преобразования последовательности ϑ длины n в последовательность ω длины N соответствует выравниванию их элементов, которое в данном случае полностью определяется последовательностью $w = (w_i, i = 1, \dots, n)$ длины n чисел $w_i \in \{1, \dots, N\}$.

Выравнивание, соответствующее выделенному пути на графе рисунка 1, имеет вид:

$$w = \begin{cases} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 \\ \vec{\beta}_3 & \vec{\beta}_4 & \vec{\beta}_6 & \vec{\beta}_8 & \vec{\beta}_9 \end{cases}$$

Порядком выравнивания будем называть количество определяемых им замен элементов. Заметим, что допустимые выравнивания имеют ровно n позиций, в которых происходят замены элементов. В связи с этим будем называть их выравниваниями порядка n . Множество всех выравниваний порядка n пары последовательностей $\langle \vartheta, \omega \rangle$ будем обозначать $W_{n,N}^n$. На множестве выравниваний $W_{n,N}^n$ пары последовательностей $\langle \vartheta, \omega \rangle$ определим распределение вероятностей $q(w) = q_1^{w_1} \prod_{i=2}^n q_i^{w_{i-1}, w_i}$.

Теорема 1. *Теорема 1. Если априорные относительно элементов последовательностей распределения q_1^k и q_i^{lk} определены в соответствии с условиями 2.1-2.3, то $\sum_{w \in W_{n,N}^n} q(w) = 1$*

Вероятность преобразования последовательности ϑ длины n в последовательность ω длины N определим следующим образом:

$$f(\omega|\vartheta) = \sum_{w \in W_{n,N}^n} q(w)\varphi(\omega|w, \vartheta) \tag{2.5}$$

где $\varphi(\omega|w, \vartheta) = \prod_{i=1}^n \eta(\alpha_k|\vec{\beta}_i, x_i = k)$ – вероятность преобразования последовательности ϑ в последовательность ω , при условии, что структура преобразования соответствует определенному выравниванию $w \in W_{n,N}^n$.

3. ПОСТАНОВКА ЗАДАЧИ МНОЖЕСТВЕННОГО ВЫРАВНИВАНИЯ ГРУППЫ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Пусть $\omega_j = (\alpha_{jt}, t = 1, \dots, N_j)$, $j = 1, \dots, M$ – группа последовательностей, состоящих из реальных аминокислот $\alpha_{jt} \in A$.

В качестве меры эволюционного сходства заданного порядка n для группы аминокислотных последовательностей $\vec{\omega} = (\omega_j, j = 1, \dots, M)$ естественно принять максимум правдоподобия гипотезы, что все они порождены независимо из одной и той же последовательности $\vartheta = (\vec{\beta}_i, i = 1, \dots, n)$, $\vartheta \in \tilde{\Omega}^n$, имеющей длину, равную принятому порядку n . При этом функция правдоподобия будет иметь вид:

$$F(\vec{\omega}|\vartheta) = \prod_{j=1}^M f_j(\omega_j|\vartheta)$$

или в логарифмической форме

$$\log F(\vec{\omega}|\vartheta) = \sum_{j=1}^M \log f_j(\omega_j|\vartheta) \tag{3.1}$$

Задачу множественного выравнивания группы последовательностей $\vec{\omega}$ поставим как задачу нахождения такой обобщенной последовательности $\vartheta \in \tilde{\Omega}^n$ длины n , которая обеспечивала бы максимум функции правдоподобия 3.1:

$$\vartheta^* = \arg \max_{\vartheta \in \tilde{\Omega}^n} \log F(\vec{\omega}|\vartheta) \quad (3.2)$$

Самую оптимальную последовательность ϑ^* , доставляющую максимум критерию 3.1, следует интерпретировать как профиль общего фрагмента заданной длины n , характеризующий эволюционную общность анализируемой совокупности аминокислотных последовательностей $\vec{\omega}$. При этом следует обратить внимание, что в соответствии с введенным в разделе 1 понятием обобщенной аминокислоты, понятие профиля, которое имело ранее лишь эвристический смысл, приобретает здесь строгую математическую основу.

Традиционная форма представления множественного выравнивания в виде записи анализируемых последовательностей друг под другом, с расстановкой пробелов между определенными позициями каждого белка, может быть получена путем выравнивания найденного обобщенного профиля с каждой из последовательностей при помощи любой из стандартных процедур локального парного выравнивания [14], обобщенной на случай выравнивания последовательностей, состоящих из элементов расширенного множества $\vec{\beta} \in \tilde{A}$, что с учетом определения обобщенной аминокислоты не представляет проблемы.

4. ИТЕРАЦИОННАЯ ПРОЦЕДУРА ДЛЯ РЕШЕНИЯ ЗАДАЧИ МНОЖЕСТВЕННОГО ВЫРАВНИВАНИЯ

Пусть $\vartheta^s = (\vec{\beta}_i^s, i = 1, \dots, n)$ – очередное приближение к точке максимума функции правдоподобия 3.1.

Применяя для максимизации функции правдоподобия 3.1 EM-процедуру [15], предложенную М.И. Шлезингером, нетрудно убедиться, что каждая компонента очередного приближения ϑ^{s+1} может быть найдена независимо от остальных компонент по следующей формуле:

$$\beta_i^{l,s+1} \propto \sum_{j=1}^M \sum_{k=1}^{N_j} p(w_i = k | w \in W_{nN_j}^n, \omega_j, \vartheta^s) \zeta(l | \alpha_{jt}, \vec{\beta}^s), \quad l = 1, \dots, 20, \quad (4.1)$$

где $p(w_i = k | w \in W_{nN_j}^n, \omega_j, \vartheta^s)$ – распределение вероятностей, характеризующее апостериорные свойства скрытого марковского процесса преобразования последовательностей $(X|Y)$. Соответствующие вероятности, называемые в теории скрытых марковских процессов интерполяционными, могут быть найдены по стандартной схеме [17] на основе заданных априорных свойств процесса X преобразования последовательностей;

Распределение $\zeta(l | \alpha_{jt}, \vec{\beta}^s)$, $l = 1, \dots, 20$ в 4.1 определяется соотношением $\zeta(l | \alpha_{jt}, \vec{\beta}^s) \propto \beta^{l,s} \psi(\alpha_{jt} | \alpha^l)$.

Следует отметить, что данная процедура легко реализуется алгоритмически и достаточно быстро сходится к точке глобального максимума из любого начального приближения $\vartheta^0 = (\vec{\beta}_i^0, i = 1, \dots, n)$.

ЗАКЛЮЧЕНИЕ

В данной работе предлагается принципиально новая, вероятностная постановка задачи измерения группового сходства белков, в основе которой лежит вероятностная модель эволюции составляющих их аминокислотных последовательностей. Предложенная мера сходства группы последовательностей, определяемая как максимум правдоподобия гипотезы о том, что все белки из рассматриваемой группы были получены в результате независимого случайного преобразования одной и той же неизвестной последовательности заданной длины n , состоящей из обобщенных аминокислот, является более адекватной с точки зрения молекулярной биологии, по сравнению с существующими методами, решающими ту же задачу.

Обобщенная последовательность, обеспечивающая максимум правдоподобия, фактически, является моделью, или профилем, анализируемой группы последовательностей и может быть легко использована для получения традиционной графической интерпретации множественного выравнивания последовательностей.

В дальнейшем предполагается детальное исследование предложенной процедуры и ее сравнение с существующими процедурами множественного выравнивания на материале известных баз данных VIDA и BLOCKS, с целью подтверждения ее теоретического превосходства над известными алгоритмами.

Работа выполнена при поддержке грантов РФФИ № № 6-01-00412, 06-01-00679, гранта INTAS YSF 06-1000014-6563.

СПИСОК ЛИТЕРАТУРЫ

1. *Attwood T.K.* The PRINTS database: a resource for identification of protein families. // *Brief Bioinformatics*, 3, 2002, pp. 252-263.
2. *Rost B. et al.* PHD-an automatic server for protein secondary structure prediction. // *Computational applications in biosciences*. 10, 1994, pp. 53-60.
3. *Goebel U. et al.* Correlated mutations and residue contacts in proteins. // *Proteins*, 18, 1994, pp. 309-317.
4. *Saitou N., Nei M.* The neighbor-joining method: a new method for reconstructing phylogenetic trees. // *Molecular biology*, 212, pp. 403-428.
5. *Edgar R.C., Batzoglou G.* Multiple sequence alignment // *Current opinion in structural biology*, 16, 2006, pp. 368-373.
6. *Wallance I.M., Blackshields G., Higgins D.G.* Multiple sequence alignments // *Current opinion in structural biology*, 15, 2005, pp. 261-266.
7. *Notredame C.* Recent evolution of multiple sequence alignment algorithms // *PLoS Computational biology*, vol.3, 2007, pp. 1405-1408.
8. *Pei J., Kim B.-H., Grishin N.V.* PROMALS3D: A tool for multiple protein sequence and structure alignments // *Nucleic acids*, vol.33, No.7, 2008, pp. 2295-2300.
9. *Pei J., Grishin N.V.* MUMMALS: Multiple sequence alignment improved by using hidden Markov models with local structural information // *Nucleic acids*, vol.34, 2006, pp. 4364-4374.

10. *Thompson J., Higgins D., Gibson T.* CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. // *Nucleic acids*, vol.22, 1994, pp. 4673-4690.
11. *Dayhoff M.O., Schwartz R.M., Orcutt B.C.* A model of evolutionary change in proteins. // In: *Atlas of proteins sequences and structure*. National biomedical research foundation, Vol.5, 1978, pp. 345-352.
12. *Айзерман М.А., Браверман Э.М., Розоноэр Л.И.* Метод потенциальных функций в теории обучения машин. // М.: Наука, 1970, 384 с.
13. *Моттль В.В.* Метрические пространства, допускающие введение линейных операций и скалярного произведения. // *ДАН*, 2003, том 67, №1.
14. *Smith T.F., Watterman M.S.* Identification of common molecular sequences. // *Molecular Biology*, 1981, 147, pp.195-197.
15. *Шлезингер М.И.* О самопроизвольном различении образов. // *Читающие автоматы*. – Киев: Наукова думка. 1965, С. 38–45.
16. *Bishop C.M.* *Pattern Recognition and Machine Learning*. // Springer, 2006.
17. *Моттль В.В., Мучник И.Б.* Скрытые марковские модели в структурном анализе сигналов. // М.: ФИЗМАТЛИТ, 1999б 352 с.

Статья поступила в редакцию 30.04.2008