

**ПРИМЕНЕНИЕ ФОРМАЛИЗМА ТРАНСФЕР-МАТРИЦ К АНАЛИЗУ
МНОГОКОМПОНЕНТНОГО СВЯЗЫВАНИЯ БИОЛОГИЧЕСКИ АКТИВНЫХ МОЛЕКУЛ С ДНК**

Бучельников А.С.

Белгородский государственный национальный исследовательский университет

ул. Победы, 85, г. Белгород, 308015, РФ

e-mail: buchelnikov@bsu.edu.ru

Аннотация. Статистико-механический формализм трансфер-матриц впервые использован для получения обобщенного выражения классической модели МакГи – фон Хиппеля для случая некооперативного и неспецифического связывания лигандов различных типов с решеточным биополимером.

Ключевые слова: биополимер, лиганд, константа ассоциации, статистическая сумма, трансфер-матрица.

**APPLICATION OF TRANSFER MATRIX FORMALISM FOR ANALYSIS
OF MULTICOMPONENT BINDING OF BIOLOGICALLY ACTIVE MOLECULES TO DNA**

Buchelnikov A.S.

Belgorod State National Research University

Pobedy str. 85, Belgorod, 308015, Russia

e-mail: buchelnikov@bsu.edu.ru

Abstract. The generalized derivation of classical McGhee – von Hippel equation for the first time was performed by means of statistical-mechanical transfer matrix formalism. The equation is applicable for study of non-cooperative and non-specific binding of ligands of various types to a lattice biopolymer.

Keywords: biopolymer, ligand, association constant, partition function, transfer matrix.

Исследование нековалентного связывания малых молекул (лигандов) с биополимерами является классической задачей молекулярной биофизики. Важным элементом такого исследования является построение статистико-термодинамической модели комплексообразования, устанавливающей связь между экспериментальными данными и параметрами, характеризующими сродство лиганда к биополимеру (равновесные константы комплексообразования, параметры исключения, кооперативности и т.д.). Особенностью вывода моделей комплексообразования (адсорбции) лигандов на биополимере является использование методов статистической термодинамики. Разные исследовательские группы отдают разное предпочтение тем или иным методам, однако, судя по данным литературы, наиболее частое и систематическое использование отмечается в отношении комбинаторных (или алгебраических) методов (например, [1,2]), метода производящих последовательность функций (sequence generating function [3]) и алгоритмических методов [4]. Вместе с тем, в статистической физике хорошо известен еще один метод – метод трансфер-матриц (transfer matrix, ТМ), который до недавнего времени оставался недооцененным в области моделирования процессов комплексообразования, и лишь несколько лет назад был «реанимирован» для активного использования группой Тейфа [5].

В настоящей работе нами впервые проведен анализ возможности использования метода ТМ для получения обобщенного выражения классической модели МакГи – фон Хиппеля [1] для случая некооперативного и неспецифического связывания m различных типов лигандов с решеточным биополимером.

Формализм трансфер-матриц основан на представлении исследуемой системы в качестве регулярной решетки, каждая ячейка которой характеризуется определенным статистическим весом, соответствующим определенному состоянию (например, занятый лигандом или свободный сайт ДНК). Квадратная матрица, состоящая из таких статистических весов, и называется трансфер-матрицей, а ее след (сумма диагональных элементов) представляет собой статистическую сумму описываемой ею системы, т.е. состояние ячеек решетки [6,7].

Ключевая идея записи трансфер-матрицы проста [5]: квадратная матрица \mathbf{M} заполняется элементами M_{ij} , равными статистическим весам соответствующих ячеек решетки, находящихся в i -м состоянии, за которыми следуют ячейки, находящиеся в j -м состоянии. Физический смысл этого утверждения заключается в том, что элемент трансфер-матрицы отражает условную вероятность того, что данная ячейка решетки находится в состоянии i в то время как непосредственно следующая за ней ячейка находится в состоянии j . Запрещенные состояния ячеек соответствуют статистическому весу, равному нулю.

Трансфер-матрица \mathbf{M} содержит статистические веса всех возможных состояний для одного звена решетки. Тогда возведение ее в степень N благодаря свойствам матричного произведения даст исчерпывающую информацию обо всех возможных состояниях одномерной решетки длиной N звеньев. Далее, следуя данному выше определению, получаем статистическую сумму Z в виде следа N -ной степени матрицы:

$$Z = \text{Tr } \mathbf{M}^N.$$

Применяя преобразование подобия к трансфер-матрице, приходим к выводу, что статистическая сумма равна сумме N -ных степеней всех собственных значений трансфер-матрицы \mathbf{M} :

$$Z = \text{Tr } \mathbf{M}^N = \text{Tr} \left(\mathbf{T} \mathbf{\Lambda} \mathbf{T}^{-1} \right)^N = \text{Tr } \mathbf{\Lambda}^N = \sum_{i=1}^p \lambda_i^N \quad (1)$$

где \mathbf{T} – вспомогательная матрица преобразования, Λ – диагональная матрица, состоящая из элементов λ_i , которые являются собственными значениями матрицы \mathbf{M} .

Основным свойством статистической суммы является то, что ее частные производные дают информацию о конкретных характеристиках системы. Так, например, если исследуется адсорбция лиганда на решетке либо линейная агрегация лигандов, то производная статистической суммы такой системы по натуральному логарифму мономерной концентрации дает общую концентрацию вещества в системе [7,8].

Рассмотрим конкурентное неспецифическое и некооперативное связывание m разнотипных протяженных лигандов с полимерной решеткой, состоящей из N одинаковых сайтов связывания лигандов. Понятие «протяженный лиганд» означает, что при связывании с решеткой такой лиганд «накрывает» собой сразу несколько сайтов на полимере. Пусть n_i – число сайтов решетки, занимаемой лигандом i -го типа; K_i – равновесная микроскопическая константа связывания лиганда i -го типа с решеткой; C_i – концентрация мономеров лиганда i -го типа. Квадратная

трансфер-матрица \mathbf{M} порядка $p = \sum_{i=1}^m n_i + 1$ имеет следующий вид:

$$\mathbf{M} = \begin{pmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 1 & 0 & \dots \\ 0 & 0 & (K_1 C_1)^{1/n_1} & 0 & \dots & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & (K_1 C_1)^{1/n_1} & \dots & 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \dots \\ 0 & 0 & 0 & 0 & \dots & (K_1 C_1)^{1/n_1} & 0 & 0 & \dots \\ (K_1 C_1)^{1/n_1} & (K_1 C_1)^{1/n_1} & 0 & 0 & \dots & 0 & (K_1 C_1)^{1/n_1} & 0 & \dots \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & (K_2 C_2)^{1/n_2} & \dots \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots \\ (K_2 C_2)^{1/n_2} & (K_2 C_2)^{1/n_2} & 0 & 0 & \dots & 0 & (K_2 C_2)^{1/n_2} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (2)$$

Каждый элемент матрицы (2) представляет собой статистический вес сайта полимерной решетки при условии, что за ним следует сайт в одном из разрешенных состояний. Очевидно, в рассматриваемом случае существует несколько таких состояний:

- 1) незанятый лигандом сайт; статистический вес такого сайта принимается равным 1;
- 2) сайт, занятый одной из n_1 частей протяженной молекулы первого типа; исходя из определения произведения вероятностей, статистический вес данного сайта равен корню n_1 -ной степени из произведения $K_1 C_1$;
- 3) сайт, занятый одной из n_2 частей протяженной молекулы второго типа; аналогично находится как $(K_2 C_2)^{1/n_2}$ и т.д.

Номер строки/столбца матрицы (2) соответствует последовательно: незанятому сайту; сайту, занятому первой частью молекулы первого типа; ...; сайту, занятому последней n_1 -ной частью молекулы первого типа; сайту, занятому первой частью молекулы второго типа и т.д.

Согласно выражению (1) для нахождения статистической суммы системы требуется знать спектр (совокупность собственных значений) матрицы, который находится из характеристического уравнения вида:

$$\mathbf{M}\mathbf{c} = \lambda\mathbf{c}, \quad (3)$$

где $\mathbf{c} = (1 \ c_1 \ \dots \ c_{p-1})^T$ – собственный вектор матрицы \mathbf{M} . Матричное уравнение (3) эквивалентно системе p линейных уравнений:

$$\begin{cases} 1 + c_1 + c_{n_1+1} + \dots = \lambda \\ (K_1 C_1)^{1/n_1} c_2 = \lambda c_1 \\ \dots \\ (K_1 C_1)^{1/n_1} c_{n_1} = \lambda c_{n_1-1} \\ (K_1 C_1)^{1/n_1} (1 + c_1 + c_{n_1+1} + \dots) = \lambda c_{n_1} \\ (K_2 C_2)^{1/n_2} c_{n_1+2} = \lambda c_{n_1+1} \\ \dots \\ (K_2 C_2)^{1/n_2} c_{n_1+n_2} = \lambda c_{n_1+n_2-1} \\ (K_2 C_2)^{1/n_2} (1 + c_1 + c_{n_1+1} + \dots) = \lambda c_{n_1+n_2} \\ \dots \end{cases}. \quad (4)$$

Упростим систему (4) и перепишем ее в другом виде

$$\left\{ \begin{array}{l} 1 + c_1 + c_{n_1+1} + \dots + c_{n_{m-1}+1} = \lambda \\ (K_1 C_1)^{1/n_1} c_{i+1} = \lambda c_i, \quad (i = 1, 2, \dots, n_1 - 1) \\ (K_1 C_1)^{1/n_1} = c_{n_1} \\ (K_2 C_2)^{1/n_2} c_{j+1} = \lambda c_j, \quad (j = n_1 + 1, n_1 + 2, \dots, n_1 + n_2 - 1) \\ (K_2 C_2)^{1/n_2} = c_{n_1+n_2} \\ \dots\dots\dots \\ (K_m C_m)^{1/n_m} c_{j+1} = \lambda c_j, \quad \left(j = \sum_{i=1}^m n_i + 1 - n_m, \sum_{i=1}^m n_i + 2 - n_m, \dots, \sum_{i=1}^m n_i - 1 \right) \\ (K_m C_m)^{1/n_m} = c_{n_1+n_2+\dots+n_m} \end{array} \right. \quad (5)$$

Из каждого второго уравнения системы (5) имеем

$$(K_k C_k)^{(n_k-1)/n_k} c_t = \lambda^{n_k-1} c_{t-n_k+1}, \quad t = \sum_{i=1}^k n_i.$$

С учетом каждого третьего уравнения системы (5) окончательно получим

$$K_k C_k = \lambda^{n_k-1} c_{t-n_k+1}, \quad t = \sum_{i=1}^k n_i. \quad (6)$$

Подставляя выражения (6) в первое уравнение системы (5), будем иметь после несложных преобразований

$$\lambda - 1 = K_1 C_1 \lambda^{1-n_1} + K_2 C_2 \lambda^{1-n_2} + \dots + \sum_{i=1}^m K_i C_i \lambda^{1-n_i}. \quad (7)$$

Согласно [6,7] доля сайтов решетки длиной N , занятых молекулой j -го типа, равна $n_j \theta_j$, где θ_j может быть найдено с помощью статистической суммы Z :

$$\theta_j = \frac{1}{N} \frac{\partial \ln Z}{\partial \ln C_j}.$$

Подставляя (2) в последнее выражение, получим

$$\theta_j = \frac{1}{N} \frac{\partial \ln \left(\sum_{i=1}^p \lambda_i^N \right)}{\partial \ln C_j}.$$

В приближении решетки бесконечной длины можем принять

$$\lim_{N \rightarrow \infty} Z \approx \lambda_{\max}^N,$$

где λ_{\max} – наибольшее собственное значение (спектральный радиус) матрицы \mathbf{M} .

Отсюда

$$\theta_j = \frac{\partial \ln \lambda_{\max}}{\partial \ln C_j} = \frac{C_j}{\lambda_{\max}} \frac{\partial \lambda_{\max}}{\partial C_j}. \quad (8)$$

Продифференцируем (7) по λ_{\max} и по C_j , а затем найдем производную (8):

$$\begin{aligned} d\lambda &= \sum_{i=1}^m (1-n_i) K_i C_i \lambda^{-n_i} d\lambda + K_j \lambda^{1-n_j} dC_j, \\ \lambda \theta_j &= K_j C_j \lambda^{1-n_j} + \theta_j \sum_{i=1}^m (1-n_i) K_i C_i \lambda^{1-n_i}. \end{aligned}$$

Отсюда с использованием уравнения (7) получим

$$\frac{K_j C_j}{\theta_j} \lambda^{1-n_j} = \lambda - \sum_{i=1}^m K_i C_i \lambda^{1-n_i} + \sum_{i=1}^m n_i K_i C_i \lambda^{1-n_i} = 1 + \sum_{i=1}^m n_i K_i C_i \lambda^{1-n_i}. \quad (9)$$

Последние равенства можем переписать в следующем виде:

$$\frac{K_1 C_1}{\theta_1} \lambda^{1-n_1} = \frac{K_2 C_2}{\theta_2} \lambda^{1-n_2} = \dots = \frac{K_m C_m}{\theta_m} \lambda^{1-n_m}. \quad (10)$$

В общем случае система уравнений (10) может быть записана в виде

$$K_i C_i \lambda^{1-n_i} = \frac{\theta_i}{\theta_j} K_j C_j \lambda^{1-n_j}, \quad i = \overline{1, m}.$$

Домножая обе части уравнения на n_i , получим

$$n_i K_i C_i \lambda^{1-n_i} = n_i \theta_i \frac{K_j C_j \lambda^{1-n_j}}{\theta_j}, \quad i = \overline{1, m}. \quad (11)$$

Суммируя (11) по всем i , будем иметь

$$\sum_{i=1}^m n_i K_i C_i \lambda^{1-n_i} = \frac{K_j C_j}{\theta_j} \lambda^{1-n_j} \sum_{i=1}^m n_i \theta_i. \quad (12)$$

Подставляя (12) в выражение (9), получим

$$\frac{K_j C_j}{\theta_j} \lambda^{1-n_j} \left(1 - \sum_{i=1}^m n_i \theta_i \right) = 1,$$

откуда

$$K_j C_j \lambda^{1-n_j} = \frac{\theta_j}{1 - \sum_{i=1}^m n_i \theta_i}, \quad j = \overline{1, m}. \quad (13)$$

Просуммируем выражения типа (13) по всем j , тогда после несложного преобразования получим

$$\lambda = \frac{1 - \sum_{i=1}^m (n_i - 1) \theta_i}{1 - \sum_{i=1}^m n_i \theta_i}. \quad (14)$$

В результате, подставляя (14) в (13) и переходя к координатам Скэтчарда, имеем систему уравнений МакГи – фон Хиппеля [1] для конкурентного некооперативного связывания m протяженных лигандов с полимерной решеткой, записанную в максимально общем виде:

$$\frac{\theta_j}{C_j} = K_j \left(1 - \sum_{i=1}^m n_i \theta_i \right) \left(\frac{1 - \sum_{i=1}^m n_i \theta_i}{1 - \sum_{i=1}^m (n_i - 1) \theta_i} \right)^{n_j - 1}. \quad (15)$$

Уравнение (15) представляет собой классическое уравнение МакГи – фон Хиппеля связывания лигандов с одномерной решеткой, моделирующей биополимер, выведенное в общем случае путем применения формализма трансфер-матриц. Этот результат расширяет возможности по моделированию систем лиганд-полимер, позволяя, в частности, использовать аппарат матричной алгебры, эффективно реализуемый в настоящее время в различных пакетах прикладной математики (например, MATLAB).

Список литературы / References:

1. McGhee J.D., von Hippel P.H. Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *J. Mol. Biol.*, 1974, vol. 86, pp. 469-489.
2. Нечипуренко Ю.Д. Анализ связывания лигандов с нуклеиновыми кислотами. *Биофизика*, 2014, т. 59, с. 12-36. [Nechipurenko Yu.D. Analysis of the binding of ligands to nucleic acids. *Biophysics*, 2014, vol. 59, pp. 12-36 (In Russ.)]
3. Woodbury C.P. Jr. Matrix polynomial extension of the sequence-generating function method for macromolecular binding. *J. Chem. Phys.*, 1990, vol. 92, pp. 5127-5135.
4. Beshnova D.A., Bereznyak E.G., Shestopalova A.V., Evstigneev M.P. A novel computational approach 'BP-STOCH' to study ligand binding to finite lattice. *Biopolymers*, 2011, vol. 95, pp. 208-216.
5. Teif V.B. General transfer matrix formalism to calculate DNA-protein-drug binding in gene regulation: application to OR operator of phage λ . *Nucleic Acids Res.*, 2007, vol. 35, Art. No. e80.
6. Poland D., Scheraga H.A. Theory of helix-coil transitions in biopolymers. New York: Academic Press, 1970. 797 p.
7. Tsuchiya T., Szabo A. Cooperative binding of n-mers with steric hindrance to finite and infinite one-dimensional lattices. *Biopolymers*, 1982, vol. 21, pp. 979-994.
8. Buchelnikov A.S., Evstigneev V.P., Evstigneev M.P. General statistical-thermodynamical treatment of one-dimensional multicomponent molecular hetero-assembly in solution. *Chem. Phys.*, 2013, vol. 421, pp. 77-83.