

© Коллектив авторов

## ПРОГНОЗИРОВАНИЕ ТЕРМОСТАБИЛЬНОСТИ БЕЛКОВ ПО ИХ ПЕРВИЧНОЙ СТРУКТУРЕ: СОВРЕМЕННОЕ СОСТОЯНИЕ И ФАКТОРЫ РАЗВИТИЯ

*Д.В. Гришин\*, М.В. Покровская, О.В. Подобед, Ю.А. Гладилина, В.С. Покровский,  
С.С. Александрова, Н.Н. Соколов*

Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича,  
119121, Москва, ул. Погодинская, д. 10, стр. 8; эл. почта: molbiol\_ibm@inbox.ru

Конструирование белков и пептидов с заданными свойствами, включающими устойчивость к высоким температурам, а также оптимизация их аминокислотного состава является важной и сложной задачей, как в различных отраслях фундаментальных наук, так и биомедицины и биотехнологии. При этом возникает вопрос: какой метод целесообразнее выбрать на предварительной стадии исследований для того, чтобы без лишних затрат времени и средств оценить влияние планируемых аминокислотных замен на термостабильность итоговой белковой конструкции. В представленном кратком обзоре проведена классификация существующих основных практических и теоретических подходов к исследованию и прогнозированию термостабильности нативных и рекомбинантных полипептидов. Особое внимание уделено прогностическому потенциалу статистических методов изучения термодинамических показателей первичной структуры белка и оценены их перспективы.

**Ключевые слова:** полипептиды, прогнозирование термостабильности белков, первичная структура белка, молекулярная динамика, компьютерные алгоритмы, термодинамические свойства белков

DOI 10.18097/PBMC20176302124

### ВВЕДЕНИЕ

Термостабильные белки всё чаще применяются в фундаментальных и прикладных исследованиях. Так, в сельском хозяйстве для улучшения питательности и переваримости кормов при их высокотемпературной обработке используются термостабильные ферменты. С этой целью для улучшения утилизации организмом животных различных компонентов грубых растительных кормов создаются рекомбинантные ферменты. К их числу относятся амилазы, способствующие расщеплению крахмала;  $\beta$ -глюканазы, расщепляющие глюканы; термостабильные фитазы, влияющие на усвояемость фосфатов; протеазы, гидролизующие крупные белки корма и др. [1-3]. В последнее время добавление в рационы сельскохозяйственных животных биоактивных короткоцепочечных пептидов рассматривается в качестве перспективного способа иммуномодуляции, улучшения роста, эффективности усвоения питательных веществ, улучшения работы желудочно-кишечного тракта и, как следствие, увеличения продуктивности [4-6]. В рамках этого перспективным является разработка балансирующих пептидов кормового назначения, обладающих заданным стехиометрическим составом незаменимых для разных групп сельскохозяйственных животных аминокислот и термостабильностью в качестве свойства, направленного на значительное упрощение и удешевление процедуры очистки данных белков.

Термостабильные ферменты также нашли своё применение и в пищевой промышленности. Например, для предотвращения появления канцерогенов в конечном продукте используются термостабильные L-аспарагиназы. Цель их введения в продукцию – это предотвращение развития нежелательной

реакции Майяра. Реакция представляет собой высокотемпературную каталитическую трансформацию свободного L-аспарагина в канцерогенный акриламид, при участии реактивной карбонильной группы некоторых моносахаридов, присутствующих в пище. Добавление же к пищевому сырью, подвергающемуся термической обработке, термостойкой L-аспарагиназы приводит к блокированию данной реакции за счёт удаления свободного L-аспарагина – ключевого компонента реакции [7]. В отличие от кормопроизводства, в ряде случаев, в пищевой промышленности и медицине, термостабильность белка может стать мешающим фактором, так как конечный продукт трудно будет избавить от присутствия следовых количеств данного белка, особенно если это белок-носитель или адьювантный белок.

В фундаментальных биологических и биомедицинских исследованиях термостабильные белки часто используются в качестве различного рода биосенсоров [8, 9]. Большое значение представляет структурная оптимизация таких ферментов нуклеинового обмена как, например, ДНК-полимеразы, использующиеся в полимеразной цепной реакции (ПЦР), широко применяемой в молекулярно-генетических исследованиях [8, 9]. При этом оптимизация подобных ферментов сводится, прежде всего, к повышению их термостабильности, специфичности и процессивности [9, 10].

Многочисленные исследования направлены на поиск принципов термостабильности белков и их применения к предсказанию этого явления [11, 12].

Часто при создании рекомбинантных белков и пептидов, ценных с промышленной или биомедицинской точки зрения, требуется улучшить

их структурно-функциональную организацию путём внесения в аминокислотную последовательность точечных и/или кластерных замен, сохранив при этом такой важный параметр, как термостабильность исходного нативного белка. На начальном этапе подобных исследований возникает главный вопрос, как с наибольшей степенью достоверности оценить влияние планируемых замен на стабильность белковой молекулы. Экспериментальные методы исследования и прогнозирования термостабильности белков могут предоставить весьма подробную информацию о связи структурной модификации молекулы и соответствующих ей изменений термостабильности.

К подобным *эмпирическим* стратегиям можно отнести методы дифференциальной сканирующей калориметрии, КД-спектроскопии, а также направленной эволюции, такие как клеточный и фаговый дисплей, сайт-направленный мутагенез и т.п. [13, 14]. В настоящее время, наиболее часто применяется сайт-направленный мутагенез. Целью его является повышение жесткости белковой цепи за счёт введения остатков пролина; снятие молекулярных напряжений при точечном введении остатков глицина [15]; конструирование белков с повышенной плотностью поверхностных заряженных аминокислотных остатков, стабилизирующих белок в результате снижения диэлектрической проницаемости воды при увеличении температуры [16], а также модификация структуры белков для увеличения ароматических стэкинг-взаимодействий, либо для введения стабилизирующих дисульфидных связей [17, 18]. Несмотря на ряд успешных работ следует отметить, что данная группа методов не имеет прогностического потенциала и лимитируется высокой трудоёмкостью. Кроме того, изменения в стабильности могут быть обусловлены различными взаимодействиями, дающими как положительный, так и отрицательный вклад в свободную энергию, который сложно оценить экспериментально [9].

Для преодоления данных ограничений применяют теоретические методы, которые позволяют более оперативно и подробно рассмотреть термодинамику белков, а также другие аспекты влияния замен на термостабильность. Задачей настоящей работы стала, в том числе, и классификация существующих теоретических подходов к исследованию термостабильности белков. По каждому классу методов сделаны соответствующие выводы о применимости, ограничениях и получаемых с их помощью результатах.

Теоретические методы можно разделить на следующие основные группы [9, 19]:

1) сравнительное исследование белков термофильных и мезофильных организмов для прямого выявления закономерностей, ведущих к их температурной стабилизации [20-24];

2) методы молекулярной динамики – теоретические методы, принимающие электростатические и гидрофобные взаимодействия в качестве наиболее существенных среди всех взаимодействий, влияющих на термостабильность [25, 26];

3) компьютерные алгоритмы, учитывающие физические аспекты, влияющие на температурные характеристики белков [9, 27- 32].

## 1. СРАВНИТЕЛЬНОЕ ИССЛЕДОВАНИЕ БЕЛКОВ ТЕРМОФИЛЬНЫХ И МЕЗОФИЛЬНЫХ ОРГАНИЗМОВ

Данный метод является исторически первым и основывается на использовании спектра белков из различных термофильных и мезофильных организмов, для которых оцениваются до 13 структурных параметров [24]. Все параметры объединяются в группы по следующим признакам: вогнутость белковой молекулы (рассчитывается с помощью программы *MolecularSurfacePackage* [33]); число водородных связей (рассчитывается с помощью пакета молекулярного моделирования *WHAT IF* [34]); число ионных пар, при условии, что расстояние между заряженными радикалами аминокислот не превышает 4-8 Å; параметры вторичной структуры (доля спиралей, бета слоёв и нерегулярных областей с помощью программы *DSSP* [35]); отношение площадей поверхности полярных и неполярных областей для внешних и внутренних областей белка (для расчёта также используется программа *WHAT IF* [34]).

Исходя из рассчитанных значений для всего набора белков, делаются выводы о корреляции каждого из параметров с рядом характеристик, в том числе и оптимальной температурой роста организма-хозяина [24]. Данный метод предоставляет обширную, но, в основном, не количественную информацию и более подходит для изучения уже существующих белков, нежели чем для прогнозирования изменений термостабильности проектируемых молекул [9].

## 2. МЕТОДЫ МОЛЕКУЛЯРНОЙ ДИНАМИКИ

Методы молекулярной динамики (МД) нацелены на поиск термостабильных областей в исследуемых белках [16; 36]; они позволяют вычислять классические траектории отдельных атомов и полимерных цепей, а также исследовать динамику взаимодействия частиц в конденсированных системах на молекулярном уровне [19, 37]. Эти методы особенно эффективны для исследования биологических макромолекул в тех масштабах, где квантовые эффекты менее существенны, чем электростатические и гидрофобные взаимодействия, которые, в данном случае, зачастую называются доминирующими эффектами [19, 38-40].

В случае, если в качестве доминирующего принимается, например, электростатическое взаимодействие, то электростатическая энергия каждого ионизационного состояния определяется с помощью решения вариантов уравнения Пуассона–Больцмана [41]. Основываясь на таких расчётах, некоторые авторы полагают, что в белках гипертермофильных организмов вклад

электростатических взаимодействий в свободную энергию фолдинга более существенен, чем в их мезофильных аналогах и всецело зависит от значения pH среды [9, 25].

Следует отметить, что основные вычислительные затраты МД-моделирования прежде всего связаны с адекватным расчётом всех возможных электростатических и/или гидрофобных взаимодействий. С появлением подобных методов и программ для обработки данных появились определённые успехи в области предсказания термостабильности, однако такое направление по-прежнему имеет некоторые практические ограничения, к основным из которых относится необходимость адекватного выбора потенциала взаимодействия и оптимизации его параметров, а также необходимость учёта всех возможных конфигураций изучаемой системы. Для этой цели создаются и совершенствуются специализированные процессоры и мощные вычислительные кластеры. Так, на основе подобного вычислительного комплекса для молекулярно-динамических расчётов CSD ACC RIKEN (Япония) проводится успешное моделирование физических и биологических конденсированных систем [19, 40, 42-46].

### 3. КОМПЬЮТЕРНЫЕ АЛГОРИТМЫ, УЧИТЫВАЮЩИЕ ФИЗИЧЕСКИЕ АСПЕКТЫ, КОТОРЫЕ ОКАЗЫВАЮТ ВЛИЯНИЕ НА ТЕМПЕРАТУРНЫЕ ХАРАКТЕРИСТИКИ БЕЛКОВ

Большинство из рассмотренных выше методов позволяют дать достаточно надежное качественное предсказание влияния мутации на термостабильность белка, поскольку учитывают реальные физические внутрибелковые взаимодействия [47]. Однако подобные методы зачастую весьма ресурсоёмки, не позволяют произвести прикидочные расчёты исходя только лишь из аминокислотной последовательности; они не дают точного численного значения оцениваемой величины. Таким образом, задача простого численного расчета эффектов влияния мутаций в теле некоторого полипептида на его термостабильность является важнейшей задачей для современной биотехнологической науки в целом и биоинформатики, в частности.

К группе методов, направленных на решение данной проблемы, можно отнести компьютерные алгоритмы, учитывающие физические аспекты и температурные характеристики белков.

Среди этой группы, с нашей точки зрения, можно выделить три перспективных метода:

- изучение термодинамической устойчивости белковой молекулы на основании изменения энергии Гиббса ( $\Delta G$ );
- метод расчёта локальной структурной энтропии LSE (*local structural entropy*);
- статистический метод вычисления/прогнозирования значения TI или Tm Index.

#### 3.1. Изучение термодинамической устойчивости белковой молекулы на основании вычисления/прогнозирования изменения свободной энергии Гиббса ( $\Delta G$ )

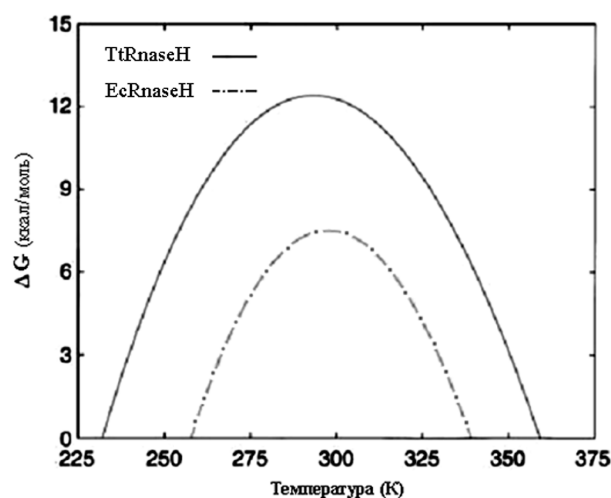
Изучение термодинамической устойчивости белковой молекулы на основании вычисления/прогнозирования изменения свободной энергии Гиббса ( $\Delta G$ ) производится для нативной (N) и денатурированной (D) форм белковой молекулы [48], находящихся в некотором равновесии:  $N \leftrightarrow D$ . То же самое может быть реализовано для мутантной (M) и нативной (N) форм белка  $N \leftrightarrow M$ .

Базовая формула для подобных расчётов выглядит следующим образом:

$$\Delta G(T_s) = \Delta H(T_s) = \Delta H_G - (T_G - T_s)\Delta C_p \quad (1)$$

где:  $\Delta G(T_s)$  – энергия Гиббса молекулы белка при температуре максимальной стабильности,  $T_s$  – температура максимальной стабильности белка или температура, при которой энтропия белковой молекулы равна нулю,  $T_G$  или  $T_m$  – (*melting temperature*), то есть температура, при которой изучаемый белок денатурировал на 50%,  $\Delta C_p$  – изменение теплоёмкости при постоянном давлении,  $\Delta H_G$  – изменение энтальпии.

На рисунке 1 представлена наглядная графическая интерпретация расчётов  $\Delta G$ , на примере сравнения термостабильности рибонуклеаз Н из *E. coli* и *T. thermophilus* [48, 49].



**Рисунок 1.** Сравнение термостабильности рибонуклеаз Н из *E. coli* (EcRnaseH) и *T. thermophilus* (TtRnaseH). Кривая стабильности термостабильного белка смещена вверх относительно кривой, характеризующей мезофильный белок. Данные адаптированы из [48, 49].

#### 3.2. Метод расчета локальной структурной энтропии (LSE)

Ряд работ, основанных на термодинамическом анализе, указывает на важную роль энтропийных факторов в стабилизации термостабильных белков по отношению к их мезофильным гомологам [50, 51]. Именно поэтому LSE считается перспективным методом [52-56]. LSE является производным от рассмотренного выше метода; единственное

отличие сводится к тому, что в качестве “маркера” термостабильности выбирается не изменение свободной энергии Гиббса ( $\Delta G$ ), а величина структурной энтропии (2):

$$S = - \sum_{j=1}^8 (\bar{\pi}_j) \ln \bar{\pi}_j \quad (2)$$

где:  $\bar{\pi}_j$  – это  $j$ -й компонент структурного вектора  $\bar{p}$  изучаемой биоинформационной последовательности (подробнее см. [53]).

В процессе масштабных термодинамических расчётов авторы данной методики сформулировали постулат о том, что между энтропией и термостабильностью полипептидов существует линейная зависимость [51, 53]. Это же правило находит отражение и в графическом соотношении расчётной и фактической температур денатурации белка (см. рис. 2). Совокупность полученных данных легла в основу нового метода, базирующегося на том, что сегменты белка, одинаково хорошо существующие в различных конфигурациях вторичной структуры, имеют более высокую энтропию, чем те, которые присутствуют в основном в одной или нескольких конфигурациях на некотором участке первичной структуры белка [53].

Основополагающая задача LSE сводится к анализу структурной энтропии на первичном и вторичном уровнях организации белковой молекулы в окрестности четырёх соседних аминокислотных остатков с последовательным смещением на одну аминокислоту от N- к С-концу белковой молекулы (подробнее см. [53]).

### 3.3. Статистический метод вычисления/прогнозирования значения $TI$ ( $T_m$ Index)

В настоящее время всё большее внимание уделяется статистическим методам изучения термодинамических показателей первичной

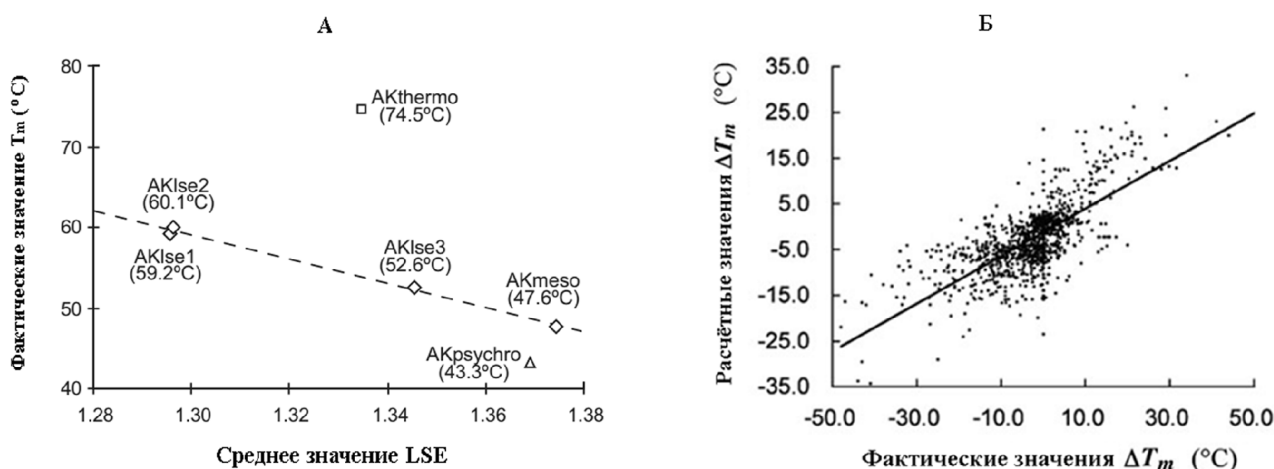
структуры белка, то есть его аминокислотной последовательности, нацеленным на оперативный расчёт и наглядный численный результат.

В рамках этого необходимо выделить сравнительно новый алгоритм, базирующийся на вычислении значения  $TI$  или  $T_m$  Index, то есть температуры, при которой изучаемый белок денатурировал на 50% с параллельным соотношением этих данных с частотами встречаемости тех или иных аминокислотных пар в первичной структуре белка [52, 53, 57].

Уникальность метода заключается в отсутствии необходимости сравнительного анализа гомологов, сложных термодинамических расчётов, а также изучения белковых структур высшего порядка. Базовая задача сводится к анализу аминокислотной последовательности как таковой.

В основе метода лежат многочисленные практические исследования 75 полноразмерных геномов термофильных и мезофильных организмов, в результате которых авторы [57] пришли к важному выводу о том, что одна аминокислота сама по себе не может рассматриваться в качестве необходимой и достаточной основы для существенного изменения термостабильности белка. В то время как наблюдается корреляция между частотой встречаемости ряда аминокислотных пар (дипептидов, дуплексов) с изменением свойств термостабильности белков при нейтральных значениях pH. Так, например, увеличение доли аминокислотных пар His-Cys, Trp-Met и Cys-Pro в последовательности белка приводит к повышению  $T_m$ , а пары Met-Met, Trp-Cys, Asp-Trp и Trp-Pro – наоборот, существенно снижают это значение [57].

Исходя из сказанного выше, Ку и соавт. [57] разработали алгоритм удобной аналитической программы “ $T_m$  Predictor” (<http://tm.life.nthu.edu.tw/>). В основу данного программного средства положен массив данных средневзвешенных значений  $T_m$  или иначе  $P_{index}(xy)$  для 400 возможных парных



**Рисунок 2.** А - линейная зависимость между LSE и температурной стабильностью полипептидов, выделенных из психро- (*AKpsychro*), мезо- (*AKmeso*) и термофильных (*AKthermo*, *AKlse*) микроорганизмов; (данные адаптированы из [51]). Б - зависимость расчётных (ось ординат) и фактических (ось абсцисс) значений  $\Delta T_m$  (°C), полученных для 1153 репрезентативных белков и пептидов (коэфф. линейной регрессии  $r=0,721$ ;  $p=0,143 \times 10^{-3}$ ); (данные адаптированы из [53]).

комбинаций 20-ти аминокислот различных белков, рассчитанных при помощи системы математических выражений (уравнение 3, 4; рисунок 3):

$$TI = \frac{\left(\frac{100}{L}\right) \sum_{i=1}^{L-1} P_{index}(x_i y_{i+1}) - 9372}{398} \quad (3)$$

где:

$$P_{index}(xy) = [P_{rev-High-Tm}(xy) - P_{rev-Low-Tm}(xy) + 1] \times 100 \quad (4)$$

$P_{rev-high-Tm}(xy)$  – статистический показатель частоты встречаемости аминокислотной пары “xy” в белке L и влияния её на увеличение термостабильности;  $P_{rev-Low-Tm}(xy)$  – статистический показатель частоты встречаемости аминокислотной пары “xy” в белке L и влияния её на снижение термостабильности; 100 – коэффициент масштабирования; 9372 и 398 – безразмерные эмпирические коэффициенты; L – количество аминокислот в полипептиде

Таким образом, результатами программных вычислений являются численные значения TI, полученные на основе простого анализа аминокислотной последовательности целевого полипептида.

Спектр значений TI, полученный для нескольких тысяч белков, позволил авторам методики объединить их в три базовые группы [57]:

TI > 1 (для белков с Tm > 65°C);

TI < 0 (для белков с Tm < 55°C);

TI 0-1 (для белков с Tm 55°C~65°C).

Эти значения обладают определённым предиктивным потенциалом, они численно и наглядно отражают уровень термостабильности белковой молекулы. Для демонстрации возможности данного способа результаты вычисления TI и Tm, полученные с помощью пакета программ “Tm Predictor” для различных белков, в том числе для тех, с которыми работают авторы настоящего обзора, сведены в таблице.

В таблице сопоставлены фактические значения термостабильности и ожидаемые показатели, предсказанные программой для белков из разных биологических источников. Представленные в таблице результаты хорошо согласуются с известными значениями термической устойчивости для целого ряда белков термофильных и мезофильных организмов.

## ЗАКЛЮЧЕНИЕ

В своей работе нам часто приходится сталкиваться с необходимостью внесения точечных или кластерных замен в аминокислотную последовательность для повышения или снижения температурного оптимума исследуемого белка или пептида. Для составления оперативного прогноза степени влияния тех или иных мутаций на термостабильность целевого белка мы использовали способ расчёта Tm Index на основе первичной структуры. Данный метод зарекомендовал себя с хорошей стороны. В большинстве случаев программа “Tm Predictor” безошибочно определяет потенциал температурной устойчивости белков. Между тем, данный алгоритм требует дальнейшего совершенствования, так как способен предсказать, условно говоря, термостабильность первого порядка, то есть термостабильность, ассоциированную с особенностями расположения аминокислотных дуплексов в первичной структуре полипептида. Термостабильность, связанную с особенностями сложного фолдинга, что особенно важно для крупных белков, данная программа предсказать не способна.

Между тем, в ряде случаев, особенности укладки белковой молекулы играют решающую роль. Наглядным примером здесь могут служить такие крупные белки как спидроин и кератин, термостабильность которых, по некоторым данным, значительно превосходит 150°C [83, 84]. Однако программный расчёт TI для элементарной,

Средневзвешенные значения Tm (P<sub>index</sub>(xy)) для дипептидов

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	100	100	<b>48.6</b>	<b>41.8</b>	100	100	100	100	126	130	150	<b>58.8</b>	100	<b>36</b>	<b>54.4</b>	100	141	168	100	<b>87</b>
C	<b>28.2</b>	100	<b>-78</b>	<b>29.3</b>	100	161	100	<b>32.1</b>	100	100	100	100	255	100	100	100	<b>-62.1</b>	100	100	100
D	100	100	100	100	142	<b>-12</b>	<b>24.7</b>	<b>55.8</b>	100	134	<b>25.1</b>	<b>52.4</b>	100	165	100	143	<b>48</b>	140	<b>-99</b>	100
E	100	<b>26.7</b>	100	110	100	178	100	100	100	<b>64.4</b>	102	143	<b>62.6</b>	<b>54.5</b>	142	<b>83.1</b>	115	100	100	<b>63.5</b>
F	100	168	100	100	<b>3.45</b>	100	248	<b>87.2</b>	153	<b>16.7</b>	100	100	<b>3.45</b>	168	100	144	100	<b>26.3</b>	197	100
G	100	100	100	174	107	124	100	150	<b>66.5</b>	<b>105</b>	<b>51.9</b>	138	100	<b>-15</b>	<b>61.3</b>	138	100	<b>2.2</b>	151	<b>67.3</b>
H	100	402	<b>93</b>	100	177	100	100	<b>-59</b>	<b>24.5</b>	<b>33.4</b>	100	<b>36</b>	189	100	<b>28.6</b>	184	100	100	100	100
I	100	<b>-43</b>	<b>35.3</b>	100	<b>-37</b>	136	244	100	100	100	100	100	100	100	100	100	<b>158</b>	100	100	<b>153</b>
K	100	<b>29</b>	100	100	105	100	100	100	141	<b>47.5</b>	100	165	<b>58.3</b>	100	137	<b>57.5</b>	100	132	100	121
L	100	<b>56.3</b>	100	132	100	<b>63.7</b>	100	100	100	133	<b>32.2</b>	100	100	<b>54.2</b>	100	100	100	100	100	173
M	164	100	226	<b>23.4</b>	100	153	100	100	100	100	<b>-209</b>	219	100	100	100	<b>46.6</b>	v 37.2	<b>37.1</b>	100	100
N	100	76	100	102	100	100	100	100	<b>3.49</b>	100	164	<b>-225</b>	100	116	166	100	100	161	100	<b>88.1</b>
P	154	159	100	100	100	<b>61</b>	100	100	100	116	100	100	139	168	<b>34</b>	100	<b>52.3</b>	<b>88.6</b>	100	100
Q	100	<b>-8.5</b>	147	156	<b>42.1</b>	132	188	<b>60.1</b>	100	<b>95.1</b>	100	100	100	253	<b>15.8</b>	<b>63.7</b>	100	<b>43.6</b>	<b>-40.6</b>	149
R	<b>62.1</b>	100	140	<b>66.3</b>	100	100	<b>25.9</b>	<b>16.3</b>	137	100	100	100	159	<b>47.5</b>	144	<b>67.6</b>	100	<b>93.6</b>	100	100
S	<b>68.8</b>	100	159	<b>68.6</b>	149	100	<b>88.5</b>	100	100	136	100	100	100	100	100	<b>44.6</b>	100	<b>55.6</b>	100	100
T	129	100	100	106	100	100	100	100	<b>34.1</b>	100	100	100	<b>62.8</b>	100	100	100	100	100	<b>-26.8</b>	100
V	100	<b>173</b>	100	100	100	<b>73.4</b>	<b>24.5</b>	100	174	158	100	100	<b>43.6</b>	100	<b>53.3</b>	100	<b>87.3</b>	146	100	<b>41.2</b>
W	100	<b>-153</b>	100	<b>25</b>	100	151	100	211	100	<b>-69</b>	306	100	<b>-84</b>	100	<b>158</b>	<b>25.5</b>	100	100	100	<b>-23</b>
Y	100	100	151	140	100	<b>11.7</b>	191	100	100	100	100	<b>88.1</b>	197	212	100	<b>40.1</b>	100	<b>30.9</b>	100	<b>21.6</b>

**Рисунок 3.** Вид матрицы средневзвешенных значений Tm или P<sub>index</sub>(xy) для парных комбинаций 20 известных аминокислот [57]. Строки обозначают первую аминокислоту дипептида, а столбцы - вторую. P<sub>index</sub>(xy) ≥ 100 означает, что пара аминокислот “xy” вносит существенный вклад в обеспечение термостабильности белка “L”. P<sub>index</sub>(xy) < 100 означает, что пара аминокислот “xy” вносит существенный вклад в снижение термостабильности белка “L”.

## ПРОГНОЗИРОВАНИЕ ТЕРМОСТАБИЛЬНОСТИ БЕЛКОВ

Таблица. Результаты вычисления значений TI и Tm, полученные с помощью пакета программ "TmPredictor" для различных белков

№	Название белка	NCBI Reference SequenceID	Фактическая термостабильность (°C)	Источник информации о фактической термостабильности	Результаты вычисления ожидаемой термостабильности в пакете программ «TmPredictor»	
					TI	Tm (°C)
1	DNA polymerase <i>Thermus aquaticus</i>	AAA27507.1	95	[58, 59]	1,65	>65
2	Sac7d <i>Sulfolobus sp.</i>	AAA80315.1	90	[60, 61]; СД*	1,49	>65
3	SSB <i>Thermotoga naphthophila</i>	KUK22843.1	80	СД*	1,62	>65
4	Hirudin <i>Poecilobdelamanillensis</i>	CAA51292.1	65	[62]	0,69	55-65
5	Ferritin <i>Pyrococcus furiosus</i>	WP_011011871.1	120	[63, 64]	1,68	>65
6	Mannose-binding lectin from <i>Dendrobium findleyanum</i>	ABU62812.1	85	[65]; СД*	1,42	>65
7	Alpha-amylase <i>Pyrococcus furiosus</i>	WP_014835154.1	98-100	[66-68]	1,33	>65
8	Chimeric lactase of <i>Th. ethanolicus</i> with dextran binding domain of <i>L. mesenteroides</i>	-	75	[69, 70]	1,01	>65
9	Hydrolase <i>Thermus thermophilus</i>	YP_144878.1	90	[71-73]	1,15	>65
10	Hydantoinease <i>Bacillus stearothermophilus</i>	AAC60487.1	80	[74-76]	1,24	>65
11	Egg white lysozyme <i>Gallus gallus</i>	AAL69327.1	55-65	[77]; СД*	0,41	55-65
12	Glutathione S-transferase <i>E. coli</i>	AIZ87861.1	50	[78, 79]	-0,38	<55
13	Histone <i>Homo sapiens</i>	AAC61625.1	36-47	[80]	-0,02	<55
14	Calmodulin, isoform E <i>Drosophila melanogaster</i>	NP_001286337.1	25-45	[81]	-3,19	<55
15	Myoglobin <i>Aplysiali macina</i>	CAA55885.1	52	[82]	-2,06	<55

Примечание: \*- собственные (не опубликованные) данные.

мономерной формы этих белков демонстрирует лишь средние параметры термической устойчивости. Это связано, по всей видимости, с тем, что термостабильность подобных белков строго ассоциирована с уровнем олигомеризации и пространственными структурами высшего порядка [85, 86]. Так, в случае паутинового белка элементарные фрагменты, не обладающие термостабильностью, объединяются вместе по принципу водородного и гидрофобного взаимодействия и образуют упорядоченную структуру, похожую на кристалл [87, 88]. Типичный размер плотно упакованных участков молекул спидроина составляет всего несколько нанометров. Внутри такой компактной структуры фрагменты уложены параллельно друг другу и сцеплены между собой водородными связями. Эти участки и обеспечивают выдающиеся показатели прочности и стабильности волокна, подобно некоторым композиционным материалам, например армированным пластмассам, устроенным по тому же принципу [87].

Таким образом, в заключение, следует отметить очевидность того, что факторы совершенствования охарактеризованных методов прогнозирования термостабильности белков лежат в плоскости развития вычислительных мощностей, протеомных исследований, пополнения международных баз данных аминокислотных последовательностей белков психро-, мезо- и термофильных организмов, усовершенствования иерархических и неиерархических алгоритмов, позволяющих одновременно учитывать влияние распределения аминокислотных дуплексов на низшем уровне молекулярной организации и их участие в формировании пространственных структур высшего порядка в процессе белкового фолдинга.

### БЛАГОДАРНОСТИ

Исследование выполнено за счёт гранта Российского научного фонда (проект №16-16-04086).

## ЛИТЕРАТУРА

- Namkung H., Leeson S. (1999) Poult. Sci., **78**, 1317-1319.
- Amerah A.M., Gilbert C., Simmins P.H., Ravindran V. (2011) WPSJ, **67**, 29-46.
- Kocher A., Choct M., Ross G., Broz J., Chung T.K. (2003) J. Appl. Poult. Res., **12**, 275-283.
- Feng X.Y., Ji C., Jiang H. (2002) J. China Agricultural University **7**, 107-113.
- Wang L., Ma Q., Cheng J., Guo B., Yue H. (2004) Asian-Australas. J. Anim. Sci., **17**, 1277-1280.
- Ravindran V. (2012) Korean J. Poult. Sci., **39**, DOI: 10.5536/KJPS.2012.39.1.053.
- Zuo S., Zhang T., Jiang B., Mu W. (2015) Extremophiles, **19**, 841-851.
- Talluri S. (2011) Intern. J. Adv. Biotechnol. Res., **2**, 190-200.
- Алемасов Н.А., Фомин Э.С. (2012) Вавиловский журнал генетики и селекции, **16**, 774-783.
- Saiki R.K., Scharf S., Faloona F., Mullis K.B., Horn G.T., Erlich H.A., Arnheim N. (1985) Science, **230**, 1350-1354.
- Eijsink V.G., Bjørk A., Geseidnes S., Sireveg R., Synstad B., van den Burg B., Vriend G. (2004) J. Biotechnol., **113**, 105-120.
- Bommarius A.S., Broering J.M., Chaparro-Riggers J.F., Polizzi K.M. (2006) Curr. Opin. Biotechnol., **17**, 606-610.
- Vermeer A.W.P., Norde W. (2000) Biophys. J., **78**, 394-404.
- Eijsink V.G., Geseidnes S., Borchert T.V., van den Burg B. (2005) Biomol. Eng., **22**, 21-30.
- Серов А.Е., Одищева Е.Р., Упоров И.В., Тишков В.И. (2005) Биохимия, **70**, 974-979.
- Полянский А.А., Косинский Ю.А., Ефремов Р.Г. (2004) Биоорган. химия, **30**, 470-480.
- Puchkaev A.V., Koo L.S., de Montellano P.R.O. (2003) Arch. Biochem. Biophys., **409**, 52-58.
- Pikkemaat M.G., Linssen A.B., Berendsen H.J., Janssen D.B. (2002) Protein Eng., **15**, 185-192.
- Холмуродов Х.Т., Алтайский М.В., Пузынин И.В., Дардин Т., Филатов Ф.П. (2003) Физика элем. частиц и атомного ядра, **34**, 474-515.
- Vieille C., Zeikus G.J. (2001) Microbiol. Mol. Biol. Rev., **65**, 1-43.
- Razvi A., Scholtz J.M. (2006) Protein Sci., **15**, 1569-1578.
- Vogt G., Woell S., Argos P.J. (1997) Mol. Biol., **269**, 631-643.
- Kumar S., Tsai C., Nussinov R. (2000) Protein Engineering, **13**, 179-191.
- Szilagyi A., Zavodszky P. (2000) Structure, **8**, 493-504.
- Xiao L., Honig B.J. (1999) Mol. Biol., **289**, 1435-1444.
- Zhu S., Elcock A.H. (2010) Chem. Theory Comput., **6**, 1293-1306.
- Gromiha M. (2003) Polymer, **44**, 4061-4066.
- Dehouck Y., Grosfils A., Folch B., Gilis D., Bogaerts P., Rooman M. (2009) Bioinformatics, **25**, 2537-2543.
- Korkegian A., Black M.E., Baker D., Stoddard B.L. (2005) Science, **308**, 857-860.
- Luo P., Hayes R.J., Chan C., Stark D.M., Hwang M.Y., Jacinto J.M., Juvvadi P., Chung H.S., Kundu A., Ary M.L., Dahiyat B. (2002) Protein Sci., **11**, 1218-1226.
- Potapov V., Cohen M., Schreiber G. (2009) Protein Eng. Des. Sel., **22**, 553-560.
- Seeliger D., De Groot B. (2010) Biophys. J., **98**, 2309-2316.
- Connolly M.J. (1993) Mol. Graphics, **11**, 139-141.
- Vriend G.J. (1990) Mol. Graphics, **8**, 52-56.
- Kabsch W., Sander C. (1983) Biopolymers, **22**, 2577-2637.
- Meharenn Y.T., Poulos T.L. (2010) Biochemistry, **49**, 6680-6686.
- Ihara S., Itoh S., Kitakami J. (1998) Phys. Rev. B., **58**, 10736-10744.
- Narumi T., Susukita R., Ebisuzaki T., McNiven G., Elmegreen B. (1999) Molec. Simulation, **21**, 401-415.
- Qi L., Sinnott S.B. (1998) Nucl. Instr. Meth. Phys. Res. B., **140**, 39-46.
- Nozawa T., Takahashi K.Z., Narumi T., Yasuoka K. (2015) J. Comput. Chem., **36**, 2406-2411. doi: 10.1002/jcc.24222. Epub 2015.
- Fogolari F., Brigo A., Molinari H. (2002) J. Mol. Recognit., **15**, 377-392.
- Sakamaki R., Sum A.K., Narumi T., Yasuoka K. (2011) J. Chem. Phys., **134**, doi: 10.1063/1.3574038.
- Takahashi K.Z., Narumi T., Yasuoka K. (2011) J. Chem. Phys., **135**:174108. doi: 10.1063/1.3658640.
- Suenaga A., Kiyatkin A.B., Hatakeyama M., Futatsugi N., Okimoto N., Hirano Y., Narumi T., Kawai A., Susukita R., Koishi T., Furusawa H., Yasuoka K., Takada N., Ohno Y., Taiji M., Ebisuzaki T., Hoek J.B., Konagaya A., Kholodenko B.N. (2004) J. Biol. Chem., **279**, 4657-4662.
- Suenaga A., Hatakeyama M., Ichikawa M., Yu X., Futatsugi N., Narumi T., Fukui K., Terada T., Taiji M., Shirouzu M., Yokoyama S., Konagaya A. (2003) Biochemistry, **42**, 5195-5200.
- Okimoto N., Futatsugi N., Fuji H., Suenaga A., Morimoto G., Yanai R., Ohno Y., Narumi T., Taiji M. (2009) PLoSComput. Biol., **10**, doi: 10.1371/journal.pcbi.1000528. Epub 2009 Oct 9.
- Фомин Э.С., Алемасов Н.А. (2012) Вавиловский журнал генетики и селекции, **2**, 415-426.
- Kumar S., Nussinov R. (2001) Cell. Mol. Life Sci., **58**, 1216-1233.
- Hollien J., Marqusee S. (1999) Proc. Natl. Acad. Sci. USA, **96**, 13674-13678.
- Schuler B., Kremer W., Kalbitzer H.R., Jaenicke R. (2002) Biochemistry, **41**, 11670-11680.
- Bae E., Bannen R.M., Phillips G.N. Jr. (2008) Proc. Natl. Acad. Sci. USA, **105**, 9594-9597.
- Wang Q., Buckle A.M., Foster N.W., Johnson C.M., Fersht A.R. (1999) Protein Sci., **8**, 2186-2193.
- Chan C.H., Liang H.K., Hsiao N.W., Ko M.T., Lyu P.C., Hwang J.K. (2004) Proteins, **57**, 684-691.
- Sojin Moon Euiyoung Bae (2014) J. Korean Society Applied Biological Chemistry, **57**, 661-665.
- Bannen R.M., Suresh V., Phillips G.N. Jr., Wright S.J., Mitchell J.C. (2008) Bioinformatics, **24**, 2339-2343.
- Pucci F., Dhanani M., Dehouck Y., Rooman M. (2014) PLoS One, **9**, e91659.
- Ku T.H., Lu P.Y., Chan C.H., Wang T.S., Lai S.Z., Lyu P.C., Hsiao N.W. (2009) Comp. Biol. Chem., **33**, 445-450.
- Karantzeni I., Ruiz C., Liu C.C., Licata V.J. (2003) Biochem. J., **374**, 785-792.
- Ahrberg C.D., Manz A. (2016) RSC Adv., **6**, 42076-42080.
- Priyakumar U.D., Ramakrishna S., Nagarjuna K.R., Reddy S.K. (2010) J. Phys. Chem. B, **114**, 1707-1718.
- Pacheco S., Béhar G., Maillason M., Mouratou B., Pecorari F. (2014) Protein Eng. Des.Sel., **27**, 431-438.
- Otto A., Seckler R. (1991) Eur. J. Biochem., **202**, 67-73.
- Sevcenco A.M., Paravidino M., Vrouwenvelder J.S., Wolterbeek H.T., van Loosdrecht M.C., Hagen W.R. (2015) Water Res., **76**, 181-186.
- Tatur J., Hagedoorn P.L., Overijnder M.L., Hagen W.R. (2006) Extremophiles, **10**, 139-148.
- Sudmoon R., Sattayasai N., Bunyatratchata W., Chaveerach A., Nuchadomrong S. (2008) ActaBiochim. Biophys. Sin. (Shanghai), **40**, 811-818.

66. Wang P., Wang P., Tian J., Yu X., Chang M., Chu X., Wu N. (2016) *Sci. Rep.*, **6**, 22229. doi: 10.1038/srep22229.
67. Dong G., Vieille C., Savchenko A., Zeikus J.G. (1997) *Appl. Environ. Microbiol.*, **63**, 3569-3576.
68. Brown I., Dafforn T.R., Fryer P.J., Cox P.W. (2013) *Biochim. Biophys. Acta*, **1834**, 2600-2605.
69. Гришин Д.В. (2009) Антибиотики и химиотерапия, **54**, 3-7.
70. Гришин Д.В., Никитин А.В. (2009) Антибиотики и химиотерапия, **54**, 49-51.
71. Pantazaki A.A., Karagiorgas A.A., Liakopoulou-Kyriakides M., Kyriakidis D.A. (1998) *Appl. Biochem. Biotechnol.*, **75**, 249-259.
72. Salminen T., Teplyakov A., Kankare J., Cooperman B.S., Lahti R., Goldman A. (1996) *Protein Sci.*, **5**, 1014-1025.
73. Coolbear T., Daniel R.M., Morgan H.W. (1992) *Adv. Biochem. Eng. Biotechnol.*, **45**, 57-98.
74. Park J.H., Kim G.J., Lee S.G., Lee D.C., Kim H.S. (1999) *Appl. Biochem. Biotechnol.*, **81**, 53-65.
75. Lee S.G., Lee D.C., Kim H.S. (1997) *Appl. Biochem. Biotechnol.*, **62**, 251-266.
76. Mukohara Y., Ishikawa T., Watabe K., Nakamura H. (1994) *Biosci. Biotechnol. Biochem.*, **58**, 1621-1626.
77. Ow S.Y., Dunstan D.E. (2013) *Soft Matter*, **9**, 9692-9701.
78. Iizuka M., Inoue Y., Murata K., Kimura A. (1989) *J. Bacteriol.*, **171**, 6039-6042.
79. Choi Y.S., Yoo Y.J. (2013) *Biotechnol. Lett.*, **35**, 1455-1460.
80. Karantza V., Freire E., Moudrianakis E.N. (2001) *Biochemistry*, **40**, 13114-13123.
81. Masino L., Martin S.R., Bayley P.M. (2000) *Protein Sci.*, **9**, 1519-1529.
82. Staniforth R.A., Giannini S., Bigotti M.G., Cutruzzola F., Travaglini-Allocatelli C., Brunori M. (2000) *J. Mol. Biol.*, **297**, 1231-1244.
83. Hsia Y., Gnesa E., Jeffery F., Tang S., Vierra C. (2011) in: *Metal, Ceramic and Polymeric Composites for Various Uses / Spider Silk Composites and Applications* (Cuppoletti J., ed.) InTech, pp. 303-324.
84. Hedhammar M., Bramfeldt H., Baris T., Widhe M., Askarieh G., Nordling K., Aulock S., Johansson J. (2010) *Biomacromolecules*, **11**, 953-959.
85. Watt I.C., Morris R. (1969) *Kolloid-Z.u.Z. Polymere*, **229**:34. doi:10.1007/BF01542504.
86. Kakkar P., Madhan B., Shanmugam G. (2014) *Springerplus*, **3**, 2-9.
87. Базров Д. (2010) *Квант*, **4**, 28-30.
88. Santos-Pinto J.R.A., Arcuri H.A., Priewalder H., Salles H.C., Palma M.S., Lubec G. (2015) *J. Proteome Res.*, **14**, 3859-3870.

Поступила: 14. 11. 2016.  
Принята к печати: 15. 03. 2017.

#### PREDICTION OF PROTEIN THERMOSTABILITY FROM THEIR PRIMARY STRUCTURE: THE CURRENT STATE AND DEVELOPMENT FACTORS

*D.V. Grishin, M.V. Pokrovskaya, O.V. Podobed, Ju.A. Gladilina, V.S. Pokrovsky, S.S. Aleksandrova, N.N. Sokolov*

Institute of Biomedical Chemistry,  
10 Pogodinskaya str., Moscow, 119121 Russia; e-mail: molbiol\_ibm@inbox.ru

The construction of proteins and peptides with desired properties, including resistance to high temperatures, as well as optimization of their amino acid composition, is an important and complex task, which attracts much attention in various branches of the basic sciences, and also in biomedicine and biotechnology. This raises the question: what method is more relevant for the at the pilot stage of research in order to estimate the influence of the planned amino acid substitutions on the thermostability of the resultant protein construct? In this brief review we have classified existing basic practical and theoretical approaches used in studies and predicting the thermal stability of native and recombinant polypeptides. Particular attention has been paid to the predictive potential of statistical methods for studying the thermodynamic parameters of the primary protein structure and prospects of their use.

**Key words:** polypeptides, prediction of thermal stability of proteins, protein primary structure, molecular dynamics, computer algorithms, thermodynamic properties of proteins