МИНОБРНАУКИ РОССИИ

Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»

Серия «Физика и технология микро- и наносистем»

B. A. KAPACEB

ПРИНЦИПЫ ТОПОЛОГИЧЕСКОГО КОДИРОВАНИЯ ЦЕПНЫХ ПОЛИМЕРОВ И СТРУКТУРА БЕЛКОВ

Санкт-Петербург 2014

УДК 577.12 ББК Е 902 К21

Карасев В. А.

К21 Принципы топологического кодирования цепных полимеров и структура белков / СПбГЭТУ «ЛЭТИ». СПб., 2014. 240 с.

ISBN

Предложена теория топологического кодирования линейных цепных полимеров, включающая топологический код и систему физических операторов. Элементарной единицей кода является пентафрагмент полимера, конформация которого описывается с помощью четырехзвенного графа, а физическими операторами служат боковые цепи полимера, воссоздающие закодированную структуру пентафрагментов.

Введено представление о молекулярной векторной машине, согласно которому векторы, действующие на формирование водородной связи в пентафрагменте, образуют группу, а боковые цепи полимера выступают в качестве неприводимых представлений группы векторов, реализующих действие последних в процессе воссоздания топологии цепного полимера.

Применение данной теории к белкам и нуклеиновым кислотам позволяет объяснить природу генетического кода, структуры канонического набора аминокислот и проблему соответствия «триплет – аминокислота» в генетическом коде.

На основе введенных понятий созданы база данных пентафрагментов белков и библиотека десятизначных описаний вторичной структуры белков, а также предложены простые и эффективные методы прогнозирования и проектирования вторичной структуры белков.

Теория намечает пути использования цепных полимеров для создания принципиально новых устройств наноэлектроники.

Книга предназначена для специалистов, работающих в молекулярной биологии и наноэлектронике, а также для студентов и аспирантов, специализирующихся в этих областях науки и техники.

> УДК 577.12 ББК Е 902

Рецензенты: д-р биол. наук, проф., зав. лабораторией Института экспериментальной медицины РАМН В. Н. Кокряков; канд. биол. наук, доц., зав. кафедрой биохимии СПбГУ В. Е. Стефанов.

© СПбГЭТУ «ЛЭТИ», 2014

ISBN

ПРЕДИСЛОВИЕ	5
Часть 1. ТЕОРИЯ ТОПОЛОГИЧЕСКОГО КОДИРОВАНИЯ ЛИНЕЙНЫХ ЦЕПНЫХ ПОЛИМЕРОВ	8
Глава 1. ПРЕДПОСЫЛКИ РАЗРАБОТКИ ТЕОРИИ	8
 1.1. Постановка проблемы и выбор подходов к ее решению 1.2. Принципы, положенные в основу развиваемой теории 	8 12
Глава 2. ТОПОЛОГИЧЕСКИЙ КОД И ФИЗИЧЕСКИЕ ОПЕРАТОРЫ	19
 2.1. Введение понятий и выбор минимального объекта для анализа 2.2. Суперматрица конформаций четырехзвенного цепного графа 2.3. Трансформация суперматрицы в топологический код 2.4. Система физических операторов 	19 24 29 33
Глава 3. МОЛЕКУЛЯРНАЯ ВЕКТОРНАЯ МАШИНА	41
 3.1. Этапы построения молекулярной векторной машины 3.2. Свойства векторов додекаэдра 3.3. Канонический набор физических операторов	41 45 48 56
Часть 2. ПРИМЕНЕНИЕ ТЕОРИИ ТОПОЛОГИЧЕСКОГО КОДИРОВАНИЯ К БИОСТРУКТУРАМ	61
Глава 4. ГЕНЕТИЧЕСКИЙ КОД И ПРОБЛЕМА СООТВЕТСТВИЯ «ТРИПЛЕТ – АМИНОКИСЛОТА»	61
 4.1. Исследование пространственной структуры генетического кода 4.2. Топологические основания триплетного генетического кода	61 73 77
 4.5. Гриплеты генетического кода и ооковые цени аминокислог	87
к анализу эволюции биологических систем кодирования Глава 5. СТРУКТУРА КАНОНИЧЕСКОГО НАБОРА АМИНОКИСЛОТ	92 94
 5.1. Проблема классификации канонического набора аминокислот 5.2. Построение модели структуры канонического набора аминокислот 5.3. Обсуждение модели	94 95 103 106
Глава 6. МОЛЕКУЛЯРНАЯ ВЕКТОРНАЯ МАШИНА БЕЛКОВ	114
 6.1. Природа молекулярной векторной машины белков 6.2. Анализ элементов структуры молекулярной векторной машины белков 6.3. Перспективы развития модели молекулярной векторной машины белков 	114 116 125

оглавление

Часть 3. ПРАКТИЧЕСКИЕ ПРИМЕНЕНИЯ ТЕОРИИ	126
Глава 7. БАЗА ДАННЫХ ПЕНТАФРАГМЕНТОВ БЕЛКА	126
 7.1. Обоснование перехода к практическому использованию теории 7.2. Получение и сортировка пентафрагментов 	126 128
 7.3. Анализ особенностей пентафрагментов, содержащихся в базе данных 7.4. Перспективы развития базы данных пентафрагментов 	136 143
Глава 8. МЕТОД ПРОГНОЗИРОВАНИЯ ВТОРИЧНОЙ СТРУКТУРЫ БЕЛКА	149
8.1. Общие вопросы8.2. Суть метода8.3. Проблемы алгоритма прогнозирования и пути их решения	149 150 158
8.4. Сопоставление результатов прогнозирования с экспериментальными данными	159
Глава 9. КАТАЛОГ ДЕСЯТИЗНАЧНЫХ ОПИСАНИЙ ВТОРИЧНЫХ СТРУКТУР БЕЛКА	162
9.1. Постановка задачи	162
9.2. Участки β-структуры и α-спирали: примеры описания9.3. Канонические переходные участки	165
«β-структура – α-спираль» и «α-спираль – β-структура»	166
9.4. Спиральные вздутия на фоне β-структур и впадины на фоне α-спиралей	167
9.5. Изгибы β-структур и изломы α-спиралей	169
9.6. Структурные модификации канонических переходов9.7. Перспективы использования каталога	172 175
Глава 10. МЕТОД ПРОЕКТИРОВАНИЯ ПЕРВИЧНОЙ СТРУКТУРЫ БЕЛКА С ЗАДАННОЙ ВТОРИЧНОЙ СТРУКТУРОЙ	177
10.1. Современное состояние проблемы дизайна белков	177
10.2. Сущность метода проектирования первичной структуры белка 10.3. Пример проектирования первичной структуры белка	181
с заданной вторичной структурой	187
10.4. Перспективы развития предложенного подхода	196
ЗАКЛЮЧЕНИЕ	198
СПИСОК ЛИТЕРАТУРЫ	201
СОКРАЩЕНИЯ И ОБОЗНАЧЕНИЯ	214
ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ	216
ПРИЛОЖЕНИЕ. Таблицы каталога десятичных описаний вторичных структур белка (табличный материал к гл. 9)	219

ПРЕДИСЛОВИЕ

Развитие исследований в области миниатюризации электронных устройств привело к появлению нового научного направления – наноэлектроники. Развитие биологии и переход к исследованию биопроцессов на клеточном, субклеточном и молекулярном уровнях привели к появлению молекулярной биологии. Масштабы объектов молекулярной биологии и объектов наноэлектроники соизмеримы.

Сравнительное рассмотрение этих направлений выявляет схожесть проблем. В частности, в наноэлектронике актуальны проблемы конструирования топологии молекулярных электронных схем. В молекулярной биологии актуальны проблемы прогнозирования надмолекулярной (вторичной и третичной) структуры белков, а также проблемы конструирования белковых структур («дизайн белков»).

На протяжении ряда лет в Центре микротехнологии и диагностики (ЦМИД) Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» разрабатывается концепция систем сопряженных ионно-водородных связей (ССИВС¹), в рамках которой вводится представления о ССИВС как об основе биоструктур и каналах переноса энергии и зарядов в этих структурах.

Формализованный характер этих представлений и наличие общих принципов (континуальность ССИВС при построении надмолекулярных структур, перенос заряда по ССИВС через водородную связь) позволяет использовать их не только в биологии, но и в наноэлектронике (в частности, в органической наноэлектронике). На основе этих представлений был проведен анализ биомолекул как электронных модулей ССИВС. На этих же принципах возможна разработка систем электронных модулей и в наноэлектронике.

Однако создание электронных и молекулярных систем наталкивается на проблему управляемой сборки конструируемых схем в трехмерном пространстве. Одним из способов формирования молекулярных (в том числе электронных) схем является их самосборка, которая обеспечивается выстраиванием модулей в линейную цепь. Впервые эта идея применительно к молекулярной электронике была изложена в ра-

¹ Карасев В. А., Лучинин В. В. Ведение в конструирование бионических наносистем. М.: Физматлит, 2009.

боте², а применительно к биологии – в работе³. Этот подход в наших работах получил название топологического кодирования. Разработке и развитию данного подхода посвящена представляемая монография.

Книга состоит из трех частей, каждая из которых логически связана с предыдущими и в то же время может рассматриваться вполне самостоятельно.

В первой части изложены основные теоретические посылки к разработке нашего подхода (гл. 1) и общетеоретические проблемы топологического кодирования цепных полимеров. Для целей конструирования надмолекулярных структур на основе полимеров подходят не все органические полимеры, а лишь определенный их класс - «линейные цепные полимеры». Основной их особенностью является возможность формирования непрерывных ССИВС внутри основной цепи полимера. Элементарной единицей, на базе которой получены основные результаты теории, является фрагмент цепного полимера, состоящий из пяти мономерных единиц, – пентафрагмент (ПФ). В рамках теории топологического кодирования линейных цепных полимеров рассмотрены проблема построения топологического кода (гл. 2), проблема создания системы физических операторов и проблема их соответствия триплетам топологического кода (гл. 3). Разработанная в рамках данного подхода модель молекулярной векторной машины (гл. 4) обеспечивает возможность перехода от первичной структуры полимера к структурам более высокого порядка – вторичной и третичной.

Вторая часть монографии иллюстрирует возможность применения разработанных в первой части общих теоретических положений к реальным цепным полимерам, которыми являются биополимеры – белки и нуклеиновые кислоты. С позиций общей теории дается трактовка структуры генетического кода и проблемы соответствия триплетов определенным типам боковых цепей аминокислот (гл. 4). Предложена модель структуры канонического набора аминокислот на додекаэдре, основанная на принципах антисимметрии (гл. 5), и дана трактовка найденных закономерностей с позиций молекулярной векторной

² Дорфман В. Ф. О топологической разрешимости интегральных структур без совмещений // Микроэлектроника. 1975. Т. 4, вып. 3. С. 213–219.

³ Карасев В. А. Как закодировать топологию биочипа? // Биотехнология. 1998. № 3. С. 62–75.

машины белков (MBM_{δ}) (гл. 6). Как и в общей теории, основным элементом MBM_{δ} является ПФ.

Третья часть монографии посвящена возможности практического использования разработанного подхода. С целью создания базы данных ПФ белков из 2500 структур белков, взятых из Protein Data Bank, были получены и систематизированы более 600 000 ПФ (гл. 7). В монографии подробно анализируется структура созданной базы $\Pi \Phi$. В процессе их систематизации был использован способ описания конформации ПФ (Н-связей) с помощью десятизначных булевых чисел. База данных послужила основой для разработки метода прогнозирования вторичной структуры белка на основе первичной структуры (гл. 8). Для целей интерпретации прогнозируемой структуры был создан каталог десятизначных описаний вторичных структур белка (гл. 9). Предложенный далее метод проектирования первичной структуры белка по заданному описанию вторичной структуры (гл. 10) также использует каталог десятизначных описаний вторичных структур белка. Использование в этих способах десятизначных булевых чисел для описания вторичных структур на практике реализует биоинформационный подход в этой области.

В процессе написания монографии были использованы результаты работ, проведенных в ЦМИД СПбГЭТУ «ЛЭТИ».

Автор благодарен директору ЦМИД д-ру техн. наук, проф. В. В. Лучинину за постоянную поддержку развиваемого подхода и молекулярного конструирования биоорганических наносистем, а также сотрудникам ЦМИД СПбГЭТУ «ЛЭТИ», обеспечившим благоприятную творческую атмосферу для выполнения этой работы. Особая благодарность – программистам Е. Л. Демченко и А. И. Беляеву. Благодарю студентов Вязьмитинова Д. В., Кротова С. В., Лебедева А. В. и Некрасова И. В. за участие в создании базы данных ПФ белков, а также Перчаткину Д. С. и Павленко Т. В. за помощь в разработке каталога описаний вторичных структур белка.

Часть 1. ТЕОРИЯ ТОПОЛОГИЧЕСКОГО КОДИРОВАНИЯ ЛИНЕЙНЫХ ЦЕПНЫХ ПОЛИМЕРОВ

Глава 1. ПРЕДПОСЫЛКИ РАЗРАБОТКИ ТЕОРИИ

1.1. Постановка проблемы и выбор подходов к ее решению

Общность проблем самоорганизации в биологии и наноэлектронике

Проблемы проектирования надмолекулярных структур существуют как в молекулярной биологии, так и в наноэлектронике.

В первом случае помимо необходимости создания уже известных структур есть необходимость создания принципиально новых аминокислотных последовательностей, принимающих задаваемую исследователем структуру.

Во втором случае, когда наноустройства создаются на основе небелковых материалов, сущность проблемы не меняется: нужно уметь так запрограммировать последовательность элементов устройства, чтобы в результате их синтеза под действием определенных механизмов возникала работоспособная структура.

В [1] уже обсуждалось сходство данных проблем. Отмечалось, что постановка проблемы формирования технических наноструктур, независимого от внешнего вмешательства, приводит к технологической схеме, близкой по сути к схеме синтеза белка [2]. И хотя предложенные в [2] механизмы не были практически воплощены в электронике, в молекулярной биологии они оказались полностью реализованными. Общность обсуждаемых проблем самоорганизации структур в биологии и наноэлектронике заставляет искать и общие пути их решения. Именно по этой причине в данной монографии вначале предлагается теоретический подход, сформулированный в обобщенных понятиях, не зависящих от конкретных материалов, а затем этот подход иллюстрируется на примере биоструктур. Поскольку в биологии проблемы самоорганизации структур возникли существенно раньше, чем в наноэлектронике, то и попытки ответить на возникающие вопросы также появились раньше.

Анализ подходов, применяемых в биологии, к решению проблем самоорганизации

В молекулярной биологии можно выделить три подхода к решению этих проблем.

Первый подход базируется на широкоизвестных экспериментах Анфинсена по восстановлению активной структуры фермента рибонуклеазы в растворе после полного разворачивания его третичной структуры [3]. Эти эксперименты показали, что самоорганизация белка может происходить без вмешательства внешних сил и что вся информация для осуществления этого процесса заключена в первичной структуре белка. Именно после этого возникло направление, существующее и поныне и предполагающее возможность предсказания структуры белка на основе его первичной структуры [4]-[6]. Недостатком этого подхода является то, что в биоструктурах нет условий, которые нужны для такого способа самоорганизации. Основным требованием к механизму формирования надмолекулярной структуры белка является его точность и однозначность, поскольку процессы синтеза белковых структур идут в живых клетках постоянно. При этом концентрации белков в реальных живых клетках на много порядков превышают те концентрации, при которых происходит самоорганизация в экспериментах. Живые существа исчезли бы «еще до появления на свет», если бы они использовали столь неоднозначные механизмы самоорганизации. Аналогичные требования точности и однозначности предъявляются к техническим наноструктурам в процессе их массового тиражирования.

Второй подход возник после полной расшифровки генетического кода, когда в таблице кода нашли отражение соответствия аминокислот триплетам азотистых оснований, кодирующих эти аминокислоты. Этот весьма спорный подход был предложен Л. Б. Меклером [7]. Его идея оказалась подкупающе проста. В структуре нуклеиновых кислот господствует принцип комплементарности азотистых оснований: аденин (A) комплементарно спаривается с тимином (T), а гуанин (G) – с цитозином (C). В последовательности триплетов азотистых оснований записана информация о первичной структуре белка. А что, если аминокислоты, кодируемые комплементарными триплетами генетического кода, являются комплементарными в третичной структуре белка? Были предложены

9

группы взаимно комплементарных аминокислот, основанные на комплементарности триплетов, и предприняты попытки их использования для прогнозирования белковых структур по их первичной структуре [7]-[9]. Однако факт комплементарности аминокислот, связанных комплементарностью кодирующих их триплетов, до сих пор не доказан. Для его доказательства нужно выявить природу соответствий «триплет – аминокислота», из которой (если бы это было так) и была бы очевидна данная комплементарность. На сегодня природа этих соответствий не выявлена ни теоретически, ни экспериментально, так что данный подход попрежнему остается весьма спорным и уязвимым для критики. Несмотря на это, некоторые идеи, высказанные Л. Б. Меклером в процессе работы с белковыми структурами, оказались продуктивными, в частности, идея о том, что прогнозирование белка нужно вести последовательно – от Nконца к С-концу, т. е. так же, как происходит биосинтез белка. Боковые цепи аминокислот, способные образовывать водородные связи, должны формировать такие связи при создании надмолекулярной структуры [7]. Эти идеи воплотились в третьем подходе.

Развитию *третьего подхода* (именно он излагается в настоящем издании) способствовало появление данных о структуре рибосом, осуществляющих синтез белка. В [1] этому вопросу уделено большое внимание. Суть полученных результатов состоит в том, что в структуре рибосом обнаружен протяженный канал (тоннель) [10], [11]. Выход из него фланкирован двумя специальными рибосомными белками и открывается только на момент выхода очередного звена полипептидной цепи. Таким образом, тоннель вместе с частью синтезированного белка изолирован от внешней среды [12], [13]. Протяженность тоннеля составляет ~ 100 Å, а диаметр – 10...20 Å. В таком тоннеле может умещаться полипептидная цепь – как в растянутом виде, так и в виде α -спирали. Это позволяет предполагать, что по крайней мере частичное формирование вторичной структуры белка может происходить уже в процессе его синтеза, т. е. ко-трансляционно. Именно об использовании такого механизма и говорил Л. Б. Меклер, основываясь на интуитивных соображениях.

Условия практической реализации ко-трансляционного механизма сворачивания белка в живой природе существенно отличаются от условий, в которых моделируют процесс фолдинга in vitro. Весь процесс сворачивания белка, по-видимому, находится под контролем клеточных структур, что гарантирует правильную укладку.

Развивая идею ко-трансляционного механизма сворачивания белка, можно предположить существование внутри рибосомы специального «устройства», которое укладывает вновь синтезированные фрагменты белка в правильную структуру. Логика теории, изложенной в данной монографии, также приводит к идее такого «устройства».

Круг проблем, решаемых в рамках теории топологического кодирования

Эффективность того или иного подхода может быть оценена на основе связывания воедино фактов, казавшихся необъяснимыми с позиций существующих представлений. В частности, ни один из изложенных выше подходов не в состоянии ответить на следующие вопросы:

• Почему живые существа используют оптические изомеры биомолекул только одного типа (например, L-аминокислоты)?

• Почему генетический код содержит именно 64 кодирующих триплета, а не иное их число?

• Какова природа вырожденности генетического кода (одну и ту же аминокислоту кодирует несколько разных триплетов)?

• Почему в состав канонического набора входит лишь 20 аминокислот и почему представлены именно эти аминокислоты, а не другие?

• Какова природа соответствия триплетов определенным боковым цепям аминокислот?

• Какова природа минорных оснований в транспортной РНК (т-РНК)?

Последний вопрос кажется не связанным с предыдущими. А между тем, в рамках изложенного подхода функции минорных оснований также могут быть успешно объяснены.

На перечисленные вопросы нельзя ответить с позиций изложенных выше подходов, поскольку эти подходы пытаются использовать лишь видимые свойства вне их связи с принципами организации и работы клеточных систем в целом. Эти недостатки, как будет показано далее, были учтены при разработке теории топологического кодирования линейных цепных полимеров.

1.2. Принципы, положенные в основу развиваемой теории *Концепция ССИВС*

Всякая теория имеет ряд допущений, положенных в основу ее построения. Суть наших допущений вытекает из описанной в [1] концепции ССИВС. Ниже изложена краткая версия этой концепции.

Данная концепция, помимо [1], рассматривалась и в других публикациях (см., например, [14]–[22]). Она основана на допущении, что биомолекулы (аминокислоты, азотистые основания, липиды и т. п.) обладают неким системообразующим свойством, обеспечивающим их участие в формировании биоструктур, а также в переносе энергии в этих структурах.

Для выявления этого системообразующего свойства в составе биомолекул были выделены два минимальных типа сочетаний атомов – простые группы и резонансные группы.

Простой группой (R–Z) называется сочетание из двух атомов элементов-органогенов (C, N, O, P, S), содержащее одинарную σ-связь.

Резонансной группой (Q–R=X) называется сочетание из трех атомов элементов-органогенов, содержащее две σ -связи и одну π -связь, способную к перемещению и резонансу: Q–R=X $\leftarrow \rightarrow$ Q=R–X. В этом понятии в обобщенном виде используются представления о резонансе двойных связей, развитые Л. Полингом [23]. В [14]–[22] опубликована таблица, систематизирующая эти группы, расположенные в порядке возрастания молекулярной массы (табл. 1.1).

Таблица 1.1

Простые группы	Резонансные группы				
C–C	C–C=C				
C–N	C-C=N	N-C=N			
C–O	C–C=O	N-C=O	O-C=O		
C–S	C–C=S	N-C=S	O–C=S		
_	С–Р=О	N-P=O	O-P=O		
S–S	C–S=O	N-S=O	O-S=O		

Типы групп, встречающихся в биомолекулах и биоструктурах

Из табл. 1.1 можно сделать следующие выводы:

- в группах, составляющих биомолекулы, используются сочетания из пяти элементов-органогенов (C, N, O, P, S); – количество сочетаний ограниченно: возможно лишь 5 простых групп и 15 резонансных групп.

В молекулах, входящих в состав работающих биоструктур, другие группы практически не найдены.

Взаимодействие между простыми и резонансными группами через атом водорода может приводить к образованию следующих систем:

$$\begin{array}{c} R & R \\ HQ_1 - R = X_1 \dots HZ_1 \dots HQ_2 - R = X_2 \dots \dots HQ_n - R = X_n \dots HZ_m \end{array}$$
(1.1)

Предельные варианты таких систем состоят из регулярно чередующихся резонансных и простых групп или только из резонансных групп. Наилучшим образом указанные системы будут образовываться из групп, обладающих противоположными свойствами, т. е. протонодоноров и протоноакцепторов. Термин «Системы сопряженных ионноводородных связей» отражает их сходство с полисопряженными системами, донорно-акцепторный характер упомянутых взаимодействий (ионно-водородные), а также наличие в них водородных связей.

Поскольку способность простых и резонансных групп к образованию ССИВС является одним из их наиболее общих свойств, то, значит, и биомолекулы, содержащие данные сочетания, также обладают свойством образовывать между собой непрерывные ССИВС. Таким образом, был сделан вывод о том, что наиболее существенным системообразующим свойством биомолекул является их способность к образованию ССИВС, которые могут служить основой для построения надмолекулярных наноструктур и каналами для передачи энергии (заряда) в этих структурах.

Подставляя в ССИВС группы с противоположными свойствами (табл. 1.1) и далее – конкретные молекулы, можно строить разнообразные структуры на основе непрерывных ССИВС. При этом комплексы будут разными, а ССИВС могут оставаться одними и теми же. В этом состоит их универсальный характер, позволяющий конструировать на основе ССИВС разнообразные структуры.

Нами был сформулирован *первый принцип концепции ССИВС*: надмолекулярные биоструктуры должны строиться на основе непрерывных ССИВС [14]–[22]. Его можно назвать принципом континуальности (непрерывности) ССИВС при построении биоструктур.

Второй принцип концепции ССИВС, основанный на модели переноса зарядов по ССИВС и анализе особенностей работы биоструктур, формулируется так: перенос зарядов в биоструктурах должен происходить по ССИВС при участии водородных связей. Его можно назвать принципом сопряжения через водородную связь для переноса зарядов [14]–[22].

Роль принципа континуальности ССИВС в решении проблем самоорганизации цепных полимеров

Для последующего изложения наиболее важен принцип континуальности ССИВС, который может служить основой процесса самоорганизации биоструктур, в частности, белков. Это можно проиллюстрировать на примере формирования спиральных структур.

Предположим, что некий цепной полимер, содержащий в звеньях резонансные группы HQ–R=X, атомы которых обладают упомянутыми свойствами, сформировал фрагмент спиральной структуры (рисунок).



Анализ формирования фрагмента спиральной структуры на основе принципа непрерывности ССИВС

В условиях, когда цепь полимера содержит бесконечное количество звеньев, содержащих HQ–R=Х-группы, такая цепь неизбежно сформирует спираль (на рисунке она направлена справа налево), вдоль которой будут проходить непрерывные ССИВС, состоящие из HQ–R=Хгрупп. В рамках изложенного подхода способность к формированию непрерывных ССИВС интерпретируется как свойство связности. Это свойство, проявляющееся при наличии групп, обладающих противоположными свойствами, как способность к самоорганизации, изначально заложено в таких цепных полимерах [1]. Важно только, чтобы эти группы имели ограниченное число степеней свободы и могли поддаваться регулированию, как это имеет место для HN–C=O-групп белков [1].

Если такой полимер будет иметь боковые цепи, размер которых позволит им влиять на стабильность спиралей (как в сторону усиления, так и в сторону разрушения), то в структуре такой спирали появятся дефекты, которые смогут менять направление ее фрагментов и способствовать определенной укладке этих фрагментов в третичную структуру. Таким образом, наличие боковых цепей, обладающих различной степенью связности, в сочетании со свойством связности основной цепи само по себе может обеспечить самоорганизацию исходной молекулы полимера. Цепь такого полимера обладает «молекулярной памятью» и при сохранении расположения боковых цепей потенциально может воссоздавать одну и ту же надмолекулярную структуру. Необходимым условием этого является постепенное формирование структуры, что обеспечивается последовательным наращиванием длины исходной цепи.

Анализ боковых цепей аминокислот как функциональных модулей ССИВС

Из предложенной нами модели переноса зарядов возникает три следствия:

 водородной связи можно придать функцию «контакта» между звеньями цепи коммутации или модулями;

– за вход в звено или модуль можно принять атом водорода, а за выход – неподеленную пару электронов, т. е. \rightarrow HQ–R=X: \rightarrow ;

- процесс переноса заряда можно обозначать термином «сигнал».

Эти следствия были использованы нами в [1] и [15] при анализе биомолекул в качестве функциональных модулей. В процессе дальнейшего изложения материала такой анализ будет необходим в применении к боковым цепям аминокислот.

В составе аминокислот нами были выделены пассивные и активные элементы [1]. Пассивные элементы – это боковые цепи аминокислот, не способные к образованию водородных связей (аланин, валин, лейцин, изолейцин и фенилаланин). Их роль состоит в создании изолирующей неполярной среды для ССИВС. Активные элементы – это аминокислоты, полярные группы которых могут встраиваться в ССИВС. С позиций электроники боковые цепи, создающие среду для ССИВС, можно рассматривать как множество маскирующих (пассивных) элементов, а цепи, участвующие в формировании ССИВС, как «окна» (активные элементы) [2].

В зависимости от молекулярной функции в составе ССИВС активные элементы были разделены на четыре группы [1] (табл. 1.2):

1. Инициаторы ССИВС: пролин (Pro), метионин (Met).

2. Молекулярные клапаны: серин (Ser), треонин (Thr), аспарагиновая кислота (Asp), глютаминовая кислота (Glu), аспарагин (Asn), глютамин (Glu), лизин (Lys), аргинин (Arg).

3. Элементы задержки сигнала: гистидин (His), тирозин (Tyr).

4. Элементы инверсии сигнала: триптофан (Тгр), тирозин (Туг).

Рассмотрим эти аминокислоты более подробно, используя термины «сигнал», «вход» и «выход».

1. Инициаторы ССИВС. К ним отнесены Рго и Меt. Обычно именно с Рго начинается формирование α-спиральных фрагментов в белках, поскольку атом азота его N–C=O-группы встроен в пятичленный цикл. Атом серы в Меt с двумя неподеленными парами электронов также может служить местом прикрепления ССИВС. Не случайно, именно с Меt начинается синтез всех белков.

2. Молекулярные клапаны. Аминокислоты этой группы подразделяются на две подгруппы:

а) элементы, имеющие 1 вход и 2 или более выхода (Ser, Thr с простой группой С–ОН и Asp, Glu с резонансной НО–С=О-группой). Они могут использовать не все возможные выходы (вторая строка в этой подгруппе), а также образовывать вилочковые водородные связи (третья строка), т. е. действовать как аналог элемента ИЛИ [1], [22];

б) элементы, имеющие 2 входа и 2 или более выхода (Asn, Gln c HN–C=O-группой, Lys – с простой C–NH₂-группой и Arg – с резонансными HN–C=N-группами). Особенностью этих элементов является возможность их модификации, например путем метилирования (третья строка в этой подгруппе). Количество возможных входов при этом уменьшается.

3. Элементы задержки сигнала. К ним отнесены боковые цепи аминокислот, имеющих 1 вход и 1 выход (His и Tyr). Необходимость в

Таблица 1.2 Полярные аминокислоты как функциональные элементы ССИВС

Элементы инициации (0 входов – 1 выход)				
O = C	$ X_1 = R - Q_1 H $			
$-X_1 = R - Q_1 - O = C$ Pro				
Молекулярн	ные клапаны			
1 вход —	2 выхода			
Ser, Thr	Asp, Glu			
1 HQ ₁ -R=X ₁ \rightarrow HO; HQ ₂ -R=X ₂ 2 HQ ₃ -R=X ₃ 3	1 HQ ₁ -R=X ₁ \rightarrow HO-C=O: $\begin{array}{c} \checkmark HQ_2-R=X_2 & 2\\ \downarrow & HQ_3-R=X_3 & 3 \end{array}$			
$1 \text{ HQ}_1 - R = X_1 \rightarrow \text{HO}_2 + R = X_2 + 2$	1 HQ ₁ -R=X ₁ \rightarrow HO-C=O: HQ_2 -R=X ₂ 2			
$\begin{array}{c} 1 HQ_1 - R = X_1 \\ 2 HQ_2 - R = X_2 \end{array} HO: \longrightarrow HQ_3 - R = X_3 3 \\ \end{array}$	$\begin{array}{c} 1 HQ_1 - R = X_1 \\ 2 HQ_2 - R = X_2 \end{array} HO - C = \ddot{O}: \longrightarrow HQ_3 - R = X_3 3 \\ \downarrow \end{array}$			
2 и более входа –	1 и более выходов			
Asn, Gln	Lys, Arg			
$1 HQ_1-R=X_1 HN-C=O: HQ_3-R=X_3 3$ 2 HQ_2-R=X_2 H HN-C=O: HQ_4-R=X_4 4	$1 \rightarrow HQ_1 - R = X_1 \rightarrow HN : \rightarrow HQ_3 - R = X_3 \rightarrow HQ_2 - R = X_2 \rightarrow HQ_2 - R = X_2 \rightarrow HQ_3 - R = X_3 \rightarrow HQ_3 \rightarrow HQ_3 - R = X_3 \rightarrow HQ_3 \rightarrow HQ$			
$1 HQ_1 - R = X_1 HN - C = \ddot{O} : \rightarrow HQ_3 - R = X_3 3$	$1 \rightarrow HQ_1 - R = X_1 \xrightarrow{H} H_N \xrightarrow{H} X_2 = R - Q_2 H \rightarrow 2$			
	• • • HQ ₃ -R=X ₃ -• 3			
Asn, Gln	Lys			
$1 HQ_1 - R = X_1 H_3C$	$1 HQ_1 - R = X_1 H_3C$ $N' \rightarrow HQ_2 - R = X_2 \rightarrow$			
2 $HQ_2-R=X_2$ H^{1} $HQ_4-R=X_4$ 4	$2 \rightarrow HQ_2 - R = X_2 \xrightarrow{H_1} H$			
Элементы задержки си	гнала (1 вход – 1 выход)			
His	Tyr			
Н	$I \rightarrow HQ_1 - R = X_1 \rightarrow HQ_2 - R = X_2 \rightarrow Z$			
$\rightarrow HQ_1 - R = X_1 \rightarrow HN \qquad $	HC			
٠́	ĺ ↓ ↓			
Элементы инверсии сигнала (1 вход-выход)				
Tyr	Тгр			
$1 \rightarrow HQ_1 - R = X_1 \rightarrow H : O - C_{N}$	$1 \rightarrow HQ_1 - R = X_1 \rightarrow HN$			

элементах задержки сигнала появляется в тех случаях, когда должны быть согласованы во времени два и более параллельных процесса. Наиболее подробно в качестве этого элемента рассмотрена аминокислота His, которая включается в состав ССИВС последовательно. Аминокислота Туг также может рассматриваться в качестве элемента задержки сигнала. Однако, в отличие от His, Туг включен в цепь не всем циклом, а лишь своей C=C-OH-группой, т. е. не последовательно, а параллельно цепи. Атом водорода, связанный с этой группой, образует водородную связь с группой HQ₁–R=X₁ системы 1, а неподеленная пара электронов, принадлежащая кислороду, – с группой HQ₂–R=X₂ системы 2.

4. Элементы инверсии сигнала. К ним были отнесены боковые цепи Туг и Тгр, которые могут иметь 1 вход-выход (табл. 1.2). Необходимость появления элементов инверсии сигнала диктуется необходимостью согласования во времени нескольких параллельных процессов. На основании проведенного в [1] анализа боковые цепи аминокислот Туг и Тгр можно рассматривать как элементы НЕ.

В сочетании с рассмотренными в [1] и [22] узлами ИЛИ, основанными на вилочковой водородной связи, на основе этих двух типов аминокислот был предложен простейший мультивибратор.

Глава 2. ТОПОЛОГИЧЕСКИЙ КОД И ФИЗИЧЕСКИЕ ОПЕРАТОРЫ

2.1. Введение понятий и выбор минимального объекта для анализа

Определение термина «линейный цепной полимер» и выделение круга объектов, охватываемых теорией

Излагаемая теория топологического кодирования цепных полимеров описывает линейные цепные полимеры [24]–[30]. Из биоструктур к таким полимерам относятся белки и нуклеиновые кислоты, которые, как и всякие полимеры, состоят из отдельных звеньев. Полимеры, состоящие из звеньев, связанных в последовательную цепь, называются цепными. Структура звена имеет принципиальное значение для построения нашей теории. Ориентируясь на биополимеры, рассмотрим, как должно выглядеть звено линейного цепного полимера.



Рис. 2.1. К определению термина «линейный цепной полимер»: 1 – альфа-атомы; 2 – соединительная группа; 3 – боковая цепь; 4 – отдельное звено

Как следует из рис. 2.1, это звено (4) должно состоять из двух альфа-атомов i и i-1 (1) и соединительной группы (2). К альфаатомам прикрепляются различные боковые цепи (3). Вместо букв Q, R, X и Q₁, R, X₁ можно подставить конкретные атомы элементов-органогенов: C, N, O, P, S. Например, если в соединительную группу подставить Q \rightarrow N, R \rightarrow C, X \rightarrow O, то она превратится в HN–C=O-группу (пептидную связь белков), а если Q \rightarrow O, R \rightarrow P, X \rightarrow O, то – в HO–P=O-группу нуклеиновых кислот.

Для построения теории подходят не любые соединительные группы, а только такие, которые содержат резонансные группы HQ–R=X, способные образовывать между собой непрерывные ССИВС [1], [28]. В таблице приведены примеры подобных групп и содержащих их полимеров. Видно, что количество подходящих групп и полимеров весьма ограничено.

Примеры соединительных групп, способных образовывать ССИВС и содержащих эти группы цепных полимеров

Соединительные группы	Полимеры
O=C-NH O=C-NH	Пептиды и белки
S=C-NH S=C-NH	Тиополипептиды
N=C-NH N=C-NH	Полиимидазолы
O=P-OH O=P-OH	Нуклеиновые кислоты
O=P-NH O=P-NH	Аминофосфатный аналог нуклеиновых кислот
O=S-NH O=S-NH	Аминосульфатный аналог нуклеиновых кислот
O=Si-NHO=Si-NH	Аминосиликатный аналог нуклеиновых кислот

В наше определение также должно быть включено слово «линейные». Оно ограничит круг цепных полимеров только такими, которые связаны в линию, т. е. являются неразветвленными. Таким образом, наш подход основан на теоретическом рассмотрении группы неразветвленных цепных полимеров, состоящих из звеньев, содержащих резонансные группы, способные образовывать непрерывные ССИВС.

Выбор минимального объекта для теоретического анализа

В качестве такого объекта был выбран четырехзвенный фрагмент цепного полимера (рис. 2.2, *a*) [24]–[30] (ПФ), содержащий пять альфаатомов. Выбор для анализа именно такого объекта обусловлен тем, что для ряда полимеров (например, белков) в этом фрагменте образуется водородная связь между двумя соединительными группами $X_{i-1}=R-Q_iH...X_{i-4}=R-Q_{i-3}H$, что приводит к образованию цикла с фиксированными *i*-м и (*i* – 4)-м атомами (показано штриховой линией).

Аналогом цепных полимеров может служить *n*-звенный цепной граф [24]–[30]. Аналогом ПФ полимера является четырехзвенный цеп-

ной граф (рис. 2.2, δ). В этом графе вершины (i, i - 1, ..., i - 4) соответствуют атомам повторяющихся звеньев полимера, а структурные ребра – связям, соединяющим атомы звеньев в цепь. Для описания конформаций полимера и графа нами были введены дополнительно ребра связности (на рис. 2.2, δ – штриховая линия), которые соответствуют фиксированным атомам в полимере. Фиксация атомов может обеспечиваться как за счет водородных связей внутри основной цепи, так и за счет водородных связей боковых цепей с основной цепью.

Ребра связности соединяют несмежные вершины. Длина структурного ребра соответствует константе k_s и для цепного полимера является постоянной величиной, а ребра связности соответствуют константе k_c , которая может изменяться в пределах $0...2k_s$. Может существовать и несколько констант k_c . Конформации полимера и цепного графа можно описать с помощью верхних треугольных матриц. Для такого описания существенны только ребра связности. Если наличие ребер связности обозначить как 1, а их отсутствие – как 0, то матрицу, описывающую конформацию ПФ и цепного графа (рис. 2.2, *a* и *б*), можно представить в виде рис. 2.2, *в*.



Рис. 2.2. Пентафрагмент цепного полимера: *а* – структурная формула; *б* – граф; *в* – матричное описание

Пентафрагменты, в которых имеется H-связь между атомами группы $Q_iH...X_{i-4}=R$, будем называть циклическими. Соответственно, циклическими будут и их математические аналоги – графы. В матрицах, описывающих эти конформации, как мы видели, переменная $x_3 = 1$. В то же время, пентафрагменты, в которых нет H-связи между

атомами группы $Q_iH...X_{i-4}=R$, а также их математические аналоги – графы – будут называться ациклическими. В матрицах, описывающих ациклические конформации, переменная $x_3 = 0$. Общий вид матриц, пригодных для описания конформаций ПФ полимера и графа, можно представить следующим образом:

где *i*, *i* – 1, ..., *i* – 4 – вершины графа; $x_1, x_2, ..., x_6$ – переменные, принимающие значения 0 или 1. В дальнейшем будет использоваться также строчная запись : $x_1x_2x_3x_4x_5x_6$.

Примеры конформаций цепных полимеров, их графов и матричных описаний

Наиболее важные примеры рассмотрены в [1], [24]–[26] и показаны на рис. 2.3, где приведены только минимальные ПФ, описываемые матрицами из 6 переменных.

Как видно из рис. 2.3, *а*, полностью несвязная ациклическая конформация описывается матрицей из 6 переменных, которые принимают значения, равные нулю. Слоистая ациклическая конформация (рис. 2.3, δ) описывается матрицей, содержащей три единицы и три нуля ($x_3 = 0$). Спиральная конформация первого типа (рис. 2.3, ϵ) – аналог спирали 3₁₀ в белках – содержит в матричном описании только один ноль (связь *i* – (*i* – 4)). Согласно принятому определению, эта конформация отнесена нами к ациклическим ($x_3 = 0$). Наконец, матрица спиральной конформации второго типа (рис. 2.3, ϵ) – аналог α -спирали в белках – состоит только из единиц ($x_3 = 1$, циклическая конформация). Дальнейший анализ конформаций удобнее проводить на четырехзвенных цепных графах, которые были представлены в виде суперматрицы из 64 элементов [1], [24]–[30].



Рис. 2.3. Типичные конформации цепных полимеров, их графы и матричные описания: *a* – несвязная; *б* – слоистая; *в*, *г* – спиральные

2.2. Суперматрица конформаций четырехзвенного цепного графа

Принцип построения суперматрицы

Нами были рассмотрены все возможные конформации (состояния связности) четырехзвенного графа – от полностью развернутой, матрица которой состоит из нулей, до полностью связной, матрица которой состоит из шести единиц [24]–[30]. Общее их количество составляет 64. В упорядоченном виде они представляют собой блочную суперматрицу конформаций четырехзвенного графа (рис. 2.4).

Элементами суперматрицы являются матрицы с тем же порядком описания связей, что и на рис. 2.2 и 2.3. Как видно из рис. 2.4, в суперматрице имеется 4 блока с общими переменными x_3x_4 (показаны крупными цифрами). Ряды в блоках образованы матрицами с одинаковыми первыми парами переменных (x_1x_2) в последовательности 00, 10, 01, 11, а столбцы – матрицами с едиными третьими парами (x_5x_6) в порядке 00, 01, 10, 11. Полностью несвязный граф, описываемый матрицей с шестью переменными «0», расположился в левом верхнем углу суперматрицы, а граф с полностью связными вершинами, описываемый матрицей с шестью переменными «1», – в правом нижнем углу. В блоках 00 и 01 находятся ациклические конформации графа ($x_3 = 0$), а в блоках 10 и 11 – циклические ($x_3 = 1$).

Симметрия элементов суперматрицы

В пределах блоков наблюдается два типа симметрии [26], [27]. В соответствии с принципом построения таблицы графы в блоках и описывающие их матрицы расположены симметрично относительно главных диагоналей. Например, в первом блоке граф, описываемый переменными 100000, симметричен графу с описанием 000001. Второй тип симметрии, наблюдаемый в блоках, – это внутренняя симметрия графов и матриц, расположенных на главных диагоналях. Например, в первом блоке на главной диагонали располагаются графы, описываемые переменными 000000, 100001, 010010 и 110011.

Кроме симметрии внутри блоков можно выделить еще один тип симметрии, имеющий отношение ко всей суперматрице. Он содержит два типа графов и описывающих их матриц, обладающих следующим свойством: 0-элементы в матрицах одной группы соответствуют 1-элемен-



Рис. 2.4. Суперматрица конформаций четырехзвенного графа и их матричные описания

там в матрицах второй группы и наоборот, например: 000000 $\leftarrow \rightarrow$ 111111, 100000 $\leftarrow \rightarrow$ 011111, 010000 $\leftarrow \rightarrow$ 101111 и т. д. Эти две группы графов и матриц разделены жирной линией. В таблице эти графы и матрицы занимают положение, описываемое группой симметрии C₂. Этот тип симметрии можно назвать антисимметрией, а само преобразование 0 $\leftarrow \rightarrow$ 1 можно назвать преобразованием антисимметрии.

Степень свободы графов, расположенных в блоках

Детальный анализ степени свободы графов в отдельных блоках был проведен авторами [26]. Следует помнить, что точка и вершина на конце стержня имеют степень свободы, равную трем. Степень свободы внутри звена равна единице или двум, а внутри жестких фигур (например, тетраэдра) – нулю. Рассмотрим более конкретно каждый из блоков.

В блоке 00 присутствуют слабосвязные конформации, обладающие наибольшей подвижностью (преобладают степени свободы 2 и 3). Так, первый граф в этом блоке, обладает максимальной подвижностью и описывается матрицей с переменными 000000. Его аналогом в белках является полностью несвязная конформация (рис. 2.3, *a*). Единственный граф, у которого нет степеней свободы, занимает правую нижнюю клетку блока и описывается матрицей с переменными 110011.

В блоке 01 большая часть графов имеет плоскую конформацию (степень свободы 2–1). Типичным примером такой конформации является граф, описываемый матрицей 100101, который является аналогом β -структуры белков (см. рис. 2.3, δ). Только один граф (в последней клетке блока правой нижней), описываемый матрицей 110111, имеет стабильную конформацию. Данный граф является аналогом спирали 3₁₀ (см. рис. 2.3, ϵ). Как и другие графы блока, он имеет ациклическую конформацию (см. 2.1).

Блок 10 содержит графы с циклической конформацией ($x_3 = 1$). Однако отсутствие связности между вершинами i - 1 и i - 3 ($x_4 = 0$) приводит к тому, что, несмотря на замкнутый характер, степень свободы вершин графов внутри блоков достаточно велика (1 или 2). Только три конформации, описываемые матрицами 111010, 011011 и 111011, являются стабильными.

Блок 11 также содержит циклические конформации. Восемь из них являются стабильными и занимают последние три колонки и три нижних ряда (за исключением элемента 011110).

Расположение элементов и графов в таблице таково, что соседние элементы в блоках отличаются друг от друга только на одно значение (на один бит информации). Таким образом, данная система напоминает код Грея, в котором различие между соседними элементами минимально [31], и это минимизирует ошибки передачи информации. Однако приведенное

сходство носит формальный характер, поскольку мы имеем дело со структурным описанием связного графа, а не с кодированием. Также нужно иметь в виду, что двумерное расположение элементов в данной суперматрице не отражает всех возможных однобитовых переходов. Последние выявляются пространственным представлением этой структуры.

Гиперкуб B⁶ как пространственное представление суперматрицы

В [25] и [26] была представлена пространственная структура суперматрицы в форме булева гиперкуба В⁶ (рис. 2.5). Его элементы связаны между собой однобитовыми переходами. В такой структуре можно выделить четыре особенности, определяемые свойствами гиперкуба В⁶ [32]: ярусность, иерархичность, связность элементов и симметрию. Остановимся на них подробнее.

Ярусность. Структура содержит 7 ярусов, отмеченных на рис. 2.5, *а* римскими цифрами. В первом и седьмом ярусах содержится по 1 матрице, во втором и шестом – по 6, в третьем и пятом – по 15 и в четвертом – 20 матриц.

Иерархичность. Совокупность из 64 матриц распадается на два множества M_1 и M_2 по 32 элемента (рис. 2.5, *б*). Каждое множество содержит по два подмножества: в M_1 – это SM_1 и SM_2 , а в M_2 – SM_3 и SM_4 (рис. 2.5, *в*). Наконец, каждое подмножество распадается на два октета, общее количество которых равно 8 (O₁–O₈, рис. 2.5, *г*).

Связность элементов. Все матрицы из шести переменных, расположенные в вершинах гиперкуба, связаны между собой единичными (однобитовыми) переходами. При этом строки блоков 00 и 01 расположены друг над другом в виде квартетов с варьирующими переменными x_5 и x_6 в верхней части множеств M_1 и M_2 (выделены жирным шрифтом), а строки блоков 10 и 11 расположены аналогичным образом в нижней части гиперкуба.

Симметрия. Матрицы, описывающие симметричные графы, расположены в гиперкубе симметрично относительно вертикальной оси, проведенной через гиперкуб (например: 100000 и 000001, 111000 и 000111, 111110 и 011111). Расположение элементов суперматрицы, связанных преобразованием антисимметрии (0 $\leftarrow \rightarrow$ 1), описывается группой симметрии C₂ (например, матрица 100000 расположена в гиперкубе слева вверху, а антисимметричная ей матрица 011111 – справа внизу, аналогично 110000 и 001111 и т. д.). Элементы, выделенные жирным шрифтом в гиперкубе и занимающие верхнюю половину каждого множества, принадлежат к элементам верхней части суперматрицы, отделенной от нижней части жирной линией (см. рис. 2.4). В кружках, занимающих нижние половины каждого множества, находятся элементы нижней части суперматрицы.



Рис. 2.5. Пространственное представление суперматрицы, описывающей конформации четырехзвенного графа, в виде булева гиперкуба В⁶



Рис. 2.6. Пространственное представление конформаций четырехзвенного графа на булевом гиперкубе В⁶

Следует отметить, что на гиперкубе можно представить и сами конформации графов (рис. 2.6) [26]. Принципиально такая форма структуры не отличается от формы представления в виде матриц, однако по ней можно наглядно проследить плавные переходы между структурами графов и сопряженные с ними изменения степени их связности.

2.3. Трансформация суперматрицы в топологический код *Кодирование суперматрицы*

Форма представления информации о структуре графа в виде матриц, состоящих из 6 булевых переменных, не удобна для передачи, воспроизведения и тиражирования. В связи с этим возникает необходимость ее перекодирования в пригодную для этих целей линейную цепь [24]–[29]. Поскольку количество переменных в матрицах, описывающих состояния связности четырехзвенных цепных графов, равно 6 (три пары), то каждую пару можно обозначить своей буквой, например:

$$x_1 x_2 - X \qquad x_3 x_4 - Y \qquad x_5 x_6 - Z. \tag{2.2}$$

В результате образуется триплет *XYZ*. Каждая пара переменных в этом триплете может принимать четыре значения (00, 01, 10, 11). Их можно закодировать с помощью четырехбуквенного кода, присвоив каждой паре свою букву:

$$K - 00 \quad L - 01 \quad N - 10 \quad P - 11.$$
 (2.3)

Используя эти соответствия, суперматрицу состояний связности можно трансформировать в триплетный топологический код (рис. 2.7). Теперь в закодированном виде информация о структуре четырехзвенных графов будет носить характер одномерной цепи.

Таблица триплетного топологического кода

В результате трансформации мы получили триплетный топологический код (рис. 2.7). Триплеты здесь сохраняют многие свойства графов. Так, триплеты, расположенные симметрично относительно главных диагоналей блоков, кодируют симметричные конформации графа: первая буква К симметрична третьей букве К (кодирует пару переменных 00); первая буква Р симметрична третьей букве Р (кодирует пару переменных 11). Однако при этом первая буква L симметрична третьей букве N (01 $\leftarrow \rightarrow$ 10) и наоборот. Например, для триплета NKK симметричным будет триплет KKN, для PKK – KKP, в то время как NNK симметричен KNL. Триплеты, кодирующие симметричные графы на главной диагонали, для первого блока будут выглядеть так: KKK, NKL, LKN, PKP. Триплеты LKL и NKN кодируют несимметричные конформации.

Наибольшую информацию о структуре графов несет вторая буква триплета, которая является единой для всего блока триплетов. Она кодирует переменную x_3 , которая, как отмечалось ранее, описывает ациклические ($x_3 = 0$) и циклические ($x_3 = 1$) конформации графа (наличие связности между вершинами i...i - 4), и переменную x_4 (наличие связности между верщинами i - 1...i - 3 в каждом блоке). Буквы три-

2	K ↔ 00				$L \leftrightarrow 01$				
13	$K \leftrightarrow 00$	$L \leftrightarrow 01$	$N \leftrightarrow 10$	P ↔ 11	$K \leftrightarrow 00$	$L \leftrightarrow 01$	$N \leftrightarrow 10$	P ↔ 11	
00 ↓ K	0 0 0 0 0 KKK 0	0 0 0 0 0 KKL	0 0 0 0 1 KKN	0 0 0 0 1 KKP	0 0 0 1 0 KLK	0 0 0 1 0 KLL	0 0 0 1 1 KLN	0 0 0 1 1 KLP	
10 ↓ N	1 0 0 0 0 NKK	1 0 0 0 0 NKL	1 0 0 0 1 0 NKN	1 0 0 0 1 NKP	1 0 0 1 0 NLK	1 0 0 1 0 NLL	1 0 0 1 1 NLN	1 0 0 1 1 NLP	
01 ↓ L	0 1 0 0 0 LKK 0	0 1 0 0 0 LKL 1	0 1 0 0 1 LKN 0	0 1 0 0 1 LKP 1	$\begin{array}{ccc} 0 & 1 & 0 \\ & 1 & 0 \\ LLK & 0 \end{array}$	0 1 0 1 0 LLL 1	$\begin{array}{ccc} 0 & 1 & 0 \\ & 1 & 1 \\ LLN & 0 \end{array}$	0 1 0 1 1 LLP 1	
11 ↓ P	1 1 0 0 0 PKK 0	1 1 0 0 0 PKL 1	1 1 0 0 1 PKN 0	1 1 0 0 1 PKP ¹	1 1 0 1 0 PLK 0	$\begin{array}{ccc} 1 & 1 & 0 \\ & 1 & 0 \\ PLL & 1 \end{array}$	1 1 0 1 1 PLN 0	1 1 0 1 1 PLP 1	
00 ↓ K	0 0 1 0 0 KNK	0 0 1 0 0 1 KNL	0 0 1 0 1 0 KNN	0 0 1 0 1 1 KNP	0 0 1 1 0 0 KPK	0 0 1 1 0 1 KPL	0 0 1 1 1 0 KPN	0 0 1 1 1 KPP	
10 ↓ N	1 0 1 0 0 NNK 0	1 0 1 0 0 NNL 1	1 0 1 0 1 NNN 0	1 0 1 0 1 NNP 1	1 0 1 1 0 NPK 0	1 0 1 1 0 NPL 1	1 0 1 1 1 NPN 0	1 0 1 1 1 NPP 1	
01 ↓ L	0 1 1 0 0 UNK	0 1 1 0 0 1 LNL	0 1 1 0 1 LNN	0 1 1 0 1 1 LNP	0 1 1 1 0 0 LPK	0 1 1 1 0 1 LPL	0 1 1 1 1 LPN	0 1 1 1 1 LPP	
11 ↓ P	1 1 1 0 0 PNK	1 1 1 0 0 PNL	1 1 1 0 1 PNN	1 1 1 0 1 PNP	1 1 1 1 0 PPK	1 1 1 1 0 1 PPL	1 1 1 1 1 PPN	1 1 1 1 1 PPP	
2	$2 \qquad \qquad N \leftrightarrow 10$				$\mathbf{P} \leftrightarrow$	11			

Рис. 2.7. Трансформация суперматрицы состояний связности четырехзвенного цепного графа в триплетный топологический код

плетов, кодирующие антисимметричные пары ($00 \leftarrow \rightarrow 11, 10 \leftarrow \rightarrow 01$), преобразуются по правилу: К $\leftarrow \rightarrow$ Р и L $\leftarrow \rightarrow$ N. При этом антисимметричные триплеты переходят друг в друга и занимают в топологическом коде положение, описываемое группой симметрии C₂ (две компактные Г-образные фигуры, разделенные жирной линией).

Пространственное представление топологического кода в форме булева гиперкуба В⁶

Как было показано ранее, пространственным представлением суперматрицы, описывающей конформации четырехзвенного графа с помощью матриц из шести переменных, является булев гиперкуб В⁶. Очевидно, что представлением топологического кода будет структура, изоморфная этому гиперкубу (рис. 2.8).



Рис. 2.8. Пространственная структура триплетного топологического кода, изоморфная булеву гиперкубу В⁶

Как следует из рисунка, триплеты с двумя общими неизменными буквами и третьей варьирующей буквой расположились в каждом множестве друг над другом в виде квартетов. При этом квартеты верхней части множеств соответствуют верхней Г-образной фигуре таблицы кода (они выделены жирными линиями). В данной структуре сохранились все свойства гиперкуба, в частности, – иерархия структур. Симметрия триплетов имеет в гиперкубе те же особенности, что и в таблице кода. Преобразование антисимметрии в гиперкубе осуществляется заменой букв триплетов по правилу К $\leftarrow \rightarrow$ P, N $\leftarrow \rightarrow$ L. Выделенные жирными линиями триплеты верхней части гиперкуба преобразуются на основе этого правила в триплеты нижней половины гиперкуба занимая поворотно-симметричное положение (группа симметрии C₂), например: NKK $\leftarrow \rightarrow$ LPP, PKK $\leftarrow \rightarrow$ KPP, PNK $\leftarrow \rightarrow$ KLP и т. д.

Рассмотрение вопросов кодирования конформаций ПФ и четырехзвенных графов можно на этом закончить. В [24]–[29] мы уделили значительное внимание вопросам создания алгоритма кодирования n-звенного полимера и n-звенного графа. Однако, учитывая, что проблема прогнозирования вторичной структуры белка на основе декодирования последовательности триплетов была решена несколько иным способом (подробности позднее), использование алгоритма декодирования остается проблематичным. В настоящем издании мы его опускаем и переходим к обсуждению проблемы воссоздания закодированной структуры цепного полимера.

2.4. Система физических операторов

Понятие о физическом операторе

Идея воссоздания закодированной структуры базируется на способности цепного полимера к самоорганизации и формированию непрерывных ССИВС. Этот вопрос был подробно проанализирован в [26]– [29]. Согласно этим работам боковые цепи полимера могут выступать в роли регуляторов самоорганизации, способствуя или препятствуя формированию спиральных структур на основе непрерывных ССИВС.

Для воссоздания структуры графа, закодированной цепным полимером, необходимо, чтобы между кодирующим триплетом и физическим оператором, воссоздающим закодированную структуру, существовало определенное соответствие.

Как показано на рис. 2.9, *a*, цепь полимера, содержащая группы, способные к образованию непрерывных ССИВС, т. е. наделенная свойством связности, способна к образованию циклического ПФ. Описание

конформации этого цикла осуществляется с помощью матрицы из шести переменных согласно схеме (2.1) (на рис. 2.9, *a* она показана справа), причем наиболее существенной для его формирования является связность *i*-й и (*i* – 4)-й вершин, описываемая переменной x_3 (выделена жирным шрифтом). Другие группы в формировании замкнутого цикла не участвуют (кроме водородной связи Q_iH...X_{*i*-4}=R). Боковая цепь в *i*-м положении (R_i) может воздействовать на образование водородной связи групп Q_iH...X_{*i*-4}=R (показана стрелкой), причем единственной областью приложения действия боковых цепей может быть только эта связь.



Рис. 2.9. Иллюстрация понятия «физический оператор»: *а* – область действия физического оператора; *б* – оператор связности; *в* – оператор антисвязности

В зависимости от оказываемого эффекта можно выделить два типа таких операторов: операторы связности и операторы антисвязности (рис. 2.9, б и в).

Операторы связности – это такие боковые цепи полимера, которые способствуют образованию циклического ПФ и его дополнительной фиксации (например, за счет водородных связей). Для реализации этой функции боковая цепь должна иметь на конце группы, способные образовывать водородные связи. Обобщенный вид оператора связности показан на рис. 2.9, δ . Переменная x_3 в матрице равна 1, что свидетельствует об образовании циклического ПФ.

Операторы антисвязности – это боковые цепи полимера, которые препятствуют формированию циклического ПФ, т. е. способствуют образованию ациклических ПФ. Боковые цепи операторов антисвязности должны вклиниваться в область водородной связи основной

2	$\mathbf{K} \leftrightarrow 00$				L↔01			
13	$K \leftrightarrow 00$	$L \leftrightarrow 01$	$N \leftrightarrow 10$	$P \leftrightarrow 11$	$K \leftrightarrow 00$	$L \leftrightarrow 01$	$N \leftrightarrow 10$	P ↔ 11
00 ↓ K	0 0 0 0 0 KKK	0 0 0 0 0 KKL	0 0 0 0 1 KKN	0 0 0 0 1 KKP	0 0 0 1 0 KLK	0 0 0 1 0 KLL	0 0 0 1 1 KLN	0 0 0 1 1 KLP
10 ↓ N	1 0 0 0 0 NKK	1 0 0 0 0 NKL	1 0 0 0 1 0 NKN	1 0 0 0 1 NKP	1 0 0 1 0 NLK	1 0 0 1 0 NLL	1 0 0 1 1 NLN	1 0 0 1 1 NLP
01 ↓ L	0 1 0 0 0 LKK 0	0 1 0 0 0 LKL 1	0 1 0 0 1 LKN 0	0 1 0 0 1 LKP 1	$\begin{array}{ccc} 0 & 1 & 0 \\ & 1 & 0 \\ LLK & 0 \end{array}$	0 1 0 1 0 LLL 1	$\begin{array}{ccc} 0 & 1 & 0 \\ & 1 & 1 \\ LLN & 0 \end{array}$	0 1 0 1 1 LLP 1
11 ↓ P	1 1 0 0 0 PKK 0	1 1 0 0 0 PKL 1	1 1 0 0 1 PKN 0	1 1 0 0 1 PKP ¹	1 1 0 1 0 PLK 0	$\begin{array}{ccc} 1 & 1 & 0 \\ & 1 & 0 \\ PLL & 1 \end{array}$	1 1 0 1 1 PLN 0	1 1 0 1 1 PLP 1
00 ↓ K	0 0 1 0 0 KNK	0 0 1 0 0 KNL	0 0 1 0 1 KNN	0 0 1 0 1 1 KNP	0 0 1 1 0 KPK	0 0 1 1 0 1 KPL	0 0 1 1 1 KPN	0 0 1 1 1 KPP
10 ↓ N	1 0 1 0 0 NNK 0	1 0 1 0 0 NNL 1	1 0 1 0 1 NNN 0	1 0 1 0 1 NNP 1	1 0 1 1 0 NPK 0	1 0 1 1 0 NPL 1	1 0 1 1 1 NPN 0	1 0 1 1 1 NPP 1
01 ↓ L	0 1 1 0 0 UNK	0 1 1 0 0 1 LNL	0 1 1 0 1 LNN	0 1 1 0 1 LNP	0 1 1 1 0 LPK	0 1 1 1 0 1 LPL	0 1 1 1 1 LPN	0 1 1 1 1 LPP
11 ↓ P	1 1 1 0 0 PNK	1 1 1 0 0 PNL 1	1 1 1 0 1 PNN 0	1 1 1 0 1 PNP 1	1 1 1 1 0 PPK 0	1 1 1 1 0 PPL 1	1 1 1 1 1 PPN 0	1 1 1 1 1 PPP
2	$2 \qquad \qquad N \leftrightarrow 10$					P↔	11	

Рис. 2.10. Соответствие физических операторов блокам триплетов топологического кода

цепи и не допускать образования водородной связи в этой области (см. рис. 2.9, e). Переменная x_3 в матрице принимает значение, равное 0, что указывает на отсутствие H-связи в группе $Q_iH...X_{i-4}=R$.

Как было показано ранее, в суперматрице топологического кода, содержащего аналоги ПФ (четырехзвенные графы), имеется два блока с ациклическими конформациями (00 и 01) и два блока с циклическими конформациями (10 и 11). Соответственно в матрицах, описывающих конформации графов первой пары блоков, $x_3 = 0$, а в матрицах, описывающих конформации графов второй пары, $x_3 = 1$. Если вместо графов в суперматрицу подставить сами ПФ, то операторы антисвязности, групповым свойством которых является способность формировать ациклические конформации полимера, должны быть приписаны к блокам 00 и 01. В то же время операторы связности, общим свойством которых является конформации полимера, должны быть приписаны к блокам 10 и 11.

При переходе к триплетному топологическому коду (см. рис. 2.7) блокам триплетов, содержащим во втором положении K и L (K = 00, L = 01), должны соответствовать операторы антисвязности, а блокам N и P (N = 10, P = 11) – операторы связности. Это наглядно иллюстрирует рис. 2.10.

Таким образом, операторы связности соответствуют блокам триплетов, кодирующим циклические конформации графа или полимера (заштрихованы), а операторы антисвязности – блокам, кодирующим ациклические конформации.

Требования к процессу синтеза цепного полимера и физическим операторам

Чтобы физические операторы могли выполнять свою функцию, процесс синтеза цепного полимера и свойства физических операторов должны удовлетворять ряду требований [26]–[29].

Как показано на рис. 2.9, появление очередного физического оператора должно происходить путем элонгации цепи со стороны ее прикрепления к матрице. Осуществить этот процесс чисто химическим путем трудно, поскольку нужно, чтобы очередной оператор обладал способностью внедряться в область прикрепления полимера к матрице. Для реализации процесса требуется «молекулярная машина». В биоструктурах та-
кой «машиной» является рибосома, на которой и осуществляются последовательные этапы синтеза белка. Иными словами: данный механизм действия физических операторов, имеющий ко-трансляционный характер, тесно связан с механизмом синтеза цепного полимера.

К физическим операторам предъявляется два общих требования.

Хиральность канонического набора операторов. Как видно из рис. 2.9, б и в, действие физических операторов направлено назад, на предшествующую структуру, т. е. они являются ретрооператорами. Следствием этого является требование единой стереоконфигурации (т. е. хиральной чистоты), предъявляемое ко всем физическим операторам. В противном случае они не смогут выполнять роль операторов. При отсутствии хиральной чистоты одна часть операторов будет направлена в нужную сторону, а другая – в противоположную. Таким образом, первым общим требованием является единый тип стереоконфигурации физических операторов, обеспечивающих воссоздание закодированной структуры.

Согласованность размера оператора с пространственными параметрами связи, на которую он действует. Из рис. 2.9 следует, что размер физического оператора определяется параметрами водородной связи в группе $Q_iH...X_{i-4}$, на которую он воздействует. Например, боковые цепи, у которых терминальные полярные группы Q=R–XH будут расположены ближе или дальше этой связи, уже не подойдут на роль операторов связности. То же можно сказать об операторах антисвязности. Таким образом, размеры боковых цепей, выполняющих функции физических операторов, должны быть согласованы с размерами той области связей, на которую они будут действовать.

Следствием этого требования является вывод о том, что в зависимости от типа связующих групп X_i =R–Q_iH создаваемые системы физических операторов должны обладать различными параметрами боковых цепей. Например, если в качестве таких групп выступают группы O=C–NH, то это будет одна система операторов, если группы S=C–NH – другая, а если группы O=P–OH – третья. Это открывает перспективу создания разнообразных систем физических операторов, параметры которых «увязаны» с размерами области, на которую они действуют. В ч. 2 этот вывод будет проиллюстрирован на примере боковых цепей аминокислот (система операторов, воздействующих на связь N_iH…O_{i–4}=C) и минорных нуклеотидов (система операторов, воздействующих на связь O_iH…O_{i–4}=P).

Физические операторы, воссоздающие симметричные конформации и их расположение в топологическом коде

Прежде чем перейти непосредственно к рассмотрению этой проблемы, остановимся на вопросе о том, что может служить в ПФ физической основой для ребер связности в графах. В 2.1 отмечалось, что ребра связности соответствуют фиксированным атомам в полимере. При этом фиксация атомов может обеспечиваться как за счет водородных связей внутри основной цепи (связь Q_iH...X_{i-4}=R), так и за счет водородных связей боковых цепей с основной цепью. Идейной предпосылкой такого воздействия боковых цепей могут служить представления, развиваемые Г. Н. Лингом (см., например, [33]), согласно которым в белках может осуществляться передача влияния боковых цепей на основную цепь с последующим распространением этого влияния вдоль основной цепи. Не углубляясь в подробности этих идей, можно предполагать, что именно такое влияние приводит к фиксации ряда атомов ПФ, которые на четырехзвенном графе изображаются в виде ребер связности, а на матрицах соответствующие переменные имеют единичные значения.

Для рассмотрения вопроса о воссоздании симметричных конформаций графов и ПФ обратим внимание на то, что внутри блоков суперматрицы (см. рис. 2.4) симметричные конформации графов и описывающие их матрицы расположены симметрично относительно главных диагоналей. (Это вытекает из принципов построения суперматрицы.)

Рассмотрим, например, какие физические операторы должны воссоздавать эти конформации в блоке 11. Этому блоку соответствуют операторы связности (см. рис. 2.10). Как показано на рис. 2.11, *a*, более длинный оператор связности, имеющий на конце группу Q_1 –R=X₁, производит усилие, тянущее влево, в результате чего разложение сил за счет влияния этого усилия от группы X_{i-4} =R будет приводить к связности атомов i - 2...i - 4 (показано штриховой), что описывается соответствующей матрицей справа. На рис. 2.11, *б* другой оператор, содержащий аналогичную функциональную группу Q_1 –R=X₁, но имеющий более короткую боковую цепь, производит тянущее усилие, направленное вверх, что приводит к связности атомов i...i - 2 за счет влияния этого усилия от группы X_{i-4} =R. Это также описывается мат-

рицей рис. 2.11, б. Сравнение этих матриц показывает, что они описывают симметричные конформации. Таким образом, для воссоздания симметричных конформаций необходимы физические операторы, имеющие различную длину боковой цепи. При этом функциональные группы могут быть как сходными, так и различными.



Рис. 2.11. Соответствия физических операторов триплетам, кодирующим симметричные конформации графа и ПФ

С учетом этих рассуждений становится ясно, как должны быть расположены физические операторы, воссоздающие симметричные конформации. Из рис. 2.12 следует, что в пределах блоков триплетам, кодирующим симметричные конформации (они расположены симметрично относительно главных диагоналей), должны соответствовать разные физические операторы (заштрихованные и незаштрихованные клетки).

В целом, учитывая сформулированные следствия, можно сказать следующее. Система соответствий физических операторов триплетам топологического кода должна состоять из двух групп (см. рис. 2.10). Верхней паре блоков должны соответствовать операторы антисвязности, а нижней паре блоков – операторы связности. Закодированные триплетами симметричные конформации должны воссоздаваться разными физическими операторами (рис. 2.11). В ходе последующего изложения эти свойства соответствий топологического кода будут сопоставлены со свойствами биологического кода, рассматриваемого нами в качестве частного случая таких кодов.

Проведенный в данной главе анализ позволяет в общем виде ответить на следующие вопросы, поставленные ранее:

• Почему генетический код содержит 64 кодирующих триплета, а не иное их число?

2	$\mathbf{K} \leftrightarrow 00$				$L \leftrightarrow 01$				
13	$K \leftrightarrow 00$	$L \leftrightarrow 01$	$N \leftrightarrow 10$	P↔11	$K \leftrightarrow 00$	$L \leftrightarrow 01$	$N \leftrightarrow 10$	$P \leftrightarrow 11$	
00 ↓ K	0 0 0 0 0 KKK	0 0 0 0 0 KKL 1	0 0 0 0 1 KKN	0 0 0 0 1 KKP	0 0 0 1 0 KLK	0 0 0 1 0 KLL 1	0 0 0 1 1 KLN	0 0 0 1 1 KLP 1	
10 ↓ N	1 0 0 0 0 NKK	1 0 0 0 0 NKL	1 0 0 0 1 NKN 0	1 0 0 0 1 NKP	1 0 0 1 0 NLK	1 0 0 1 0 1 1 NLL	1 0 0 1 1 NLN	1 0 0 1 1 NLP	
01 ↓ L	0 1 0 0 0 LKK 0	0 1 0 0 0 LKL 1	0 1 0 0 1 LKN 0	0 1 0 0 1 LKP 1	0 1 0 1 0 LLK 0	0 1 0 1 0 LLL 1	0 1 0 1 1 LLN 0	0 1 0 1 1 LLP 1	
11 ↓ P	1 1 0 0 0 PKK 0	1 1 0 0 0 PKL 1	1 1 0 0 1 PKN 0	1 1 0 0 1 PKP 1	$ \begin{array}{r} 1 & 1 & 0 \\ 1 & 0 \\ PLK & 0 \end{array} $	1 1 0 1 0 PLL 1	1 1 0 1 1 PLN 0	1 1 0 1 1 PLP	
00 ↓ K	0 0 1 0 0 KNK	0 0 1 0 0 1 KNL	0 0 1 0 1 0 KNN	0 0 1 0 1 KNP	0 0 1 1 0 0 KPK	0 0 1 1 0 1 KPL	0 0 1 1 1 KPN	0 0 1 1 1 KPP	
10 ↓ N	1 0 1 0 0 NNK 0	1 0 1 0 0 NNL 1	1 0 1 0 1 NNN 0	1 0 1 0 1 NNP 1	1 0 1 1 0 NPK 0	1 0 1 1 0 NPL 1	1 0 1 1 1 NPN 0	1 0 1 1 1 NPP 1	
01 ↓ L	0 1 1 0 0 LNK	0 1 1 0 0 1 LNL	0 1 1 0 1 0 LNN	0 1 1 0 1 LNP	0 1 1 1 0 LPK	0 1 1 1 0 1 LPL	0 1 1 1 1 LPN	0 1 1 1 1 LPP	
11 ↓ P	1 1 1 0 0 PNK	1 1 1 0 0 PNL 1	1 1 1 0 1 PNN	1 1 1 0 1 PNP	1 1 1 1 0 PPK	1 1 1 1 0 PPL 1	1 1 1 1 1 PPN	1 1 1 1 1 PPP	
2	$N \leftrightarrow 10$				P ↔ 11				

Рис. 2.12. Соответствие физических операторов триплетам, кодирующим симметричные конформации графа (пояснения в тексте)

• Какова природа соответствия триплетов определенным боковым цепям аминокислот?

Более конкретно (применительно к биоструктурам), мы ответим на эти вопросы во второй части издания. В то же время, дальнейший анализ в нашем подходе был сосредоточен на области водородной связи ПФ: Q_iH.....X_{i-4}=R. Результаты изложены в следующей главе.

Глава 3. МОЛЕКУЛЯРНАЯ ВЕКТОРНАЯ МАШИНА

3.1. Этапы построения молекулярной векторной машины Выделение плоскостей и векторов в области Q_iH...X_{i-4}=R-связи

Как и предыдущие представления, модель «молекулярной векторной машины» (MBM) является теоретической абстракцией, возникшей из потребности разобраться в природе механизмов самоорганизации цепных полимеров. Эти механизмы носят общий характер, поэтому подход применим ко всем цепным полимерам и, в первую очередь, к биополимерам – белкам и нуклеиновым кислотам. По этой проблеме нами опубликованы работы [1] и [34]–[37]. Основное внимание будет уделено области образования водородной связи Q_iH...X_{i–4}=R.

Рассмотрим более детально область $Q_iH...X_{i-4}$ =R-связи (рис. 3.1). Как и ранее (см. 1.2 и 2.1), вместо символов конкретных элементов (N, O, P, S, C) будем писать символы типа Q, R, X с индексами. В вершинах звеньев, как правило, располагаются α -атомы элемента R, способного формировать тетраэдрическую структуру связей (например, атомы углерода, кремния, германия и т. п.). К ним прикрепляются боковые цепи полимера. Будем обозначать их символом R^{α} с индексами. Примем во внимание, что $HX_i - R_i = Q_i$ -группа в цепных полимерах может иметь частично делокализованную двойную связь, за счет чего эта группа (например, пептидная группа в белках) является плоской [38]. Ввиду частичной делокализации электронное окружение связи $HX_i - R_i = Q_i$ -группы может быть представлено в виде трех sp^2 -гибридизированных облаков (на атомах X_i , R_i и Q_i). На рис. 3.1, *а* показано одно из таких облаков для атома X_{i-4} .

Через $Q_iH...X_{i-4}$ =R-связь можно провести три взаимно перпендикулярные плоскости (рис. 3.1, *a*). Плоскость I проходит перпендикулярно плоскости листа, в результате чего каждая из лопастей *sp*²-гибридизованного облака будет находиться симметрично по разные стороны от атома X_{i-4} . Плоскость II проходит параллельно плоскости листа (перпендикулярно плоскости I) и разделяет *sp*²-гибридизованное облако на две симметричные половины – переднюю и заднюю. Плоскость III перпендикулярна двум предыдущим плоскостям и разделяет каждую из лопастей *sp*²-гибридизованного облака на две половины – верхнюю и нижнюю.



Рис. 3.1. Введение плоскостей симметрии (*a*) и анализ возможных направлений действия физических операторов в области Q_iH...X_{i-4}=R-связи (*б*-*г*)

Рассмотрим направления, в которых действуют на эту связь физические операторы (рис. 3.1, δ –г). Будем называть эти направления векторами действия (или просто векторами) и обозначать их стрелками. В пределах плоскости I возможно два варианта векторов действия на водородную связь: вдоль $Q_iH...X_{i-4}$ =R-связи и перпендикулярно ей (рис. 3.1, δ). При действии вдоль этой связи возможны два вектора. Вектор, направленный в сторону атома Q_i (стрелка вверх), будет способствовать сохранению водородной связи. Симметричный вектор будет направлен в сторону атома X_{i-4} (стрелка вниз), на разрыв водородной связи. В случае действия векторов перпендикулярно $Q_iH...X_{i-4}$ =R-связи (рис. 3.1, δ) оба вектора ориентированы в противоположные стороны (симметричны) и направлены на разрыв водородной связи.

По отношению к плоскости II возможны две пары симметричных векторов, направленных примерно параллельно этой плоскости: вверх (в сторону атома Q_i) и, после вращения вокруг плоскости III, вниз (в сторону атома X_{i-4}) (рис. 3.1, *в*). Наконец, возможен еще один вариант действия векторов, направленных примерно параллельно плоскости III (рис. 3.1, *г*): две пары симметричных векторов справа и слева от плоскости I, а также спереди и сзади от плоскости II расположены над плоскостью III, и еще две симметричные пары направлены аналогично под плоскостью III. Таким образом, анализ области $Q_iH...X_{i-4}$ =R-связи позволил выявить по крайней мере 20 векторов, связанных преобразованиями симметрии, которые могут быть воссозданы с помощью 20 физических операторов.

Задание направлений векторов

Двадцать направлений векторов можно задать с помощью какойлибо пространственной фигуры. Наиболее подходящим для этих целей оказался додекаэдр, имеющий 20 вершин (рис. 3.2, *a*). В центр додекаэдра помещен атом X_{i-4} , в одной из вершин – атом Q_i , а векторы направлены в вершины додекаэдра. При этом было сохранено разделение области $Q_iH...X_{i-4}$ =R-связи тремя плоскостями, что обеспечило сохранение принципов симметрии в положении векторов, а размеры додекаэдра определились, исходя из параметров ПФ полимера. Чтобы задать вектор, необходимо знать положения двух точек: начальной (из которой исходит вектор) и конечной (в которую он направлен). В данном случае за начальную точку всех векторов можно принять центр додекаэдра, а конечными точками будут вершины додекаэдра. Для этих вершин введены следующие обозначения (рис. 3.2, б). В плоскости I вершина, в которой расположен атом Q_i, обозначена буквой A, а симметричная ей и находящаяся под плоскостью III вершина – буквой (-А). Две другие вершины в плоскости I, в которые направлены векторы, перпендикулярные первым двум (на рис. 3.2, а они не видны), обозначены соответственно В и (-В). Вершины, симметричные относительно плоскости I, были обозначены так: слева все буквы, обозначающие вершины, имеют индекс, расположенный справа внизу, а справа – индекс, расположенный слева вверху. Так, ближние к А вершины, находящиеся за плоскостью II, обозначены A₁ и ¹A, а отдаленные (расположенные перед плоскостью II) – А2 и ²А. Вершины, расположенные под плоскостью III, будут обозначены соответственно (-A₁) и (-1А) перед плоскостью II и (-А₂), (-2А) за плоскостью II.



Рис. 3.2. Задание направлений векторов с помощью додекаэдра (a) и обозначение вершин додекаэдра (б)

Аналогичные обозначения введены для второй группы из восьми векторов. Вершины, расположенные ближе к вершине В за плоскостью II, обозначены ¹В и B_1 , а вершины, расположенные перед плоскостью II, обозначены ²В и B_2 . По отношению к вершине (–В) обозначения будут

таковы: $(-^{1}B)$ и $(-B_{1})$ перед плоскостью II; $(-^{2}B)$ и $(-B_{2})$ за плоскостью II. Таким образом, все 20 вершин, в которые направлены векторы, получили свои обозначения.

Введение представления о МВМ

Теперь практически все готово для введения представления о MBM, которая изображена на рис. 3.3. Видно, что к плоскостям, разделяющим область Q_iH...X_{i-4}=R-связи, и к додекаэдру с обозначениями



Рис. 3.3. Молекулярная векторная машина

вершин (рис. 3.2, *а* и б) добавились лишь стрелка с буквой S_{*i*}, обозначающая сменяемый физический оператор, и фрагмент (*i* + 1)-го звена со стрелкой, указывающей направление на (*i* + 1)-й R^{α} атом. В целом MBM состоит из трех частей:

 – додекаэдра, содержащего группу из 20 векторов – радиусов додекаэдра;

- канонического набора сменяемых физических операторов;

– тетраэдрического R_i^{α} -го атома, к которому прикреплены физические операторы и который задает направление растущей цепи полимера.

Рассмотрим свойства каждой из этих частей МВМ.

3.2. Свойства векторов додекаэдра

Преобразования симметрии

В результате переноса додекаэдра в область $Q_iH...X_{i-4}$ =R-связи векторы, выделенные в этой области, оказались радиусами додекаэдра, исходящими из центра и направленными к вершинам. В силу симметрии векторов, направленных в противоположные стороны, их можно рассматривать в качестве диаметров додекаэдра. Направление каждого вектора ранее было охарактеризовано двумя точками:

– точкой, из которой исходит радиус (в данном случае это центр додекаэдра, т. е. атом X_{*i*-4}, общий для всех 20 радиусов);

 – точкой, в которую направлен радиус (в данном случае это одна из вершин додекаэдра).

Положение этих вершин в принципе можно описать, используя какие-либо стандартные координаты (прямоугольные, полярные и др.). Мы использовали для этого введенные ранее буквенные обозначения.

Внутри додекаэдра векторы образуют три группы, связанные взаимными преобразованиями симметрии (см. рис. 3.2, *б*). Обозначим преобразования симметрии относительно плоскости I (отражение) буквой α, относительно плоскости II (отражение) буквой β и относительно плоскости III (вращение) – буквой γ. Тогда все преобразования векторов можно представить в виде табл. 3.1.

Таблица 3.1

Номер	Преобразования симметрии								
подгруппы	1	α	β	γ	αβ	αγ	βγ	αβγ	
1	Α			-А					
2	В			-В					
3	A ₁	¹ A	A ₂	-A ₁	^{2}A	$-^{1}A$	-A ₂	$-^{2}A$	
4	B ₁	^{1}B	B ₂	-B ₁	$^{2}\mathrm{B}$	$-^{1}B$	-B ₂	$-^{2}B$	

Подгруппы векторов, связанные преобразованиями симметрии

Подгруппы 1 и 2 локализованы в плоскости I и включают две пары взаимно перпендикулярных векторов, связанных тождественным преобразованием вектора самого в себя (1: A \rightarrow A, B \rightarrow B) и преобразованием вращения γ вокруг оси C₂ относительно плоскости III: A \rightarrow (-A) и B \rightarrow (-B)).

Подгруппа 3 включает восемь векторов, обозначенных буквой A с индексами, и связана тождественным преобразованием (1) вектора самого в себя ($A_1 \rightarrow A_1$), преобразованиями α -отражения относительно плоскости I ($A_1 \rightarrow {}^{1}A$) и β -отражения относительно плоскости II ($A_1 \rightarrow A_2$), а также и их сочетанием: $\alpha\beta$ ($A_1 \rightarrow {}^{2}A$). Эти элементы также связаны преобразованием вращения γ вокруг оси C₂ относительно плоскости III в разных сочетаниях ($\gamma : A_1 \rightarrow (-A_1), \alpha\gamma : A_1 \rightarrow (-{}^{1}A), \beta\gamma : A_1 \rightarrow (-{}^{2}A)$).

Подгруппа 4 включает восемь векторов, обозначенных буквой B с индексами и связанных теми же операциями преобразования симметрии, но сдвинутыми вниз относительно элементов подгруппы 3. Так, имеется тождественное преобразование ($B_1 \rightarrow B_1$), отражение α ($B_1 \rightarrow {}^1B$), отражение $\beta - (B_1 \rightarrow B_2)$ и их сочетание – $\alpha\beta$ ($B_1 \rightarrow {}^2B$), а также вращение γ относительно плоскости III в различных сочетаниях (γ : $B_1 \rightarrow (-B_1)$; $\alpha\gamma$: $B_1 \rightarrow (-{}^1B)$; $\beta\gamma$: $B_1 \rightarrow (-B_2)$; $\alpha\beta\gamma$: $B_1 \rightarrow (-{}^2B)$).

Все эти операции, проделанные для A₁ и B₁, можно провести для любого из векторов этих двух групп.

Применение теоретико-группового подхода

Теоретико-групповой подход оказался очень эффективным в теоретической физике [39]. Мы применили его к анализу свойств векторов МВМ. Прежде чем проводить такой анализ, напомним основные аксиомы теории групп [39].

Непустое множество G с заданной на нем бинарной операцией $G \times G \to G$ называется группой (G), если оно аксиоматически обладает следующими свойствами:

• ассоциативность: для любых a, b и c из G верно $(a \cdot b) \cdot c = a \cdot (b \cdot c);$

• наличие нейтрального элемента: в G существует элемент e, такой, что для всех a из G справедливо $e \cdot a = a \cdot e = a$;

• наличие обратного элемента: для любого a из G найдется элемент a^{-1} из G, называемый обратным, такой, что $a \cdot a^{-1} = a^{-1} \cdot a = e$.

Сопоставление этих аксиом с множеством векторов додекаэдра позволяет утверждать, что это множество – не что иное, как группа. В данном случае ассоциативность векторов следует понимать так, что результат действия любых векторов из этой группы всегда будет одним и тем же, если сохраняется последовательность их действия.

Наличие нейтрального элемента предполагает существование такого вектора, появление и действие которого не меняет типа структуры. Вектор А, действие которого всегда направлено на формирование циклического ПФ, можно рассматривать как нейтральный элемент.

Существование обратного элемента предполагает для нашей группы наличие вектора, реализация которого коренным образом меняет характер структуры. В качестве обратного элемента можно принять вектор (–А), поскольку его действие всегда направлено на разрушение циклического ПФ.

3.3. Канонический набор физических операторов

Физические операторы как представления группы векторов

Развивая теоретико-групповой подход, мы применили его к физическим операторам (боковым цепям полимера), рассмотрев их в качестве группы неприводимых представлений [1], [34], [36]. Из определения представлений группы векторов [39] имеем следующее.

Пусть есть некоторая конечная группа G с элементами $g_1, g_2, ..., g_n$. Если группа T линейных операторов Tg_i в некотором пространстве R гомоморфна группе G, то говорят, что группа T образует представ-

ление группы G. Из анализа свойств линейных операторов вытекает, что каждому элементу g_i группы можно сопоставить матрицу $||D_{rk}(g_i)||$. При этом нейтральному элементу должна быть сопоставлена нейтральная матрица, а обратному элементу – обратная.

По аналогии с элементами группы и их представлениями физические операторы можно назвать «неприводимыми представлениями» группы векторов. Термин «неприводимые» используется в том смысле, что физические операторы являются самыми простыми (проще их не может быть). Используя эту аналогию, проведем сопоставление группы векторов и группы физических операторов, воссоздающих действие этих векторов. Это сопоставление способствует формированию представлений о составе боковых цепей, входящих в канонический набор физических операторов.

Физические операторы, реализующие действие нейтрального и обратного элементов

Точками приложения векторов и физических операторов являются вершины додекаэдра. В самом деле: с одной стороны, они являются точками, в которые направлены векторы, действующие на Q_iH...X_{i-4}=Rсвязь, а с другой – они являются местами приложения физических операторов, реализующих действие этих векторов.

Вектор, являющийся в группе векторов нейтральным элементом, должен воссоздаваться физическим оператором, который не должен влиять на структуру ПФ. Это означает, что такой оператор может вообще не иметь боковой цепи, а являться «пустым» звеном в цепи полимера. По своим свойствам он должен относиться к подгруппе операторов связности.

Вектор, соответствующий обратному элементу группы, по своему действию направлен на разрыв $Q_iH...X_{i-4}$ =R-связи. Его может реализовать такой физический оператор, у которого атом Q_i в принципе не может образовать водородную связь с группой X_{i-4} =R (например, если этот атом заключен в цикл). Атом X_{i-4} может приобрести противоположное направление в результате «лобового» столкновения его электронной оболочки с электронной оболочкой этой циклической структуры. Такое столкновение может служить прототипом действия всех физических операторов антисвязности.

Свойства физических операторов в структуре МВМ

Общие требования к физическим операторам были сформулированы в гл. 2. Здесь будут рассмотрены свойства физических операторов, реализующих действие векторов в структуре MBM. Выводы будут использованы при построении модели структуры физических операторов, основой которой (как и в случае MBM) будет додекаэдр.

Длина оператора должна увеличиваться по мере изменения направления вектора в сторону атома X_{i-4} . Боковые цепи операторов, воссоздающих векторы, направленные в сторону атома Q_i (как это видно из рис. 3.3), должны быть короче операторов, воссоздающих векторы, направленные в сторону X_{i-4} .

Векторы, направленные вправо от плоскости I, должны воссоздаваться более короткими операторами. Как видно из рис. 3.3, положение R_i^{α} -атома, к которому прикрепляются физические операторы в MBM, несколько асимметрично и сдвинуто вправо относительно $Q_iH...X_{i-4}$ =Rсвязи. Это означает, что физические операторы, воссоздающие векторы и направленные вправо от плоскости I, проходящей через $Q_iH...X_{i-4}$ =Rсвязь, должны быть короче, чем операторы, реализующие действие векторов, направленных влево от этой плоскости. Данная закономерность должна проявляться в отношении всех физических операторов, что должно приводить к формированию двух групп операторов с близкими свойствами, но разной длиной боковых цепей.

Векторы, ориентированные симметрично относительно плоскости II, должны воссоздаваться близкими по размеру операторами. Как видно из рис. 3.3, расстояния до вершин, в которые направлены векторы, находящиеся за плоскостью II (направлены от нас) и перед плоскостью II (направлены к нам), отличаются не очень сильно. По этой причине физические операторы, воссоздающие эти векторы, не должны существенно отличаться по размеру. Их отличия могут состоять лишь в замене терминальных групп.

Векторы, расположенные над плоскостью III, должны воссоздаваться более короткими физическими операторами. Как видно из рис. 3.3, вершины векторов, направленных в сторону группы HQ_i-R=X_{i-1},

50

расположенные над плоскостью III, находятся значительно ближе в R_i^{α} -

атому, чем вершины векторов, направленных в противоположную сторону (находятся под плоскостью III). По этой причине все векторы, находящиеся над плоскостью III, должны воссоздаваться более короткими физическими операторами, чем векторы, находящиеся под плоскостью III. При этом, поскольку эти векторы связаны γ -операцией симметрии (вращение вокруг оси C₂), то операторы с самыми короткими боковыми цепями должны противостоять векторам с самыми длинными боковыми цепями. В то же время физические операторы, воссоздающие векторы, расположенные в экваториальной области додекаэдра, должны отличаться в меньшей степени (например, на одно звено в цепи).

Следует также отметить, что физические операторы на MBM образуют своеобразные семейства, обозначенные буквами (A), (–A), (B) и (–B) с индексами или без них. Они располагаются в MBM на одном уровне и образуют своеобразные зоны. Эти семейства должны обладать рядом общих свойств. В частности, длины их боковых цепей должны быть примерно одного порядка.

Физические операторы, воссоздающие векторы противоположного направления (например: $A_1 - (-^1A)$, $B_1 - (-^1B)$ и т. д.), должны обладать противоположными свойствами. Однако нужно принять во внимание, что наклон оси этих векторов один и тот же, поскольку они направлены в противоположные вершины додекаэдра и вместе составляют его диаметр. По этой причине мы вправе ожидать, что направление ($R_i^{\alpha} - R_i$) обоих физических операторов, воссоздающих векторы одинакового наклона, также должно быть одним и тем же.

Система физических операторов на додекаэдре

Выводы, сделанные ранее на основе анализа MBM, удалось обобщить в виде формальной модели структуры канонического набора физических операторов на додекаэдре (рис. 3.4). Через структуру додекаэдра были проведены три взаимно перпендикулярные плоскости (I, II, III). В отличие от плоскостей симметрии, проведенных через структуру додекаэдра при анализе векторов, на рис. 3.4 они являются плоскостями антисимметрии, поскольку боковые цепи операторов, расположенных относительно них, обладают сходными, но не идентичными свойствами.



Рис. 3.4. Модель структуры канонического набора физических операторов на додекаэдре на основе принципов антисимметрии: I, II, III – плоскости антисимметрии; буквами в кружках обозначены боковые цепи физических операторов

Двадцать физических операторов (они обозначены теми же буквами, что и вершины MBM) помещены в кружки в порядке увеличения размера сверху вниз. В плоскости I находятся 4 оператора (A; B; –A; –B), являющиеся своеобразными начальными структурами, причем оператору A приписываются свойства нейтрального элемента, а оператору (–A) – свойства обратного элемента. Все операторы с аналогичными индексами слева вверху и справа внизу расположены симметрично справа и слева относительно плоскости I. Операторы, расположенные справа от плоскости I (¹A; ²A; –¹A; –²A; ¹B; ²B; –¹B; –²B), должны быть короче, чем операторы с индексами справа внизу (A₁; A₂; –A₁; –A₂; B₁; B₂; –B₁; –B₂). Операторы с близкими размерами, но разными (например, гомологичными) свойствами терминальных групп расположены симметрично сзади и спереди от плоскости II. На рис. 3.4 они образуют две группы над плоскостью III, причем группа с индексом 1 находится за плоскостью II (A₁; B₁; ¹A; ¹B), а группа с индексом 2 – перед плоскостью II (A₂; B₂; ²A; ²B). Аналогичные группы, но со знаком минус и измененными индексами расположены под плоскостью III: группа с индексом 2 находится за плоскостью II (2A ; 2B ; 2A ; 2B), а группа с индексом 1 – перед плоскостью II (^{-1}A ; ^{-1}B ; ^{-1}B ; $^{-1}B_1$).

Операторы с противоположными свойствами расположены поворотно симметрично относительно плоскости III. Наибольшими различиями должны обладать боковые цепи, расположенные в вершинах верхней и нижней граней додекаэдра: $A_1 - (-A_1)$; $A_2 - (-A_2)$; $^1A - (-^1A)$; $^2A - (-^2A)$. Различия боковых цепей, расположенных симметрично в экваториальной области додекаэдра ($B_1 - (-B_1)$; $B_2 - (-B_2)$; $^1B - (-^1B)$; $^2B - (-^2B)$), должны быть не столь существенны.

В целом, построенная на додекаэдре система физических операторов может служить прототипом для создания конкретных систем физических операторов. Именно на ее основе нами была предложена модель структуры канонического набора аминокислот.

Вырожденность свойств физических операторов. Двухъярусная модель MBM

Из 2.4 следует, что 64 конформации ПФ цепного полимера должны иметь для своего воссоздания соответствующие физические операторы. С учетом выделения хотя бы двух триплетов для обозначения конца цепи, этих операторов должно быть ~ 62. Однако, на наш взгляд, это «многовато». В связи с этим нами была разработана двухъярусная модель MBM, в которой количество физических операторов равно 20, а все разнообразие направлений обеспечивается за счет свойств терминальных групп физических операторов.

Вопрос создания такой модели решается достаточно просто. Рассмотрим физический оператор, воссоздающий одну из конформаций посредством образования водородной связи между X_1 и X_{i-4} (рис. 3.5, *a*). Она описывается матрицей, в которой x_1, x_2, x_3 принимают значения 1, 0 и 1, а x_6 принимает значение 0.



Рис. 3.5. Вырожденность состояний физических операторов: a – водородная связь $X_1 - X_{i-4}$; δ – водородная связь $Q_1 - X_{i-4}$

Группа HX₁–R=Q₁, в которую входит X₁, является резонансной, вследствие чего положения двойной связи и атома водорода могут быть иными: X₁=R–Q₁H. Как видно из рис. 3.5, δ , перемещение атома водорода к Q₁ и образование водородной связи с X_{*i*-4} меняет направление вектора на Q₁ и соответственно меняет распределение силовых линий. Появляется дополнительное ребро связности ($x_6 = 1$). Нетрудно догадаться, что если x_5 будет равно нулю, то третья буква в триплетном топологическом коде согласно схеме (2.3) будет К или L, а если $x_5 = 1$, то N или P. Таким образом, вследствие резонанса терминальной группы возникает пучок направлений, которые реализует один и тот же физический оператор. Очевидно, что в зависимости от типа этой терминальной группы количество векторов в пучке (т. е. степень вырожденности состояний этого оператора) будет составлять 1...5 и более. При этом большинство операторов, содержащих на конце резонансные группы, будет, по-видимому, иметь вырожденность, равную двум.

Чтобы описать положение всех векторов, которые могут появиться вследствие вырожденности терминальных групп (т. е. расщепления исходных векторов), модели додекаэдра уже недостаточно. Однако обратим внимание, что общее направление физического оператора в вырожденных состояниях (рис. 3.5) остается неизменным. Вследствие этого многогранник, который подходил бы для этих целей, должен вписываться в структуру додекаэдра и составлять его нижележащий ярус. Одним из таких многогранников в двухъярусной модели МВМ может быть ромбоикосододекаэдр (рис. 3.6).



Рис. 3.6. Ромбоикосододекаэдр как возможный многогранник нижнего яруса молекулярной векторной машины

Напомним, что этот многогранник имеет 60 вершин, 120 ребер и 62 грани, причем 20 граней являются треугольными, 12 – пятиугольными и 30 – четырехугольными. Ромбоикосододекаэдр вписывается в додекаэдр таким образом, что его пятиугольные грани оказываются на гранях додекаэдра. Центры треугольных граней располагаются прямо под вершинами додекаэдра, что хорошо видно на рис. 3.6 для вершин додекаэдра ²A; A₂; ²B; B₂ и –В. Таким образом, центры граней ромбоикосододекаэдра, расположенные вблизи вершин додекаэдра, вполне могут служить для описания положения групп векторов, реализуемых вырожденными состояниями терминальных групп физических операторов.

3.4. Свойства тетраэдрического α-атома

Анализ векторов

Согласно исходным условиям MBM (см. 3.1) атом R_i^{α} , к которому прикрепляются боковые цепи канонического набора физических операторов, должен формировать тетраэдрическую структуру. Четыре его связи направлены в вершины тетраэдра и образованы (рис. 3.7, *a*) атомами Q_i , H_i , R_i , (атомы группы $X_i=R_i-Q_{i+1}H$) и S_i (атом боковой цепи). При этом, как показано на рис. 3.7, *a* и *б*, резонансная группа $X_{i-1}=R_{i-1}-Q_iH$ является плоской (см. 3.1), представляет собой единое



Рис. 3.7. Сопоставление тетраэдров, сформированных на основе R_i^{α} : *а* – валентными связями R_i^{α} , направленными в вершины тетраэдра; δ – векторами, направленными на R_{i+1}^{α} основной цепи и S_i^{α} боковой цепи

целое и может свободно вращаться. Через нее можно провести условную ось $R_{i-1}^{\alpha} - R_i^{\alpha}$, на которой будет вращаться атом R_i^{α} . Поскольку все связи тетраэдрического атома жесткие, то векторы, направленные на (i + 1)-й атом основной цепи (R_{i+1}^{α}) и на атом S_i боковой цепи, будут жестко фиксированы и взаимосвязаны (рис. 3.7, δ). Это означает, что всякое перемещение боковой цепи в процессе воссоздания закодированной конформации полимера будет синхронно отражаться на векторе ($R_i^{\alpha} - R_{i+1}^{\alpha}$). Таким образом, боковые цепи полимера, связанные с тетраэдрическим α -атомом, в рамках модели MBM предопределяют направление роста основной цепи.

Возможные связи группы X_i=R_i-Q_{i+1}H с основной цепью

В зависимости от возможной связи группы $X_i = R_i - Q_{i+1}H$ с основной цепью возможны разные варианты направлений, показанные на рис. 3.8. Для их описания можно использовать двоичную систему: наличие H-связи с атомом Q_i , X_i или Q_{i+1} с образованием циклической конформации ПФ можно обозначить цифрой 1, а отсутствие такой связи – цифрой 0.



Рис. 3.8. Варианты возможной водородной связи группы $X_i = R_i - Q_{i+1}H$: a - c атомом X_{i-3} ; $\delta - c$ атомом X_{i-2}

Если физический оператор S задает направление, при котором возможна водородная связь $Q_{i+1}H$ с атомом X_{i-3} (рис. 3.8, *a*), то циклическая конформация цепи остается неизменной (напомним, что предыдущая водородная связь была $Q_iH...X_{i-4}$). Водородные связи для Q_i и Q_{i+1} принимают значение 1. При этом группа $X_i=R_i-Q_{i+1}H$ ориентируется аналогично группе $X_{i-1}=R_i-Q_iH$ и также задает направление для

роста α -спирали, поэтому возможная водородная связь с X_i также может быть описана цифрой 1. Таким образом, связи аминогруппы и карбонильной группы *i*-го α -атома могут быть заданы парой переменных 11.

Если физический оператор задает направление, допускающее образование водородной связи группы $X_i = R_i - Q_{i+1}H$ с атомом X_{i-2} (рис. 3.8, б), то спиральная структура становится более крутой (аналог спирали 3_{10} в белках). При этом группа $X_i = R_i - Q_{i+1}H$ уже не будет определять рост α -спирали, поэтому связь Q_iH с X_{i-4} можно по-прежнему обозначать цифрой 1, а возможная связь атома X_i будет обозначаться цифрой 0. Таким образом, связи аминогруппы и карбонильной группы *i*-го α -атома в этом случае будут заданы парой переменных 10.

Следует отметить, что ось MBM, установленная в (i + 1)-е положение, идущее вдоль связи $Q_{i+1}H...X_{i-3}$, параллельна оси MBM, идущей вдоль связи $Q_iH...X_{i-4}$. В то же время, ось структуры MBM вдоль связи $Q_{i+1}H...X_{i-2}$ при установке ее в (i + 1)-е положение не является параллельной оси связи $Q_iH...X_{i-4}$. Нижняя ее часть смещается влево, а верхняя – вправо. Как следствие этого, левая нижняя часть вершин додекаэдра не вписывается в структуру такой спирали. Это означает, что структура MBM для такой спирали частично редуцирована, и некоторые физические операторы в этой области встречаться не будут. Это совпадает и с экспериментальными данными по встречаемости боковых цепей аминокислот в спирали 3_{10} [38]. Возможны и другие варианты связи группы $X_i = R_i - Q_{i+1}H$ с основной цепью, однако пока ограничимся этими двумя, поскольку они вполне наглядно демонстрируют связь между типом физического оператора и направлением роста полипептидной цепи.

Формирование подгрупп физических операторов в связи с заданием направлений $\mathbf{R}^{\alpha}_i - \mathbf{R}^{\alpha}_{i+1}$ -связи α-атома

Мы рассмотрели два варианта направлений и два варианта связей, которые могут возникать в процессе работы MBM. Они задаются поворотом $R_i^{\alpha} - R_{i+1}^{\alpha}$ -связи α -атома. Очевидно, что ограниченность числа направлений должна приводить к тому, что в соответствии со способностью определять то или иное направление на R_{i+1}^{α} -атом физические

операторы должны формировать несколько подгрупп. Чисто теоретически решить этот вопрос, на наш взгляд, невозможно, поскольку необходимо анализировать свойства каждого конкретного оператора. Однако некоторые аспекты этой проблемы все же можно обсудить.

Как видно из рис. 3.3, чем дальше находится функциональная группа S физического оператора от плоскостей I и III, тем выше ориентирован вектор связи $R_i^{\alpha} - R_i$, что особенно хорошо заметно при сравнении рис. 3.8, *а* и *б*. Это означает, что для задания направления $Q_{i+1}H...X_{i-3}$ связи, предопределяющей формирование α -спирали, наиболее подходящими являются операторы, расположенные слева, в экваториальной области MBM, т. е. B¹ и (-B¹). Если учесть, что операторы-антиподы (например, B₁ – (-B¹) и B¹ – (-B₁)), по-видимому, должны задавать одинаковое направление (см. 3.3), то вот уже и определилась одна из возможных подгрупп, Она расположена в экваториальной области додекаэдра (подгруппа из 8 операторов B с индексами).

Подгруппой физических операторов, задающих более крутую спираль (связь $Q_{i+1}H...X_{i-2}$), могут быть операторы, приближенные к вертикальной плоскости I, проходящей через центр MBM (см. рис. 3.8, б). В эту подгруппу будут входить все 8 операторов, воссоздающих векторы, связанные преобразованиями симметрии, т. е. ${}^{1}A - A_{1}$; $(-{}^{1}A) - (-A_{1})$; ${}^{2}A - A_{2}$; $(-{}^{2}A) - (-A_{2})$. Аналогичным образом можно рассмотреть и другие подгруппы операторов, однако, как отмечалось, более целесообразно проводить такое рассмотрение на экспериментальной основе.

Модель МВМ как основа систематизации функциональных групп физических операторов

Резонансная группа $HQ_{i-3}-R_{i-3}=X_{i-4}$, из атома X_{i-4} которой исходят все векторы MBM, может входить во вторичной структуре цепного полимера в состав ССИВС. Вследствие этого структура MBM может служить для разработки системы терминальных групп, реализующих действие векторов. При этом векторы можно рассматривать не только как направления действия боковой цепи при ко-трансляционном сворачивании полимера, но и как «сигналы», направленные из ССИВС полимера на входы этих групп (последние будут выступать в качестве логи-

ческих модулей ССИВС). Мы уже рассматривали с этих позиций боковые цепи аминокислот (см. 1.2). Использование МВМ, в которой заложены отношения симметрии векторов и антисимметрии боковых цепей, позволяет провести такой анализ более углубленно. Целесообразнее провести такую работу на конкретных боковых цепях.

Выполненный в данной главе теоретический анализ позволяет ответить на следующие вопросы, поставленные в гл. 1:

• Почему все живые существа используют оптические изомеры биомолекул только одного типа (например, L-аминокислоты)?

• Почему в состав канонического набора входит лишь 20 аминокислот и почему представлены именно эти аминокислоты, а не другие?

• Какова природа минорных оснований в т-РНК?

Ответы на эти вопросы будут даны во второй части данного издания.

Часть 2. ПРИМЕНЕНИЕ ТЕОРИИ ТОПОЛОГИЧЕСКОГО КОДИРОВАНИЯ К БИОСТРУКТУРАМ

Глава 4. ГЕНЕТИЧЕСКИЙ КОД И ПРОБЛЕМА СООТВЕТСТВИЯ «ТРИПЛЕТ – АМИНОКИСЛОТА»

4.1. Исследование пространственной структуры генетического кода

Современная таблица генетического кода

Под биоструктурами понимаются не только белки и нуклеиновые кислоты, являющиеся частным случаем цепных полимеров, но и такие структуры, как генетический код и канонический набор аминокислот, система минорных оснований транспортных РНК и др. «Формализм» нашей теории вполне позволяет рассмотреть перечисленные структуры.

Проблема поиска идеальной формы для генетического кода не нова. Результаты первых попыток, предпринятых вскоре после экспериментальной расшифровки соответствий «триплет – аминокислота», приведены в [40]. В [41] была предложена форма представления кода, которая ныне является общепринятой (рис. 4.1). В ней буквы триплетов выписаны для обозначения строк (первая буква), столбцов (вторая буква) и строк внутри горизонтальных клеток (третьи буквы), а соответствующие им аминокислоты находятся в клетках. При этом буквы расположены в последовательности U, C, A, G.

Такая последовательность букв, по мнению автора, наиболее полно выявляет особенности генетического кода [41]. Было отмечено, что столбцам, кодируемым триплетами с буквами U и C (вторая буква триплета), соответствуют главным образом гидрофобные (Phe, Leu, Ile, Val, Ala, Pro) и слабополярные (Ser, Thr) аминокислоты, в то время как столбцам, кодируемым триплетами с буквами A и G, – сильнополярные аминокислоты (Asp, Glu, Asn, Gln, His,Arg, Lys). Третье основание несущественно влияет на значение кодируемого остатка (вырожденность третьего основания триплетов). Замена основания в первом положении по сути не изменяет характер аминокислоты (позднее это свойство было названо связностью аминокислотных серий [42]). Было также уста-

2	U	С	А	G	3
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STP	STP	A
	Leu	Ser	STP	Trp	G
С	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
А	Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Рис. 4.1. Общепринятая форма представления генетического кода (по [41])

новлено, что замена в одном основании триплета часто коррелирует с минимальными изменениями в боковой цепи аминокислот. Такая корреляция, по мнению автора [41], не является случайной.

Таблица с иной последовательностью букв (A, C, G, U) была предложена в [43]. По мнению автора, такая последовательность лучше выявляет принципы генетического кода. Существуют и иные формы представления кода: круг [42], кольцо [44] и калиграмма [45], однако они практически ничего не добавляют к раскрытию структуры кода.

Развитие представлений о пространственной структуре дуплетного генетического кода

Двумерная структура кода. Возможность изучения структуры дуплетного генетического кода основана на вырожденности третьего основания в триплетах. Первоначально эти исследования были сосредоточены на рассмотрении двумерной структуры кода. Согласно [46] и [47] в каждом триплете можно выделить двухбуквенный корень (дуплет) и однобуквенное окончание (флексию). Шестнадцать дуплетов автор разделил на два октета [46], [47]. Восемь дуплетов кодирует по одной аминокислоте (GG, GU, AC, CU, UC, CG, GC, CC) и восемь дуплетов – по две аминокислоты (UU, UG, CA, AG, GA, AU, UA, AA). Взаимный переход этих двух множеств подчиняется правилу: C $\leftarrow \rightarrow$ A, G $\leftarrow \rightarrow$ U («правило Румера» [46], [47]), например: GG $\leftarrow \rightarrow$ UU, CG $\leftarrow \rightarrow$ AU, CC $\leftarrow \rightarrow$ AA. Матрицу дуплетов, предложенную в [46] на основе канонической последовательности оснований C, G, U, A, можно представить в следующем виде:

$$CC CG CU | CA$$

$$GC GG GU | GA
$$UC | UG UU UA$$

$$AC | AG AU AA$$

$$(4.1)$$$$

Эта матрица может быть переписана в ромбической форме:

CC

В верхней части ромба находятся дуплеты, кодирующие по одной аминокислоте (сильные), в нижней части – дуплеты, кодирующие по две аминокислоты (слабые). На диагонали находятся переходные дуплеты. При другой (не канонической) последовательности букв сильные и слабые дуплеты перемешиваются между собой.

Результаты, полученные в [46] и [47], в свое время «не были замечены» и были «как бы заново открыты» через 6–7 лет. Так, в [48] было предложено два субкаталога: для дуплетов монокодонов, кодирующих по одной аминокислоте, и для дуплетов гетерокодонов, кодирующих по две аминокислоты. Детальный анализ симметрии дуплетов в генетическом коде был проведен в [49]. Авторы выделили два множества дуплетов: М₁ и М₂. Дуплеты множества М₁ принадлежат к четырежды вырожденным триплетам, а дуплеты множества М₂ – к дважды вырожденным триплетам. Буквы в дуплетах расположены по вертикали – в последовательности C, U, G, A, а по горизонтали – в последовательности C, A, G, U. Чтобы исследовать структуру множеств M_1 и M_2 , авторы ввели операторы обмена дублетов (doublet exchange operators) σ_i и σ_j , которые преобразуют четыре основания A, C, U, G в парах:



Биохимически, по мнению авторов [49], операторы σ_i имеют следующее значения: $\sigma_1 = 1$ – идентичность; $\sigma_2 = \alpha$ – трансверсия между некомплементарными основаниями; $\sigma_3 = \beta$ – трансверсия между комплементарными основаниями; $\sigma_4 = \gamma$ – транзиция «пурин $\leftarrow \rightarrow$ другой пурин, пиримидин $\leftarrow \rightarrow$ другой пиримидин».

Математически σ_i образует Абелеву группу (группу Клейна-4):

$$1^2 = \alpha^2 = \beta^2 = \gamma^2 = 1. \tag{4.4}$$

Оператор σ_i действует на первое основание дуплета, а σ_j – на второе. С помощью операторов (α , α) и (γ , α) два октета дуплетов были преобразованы друг в друга.

В [50] и [51] был предложен ромбический вариант генетического словаря, основанный на комплементарности кодирующих дуплетов, который наглядно представил найденные в [46] и [47] закономерности (рис. 4.2). В качестве основы для построения данной структуры были использованы две пары дуплетов, совпадающие с каноническим порядком оснований [46] и подчиняющиеся правилам комплементарности Уотсона–Крика (С $\leftarrow \rightarrow$ G, U $\leftarrow \rightarrow$ A).

Пары дуплетов расположены на вертикальной диагонали ромба. Остальные дуплеты были получены комбинацией исходных пар. В таблице дуплеты находятся вместе с кодируемыми аминокислотами, что позволяет проводить их совместный анализ [50], [51]. Так, на рис. 4.2 четко видно формирование групп связности по второму основанию (ряд дуплетов с U во втором положении – неполярные аминокислоты, с А – полярные аминокислоты и т. д.).



Рис. 4.2. Ромбический вариант генетического словаря на основе комплементарности кодирующих дуплетов [50], [51]

По углам ромба расположены квартеты дуплетов, связанные параллельной комплементарностью, например: $CC \leftarrow \rightarrow GG$, $GC \leftarrow \rightarrow CG$ и т. д. (отмечены перекрестьями). Дуплеты, у которых во втором положении находятся комплементарные азотистые основания, образуют две замкнутые группировки (разделены штриховой линией). Кроме того, как видно из рис. 4.2, имеется две группы дуплетов, кодирующих по одной и по две аминокислоты (разделены сплошной линией). Они преобразуются друг в друга согласно правилу Румера ($C \leftarrow \rightarrow A$, $G \leftarrow \rightarrow U$) и связаны симметрией C_2 : $AC \leftarrow \rightarrow CA$, $UC \leftarrow \rightarrow GA$, $GC \leftarrow \rightarrow UA$ и т. д. Традиционный вариант таблицы кода с каноническим порядком оснований был предложен в [52].

Пространственная структура дуплетного кода. Дальнейший этап в исследовании генетического кода связан с пространственной интерпретацией дуплетного кода. Для характеристики взаимоотношений между дуплетами в [53] были использованы группа Клейна-4 и операторы обмена между основаниями. Вместо табличной формы, описанной в [49], авторы предложили использовать граф группы – четырехмерный гиперкуб (тессеракт), который наглядно представил трансверсии между некомплементарными основаниями (α) и комплементарными основаниями (β) (рис. 4.3).



Рис. 4.3. Пространственный граф дуплетного генетического кода [36]: α – переход между некомплементарными основаниями (С $\leftarrow \rightarrow$ A, G $\leftarrow \rightarrow$ U); β – переход между комплементарными основаниями (С $\leftarrow \rightarrow$ G, U $\leftarrow \rightarrow$ A)

При этом дуплеты, кодирующие по одной аминокислоте, занимают в гиперкубе компактное положение. Авторы этой работы пришли к выводу о важности «взаимоотношений» между алгебраической структурой кода и его биологической структурой. Позднее в [54] была предложена иная версия структуры дуплетного кода. В отличие от авторов [53], рассматривающих в качестве основных переходов α - и β -трансверсии, авторы [54] вводят β - и γ -переходы, поскольку α -оператор производит двубитовые изменения, а β - и γ -операторы осуществляют только однобитовые изменения (см. далее). Следует отметить, однако, что оба варианта имеют дело только с дуплетами оснований, но не касаются свойств кодируемых ими аминокислот.

Несколько иной подход к построению пространственной структуры дуплетного кода был предложен в [51]. В соответствии с топологическим понятием «окрестности точки» [55] положение данной точки среди множества других точек определяется положением ее ближайших соседей. Если множество состоит из четырех точек (азотистых оснований), то каждая из них определяется через отношение с тремя другими. Авторы представили это положение в виде трехмерного графа (тетраэдра) с азотистыми основаниями в его вершинах. Этот граф можно изобразить в виде двух диастереоизомеров: L и D. В первом графе каноническая последовательность C, G, U, A записана против часовой стрелки, а во втором – по часовой стрелке:



Как показано на этой схеме, азотистые основания могут претерпевать следующие преобразования: С $\leftarrow \rightarrow$ G, U $\leftarrow \rightarrow$ A – переходы между взаимно комплементарными пуринами и пиримидинами (сплошные линии); С $\leftarrow \rightarrow$ U, G $\leftarrow \rightarrow$ A – переходы между взаимно гомологичными основаниями (частый штрих); С $\leftarrow \rightarrow$ A, G $\leftarrow \rightarrow$ U – переходы между некомплементарными основаниями (редкий штрих). Они соответствуют β-, γ- и α-операциям, введенным в [49].

Структуру генетического кода можно также построить в виде двух вариантов (диастереоизомеров). Авторы [51] предложили структуру, основанную на одном из стереоизомеров (схема 4.5, *a*). При этом была использована последовательность оснований C, G, U, A, которая сохраняет симметрию C₂ между подмножествами M₁ и M₂, а также минимум структурных изменений в структуре аминокислот в случае переходов между ближайшими дуплетами в ромбическом варианте генетического словаря [50], [51]. В соответствии с понятием «окрестность точки» каждый дуплет может быть связан единичным переходом с шестью ближайшими дуплетами: по три связи для первого и второго оснований:

67



В результате последовательного построения авторы [51] получили структуру (рис. 4.4), изоморфную булеву гиперкубу В⁴ [32]. Она является шестимерным симплексом, поскольку помимо основных переходов, характерных для гиперкуба В⁴, в ней наложены переходы между вершинами, расположенными по диагоналям параллелограммов.



Рис. 4.4. Пространственная структура дуплетного генетического кода, изоморфная булеву гиперкубу В⁴ [51]. Сплошные линии – переходы С ←→ G, U ←→ A (β);частый штрих – переходы С ←→ U, G ←→ A (γ); редкий штрих – С ←→ A, G ←→ U (α)

Как следует из рис. 4.4, эта структура состоит из двух групп дуплетов: восемь дуплетов кодирует по одной аминокислоте (на рисунке кружки показаны утолщенными линиями, а буквы выделены жирным шрифтом), и еще восемь дуплетов кодирует по две аминокислоты. Основания, связанные преобразованием Румера (С $\leftarrow \rightarrow$ A, G $\leftarrow \rightarrow$ U), занимают на плоском изображении гиперкуба симметричные положения (группа симметрии C_2), например: CC $\leftarrow \rightarrow$ AA, GC $\leftarrow \rightarrow$ UA, AC $\leftarrow \rightarrow$ CA и т. д. Как и на других вариантах четырехмерного представления дуплетного кода [53], [54], на этом варианте структуры показаны элементы, связанные единичными переходами, отличающимися на один бит информации: сплошные линии (β) и частый штрих (γ). В дополнение к этому на рис. 4.4 показан и третий тип переходов – редкий штрих (α). Таким образом, в отличие от предыдущих структур [53], [54], на данной структуре наглядно представлены все три типа возможных переходов азотистых оснований. В данной структуре видно, что квартеты дуплетов объединяют структурно близкие аминокислоты. Так, большая часть аминокислот, кодируемая дуплетами AC, UC, UG и AG, содержит С-ОН- и С-SH-группы; дуплеты GG, CG, UG и AG включают все известные кодоны, кодирующие аргинин; дуплеты GU, CU, UU и AU кодируют неполярные аминокислоты; а дуплеты AA, GA, CA и UA – только полярные. Предложенная структура показывает, что эти квартеты дуплетов формируют замкнутые циклы.

Пространственная структура триплетного генетического кода

Первые попытки предложить пространственную интерпретацию триплетного кода в виде икосаэдра [56] и тетраэдра [57] оказались неудачными, поскольку в них отсутствовала однозначная связь между элементами кода и элементами предложенной геометрической структуры. Прогресс в исследовании пространственной структуры триплетного кода был достигнут лишь в конце XX в. [51], [54], [58], [59].

Согласно [59] существующая таблица генетического кода (см. рис. 4.1) позволяет легко и быстро находить аминокислоты, соответствующие тому или иному триплету, однако она не дает пространственного представления о триплетах и их взаимосвязях, поскольку является плоской, а триплеты заданы в ней в неявном виде. Чтобы получить пространственное представление, автор [59] предложил идею «последовательных пространств». Для этого пуриновые основания (А или G) были обозначены буквой R, а пиримидиновые (C, U или T) – буквой Y. В результате вместо 64 триплетов была получена группа из 8 триплетов, в которой вместо четырех азотистых оснований фигурируют буквы R и Y: RRR, RRY, RYR, YRR, RYY, YRY, YYR, YYY. Эти триплеты были расположены в вершинах большого куба (на рис. 4.5 они выделены жирным шрифтом – первое трехмерное пространство).



Рис. 4.5. Пространственная структура триплетного генетического кода [59]

Если теперь придать каждому триплету свое реальное значение (например, для RRR: GGG, GGA, GAG, AGG, GAA, AGA, AAG, AAA), то для каждой вершины возникает еще одно трехмерное пространство. Таким образом, общее количество измерений в этом гиперкубе составляет 6: 3 измерения в большом кубе и 3 – в малых кубах. В результате, все 64 триплета расположились в 8 малых кубах.

Иной подход к пространственному представлению триплетного кода был предложен авторами [54] и [58]. По их мнению, возможна бинарная интерпретация генетического кода как кода Грея. В рамках этой интерпретации с генетическим кодом может быть связано множество различных кодов Грея, в зависимости от порядка выделения одного бита в кодовом слове. Для обоснования своего выбора авторы анализировали значение химического типа азотистых оснований, обозначив буквой R – A, G – пурины, а буквой Y – C, U – пиримидины и характер водородных связей: W (слабый (weak) – A, U), две водородные связи и S (сильный (strong) – G, C), три водородные связи в парах Уотсона-Крика (схема 4.7). Они приняли, что первый бит в бинарном кодировании – это тип химической связи (R – 0, Y – 1), а второй бит – характер водородной связи (W - 0, S - 1), и приписали следующие соответствия: A = 00, G = 01, U = 10, C = 11. Следует отметить, что ранее предлагались и другие варианты соответствий букв триплетов и бинарных пар [44], [53], [60].



Пространственным представлением кода из 6 переменных является булев гиперкуб В⁶. Используя выбранные соответствия, авторы [54] и [58] трансформировали гиперкуб в триплетный генетический код (рис. 4.6, a, δ).

Вершины гиперкуба были обозначены соответствующими кодонами (рис. 4.6, *a*). Расположение аминокислот с использованием однобуквенных обозначений было показано авторами на другой структуре (рис. 4.6, δ). Для наглядности мы заменили однобуквенные обозначения, примененные в [54], на трехбуквенные.

Авторы подчеркивают принципиальную важность единичных изменений азотистых оснований на гиперкубе, поскольку они учитывают двойственный характер каждого основания. Следует отметить, что



Рис. 4.6. Пространственная структура триплетного генетического кода [54], [58]: *а* – расположение триплетов в вершинах гиперкуба; *б* – положение аминокислот, соответствующих этим триплетам
на этой структуре, в отличие от [59], довольно четко просматривается возможность перехода триплетов из одной группы в другую по правилу Румера. Одна часть триплетов (до преобразования) показана обычным шрифтом (рис. 4.6, *a*), а другая часть (после преобразования) выделена жирным шрифтом. Надо отдать должное авторам [54] и [58]: приоритет в построении пространственной структуры генетического кода принадлежит именно им. Однако нами такая структура была предложена на принципиально иных основаниях.

4.2. Топологические основания триплетного генетического кода

Трансформация суперматрицы четырехзвенного графа в триплетный генетической код

В гл. 2 было показано, что в основание топологического кода может быть положено описание конформаций четырехзвенного цепного графа – аналога пентафрагмента линейного цепного полимера. Эти конформации были описаны с помощью верхних треугольных матриц из 6 переменных и составили суперматрицу из 64 элементов. В 2.3 эта суперматрица с помощью четырех букв латинского алфавита – K, L, N и P – была трансформирована в триплетный топологический код. Очевидно, что вместо приведенных выше букв можно использовать буквы C, G, U, A:

$$C - 00 \quad G - 10 \quad U - 01 \quad A - 11.$$
 (4.8)

Используя эти соответствия, мы трансформировали суперматрицу, описывающую конформации четырехзвенного графа (см. рис. 3.3), в триплетный генетический код (рис. 4.7) [1], [20], [24]–[30]. В клетки таблицы были вписаны также кодируемые триплетами аминокислоты.

Видно, что вторые пары состояний, общие для каждого из четырех блоков, трансформировались во вторую букву триплетов в последовательности C, U, G, A. Такой же порядок имеют третьи буквы триплетов, в то время как первые буквы, кодирующие симметричные третьим парам состояния, расположены в ином порядке: C, G, U, A, что обусловлено данными правилами преобразования. Отметим также, что в результате проведенной на основе схемы (4.8) трансформации была получена характерная для генетического кода симметрия C_2 [46], [47] двух групп кодонов: квартетов триплетов, кодирующих по одной

2		$\mathbf{C} \leftrightarrow$	00		U ↔ 01				
13	$C \leftrightarrow 00$	$C \leftrightarrow 00 U \leftrightarrow 01$		A⇔11	$C \leftrightarrow 00$	$U \leftrightarrow 01$	$G \leftrightarrow 10$	A⇔11	
00 ↓ C	0 0 0 0 0 CCC 0 Pro	0 0 0 0 0 CCU 1 Pro	0 0 0 0 1 CCG 0 Pro	0 0 0 0 1 CCA 1 Pro	0 0 0 1 0 CUC 0 Leu	0 0 0 1 0 CUU 1 Leu	0 0 0 1 1 CUG 0 Leu	0 0 0 1 1 CUA 1 Leu	
10 ↓ G	1 0 0 0 0 GCC 0 Ala	1 0 0 0 0 GCU 1 Ala	1 0 0 0 1 GCG 0 Ala	1 0 0 0 1 GCA 1 Ala	1 0 0 1 0 GUC 0 Val	1 0 0 1 0 GUU 1 Val	1 0 0 1 1 GUG 0 Val	1 0 0 1 1 GUA 1 Val	
01 ↓ U	0 1 0 0 0 UCC 0 Ser	0 1 0 0 0 UCU 1 Ser	0 1 0 0 1 UCG 0 Ser	0 1 0 0 1 UCA 1 Ser	0 1 0 1 0 UUC 0 Phe	0 1 0 1 0 UUU 1 Phe	0 1 0 1 1 UUG 0 Leu	0 1 0 1 1 UUA 1 Leu	
11 ↓ A	1 1 0 0 0 ACC 0 Thr	1 1 0 0 0 ACU 1 Thr	1 1 0 0 1 ACG 0 Thr	1 1 0 0 1 ACA 1 Thr	1 1 0 1 0 AUC 0 Ile	1 1 0 1 0 AUU 1 Ile	1 1 0 1 1 AUG 0 Met	1 1 0 1 1 AUA 1 Ile	
00 ↓ C	0 0 1 0 0 CGC 0 Arg	0 0 1 0 0 CGU 1 Arg	0 0 1 0 1 CGG 0 Arg	0 0 1 0 1 CGA 1 Arg	0 0 1 1 0 CAC 0 His	0 0 1 1 0 CAU 1 His	0 0 1 1 1 CAG 0 Gln	0 0 1 1 1 CAA 1 Gln	
10 ↓ G	$\begin{array}{ccc} 1 & 0 & 1 \\ GGC & 0 \\ Gly \end{array}$	1 0 1 GGU 0 Gly	1 0 1 GGG ⁰ 1 Gly	1 0 1 GGA 1 Gly	1 0 1 GAC ^{1 0} Asp	1 0 1 GAU ¹ 0 Asp	$\begin{array}{ccc}1&0&1\\GAG&1&1\\Glu&0\\Glu&\end{array}$	1 0 1 GAA 1 1 Glu	
01 ↓ U	0 1 1 0 0 UGC 0 Cys	0 1 1 0 0 UGU 1 Cys	0 1 1 0 1 UGG 0 Trp	0 1 1 0 1 UGA 1 T	0 1 1 1 0 UAC 0 Tyr	0 1 1 1 0 UAU 1 Tyr	0 1 1 1 1 UAG 0 T	0 1 1 1 1 UAA 1 T	
11 ↓ A	1 1 1 0 0 AGC 0 Ser	1 1 1 0 0 AGU 1 Ser	1 1 1 0 1 AGG 0 Arg	1 1 1 0 1 AGA 1 Arg	$\begin{array}{ccc} 1 & 1 & 1 \\ & 1 & 0 \\ AAC & 0 \\ Asn \end{array}$	1 1 1 1 0 AAU 1 Asn	1 1 1 AAG 0 Lys	1 1 1 AAA 1 Lys	
2		$\mathbf{G} \leftrightarrow$	10			A↔	11		

Рис. 4.7. Трансформация суперматрицы, описывающей конформации четырехзвенного графа, в триплетный генетический код

аминокислоте (выделены жирным шрифтом) и по две. Они отделены друг от друга утолщенной линией. Такая симметрия реализуется еще лишь в одном случае [54], когда имеют место соответствия 11 – C, 01 – G, 10 – U, 00 – A, но они в рамках нашего подхода входят в противоречие со свойствами кодируемых аминокислот.

Поскольку пространственной структурой суперматрицы триплетного топологического кода является булев гиперкуб B^6 (см. рис. 2.8), то триплетный генетический код, полученный на основе соответствий схемы (4.8), должен иметь структуру, изоморфную этому гиперкубу. Таким образом, к описанной ниже форме представления генетического кода можно прийти, исходя из изложенной в гл. 2 общей теории топологического кодирования цепных полимеров.

Модель пространственной структуры триплетного генетического кода

Построение пространственной структуры генетического кода в форме гиперкуба В⁶ [1], [20], [24]–[30] проводилось в соответствии с понятием «окрестность точки», примененным ранее к дуплетному генетическому коду [27], [51].

Процесс построения привел к форме гиперкуба, описанной в [32]. Отметим, однако, что при построении структуры суперматрицы на рис. 2.5 были изменены положения начальных вершин с нулевыми переменными: в работе [32] они находятся внизу, а на рис. 2.5 – вверху, что обеспечивает бо́льшую компактность групп вершин, обладающих рядом общих свойств. При последующих трансформациях эти изменения были сохранены (см. рис. 2.6 и 2.8). Эта версия структуры генетического кода (рис. 4.8 и 4.9) существенно сложнее структур, предложенных авторами [54], [58], [59] (см. рис. 4.5, 4.6). Рассмотрим ее более подробно.

В 64 вершинах гиперкуба были расположены как триплеты, так и кодируемые ими аминокислоты. Связь между триплетами была задана единичными переходами оснований, благодаря которым ближайшие соседние триплеты отличаются друг от друга на одно основание. На рис. 4.8 показаны единичные переходы оснований типа β и γ , а на рис. 4.9 – переходы типа α^* . В них сохраняются основные свойства булева гиперкуба – ярусность и иерархичность (см. 2.2).

Триплеты образуют семь ярусов, из которых два содержат по 1 триплету, два – по 6, два – по 15 и один (центральный) – 20 триплетов. Структура кода является иерархической и распадается на два мно-

^{*} Такой иллюстративный прием применен в связи с технической трудностью четко показать все переходы на одном рисунке.

жества (M_1 и M_2) по 32 триплета и четыре подмножества (SM_1 – SM_4) по 16 триплетов и 8 октетов (O_1 – O_8).



Рис. 4.8. Пространственная структура триплетного генетического кода [27], [51]. Сплошные линии – единичные переходы типа β (С $\leftarrow \rightarrow$ G, U $\leftarrow \rightarrow$ A); штриховые линии – переходы типа γ (С $\leftarrow \rightarrow$ U, G $\leftarrow \rightarrow$ A). Красным цветом показаны квартеты триплетов, кодирующие по одной аминокислоте, синим – по две аминокислоты

Базируясь на канонической последовательности азотистых оснований, такая структура выявляет разделение аминокислот на две группы. Восемь квартетов триплетов в верхней части структуры (триплеты выделены жирным шрифтом и красными линиями) кодируют по одной аминокислоте и еще восемь квартетов – по два значения (показаны синими линиями). Каждый триплет верхней группы связан с триплетом нижней группы симметрией C_2 и преобразованием Румера (С $\leftarrow \rightarrow$ A, G $\leftarrow \rightarrow$ U).



Рис. 4.9. Пространственная структура триплетного генетического кода [51]. Сплошные линии – единичные переходы типа α (С $\leftarrow \rightarrow$ A, G $\leftarrow \rightarrow$ U)

Таким образом, в настоящее время имеется три варианта структуры генетического кода, изоморфных булеву гиперкубу В⁶ [51], [54], [59] и построенных, исходя из разных посылок. Это дает уверенность в том, что именно данная структура должна соответствовать реальной пространственной структуре триплетного генетического кода.

4.3. Триплеты генетического кода и боковые цепи аминокислот

Положение боковых цепей аминокислот в структуре триплетного генетического кода

Как отмечалось в 2.4, боковые цепи полимера могут вести себя как физические операторы, воздействие которых на область формирования водородной связи может обеспечить воссоздание закодированной структуры. Было выделено два типа таких операторов: операторы связности, способствующие формированию циклических структур, и операторы антисвязности, препятствующие этому процессу. Нетрудно видеть, что если на рис. 2.9, δ символы Q, R, X в боковой цепи заменить на атомы O, C и N, то получится боковая цепь аспарагина, выступающего в роли оператора связности. В то же время, если серые шары в боковой цепи на рис. 2.9, ϵ считать атомами C, то получится боковая цепь валина, играющего роль оператора антисвязности. Таким образом можно легко перейти от общих представлений о роли физических операторов в системе топологического кодирования цепных полимеров к роли аминокислот в белках.

2		$C \leftrightarrow$	00		$\mathbf{U} \leftrightarrow 01$				
13	$C \leftrightarrow 00$	$U \leftrightarrow 01$	$G \leftrightarrow 10$	$A \leftrightarrow 11$	$C \leftrightarrow 00$	$U \leftrightarrow 01$	$G \leftrightarrow 10$	$A \leftrightarrow 11$	
00	0 0 0	0 0 0	0 0 0	000	$ \begin{array}{c} 0 & 0 & 0 \\ & 1 & 0 \end{array} $	$ \begin{array}{c} 0 & 0 & 0 \\ & 1 & 0 \end{array} $	000	$ \begin{array}{c} 0 & 0 & 0 \\ & 1 & 1 \end{array} $	
$\begin{vmatrix} \uparrow \\ \mathbf{C} \end{vmatrix}$	CCC 0	CCU ¹	CCG Ô	CCA 1	CUC ⁰	CUU ¹ 1	CUG ¹ 0	CUA ¹	
10	1 0 0	1 0 0	100	100	1 0 0	1 0 0	1 0 0	1 0 0	
10 \$	$GCC \stackrel{0}{} \stackrel{0}{} \stackrel{0}{} \stackrel{0}{} \stackrel{0}{}$	$\begin{array}{c} 0 & 0\\ GCU & 1 \end{array}$	$\begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 $	$\begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 $	$\begin{array}{c} 1 & 0 \\ \text{GUC} & 0 \end{array}$	$\begin{array}{c} 1 & 0 \\ \text{GUU} & 1 \end{array}$	$\begin{array}{c}1 \\ GUG \end{array}$	$\begin{array}{c}1 \\ \text{GUA}\end{array}$	
G	Ala	Ala	Ala	Ala	Val	Val	Val	Val	
01 ↑	$ \begin{array}{cccc} 0 & 1 & 0 \\ & 0 & 0 \end{array} $	$ \begin{array}{c} 0 \ 1 \ 0 \\ 0 \ 0 \end{array} $	$ \begin{bmatrix} 0 & 1 & 0 \\ & 0 & 1 \end{bmatrix} $	$ \begin{bmatrix} 0 & 1 & 0 \\ & 0 & 1 \end{bmatrix} $	$ \begin{array}{ccc} 0 & 1 & 0 \\ & 1 & 0 \end{array} $	$ \begin{array}{c} 0 \ 1 \ 0 \\ 1 \ 0 \end{array} $	$ \begin{array}{c} 0 \ 1 \ 0 \\ 1 \ 1 \end{array} $	$ \begin{array}{cccc} 0 & 1 & 0 \\ & 1 & 1 \end{array} $	
↓ U	UCC 0 Ser	UCU 1 Ser	UCG 0 Ser	UCA 1 Ser	UUC 0 Phe	UUU 1 Phe	UUG 0	UUA 1	
11		1 1 0			1 1 0	1 1 0	1 1 0	1 1 0	
\$	ACC 0 0	ACU 1	ACG^{0}	ACA 1	AUC_{0}^{1}	AUU ¹ 1	AUG ¹ 0	AUA ¹ 1	
A	Thr	Thr	Thr	Thr	Ile	Ile	Met	Ile	
00 ↑							001		
Ċ	CGC 0	CGU / I	CGG 0	CGA I	CAC 0 His	CAU 1 His	CAG 0 Gln	Gln	
10	1/1/0/1	/10/1	101	101	101	101	101	101	
10	GGC ⁰ 0	GGU^{0}	GGG^{0}_{0}	GGA 1	GAC^{1} 0	GAU ¹	GAG^{1}_{0}	GAAL	
G	Gly///	Gly//	Gly//	Gly//	Asp///	Asp //	Glu //	Glu	
01	$011 \\ 00$	$011 \\ 00$			$011 \\ 10$		$\binom{0}{1}_{1}^{1}_{1}^{1}_{1}$	$\begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\$	
$ \stackrel{\uparrow}{\mathbf{U}} $	UGC/0	UGU/1	UGG 0	UGA 1	UAC 0	UAU 1	UAG 0	UAA 1	
11	111	$\frac{Cys}{1/1/1}$	110	VIII	$\frac{\gamma_{1}}{111}$	191	1/1/1	1/1/1	
\uparrow	AGC	AGU	AGG	AGA	AAC	AAU	AAG	AAA	
Α	Ser///	Ser///	Arg//	Arg//	Asn//	Asn///	Lys///	Lys/	
2		$\mathbf{G} \leftrightarrow$	10			$\mathbf{A} \leftrightarrow$	11		

Рис. 4.10. Положение аминокислот, соответствующих операторам связности (заштрихованы) и операторам антисвязности в генетическом коде

Перенося эти выводы на генетический код, можно утверждать, что боковые цепи аминокислот, рассматриваемые в качестве операторов связности^{*}, должны воссоздавать циклические конформации ПФ белка. В матрицах, описывающих эти конформации и закодированных триплетами, $x_3 = 1$, поэтому такие аминокислоты должны быть приписаны к блокам триплетов, кодируемых буквами G и A таблицы генетического кода. В то же время, операторы антисвязности (неполярные боковые цепи, не способные образовывать водородные связи) должны соответствовать блокам C и U (в матрицах переменная $x_3 = 0$). Как видно из рис. 4.10, блокам G и A таблицы генетического кода соответствуют преимущественно полярные аминокислоты, способные выполнять функции операторов связности, а блокам C и U – неполярные и слабополярные аминокислоты, пригодные на роль операторов антисвязности [27].

Боковые цепи аминокислот и требования, предъявляемые к физическим операторам

Согласно требованиям, изложенным в 2.4, набор физических операторов должен обладать хиральностью, т. е. единым типом стереоконфигурации. Действие операторов должно быть направлено назад, на уже синтезированную структуру (ретродействие). Размер операторов должен быть согласован с пространственными параметрами связи, на которую они действуют. Рассмотрим с этих позиций боковые цепи аминокислот.

Все боковые цепи аминокислот белков, как известно, относятся к L-ряду и являются хирально чистыми. Это означает, что они вполне соответствуют требованию хиральной чистоты физических операторов. С этих позиций находит объяснение трудно объяснимый до настоящего времени факт хиральной чистоты аминокислот в белках [38].

Ретродействие боковых цепей аминокислот как физических операторов было специально проанализировано нами в [27]. Если справедливо наше требование, то в спиральных фрагментах белков боковые цепи полярных аминокислот должны часто образовывать водородные связи с О_{*i*-4}=С. В результате анализа более 50 белковых струк-

^{*} Они должны быть способны образовывать водородные связи с основной цепью.

тур были найдены циклические фрагменты, в которых боковые цепи аминокислот образуют водородные связи с основной цепью (табл. 4.1). Конформации этих аминокислот изображены в том виде, в котором они были найдены в белках. Все они, как видно из табл. 4.1, образуют водородные связи с группой O_{i-4} =С. Количество таких циклов составляет более 90 % от их общего числа.

Таблица 4.1



Боковые цепи полярных аминокислот белков как возможные физические операторы связности Из табл. 4.1 видно, что размер полярных аминокислот не превышает размера области действия физических операторов, поскольку все они образуют водородные связи с основной цепью. Из приведенных аминокислот только Тур обладает некоторыми особенностями. Вследствие массивности его боковой цепи образование водородной связи с O_{i-4} =С приводит к разрыву водородной связи внутри цикла (табл. 4.1).

С позиции нашего требования были рассмотрены также боковые цепи неполярных аминокислот [27]. Поскольку действие операторов антисвязности никак не может быть зафиксировано, то мы представили их гипотетическое действие, изобразив боковые цепи с сохранением масштаба (табл. 4.2).

Таблица 4.2



Боковые цепи неполярных аминокислот как возможные операторы антисвязности

Из приведенной таблицы видно, что все неполярные боковые цепи без больших натяжек можно представить в виде операторов антисвязности. То, что среди операторов антисвязности находятся боковые цепи Ser и Thr, обусловлено тем, что вследствие небольшого размера они могут образовывать (и по нашим наблюдением часто образуют) водородную связь не с (i - 4)-м карбонилом, а с (i - 3)-м. По этой причине может образоваться не только четырехзвенный цикл (табл. 4.1), но и трехзвенный (табл. 4.2), который соответствует в нашем определении (см. 2.1) ациклической конформации.

Как видно из табл. 4.1 и 4.2, размер боковых цепей (как операторов связности, так и операторов антисвязности) находится в пределах области связи $N_iH...O_{i-4}=C$, что также соответствует общим требованиям к физическим операторам, сформулированным в 2.4. С учетом этого становится понятным, почему природа включила в состав канонического набора аминокислот только те боковые цепи аминокислот, размер которых согласован с размером области связи $N_iH...O_{i-4}=C$. Все боковые цепи, выходящие за пределы заданных параметров и имеющие слишком большую длину, не смогут работать в качестве физических операторов.

Природа соответствия «триплет – аминокислота» в генетическом коде

Вопрос о том, какими должны быть соответствия между аминокислотами и кодирующими их основаниями, был впервые поставлен Гамовым при теоретическом рассмотрении этой проблемы [61]. Однако практическая расшифровка генетического кода показала, что соответствия «триплет – аминокислота» не столь регулярны, как в «алмазном» коде Гамова [61]. Наиболее часто цитируется работа Ф. Крика [62], в которой рассматриваются две альтернативные гипотезы: гипотеза «замороженной случайности»^{*} и гипотеза стереохимического соответствия между структурой аминокислот и триплетами азотистых оснований.

Слабость первой гипотезы состоит в том, что в ней не заданы критерии отбора наиболее удачных соответствий и поэтому понять, например, природу связных серий аминокислот в генетическом коде [42] не представляется возможным. Несостоятельной можно считать и вторую гипотезу, поскольку все попытки показать стереохимические соответствия между структурой аминокислот и триплетами азотистых оснований окончились неудачей.

^{*} Согласно ей эти соответствия возникли случайно и как наиболее удачные были закреплены естественным отбором.

Известны работы, обращающие внимание на соответствие аминокислот корневым дуплетам в генетическом коде и предполагающие связь этих соответствий с вырожденностью генетического кода [46], [47]. Эти взгляды развивались и в [45], [63], [64]. Близких взглядов придерживаются также авторы [65], вводящие три варианта порядковых номеров для характеристики положения аминокислот в генетическом коде. Однако приведенные работы лишь констатируют наличие такой связи, но не вскрывают причины этих соответствий.

В рамках нашей теории природа соответствий «триплет – аминокислота» находит принципиально иное толкование. Для обсуждения этой проблемы кратко повторим ее основные положения.

Боковые цепи, осуществляющие воссоздание циклических конформаций ПФ белка, должны содержать группировки, способные образовывать водородные связи с O_{i-4} =С-группой основной цепи. Согласно схеме (4.8) пары переменных в описывающих эти конформации матрицах трансформируются в буквы названий азотистых оснований триплетов на основе соответствия G –10, A – 11. Второе положение триплета при этом кодирует переменные x_3 и x_4 (схема (2.2)), а наличие цикла в ПФ характеризует переменная $x_3 = 1$. Таким образом, общая принадлежность боковых цепей, осуществляющих воссоздание циклических конформаций ПФ, – блоки триплетов с G- и A-основаниями во втором положении.

Аналогичным образом боковые цепи, препятствующие воссозданию циклических конформаций ПФ белка, должны содержать группировки, не способные к образованию водородных связей с $O_{i-4}=C$ группой основной цепи, т. е. иметь неполярные боковые цепи. Переменная $x_3 = 0$ характеризует ациклические ПФ, входящие в блоки C – 00 и U – 01. Таким образом, этим блокам триплетов должны соответствовать неполярные боковые цепи аминокислот.

Нетрудно видеть, что в таблице генетического кода (см. рис. 4.10) в общих чертах наблюдается именно такое соответствие боковых цепей аминокислот блокам триплетов азотистых оснований. Однако помимо общих рассуждений в рамках нашего подхода вполне возможен конкретный анализ положения отдельных аминокислот. *Блок С* $\leftarrow \rightarrow 00$. Рассмотрим боковую цепь пролина (см. табл. 4.2, Pro), иминогруппа которого заключена в пятичленный цикл. Согласно нашему подходу это означает, что данная аминокислота должна всегда воссоздавать только ациклические конформации в белке. При этом Pro должен располагаться в первой строке таблицы топологического кода. И действительно, как видно из рис. 4.10, Pro кодируют триплеты ССС, ССU, ССG, ССА, которые занимают эту строку в таблице генетического кода. Эти триплеты, как видно из таблицы кода, кодируют матрицы, первые четыре переменные которых равны нулю, т. е. они описывают открытые конформации ПФ (или их графов), имеющие степень связности, близкую к нулевой. Именно по этой причине Pro соответствует триплетам первой строки генетического кода.

В этот блок входят также слабополярные боковые цепи Ser и Thr. Как видно из табл. 4.2, эти аминокислоты образуют водородную связь с ($O_{i-3}=C$)-группой основной цепи. При этом четырехзвенный цикл в ПФ не образуется, что объясняет, почему эти аминокислоты попали в блок триплетов, кодирующих открытые конформации ПФ. Из таблицы кода видно, что как группа триплетов UCC, UCU, UCG, UCA, которым соответствует Ser, так и группа триплетов ACC, ACU, ACG, ACA, которым соответствует Thr, кодируют матрицы, которые описывают открытые конформации ПФ ($x_3 = 0, x_4 = 0, \tau. e. C \leftarrow \rightarrow 00$).

Блок U $\leftarrow \rightarrow 01$. В 2.2 было показано, что блок 01 содержит матрицы, описывающие в основном конформации, соответствующие β-структуре. При трансформации в топологический генетический код этот блок соответствует блоку U $\leftarrow \rightarrow 01$. Боковые цепи аминокислот Val, Ile и Leu в β-положении имеют два заместителя, что препятствует формированию замкнутого четырехзвенного цикла. Типичным в этом отношении является боковая цепь Val, который, как известно, обладает наибольшим потенциалом при формировании β-структур [38]. По этой причине Val, а также Ile и Leu должны быть приписаны к блоку триплетов, содержащих U (01) во втором положении. Сопоставление этих предположений с таблицей генетического кода (см. рис. 4.10) показывает, что это действительно так. Аналогичные рассуждения будут справедливы для Met и Phe.

Блок G $\leftarrow \rightarrow 10$. Согласно 2.2 этот блок конформаций ПФ и четырехзвенных графов содержит слабосвязные циклические конформации. В генетическом коде триплетам этого блока соответствуют такие боковые цепи аминокислот, как аргинин (Arg), кодируемый шестью триплетами, а также близкие по структуре боковые цепи цистеина (Cys) и серина (Ser). Мы видели, что Ser кодируется триплетами блока $C \leftarrow \rightarrow 00$, и объяснили, почему боковая цепь этой аминокислоты оказалась приписанной к этому блоку. Однако, как следует из табл. 4.1, боковая цепь этой аминокислоты может образовывать водородные связи и с О_{*i*-4}=С-группой основной цепи, т. е. работать в качестве операторов связности. В рамках нашего подхода именно благодаря этому боковые цепи Ser кодируются триплетами не только блока C $\leftarrow \rightarrow 00$, но и блока $G \leftarrow \rightarrow 10$, триплеты которого кодируют циклические конформации. Аналогичны причины нахождения в этом блоке боковых цепей Arg, который, благодаря гуанидиновой группировке в его боковой цепи, имеет возможность образовывать разнообразные водородные связи с О_{*i*-4}=С-группой основной цепи. Более детально эта боковая цепь будет рассмотрена далее. Кроме того, Cys и Trp, вероятно, также образуют слабосвязные циклы. Что касается циклов, образуемых глицином (Gly), который кодируется триплетами GGC, GGU, GGG и GGA, то их подвижность кажется очевидной, поскольку, в отличие от других аминокислот, боковые цепи которых могут снижать подвижность циклических ПФ, эта аминокислота не имеет боковой цепи, хотя и способна к формированию циклического ПФ.

Блок $A \leftrightarrow 11$. Этот блок конформаций ПФ и четырехзвенных графов, как было показано в 2.2, содержит циклические конформации (второе основание триплета – А – кодирует пару переменных $x_3x_4 = 11$, причем $x_3 = 1$), которые должны воссоздаваться операторами связности. И действительно, практически все боковые цепи аминокислот, соответствующие триплетам этого блока (His, Gln, Asp, Glu, Asn, Lys), способны к образованию водородных связей с O_{*i*-4}=C-группой основной цепи, что хорошо видно из сопоставления рис. 4.10 и табл. 4.1. Единственной аминокислотой, обладающей некоторыми особенностями, как отмечалось ранее, является боковая цепь Тур.

В целом можно сказать, что соответствие боковых цепей аминокислот триплетам генетического кода вполне удовлетворительно описывается, исходя из рассмотрения их в качестве физических операторов связности и антисвязности. В то же время, высказанные ранее гипотезы о происхождении соответствий аминокислот триплетам генетического кода кажутся весьма далекими от реального положения дел, выявленного с помощью нашего подхода.

Расположение в генетическом коде аминокислот, воссоздающих симметричные конформации белка

Дополнительным критерием, с учетом которого должны быть расположены боковые цепи аминокислот в генетическом коде, может служить воссоздание этими цепями симметричных конформаций ПФ-белка.

2		$\mathbf{C} \leftrightarrow$	00		$\mathbf{U} \leftrightarrow 01$				
13	$C \leftrightarrow 00$	$U \leftrightarrow 01$	$G \leftrightarrow 10$	A↔11	$C \leftrightarrow 00$	$U \leftrightarrow 01$	$G \leftrightarrow 10$	$A \leftrightarrow 11$	
00 ↓ C	0 0 0 0 0 CCC 0 Pro	0 0 0 0 0 CCU 1 Pro	0 0 0 0 1 CCG 0 Pro	0 0 0 0 1 CCA 1 Pro	0 0 0 1 0 CUC 0 Leu	0 0 0 1 0 CUU 1 Leu	0 0 0 1 1 CUG 0 Leu	0 0 0 1 1 CUA 1 Leu	
10 ↓ G	1 0 0 0 0 GCC 0 Ala	1 0 0 0 0 GCU 1 Ala	1 0 0 0 1 GCG 0 Ala	1 0 0 0 1 GCA 1 Ala	1 0 0 1 0 GUC 0 Val	1 0 0 1 0 GUU 1 Val	1 0 0 1 1 GUG 0 Val	1 0 0 1 1 GUA 1 Val	
01 ↓ U	0 1 0 0 0 UCC 0 Ser	$\begin{array}{c} 0 \ 1 \ 0 \\ 0 \ 0 \\ UCU \ 1 \\ Ser \end{array}$	$\begin{array}{c} 0 \ 1 \ 0 \\ 0 \ 1 \\ UCG \ 0 \\ Ser \end{array}$	0 1 0 0 1 UCA 1 Ser	0 1 0 1 0 UUC 0 Phe	0 1 0 1 0 UUU 1 Phe	0 1 0 1 1 UUG 0 Leu	010 11 UUA 1 Leu	
11 ↓ A	$\begin{array}{ccc} 1 & 1 & 0 \\ 0 & 0 \\ ACC & 0 \\ Thr \end{array}$	1 1 0 ACU 1 Thr	$\begin{array}{c}1&1&0\\&0&1\\ACG&0\\Thr\end{array}$	1 1 0 ACA 1 Thr	$\begin{array}{ccc}1&1&0\\1&0\\AUC&0\\Ile\end{array}$	1 1 0 AUU 1 Ile	$ \begin{array}{cccc} 1 & 1 & 0 \\ AUG & 0 \\ Met \end{array} $	1 1 0 AUA 1 Ile	
00 ↓ C	0 0 1 0 0 CGC 0 Arg	0 0 1 0 0 CGU 1 Arg	0 0 1 0 1 CGG 0 Arg	001 01 CGA1 Arg	0 0 1 1 0 CAC 0 His	0 0 1 1 0 CAU 1 His	$\begin{array}{c} 0 & 0 & 1 \\ 1 & 1 \\ CAG & 0 \\ Gln \end{array}$	0 0 1 1 1 CAA 1 Gln	
10 ↓ G	$\begin{array}{ccc}1&0&1\\GGC&0\\Gly\end{array}$	$\begin{array}{c}1&0&1\\GGU&0\\Gly&1\end{array}$	1 0 1 GGG 0 Gly	$\begin{array}{ccc}1&0&1\\GGA&1\\Gly\end{array}$	$\begin{array}{ccc}1&0&1\\GAC^{l}&0\\Asp\end{array}$	$\begin{array}{c c}1&0&1\\GAU^1&0\\Asp\end{array}$	$\begin{array}{c}1&0&1\\GAG^1&1\\GIu\\Glu\end{array}$	1 0 1 GAA ¹ 1 Glu	
01 ↓ U	0 1 1 0 0 UGC 0 Cys	$egin{array}{ccc} 0 & 1 & 1 \\ 0 & 0 \\ UGU & 1 \\ Cys \end{array}$	0 1 1 0 1 UGG 0 Trp	$ \begin{array}{c} 0 1 1 \\ 0 1 \\ UGA 1 \\ T \end{array} $	0 1 1 1 0 UAC 0 Tyr	0 1 1 1 0 UAU 1 Tyr	$\begin{array}{ccc} 0 & 1 & 1 \\ 1 & 1 \\ UAG & 0 \\ T \end{array}$	011 11 UAA 1 T	
11 ↓ A	$\begin{array}{ccc}1&1&1\\AGC&0\\Ser\end{array}$	1 1 1 0 0 AGU 1 Ser	1 1 1 0 1 AGG 0 Arg	I I I O I AGA I Arg	1 1 1 1 0 AAC 0 Asn	$ \begin{array}{c} 1 1 1 \\ 1 0 \\ AAU 1 \\ Asn \end{array} $	1 1 1 AAG 0 Lys	l l l J AAA l Lys	
2	$G \leftrightarrow 10$				A ↔ 11				

Рис. 4.11. Положение в таблице генетического кода аминокислот, воссоздающих симметричные конформации белка

Как было показано в 2.4, для воссоздания симметричных конформаций цепного полимера, расположенных симметрично по разные стороны от главных диагоналей блоков и описываемых симметричными матрицами, необходимы различные физические операторы (см. рис. 2.12). Мы рассмотрели, как этот вывод проявляется в генетическом коде. Для этой цели были проанализированы боковые цепи аминокислот, которые располагаются симметрично по отношению к главным диагоналям генетического кода. Если они оказывались разными, то один из элементов пары заштриховывали (рис. 4.11). Как следует из рисунка, в блоке U находится 6 триплетов, кодирующих Leu, но ни одна из боковых цепей Leu не расположена симметрично по отношению к главной диагонали блока. То же можно сказать о блоке G, в котором находится 6 триплетов, кодирующих Arg.

В блоке А близкие по размеру аминокислоты Asp и Asn, а также Glu и Gln также находятся в несимметричных по отношению к главным диагоналям положениях. Таким образом, расположение боковых цепей аминокислот в таблице генетического кода подтверждает положение, предсказанное на основе нашей теории.

4.4. Перспективы дальнейшего изучения структуры генетического кода

Краткий обзор состояния исследований структуры генетического кода

Появление сразу двух вариантов пространственных структур генетического кода [51], [54] повлияло, по-видимому, на характер исследований в этой области. В предшествующий период, начиная с установления соответствий «триплет – аминокислота» [40], исследователи в основном занимались проблемой эволюции генетического кода [66], [67]. Такой анализ, с учетом современных данных по расшифровке геномов различных видов, проводится и поныне [68], [69]. Предложенные структуры кода представляют собой удобный математический объект, поэтому акцент в исследованиях сместился в сторону математических подходов. Рассмотрим кратко несколько аспектов, затронутых в этих исследованиях. Необходимо отметить, что время от времени появляются работы, в которых (в силу разных причин) практически дублируются варианты структуры кода, найденные ранее другими авторами. Например, в [70] предложен вариант двумерной структуры генетического кода, в которой наблюдается симметрия двух групп аминокислот, отличающихся по степени вырожденности третьего основания в триплетах (см. 4.1). Однако, в отличие от [47], [50], [52], в [70] используются принципы, основанные на бинарной интерпретации пуринов (1) и пиримидинов (0). На наш взгляд, такая классификация существенно хуже, чем в [47], [50] и [52], поскольку блоки, содержащие во втором положении одну и ту же букву, в такой классификации разбиты на несколько частей (например, блоки С и А). Круговая форма представления генетического кода, описанная в [42], была вновь использована в [71].

Арифметический подход к анализу структуры генетического кода последовательно проводил В. И. Щербак [63], [64], [72]. Если сопоставить результаты этих работ с требованиями к физическим операторам и боковым цепям аминокислот как физическим операторам (см. 2.4 и 4.3), то все арифметические закономерности, выявленные В. И. Щербаком, укладываются в пределы тех весов боковых цепей, которыми эти боковые цепи как физические операторы должны обладать. Так что, на наш взгляд, ничего принципиально нового этот подход не выявляет.

В [73] и [74] развивается подход к генетическому коду с точки зрения квантовой теории. В частности, авторы [74] на основе этого подхода получили четыре класса кристаллических кодонов, которые можно рассматривать как своеобразную их классификацию. Отметим, однако, что эта классификации никак не соотносится с физико-химическими свойствами аминокислот, попавших в одну группу. Пока эффективность предложенного подхода ни в чем себя не проявила.

По-прежнему излюбленным объектом математиков является дуплетный генетический код. Хотя принципиальных находок в этой области не было, для специалистов могут представлять интерес исследования авторов [75]–[77]. Попытки дать математическое описание генетического кода с позиции палиндромной симметрии (преобразование сохранения вырожденности в генетическом коде) приводятся в [78] и [79]. Эти работы идейно близки другим, в которых наблюдается симметрия дуплетов. Алгебраические подходы к анализу двумерной структуры генетического кода развиваются в работах С. В. Петухова [60], [80], [81]. Они вносят существенный вклад в исследования генетического кода, которые важны еще и потому, что последний является ключом к биоинформатике как к науке.

Наличие альтернативных вариантов пространственной структуры триплетного кода позволяет со стороны оценить, какой из вариантов лучше. В настоящее время стало известно еще несколько вариантов пространственной структуры генетического кода. В этом плане в первую очередь интересны работы [82] и [83], в которых предлагается сферическая модель структуры генетического кода в виде квази-28-гона (икосикаиоктагона). По сути речь идет о структуре канонического набора аминокислот, построенной на основе кода, и о четырехмерном представлении дуплетного генетического кода. На наш взгляд, при построении модели канонического набора аминокислот была применена искусственная группировка боковых цепей аминокислот. В то же время структура дуплетного кода практически аналогична тем, которые были получены в [51], [53], [54], но имеет иное расположение дуплетов.

Структура триплетного генетического кода была предложена в [84]. Эта форма представления довольно близка той, что предложили авторы [54], хотя графическое исполнение оставляет желать лучшего. К сожалению, по приведенным рисункам трудно составить полное представление о найденных авторами закономерностях.

В [80] была описана пространственная структура генетического кода, изоморфная булеву гиперкубу B^6 . Внешне она похожа на структуру [54], хотя расположение триплетов в ней несколько иное. Не исключено, что в момент разработки этой структуры предыдущие варианты были автору неизвестны, поскольку он на них не ссылается. Наконец, в [85] была предложена булева решетка для генетического кода на основе диаграммы Хассе, которую также можно рассматривать в качестве структуры генетического кода. Она похожа на структуру, предложенную нами [51]. Однако в ней использованы другие соответствия пар переменных буквам триплетов, и не получилось компактного расположения серий триплетов (см. 4.2).

Роль последовательности букв триплетов и соответствия пар переменных буквам триплетов в представлении структуры генетического кода

Хотя эти вопросы уже затрагивались ранее (см. 4.1 и 4.2), мы специально выделили этот блок текста, учитывая их важность. Остановимся сначала на роли последовательности букв триплетов. В классической таблице генетического кода используется последовательность, приведенная в [41] (U, C, A, G), а в работе [43] используется последовательность A, C, G, U. Однако ни та, ни другая последовательность не выявляет симметрию кода, связанную с вырожденностью третьего основания триплетов. Только в [46], [47], [50] последовательность C, G, U, A, названная Ю. Б. Румером канонической, проявилась для двух групп дуплетов, кодирующих четырежды и дважды вырожденные по третьему основанию аминокислоты [46], [47]. В ромбическом варианте генетического словаря [50] данная последовательность, описанная как принцип комплементарности (CC – GG, UU – AA), была использована в качестве основы для его построения и выявила ту же симметрию. В [75] и [76] на основе анализа всех вариантов таких последовательностей было выявлено несколько симметричных вариантов. Однако только каноническая последовательность С, G, U, A приводит к компактному расположению двух групп дуплетов. На наш взгляд, в природе не бывает случайностей (тем более в таких универсальных структурах, как генетический код), поэтому игнорировать эту симметрию ни в коем случае нельзя. Пренебрежение ею приводит к появлению неудачных вариантов структур генетического кода.

Второй причиной, нарушающей упорядоченность структуры кода, может служить неправильное соответствие пар переменных буквам триплетов при трансформации в булев гиперкуб. Разнообразные варианты таких соответствий были обобщены в [86] и [87]. В табл. 4.3 приведены данные по всем известным к настоящему времени вариантам. Анализу всех возможных 24 вариантов соответствий посвящена работа [87].

Как следует из [87], во всех случаях необходимо согласовать множество $N = \{C, U, A, G\}$ из четырех нуклеотидных оснований РНК с множеством $N2 \times N2 = \{00, 01, 10, 11\}$ из четырех двоичных пар переменных. Авторы проанализировали математические преобразования,

Таблица 4.3

Последовательность			ГЬ	Сонцики				
00	01	10	11	ССЫЛКИ				
U	С	Α	G	Swanson [44]				
Α	G	U	С	Jiménez-Montaño et al. [54]; Klump [59]				
С	U	G	Α	Karasev and Soronkin [51]				
U	С	G	Α	Stambuk [88]				
С	А	U	G	Stambuk et al [88]; José et al. [89]; He et al. [90]				
G	А	U	С	Sánchez et al. [91]–[93]				
G	U	Α	С	Sánchez et al. [92]				
A	U	С	G	Jiménez-Montaño et al. [94]				

Варианты соответствий пар переменных буквам триплетов

которые переводят любую модель в каждую из остальных. С биологической точки зрения выбор соответствий может быть обусловлен тремя критериями:

 наличием симметрии двух групп дуплетов в исходном дуплетном коде;

– антисимметрией пар переменных ($00 \leftarrow \rightarrow 11, 01 \leftarrow \rightarrow 10$);

 – соответствием боковых цепей аминокислот тем матричным описаниям структур, которые они воссоздают в качестве физических операторов связности и антисвязности.

Во многих упомянутых случаях этот выбор, по нашему мнению, определяется личными предпочтениями авторов и не связан с какимилибо фундаментальными основами генетического кода. В [54] и [87] при выборе последовательности A, G, U, C были использованы первые два критерия. При этом оказалось учтенным количество водородных связей, образуемых в парах C – G и U – A (3 – сильные связи и 2 – слабые связи, см. 4.1).

В рамках нашего подхода исходной последовательностью при построении дуплетного и триплетного кодов была каноническая последовательность C, G, U, A, которая соответствует первому критерию. При выборе соответствий букв триплетов парам переменных (C – 00, U – 01, G – 10, A – 11) наблюдается антисимметрия пар переменных $00 \leftrightarrow 11$ и $01 \leftrightarrow 10$, которая совпадает с преобразованием Румера для двух групп дуплетов и триплетов (C $\leftarrow A$, G $\leftarrow U$), что соответствует второму критерию. При этом триплетам генетического кода, содержащим в качестве второго основания C или U (они кодируют переменные 00 и 01), соответствуют боковые цепи, воссоздающие ациклические конформации, а если вторым основанием является G или A (переменные 10 или 11), то этому соответствуют циклические конформации (см. 4.3). Такое положение дел удовлетворяет третьему из сформулированных ранее критериев. Таким образом, совпадение всех трех критериев привело к единственной из всех возможных вариантов правильной пространственной структуре генетического кода (см. рис. 4.8), в которой нашли отражение основные его закономерности. Такую структуру можно назвать канонической.

4.5. Применение пространственной структуры генетического кода к анализу эволюции биологических систем кодирования

В связи с расшифровкой геномов различных биологических видов обнаруживается все больше примеров отклонения генетических кодов от универсального кода [69]. Факт установления канонической пространственной структуры генетического кода может служить основой для анализа эволюции биологических систем кодирования.

Перспективы исследований, связанных с использованием структуры [54] для проведения эволюционного анализа, весьма прозрачны. Для этого необходимо просто нанести на гиперкуб наблюдаемые отклонения от универсального кода. В качестве примера на рис. 4.12 приведена объединенная структура отклонений положения инициирующих триплетов от универсального кода. Видно, что практически все разнообразные отклонения находятся в нижней части гиперкуба. При этом наблюдается параллельность отклонений – как в множествах (M₁ и M₂), так и в подмножествах (SM₁, SM₂ и SM₃, SM₄), расположенных в пределах этих множеств. Основной вывод по результатам предварительного анализа состоит в том, что все изменения значений в генетическом коде связаны с единичными изменениями оснований, т. е. обусловлены точечными мутациями.

В заключение этой главы подведем краткие итоги. Применение теории топологического кодирования линейных цепных полимеров к генетическому коду позволило показать, что его можно рассматривать в качестве топологического кода. В основание генетического кода положено описание конформаций пентафрагментов белка или его анало-



Рис. 4.12. Объединенная структура отклонений инициирующих кодонов от универсального генетического кода

гов – четырехзвенных цепных графов, закодированное с помощью четырех букв триплетов. Соответствия триплетов боковым цепям аминокислот в генетическом коде обусловлены участием боковых цепей в качестве физических операторов связности и антисвязности в воссоздании циклических и ациклических конформаций ПФ.

Глава 5. СТРУКТУРА КАНОНИЧЕСКОГО НАБОРА АМИНОКИСЛОТ

5.1. Проблема классификации канонического набора аминокислот

Интерес к классификации аминокислот канонического набора в последние годы существенно возрос. С одной стороны, это объясняется тем, что природа канонического набора до сих пор остается неясной [38]. С другой стороны, от решения этого вопроса зависит решение и других вопросов, в частности, связанных с самоорганизацией белковых молекул в трехмерные структуры.

Существует несколько подходов к классификации аминокислот. Наиболее известным подходом является классификация на основе физико-химических свойств [95]. Однако следует отметить, что боковые цепи аминокислот в отношении их принадлежности к определенным классам органических соединений весьма разнородны. Так, Asp и Glu являются карбоновыми кислотами; Lys – это первичный амин; из пяти аминокислот, содержащих циклы, Pro – алифатическая иминокислота, а Phe, His, Tyr и Trp – ароматические соединения. При этом физикохимический подход не может вскрыть принципы, по которым боковые цепи аминокислот вошли в состав белков в процессе эволюции.

Большинство современных работ по классификации аминокислот ориентируется на генетический код как на их природную основу. Характерны в этом отношении подходы, рассматривающие комплементарность аминокислот, закодированных комплементарными триплетами (см. обзор [96]). Меклер и Идлис, предпринявшие последовательно такой анализ, выделили три группы связных аминокислот на основе этого принципа [7]. Другая классификация, также основанная на генетическом коде, рассматривает аминокислоты как физические операторы, воссоздающие закодированную структуру [24]. Попытка построить периодическую систему аминокислот, ориентируясь на свойства генетического кода, была предпринята в [97]. К триплетам генетического кода были применены методы теории групп [98]. На основе этих методов было выделено четыре группы триплетов, кодирующих аминокислоты с разными свойствами (гидрофобные, слабогидрофобные, гидрофильные и пролин), отнесенные к групповой симметрии SU(4). Еще один подход, также основанный на теории групп, описан в [99].

В качестве классификаций аминокислот можно рассматривать пространственные модели генетического кода, основанные на единичных заменах азотистых оснований в дуплетах и триплетах [51], [54], хотя аминокислоты в них играют подчиненную роль.

Существуют классификации аминокислот, появление которых связано с потребностями в разработке методов предсказания белковых структур. Так, в [100] предложена классификация, основанная на синтезе физико-химических и мутационных свойств аминокислот, которая представлена в форме диаграммы Венна. В ней выделяются группы аминокислот, имеющих консервативный характер в белках. На основе анализа различных типов структурных потенциалов аминокислот и матриц замещения в [101] предложены классификации аминокислот, полезные для сравнения эффективности методов предсказания белковых структур. Известны также классификации аминокислот на основе иерархических принципов [102]. Предпринимаются попытки классификации аминокислот, основанные на описании эволюции белков как марковского процесса и построении матриц скорости замен аминокислот в белках [103].

Аминокислоты также классифицируются по функциональным свойствам [1], [18]. При этом выделяются *пассивные элементы* (неполярные боковые цепи), формирующие изолирующее гидрофобное ядро белков, и *активные элементы* (боковые цепи, способные образовывать водородные связи), которые имеют различное количество входов и выходов и могут участвовать в процессах переноса зарядов в белках. Однако вышеупомянутые подходы, как правило, не решают проблемы состава канонического набора или количества включенных аминокислот.

5.2. Построение модели структуры канонического набора аминокислот

Структура боковых цепей канонического набора аминокислот

Прежде чем анализировать предпосылки к построению модели и саму модель, необходимо вспомнить структуры боковых цепей. В табл. 5.1 боковые цепи приведены с названиями на английском языке и сокращенными трехбуквенными обозначениями латиницей, что является

Таблица 5.1



Примечание. Боковые цепи прикреплены к α-углеродному атому, обозначенному кружком.

общепринятым. Существуют также и однобуквенные обозначения, но поскольку мы их в данной книге не употребляем, то в таблице они не приведены. Как следует из таблицы, боковые цепи можно разделить на три основные группы: неполярные и слабополярные (Pro, Gly, Ala, Val, Leu, Ile, Ser, Thr), полярные (Asp, Glu, Lys, Arg), нейтральные и серосодержащие (Asn, Gln, Cys, Met) и циклические (Phe, Tyr, His, Trp). Справедливости ради отметим, что к циклическим боковым цепям относится также и пролин, который иногда почему-то относят к неполярным боковым цепям. С химической точки зрения пролин является иминокислотой, поскольку его аминогруппа заключена в цикл.

Предпосылки построения модели

Хотя в настоящее время наличие в биоструктурах 21-й и 22-й аминокислот (селеноцистеина и пирролизина) вполне доказано [104], локальный характер их распространения не позволяет считать их универсальными, и магическое число 20 для аминокислот, присутствующих в белках, по-прежнему остается в силе. При выборе исходной структуры, на которой могли бы разместиться 20 аминокислот, необходимо вспомнить, что додекаэдр (одно из тел Платона) имеет 20 вершин. Вполне логично было бы расположить 20 аминокислот в 20 вершинах этого многогранника, но проблема состоит в том, что не известны принципы, по которым они должны в этих вершинах располагаться.

Анализ табл. 5.1 позволяет отметить три особенности канонического набора аминокислот.

Первая особенность состоит в том, что большинство боковых цепей аминокислот представлено в наборе в виде пар, близких по свойствам, но отличающихся по длине, например: Asp – Glu, Asn – Gln, Arg – Lys, Phe – Туг. В какой-то степени это распространяется и на другие боковые цепи.

Вторая особенность заключается в наличии боковых цепей аминокислот близкой структуры, но с отличающимися свойствами. Например, Ser и Cys по структуре боковых цепей одинаковы и отличаются только терминальными атомами С–О и С–S соответственно. То же можно сказать про боковые цепи Asp–Asn и Glu–Gln, которые в парах одинаковы, а различия касаются лишь терминальных групп. *Третья особенность* состоит в наличии боковых цепей с противоположными свойствами. Наиболее ярко она выражена для пар боковых цепей Asp⁻ – Lys⁺, Glu⁻ – Arg⁺, однако такие свойства можно найти и у других аминокислот.

До сих пор перечисленные групповые свойства аминокислот не нашли какого-либо рационального объяснения. Учитывая наш предыдущий опыт в выделении и анализе свойства антисимметрии в суперматрице, описывающей конформации ПФ и четырехзвенных цепных графов, отмеченные свойства можно рассматривать как проявление принципа антисимметрии. Именно этот принцип и был положен в основу построения на додекаэдре модели структуры канонического набора аминокислот [35]–[37], [105], [106].

Необходимо отметить, что эта модель развивалась совершенно независимо от теории топологического кодирования [105] и предшествовала разработке модели молекулярной векторной машины [35]–[37]. Вследствие этого один из принципов, на основе которого можно было бы располагать боковые цепи аминокислот на додекаэдре (в порядке возрастания молекулярной массы боковых цепей), не был правильно задан. В первоначальной версии модели боковые цепи располагались в порядке возрастания молекулярной массы снизу вверх [105].

Промежуточные этапы построения модели

Несмотря на независимую от теории топологического кодирования изначальную разработку этой модели, она уже содержала принципы антисимметрии при выделении подгрупп аминокислот [105], [106].

Выделение плоскостей антисимметрии на додекаэдре. Процедуры выделения плоскостей антисимметрии рассматривались ранее (см. 3.1) при построении модели МВМ. Для удобства последующего изложения и обсуждения результатов кратко повторим их.

Как показано на рис. 5.1, через структуру додекаэдра можно провести три плоскости (I, II и III), которые разделят додекаэдр на симметричные части. Плоскость I пройдет перпендикулярно плоскости листа и разделит додекадр на правую и левую половины. Плоскость II пройдет параллельно плоскости листа и разделит додекаэдр на переднюю и заднюю части. Наконец, плоскость III, перпендикулярная первым двум плоскостям, разделит додекаэдр на верхнюю и нижнюю части. При вращении вокруг оси, лежащей на пересечении плоскостей II и III, эти части полностью совмещаются своими вершинами.



Рис. 5.1. Выделение плоскостей антисимметрии и обозначение вершин на додекаэдре

Обозначения вершин додекаэдра. Для вершин додекаэдра были введены следующие обозначения (рис. 5.1). В плоскости I самая верхняя вершина обозначена символом A, а поворотно симметричная ей и находящаяся под плоскостью III вершина – символом (–А). Для двух других вершин, расположенных в плоскости I, перпендикулярных первым двум, были соответственно использованы обозначения B и (–В). Обе пары вершин связаны между собой преобразованием вращения, которое можно обозначить букой γ .

Вершины, симметричные относительно плоскости I, были обозначены так: слева все буквы, обозначающие вершины, имеют индекс, расположенный справа внизу, а справа – слева вверху. Так, ближние к А вершины, находящиеся за плоскостью II, обозначены A_1 и ¹A, а отдаленные, расположенные перед плоскостью II, обозначены A_2 и ²A. Для вершин, расположенных под плоскостью III, эти обозначения будут таковы: (- A_1) и (- 1A) перед плоскостью II и (- A_2) и (- 2A) за плоскостью II. Аналогично обозначены вершины для второй группы. Вершины, расположенные ближе к вершине В за плоскостью II, обозначены ¹В и В₁. Вершины перед плоскостью II имеют обозначения ²В и В₂, а по отношению к вершине (–В) соответственно (–¹В) и (–В₁) перед плоскостью II и (–²В) и (–В₂) за плоскостью II. Все вершины, расположенные по обе стороны от плоскости I, можно преобразовать друг в друга зеркальным отражением. Такое преобразование было названо преобразованием отражения и обозначено буквой α .

Все вершины, расположенные за плоскостью II, можно спроецировать на вершины, расположенные перед плоскостью II. Такое преобразование было названо преобразованием незеркального отражения. Можно назвать его также преобразованием проекции и обозначить буквой β.

Все вершины, расположенные над плоскостью III, можно преобразованием вращения (γ) перевести в вершины, расположенные под плоскостью III. Однако для подгруппы вершин, обозначенных буквой А с индексами, и подгруппы вершин, обозначенных буквой В с индексами, все преобразования могут осуществляться только в пределах своей подгруппы. Таким образом, все 20 вершин получили свои наименования и разделились на четыре подгруппы (табл. 5.2).

Таблица 5.2

Номер	Преобразования симметрии								
подгруппы	1	α	β	γ	αβ	αγ	βγ	αβγ	
1	Α			- A					
2	В			- B					
3	A ₁	¹ A	A ₂	-A ₁	² A	$-^{1}A$	-A ₂	$-^{2}A$	
4	B ₁	^{1}B	B ₂	-B ₁	^{2}B	$-^{1}B$	-B ₂	$-^{2}B$	

Подгруппы вершин, связанные преобразованиями симметрии

Как следует из табл. 5.2, 2 подгруппы из двух элементов (А и –А, В и –В) связаны преобразованием вращения (γ), а 2 подгруппы из восьми элементов каждая, обозначенных буквами А и В с индексами в разных положениях, связаны преобразованиями α, β и γ.

Выделение боковых цепей аминокислот в соответствии с подгруппами антисимметрии

Прежде чем строить модель структуры боковых цепей аминокислот на додекаэдре, было необходимо убедиться, что все боковые цепи можно распределить в соответствии с принципами антисимметрии. Как было отмечено в 5.1, среди аминокислот можно выделить четыре типа антисимметрии (табл. 5.3).

Таблица 5.3

А. Квазизеркальная	Б. Незеркальная	В. Антисимметрия	Г. Антисимметрия
антисимметрия	антисимметрия	антиподов	комплементарности
Thr-Ser	Ser-Cys	Ser-His	Ser-Trp
Met–Cys	Thr-Met	Thr–Trp	Thr–His
Glu–Asp	Asp–Asn	Cys–Phe	Cys–Tyr
Gln–Asn	Glu–Gln	Met–Tyr	Met–Phe
Lys–Arg	Arg–Val	Asp–Arg	Asp–Lys
Ile–Val	Lys–Ile	Glu–Lys	Glu–Arg
Trp–His	His-Phe	Asn–Val	Asn–Ile
Tyr–Phe	Trp–Tyr	Gln–Ile	Gln–Val
_	_	Gly–Pro	Gly–Pro
_	_	Ala–Leu	Ala–Leu

Распределение боковых цепей аминокислот в соответствии с типами антисимметрии

Квазизеркальная антисимметрия связывает аминокислоты, близкие по физическим свойствам, но различные по длине. В этой подгруппе все аминокислоты в парах имеют сходные свойства, но при этом справа расположены более короткие (т. е. менее массивные) боковые цепи, чем слева: Ser и Thr содержат гидроксильные группы; Cys и Met – атомы серы; Asp и Glu – карбоксильные группы; Asn и Gln – амидные группы и т. д.

Незеркальная антисимметрия связывает в пары аминокислоты, имеющие близкую длину, но различные по свойствам терминальные группы. Так, (Ser – Cys) – гидроксигруппа и сульфогруппа; (Asp – Asn) и (Glu – Gln) – карбоксильная и амидная группы; (Arg – Val) и (Lys – Ile) – щелочная и гидрофобная группы; (His – Phe) и (Trp – Tyr) – пятичленные и шестичленные кольца в боковых цепях.

Антисимметрия антиподов связывает боковые цепи, противоположные по длине и свойствам. В этой подгруппе аминокислот выделяются пары: (Ser – His), (Thr – Trp), (Cys – Phe), (Met – Tyr) – короткие цепи противостоят массивным; (Asp – Arg), (Glu – Lys) – боковые цепи с противоположными зарядами; (Asn – Val), (Gln – Ile) – боковые цепи с противоположными свойствами (гидрофильные – гидрофобные).

Антисимметрия комплементарности включает в себя пары аминокислот, имеющие максимальные различия (в парах они как бы компенсируют друг друга). Например, в паре (Ser – Trp) находятся короткая цепь и самая массивная цепь с пятичленным кольцом, тогда как в паре (Thr – His) – более массивная короткая цепь и менее массивная с пятичленным кольцом. В паре (Asp – Lys) находятся самая короткая цепь с отрицательно заряженной группой и самая длинная цепь с положительно заряженной группой, а в паре (Glu – Arg) – более длинная цепь с отрицательно заряженной группой находится вместе с более короткой цепью с положительно заряженной группой и т. д.

Две пары аминокислот канонического набора обладают особыми свойствами. В паре (Pro – Gly) аминокислота Gly не содержит боковой цепи, тогда как Pro является иминокислотой и его пептидная группа заключена в цикл. В паре (Ala – Leu) самая короткая неполярная боковая цепь находится вместе с самой длинной неполярной цепью. Эти две пары аминокислот входят в две подгруппы: в подгруппу антисимметрии антиподов и в подгруппу комплементарной антисимметрии (см. табл. 5.3).

С учетом приведенных типов антисимметрии боковые цепи аминокислот были помещены в вершины додекаэдра (рис. 5.2, *a*) в порядке возрастания их молекулярной массы сверху вниз, причем более короткие цепи поместили справа от плоскости I, а более длинные – слева. Выбор в пользу возрастания массы боковых цепей сверху вниз был обусловлен участием системы боковых цепей аминокислот в качестве элемента MBM (этот вопрос будет рассмотрен в следующей главе).

5.3. Обсуждение модели

Основу построенной нами модели структуры канонического набора аминокислот на додекаэдре (рис. 5.2, *a*) составляет обобщенная модель структуры канонического набора физических операторов, изложенная в 3.3. Тот факт, что антисимметрия физических операторов (им присвоены обозначения вершин додекаэдра) и антисимметрия боковых цепей аминокислот (см. табл. 5.3) совпали, позволил разместить их в соответствующих вершинах додекаэдра. При этом боковые цепи аминокислот помещены в кружки в вершинах додекаэдра в порядке увеличения их размера сверху вниз. В самом верху находится глицин^{*}, не имеющий, как известно, боковой цепи. Все остальные боковые цепи также приведены с α -углеродными атомами, расположенными вверху, над цепью.

^{*} Он обозначен кружком с треугольником, соответствующим его α-углеродному атому.



Рис. 5.2. Модель структуры канонического набора боковых цепей аминокислот на додекаэдре: а – общий вид модели; б – аминокислоты, связанные квазизеркальной антисимметрией (близкие по физическим свойствам, но различные по длине);
в – аминокислоты, связанные незеркальной антисимметрией (близкие по длине, но гомологичные по свойствам); г – аминокислоты, обладающие антисимметрией антиподов (противоположные по длине и свойствам). Подгруппы аминокислот, связанные преобразованиями антисимметрии, обозначены одним цветом:
1 и 2 – серым; 3 – розовым; 4 – голубым. І, ІІ, ІІІ – плоскости антисимметрии. Боковые цепи аминокислот показаны в кружках, под ними приведены трехбуквенные обозначения аминокислот

Согласно модели, приведенной в 3.3, в плоскости I должны находиться четыре оператора с особыми свойствами. Мы расположили в пределах этой плоскости аминокислоты Gly (над плоскостью III) и Pro (под плоскостью III). Перпендикулярно этим аминокислотам в плоскости I помещены боковые цепи Ala и Leu.

Как было показано ранее, среди аминокислот можно выделить пары боковых цепей, близких по свойствам, но различных по длине (квазизеркальная антисимметрия). В соответствии с моделью, приведенной в 3.3, справа от плоскости I были помещены аминокислоты с короткой цепью, а слева (симметрично) – с более длинной цепью (см. рис. 5.2, a): Ser–Thr, Cys–Met, Asp–Glu, Asn–Gln, Arg–Lys, Val–Ile, His–Trp, Phe–Tyr. На рис. 5.2, δ кружки, в которых находятся боковые цепи, справа закрашены светлее, чем слева.

Согласно модели из 3.3 все операторы с близкими размерами, но гомологичными свойствами расположены симметрично сзади и спереди от плоскости II. Боковые цепи, близкие по размерам, но гомологичные по свойствам и обладающие незеркальной антисимметрией, также удалось расположить таким образом. За плоскостью II находятся Ser– Thr, Asp–Glu, Val–Ile, Phe–Tyr (на рис. 5.2, *в* они обозначены кружками светло-розового цвета). Перед плоскостью II расположены Cys–Met (гомологи Ser и Thr), Asn–Gln (гомологи Asp и Glu), Arg–Lys (они рассматриваются как гомологи Val и Ile) и His–Trp (гомологи Phe и Tyr). На рис. 5.2, *в* гомологи показаны кружками темно-розового цвета.

Модель, рассмотренная в 3.3, предполагает, что операторы с противоположными свойствами должны быть расположены поворотно симметрично относительно плоскости III. Расположение боковых цепей аминокислот на основе антисимметрии антиподов в практически сформированной модели структуры канонического набора боковых цепей на додекаэдре также удовлетворяет этому требованию. В самом деле, наибольшими различиями обладают боковые цепи, расположенные в вершинах верхней и нижней граней додекаэдра: Ser–His, Cys– Phe, Thr–Trp, Met–Tyr. Боковые цепи, расположенные в экваториальной области, имеют более близкие размеры, но обладают противоположными свойствами: Asp[–]–Arg⁺ и Glu[–]–Lys⁺. Нейтральные боковые цепи, способные образовывать водородные связи, оказались при этом симметричными неполярным боковым цепям: Asn–Val, Gln–Ile. На рис. 5.2, г положение более коротких цепей показано светло-голубыми кружками, а противоположных по длине и физическим свойствам бо-ковых цепей – темно-голубыми кружками.

Наконец, если противоположные вершины додекаэдра мысленно соединить диаметрами, то в парах оказываются боковые цепи, выделенные в подгруппу комплементарной антисимметрии: Ser–Trp, Thr–His, Asp–Lys, Glu–Arg и т. д. (см. табл. 5.3).

Отметим, что особые свойства пар Pro–Gly и Ala–Leu также сохранились: эти две пары боковых цепей при вращении верхней и нижней половин додекаэдра совмещаются между собой (антисимметрия антиподов), и они же находятся в противоположных вершинах додекаэдра, связанных диаметрами (антисимметрия комплементарности).

По аналогии с симметрией векторов (см. табл. 3.1) и симметрией подгрупп вершин додекаэдра (см. табл. 5.2) преобразования антисимметрии для аминокислот можно описать в табличной форме (табл. 5.4). При этом преобразование боковой цепи в саму себя обозначено цифрой 1, переход через плоскость I – буквой α, переход через плоскость II – буквой β, а переход через плоскость III – буквой γ. Переходы, осуществляемые через несколько плоскостей, были обозначены соответственно буквами αβ, αγ, βγ, αβγ.

Таблица 5.4

Номер	Преобразования антисимметрии								
подгруппы	1	α	β	αβ	γ	αγ	βγ	αβγ	
1	Gly	-	—	—	Pro	_	_	_	
2	Ala	_	_	_	Leu	_	_	_	
3	Ser	Thr	Cys	Met	His	Trp	Phe	Tyr	
4	Asp	Glu	Asn	Gln	Arg	Lys	Val	Ile	

Преобразования антисимметрии боковых цепей аминокислот

Из табл. 5.4 следует, что все аминокислоты распадаются на четыре подгруппы, выделенные также на рис. 5.2, *а* цветом. Подгруппы 1 и 2 включают по две пары аминокислот (Gly–Pro и Ala–Leu), находящихся в плоскости I и связанных вращением относительно плоскости III (преобразование γ). На рис. 5.2, *а* кружки с этими аминокислотами выделены серым цветом. Подгруппа 3 включает 8 аминокислот, связанных взаимными переходами и расположенных вблизи плоскости I додекаэдра (на рис. 5.2, *а* эти кружки окрашены розовым цветом). Наконец, в подгруппу 4 входят аминокислоты, локализованные в экваториальной области додекаэдра, вблизи плоскости III (на рис. 5.2, *а* кружки окрашены голубым цветом).

Следует отметить, что построенная на додекаэдре система канонического набора аминокислот, содержащая суперпозицию всех выделенных нами типов антисимметрии, является единственно возможной, и любые перемещения боковых цепей в этой структуре оказываются невозможными без нарушения того или иного типа антисимметрии. В целом, сам факт построения системы боковых цепей аминокислот на додекаэдре на принципе антисимметрии, положенном в основу разработанной формальной модели (см. 3.3), свидетельствует об адекватности принципа построения этой модели свойствам канонического набора аминокислот. Это позволило нам использовать систему аминокислот при построении молекулярной векторной машины белков (MBM_б).

5.4. Применения модели

Использование модели для построения системы модулей

Основные идеи систематизации функциональных групп боковых цепей аминокислот как функциональных модулей ССИВС были сформулированы в 1.2. Использование модели структуры канонического набора аминокислот на додекаэдре позволяет углубить понимание этого вопроса. Для этого функциональные группы боковых цепей аминокислот в составе ССИВС были помещены в вершины додекаэдра (рис. 5.3). Атомы водорода, на которые ориентированы стрелки, рассматриваются как входы (см. 1.2), а неподеленные пары электронов (для простоты они на рисунке не показаны) соответствуют выходам и обозначены стрелками, исходящими из модулей.

Функции аминокислот Gly и Pro. Аминокислота Gly не содержит боковой цепи, а ее аминогруппа участвует в образовании пептидной связи основной цепи белка (HN–C=O-группы). Эта группа практически не влияет и на процесс переноса заряда по системе HN–C=Oгрупп. Аминокислота Pro наоборот, как видно из рис. 5.3, не содержит входов, так как атом азота ее N–C=O-группы включен в цикл. В то же время атом кислорода, содержащий неподеленные пары электронов, может служить основой для инициации ССИВС (см. 1.2). Цикл Рго обусловливает прерывание цепи HN–C=O-групп в белке, т. е. функция Рго является обратной по отношению к функции Gly.



Рис. 5.3. Система функциональных групп аминокислот как модулей ССИВС, построенная на додекаэдре. Кружки с группами, связанными взаимными преобразованиями, обозначены цветами: пара Gly–Pro – белым; подгруппа 3 – светло-серым; подгруппа 4 – темно-серым

Функции других неполярных боковых цепей. Неполярные боковые цепи (Leu, Ala, Val, Ile), которые не показаны на рис. 5.3, на наш взгляд, имеют функции, аналогичные Pro, хотя направление их действия несколько отличается от направления действия Pro. Согласно 1.2 они играют роль пассивных элементов, создающих гидрофобную среду для ССИВС. Функции боковых цепей – активных модулей ССИВС. В 3.2 (при рассмотрении боковых цепей аминокислот как системы модулей) были выделены лишь 3 элемента с основными функциями: элемент ИЛИ, элемент НЕ и элемент задержки.

На структуре додекаэдра эти боковые цепи можно анализировать на основе количества входов и выходов. Как видно из рис. 5.3, при перемещении сверху вниз количество входов и выходов постепенно изменяется. В частности, С–ОН-группы Ser и Thr и C-SH-группа Cys имеют 1 вход и 2 выхода; О=С–ОН-группы Asp и Glu имеют 1 вход и 3 выхода, О=С–NH₂-группы Asn и Gln имеют 2 входа и 2 выхода.

Особенностью боковых цепей Lys и Arg является то, что Lys представлен простой группой C–NH₂, а Arg – системой, состоящей из резонансных групп N–C=N. Для Lys можно выделить 2 входа и 1 выход. В то же время, гуанидиновая группировка Arg потенциально имеет 4 входа и 1 выход (рис. 5.3). Однако анализ структур, содержащих Arg, показал, что атомы водорода гуанидиновой группировки часто образуют водородные связи парами:

$$1 \longrightarrow HQ_{1} \xrightarrow{R} X^{\Gamma} H^{N} H^{N} H^{A} X_{2} \xrightarrow{Q_{2}} Q_{2} H^{A} Z$$

$$(5.1)$$

В этом случае гуанидиновая группировка связывает лишь две системы, работающие на входы, и, как видно из схемы (5.1), выполняет функции элемента, имеющего 2 входа и 1 выход (аналогично С– NH₂-группе лизина).

Функции циклических боковых цепей справа и слева от плоскости I также близки. Боковые цепи Trp и His имеют по одному входу и одному выходу в пятичленный цикл (у Trp добавляется еще шестичленный цикл). Аналогично боковые цепи Tyr и Phe также могут иметь один вход и один выход, но в шестичленный цикл. При этом особенностью Phe является отсутствие в цикле гетероатома (связанного с атомом водорода), которому можно было бы приписать функцию входа. Эту функцию, по-видимому, может выполнять какой-либо атом углерода этого цикла.
Рассмотрение количества входов и выходов и их функций при переходе через плоскости имеет следующие особенности. При переходе через плоскость I функции модулей справа и слева от этой плоскости в большинстве случаев не меняются. При переходе через плоскость II в модулях, расположенных в верхней части додекаэдра (над плоскостью III), происходит замена гетероатомов (С–О \rightarrow С–S; О–С=О \rightarrow N–С=О), что сопровождается изменением их свойств. Например, при переходе от Thr к Met остаются только выходы, а при переходе от Ser к Cys увеличивается подвижность атома водорода. Модули Asp и Glu имеют 1 вход и 3 выхода, а симметричные относительно плоскости II модули Asn и Gln содержат 2 входа и 2 выхода. В нижней части додекаэдра также происходит изменение свойств модулей относительно плоскости II. Модулям Lys и Arg, содержащим по 2 выхода и по 1 входу, соответствуют боковые цепи IIе и Val, не имеющие активных групп, а пятичленные модули Trp и His меняются на шестичленные – Туг и Phe.

По отношению к плоскости III (см. рис. 5.3), имеющей ось вращения второго порядка (C_2), модули верхней и нижней частей додекаэдра обладают поворотной симметрией. При этом модулям с простыми группами С–О и С–S, имеющим 1 вход и 2 выхода, соответствуют при переходе через плоскость III циклические модули с 1 входом и 1 выходом, обладающие функциями элементов инверсии (HE) и элементов задержки (см. 1.2), например: Ser–His, Cys–Phe, Thr–Trp. В то же время, модулям Glu и Asp, имеющим 1 вход и 3 выхода, соответствуют модули Lys и Arg, имеющие 2 входа и 1 выход. Модулям Gln и Asn с 2 входами и 2 выходами соответствуют боковые цепи IIe и Val, не имеющие активных группировок. Иными словами: при переходе через плоскость III происходит инверсия функций модулей.

Таким образом, использование додекаэдра для систематизации боковых цепей как функциональных модулей ССИВС позволило выявить особенности поведения этих модулей относительно трех плоскостей антисимметрии. Переход через плоскость I не сопровождается изменением их функций. При переходе через плоскость II эти функции изменяются постепенно, а при переходе через плоскость III происходит их инверсия. Особенности канонического набора боковых цепей аминокислот как функциональных модулей, выявленные с помощью этой модели, могут быть использованы как при конструировании на основе этих модулей логических схем, так и при создании других систем модулей, пригодных для конструирования устройств бионической наноэлектроники.

Взгляд на природу минорных оснований РНК

К фактам, не нашедшим пока исчерпывающего объяснения, относится наличие в составе т-РНК минорных оснований. Особенно много их присутствует в областях изгиба т-РНК, в петлях. С позиции теории топологического кодирования цепных полимеров эти факты можно объяснить следующим образом.

Как следует из табл. 2.1, нуклеиновые кислоты (как и белки) можно отнести к классу линейных цепных полимеров. Но, в отличие от белков, в нуклеиновых кислотах соединительными группами (вместо пептидных групп) являются НО–Р=О-группы. Аналогами боковых цепей аминокислот в нуклеиновых кислотах могут служить нуклеотиды. Для объяснения роли минорных оснований в составе нуклеиновых кислот нами была выдвинута следующая гипотеза [28]. В составе нуклеотидов, как и в составе боковых цепей аминокислот, можно выделить два типа физических операторов:

 – операторы связности^{*}, которые способствуют формированию спиральной структуры нуклеиновых кислот;

– операторы антисвязности^{**}, которые препятствуют формированию спиральной структуры.

Рис. 5.4 иллюстрирует работу неметилированных и метилированных нуклеотидов в качестве физических операторов.

Принципы антисимметрии, положенные в основу модели структуры канонического набора аминокислот, могут быть использованы и при построении системы нуклеотидов, включающей немодифицированные и модифицированные минорные нуклеотиды. Минорные основания входят в состав нуклеотидов т-РНК (в среднем 10...12 основа-

^{*} Нуклеотиды с азотистыми основаниями, способными образовывать Н-связи с фосфатной группой.

^{**} Нуклеотиды с метилированными азотистыми основаниями, которые не способны образовывать Н-связи с фосфатной группой.



Puc. 5.4. Гипотетический фрагмент «α-спирали» из полинуклеотидной цепи: *a* – азотистое основание (гуанин) в качестве оператора связности; *б* – метилированное азотистое основание (N²-диметилгуанин) в качестве оператора антисвязности. Атомы кислорода цепи, не участвующие в водородной связи, – заштрихованные кружки; R₂–R₅ – азотистые основания; входы и выходы ССИВС показаны стрелками

ний на молекулу). Они представлены метилированными основаниями, изомерами и аналогами пуринов и пиримидинов. Считается, что минорные основания выполняют 2 функции:

- делают т-РНК устойчивыми к действию нуклеаз цитоплазмы;

 – поддерживают третичную структуру молекулы, поскольку не могут участвовать в образовании комплементарных пар [107].

С позиции теории топологического кодирования цепных полимеров к этому можно добавить возможные функции минорных оснований как физических операторов связности и антисвязности.

В настоящее время известно более 50 модифицированных азотистых оснований в составе транспортных РНК. Однако их систематизация [107], хотя и учитывает отнесение к одному из биологических царств (Archea, Bacteria, Eucariota), не принимает во внимание их видовую специфичность. Можно предположить, что необходимым условием построения систем нуклеотидов как физических операторов является использование для построения модели наборов нуклеотидов, полученных из одного вида (например: E.coli, S.cerevisiae и т. д.). При построении системы нуклеотидов нужно также учитывать возможность использования другого многогранника, поскольку с точки зрения квантовой химии описание орбиталей для HO–P=O-групп несколько сложнее, чем для пептидных связей, и может быть выделено другое количество векторов действия. Мы предприняли попытку построения такой структуры (в первом приближении) для нуклеотидов т-PHK E.coli (табл. 5.5) с использованием номенклатуры вершин додекаэдра.

Таблица 5.5

A ₁		¹ A
3-метилцитозин	А	5-метилцитозин
A ₂	Цитозин	² A
лизидин		5-гидроксиметилцитозин
B ₁		¹ B
2-тиоурацил	В	4-тиоурацил
B ₂	Урацил	$^{2}\mathrm{B}$
псевдоуридин		5-метилурацил
-B ₂		$-^{2}B$
1-метиладенин	-В	6-метиладенин
-B ₁	Аденин	$-^{1}B$
пример не найден		6,6-диметиладенин
-A ₁		$-^{1}A$
1-метилгуанин	-A	2,2-диметилгуанин
-A ₂	Гуанин	2A
7-метилгуанин		2-метилгуанин

Предварительная классификация минорных оснований транспортных РНК (для E.coli)

Основания, находящиеся в центре таблицы (предполагаемой плоскости I многогранника), – это исходные основания (С, U, G, A), образующие двойную спираль в нуклеиновых кислотах. Они расположены по принципу поворотной симметрии (вершины A; B;–B; –A). В модели канонического набора аминокислот аналогом этих четырех оснований в белках являются аминокислоты Gly, Pro, Ala, Leu, также находящиеся в плоскости I (см. рис. 5.2). Имеются структурные аналогии между нуклеиновыми кислотами и белками: двойная спираль ДНК и тройная спираль коллагена [108]. Интересно, что каждая третья аминокислота в тройной спирали коллагена – Gly, а многие другие аминокислоты в этой спирали представлены Pro и его близким аналогом – гидроксиPro. Несмотря на сугубо предварительный характер данной классификации, нельзя не отметить, что общее количество выделенных оснований близко к 20, а количество модификаций для каждого основания близко к 4. Кроме того, в составе набора можно выделить основания, потенциально пригодные и на роль операторов связности^{*}, и на роль операторов антисвязности^{**}. Можно надеяться, что более детальная проработка этой модели позволит уточнить как число модификаций для E.coli, так и их положение в данной системе (в том числе и на структуре выбранного для этих целей многогранника).

Таким образом, с использованием разработанных в гл. 3 и 4 теоретических положений и на основе принципов антисимметрии предложена модель пространственной структуры канонического набора аминокислот. Расположение боковых цепей аминокислот на додекаэдре является суперпозицией всех выделенных типов антисимметрии и единственным вариантом расположения, который не нарушает ни одного типа антисимметрии. Показано, что предложенная модель может быть применена для уточнения логических функций боковых цепей аминокислот в составе ССИВС. На примере построения системы нуклеотидов, содержащих главные и минорные азотистые основания, показана возможность использования данной модели в качестве прототипа для построения систем физических операторов, обеспечивающих самоорганизацию линейных цепных полимеров.

В то же время, в пределах данной системы оказалось невозможным понять саму природу канонического набора аминокислот. Это было сделано лишь тогда, когда она стала составной частью молекулярной векторной машины белков, которая будет описана в следующей главе.

^{*} То есть способные образовывать Н-связи, например неметилированные и монометилированные по аминогруппе основания.

^{**} В частности, дважды метилированные по аминогруппе.

Глава 6. МОЛЕКУЛЯРНАЯ ВЕКТОРНАЯ МАШИНА БЕЛКОВ

6.1. Природа молекулярной векторной машины белков

Необходимость разработки модели молекулярной векторной машины белков

Не всегда разработка новых теоретических положений происходит синхронно с появлением новых экспериментальных результатов. Часто теория намного опережает возможности эксперимента и довольно долго ждет своего часа. В этом смысле теории топологического кодирования «повезло». Объективный взгляд на биосинтез белка и самоорганизацию его надмолекулярной структуры (включая вторичную) приводит к мысли о существовании в клетках отработанного в процессе эволюции механизма, обеспечивающего точное воссоздание структуры белка. Иначе и быть не может, поскольку в противном случае биосистемы моментально погибли бы вследствие ошибок самоорганизации и неправильной работы биоструктур.

Развивавшиеся до недавнего времени подходы к изучению процессов самоорганизации in vitro (в разбавленных растворах) кажутся с этих позиций наивными и далекими от реального положения дел. Краху этих подходов способствовало появление данных о ко-трансляционном механизме самоорганизации белка в структуре рибосомы (см. 1.1).

Развивая идеи ко-трансляционного механизма сворачивания белка логически нетрудно прийти к предположению о существовании согласованности всех элементов системы биосинтеза^{*}, этапов переноса информации и процесса воссоздания надмолекулярной структуры белка внутри рибосомы. Нетрудно также прийти к идее существования специального «устройства», работающего на определенных принципах, которое осуществляет правильную укладку структуры вновь синтезированных фрагментов белка. Логика теории топологического кодирования цепных полимеров также приводит к идее подобного «устройства». Таким образом, появление в процессе развития нашей теории идеи молекулярной векторной машины имеет четкую структурную направленность – это начальные этапы биосинтеза белковой молекулы при участии рибосомы.

^{*} Включая генетический код и канонический набор аминокислот.

Переход от общей модели МВМ к модели МВМ_б

Общая модель MBM появилась как результат анализа области формирования водородной связи $Q_iH...X_{i-4}$ =R в пентафрагментах цепного полимера [1], [34]–[37]. Чтобы осуществить переход к MBM₆, достаточно заменить абстрактные атомы Q, R и X на конкретные атомы элементов-органогенов N, C и O [37]. Проведенный нами анализ области N_iH...O_{i-4}=C-связи выявил те же 20 векторов, что и в общей модели MBM. Описание преобразований симметрии векторов MBM (см. табл. 3.1) совпадает с описанием преобразований симметрии вершин додекаэдра (см. табл. 5.2) и описанием преобразований антисимметрии канонического набора боковых цепей аминокислот в модели на додекаэдре (см. табл. 5.4). Вследствие этого стало возможным вместо формальных обозначений вершин додекаэдра, в которые направлены векторы MBM, подставить названия аминокислот, расположенных в соответствующих вершинах додекаэдра в модели структуры канонического набора аминокислот (рис. 6.1).

Вершина, к которой приписана аминокислота Gly, была помещена в атом азота группы N_iH, которая соответствует NH-группе Gly, тогда как α_i -углеродный атом основной цепи совпал с α -атомом Gly. В результате переноса структуры додекаэдра все 20 выделенных ранее векторов (10 диаметрально противоположных пар) оказались радиусами додекаэдра, причем одна из точек каждого вектора соответствует атому кислорода O_{i-4}, а вторая – той или иной вершине додекаэдра. Таким образом, возникла модель MBM_б (рис. 6.1), состоящая, как и общая модель MBM (см. 3.1), из трех элементов:

– системы векторов внутри додека
эдра, исходящих из атома O_{i-4} как из центра;

 – канонического набора боковых цепей аминокислот, сменяемых у α_i-углеродного атома в процессе наращивания полипептидной цепи;

- тетраэдрического α_{*i*}-углеродного атома.

Анализ свойств этих элементов упрощается тем, что аналогичный анализ уже был проведен для общей модели MBM (см. 3.2–3.4).

Отдельного рассмотрения заслуживает еще один элемент MBM_б, кажущийся само собой разумеющимся: это сам пентафрагмент. На формальной модели (см. 3.1) сами пентафрагменты, в принципе, не



Рис. 6.1. Модель молекулярной векторной машины белков

могли быть проанализированы. В то же время, в случае конкретных белков существует потребность в таком анализе. В конце этой главы будут рассмотрены пути решения данной проблемы.

6.2. Анализ элементов структуры молекулярной векторной машины белков

Свойства векторов додекаэдра

Свойства векторов додекаэдра, в принципе, уже были рассмотрены в 3.2. Единственный аспект, на который хотелось бы обратить внимание, это интерпретация свойств векторов, соответствующих нейтральному и обратному элементам. Рассмотрение этих векторов с позиции теоретико-группового подхода позволяет говорить о том, что действие вектора – нейтрального элемента, направленного на вершину Gly, расположенную в атоме N_i , – всегда должно приводить к формированию циклического четырехзвенного фрагмента белка. В то же время, действие противоположного вектора – обратного элемента – всегда должно быть направлено на разрушение четырехзвенного фрагмента. Этот вектор направлен на вершину, соответствующую аминокислоте Pro. Ниже рассмотрены особенности физических операторов боковых цепей аминокислот, реализующих действие этих векторов.

Боковые цепи аминокислот как неприводимые представления группы векторов

Как отмечалось в 3.3, каждому вектору группы из 20 векторов должен соответствовать один физический оператор, воссоздающий его действие. Взгляд на боковые цепи аминокислот как на физические операторы позволяет рассматривать их в качестве неприводимых (т. е. простейших) представлений группы векторов додекаэдра. Анализ боковых цепей с этих позиций позволяет в значительной степени объяснить природу канонического набора аминокислот.

Особые свойства аминокислот Gly и Pro. Как отмечалось в 3.3, вектор, являющийся в группе векторов нейтральным элементом, должен воссоздаваться физическим оператором, не влияющим на структуру четырехзвенного фрагмента. Это означает, что такой оператор может вообще не иметь боковой цепи, а являться «пустым» звеном в цепи полимера. По свойствам он должен относиться к подгруппе операторов связности. Особые свойства аминокислоты Gly идеально соответствуют этим требованиям. Из всего набора аминокислот только Gly не имеет боковой цепи. Ее отсутствие приводит к тому, что связь $N_iH...O_{i-4}=C$ оказывается нерегулируемой, а это является свойством нейтрального элемента. Известна большая конформационная подвижность Gly [38].

Вектор, соответствующий обратному элементу группы, имеет противоположное действие, приводящее к разрыву связи $Q_iH...X_{i-4}=R$ (см. 3.3). Его может реализовать такой физический оператор, у которого атом Q_i в принципе не может образовать водородную связь с группой $X_{i-4}=R$. Как отмечалось в 3.3, такая ситуация возможна, например,

если атом Q_i заключен в циклическую структуру. В MBM_б этот вектор направлен в вершину, соответствующую аминокислоте Pro. Особенность этой аминокислоты состоит в том, что атом N_i включен в пятичленный цикл и ни при каких условиях не может образовать водородную связь с группой O_{i-4} =С. В процессе работы этой аминокислоты в качестве физического оператора атом O_{i-4} может приобрести направление, обеспечивающее разрушение циклической структуры четырехзвенного фрагмента в результате «лобового» столкновения его электронной оболочки с электронной оболочкой CH₂-групп пятичленного цикла.

Длина боковой цепи должна увеличиваться по мере изменения направления вектора в сторону атома O_{i-4} . Боковые цепи аминокислот как физические операторы, воссоздающие векторы, направленные в сторону атома N_i (см. рис. 6.1), должны быть короче боковых цепей, воссоздающих векторы, направленные в сторону O_{i-4} . Расположение аминокислот на модели структуры канонического набора аминокислот в порядке увеличения размера обеспечило реализацию этого требования MBM.

Боковые цепи аминокислот, воссоздающие векторы, направленные вправо от плоскости I, должны быть более короткими. Положение C_i^{α} -атома, к которому прикрепляются боковые цепи аминокислот в качестве физических операторов MBM₆ (см. рис. 6.1), несколько асимметрично и сдвинуто вправо относительно связи N_iH...O_{i-4}=C (ср. с 3.3). Это означает, что боковые цепи, воссоздающие векторы, направленные вправо от плоскости I, должны быть короче, чем боковые цепи, реализующие действие векторов, направленных влево от этой плоскости. Данная закономерность должна проявляться в отношении всех боковых цепей, что должно приводить к формированию из них двух групп, обладающих близкими свойствами, но разной длиной. Именно поэтому канонический набор аминокислот состоит из двух групп, связанных преобразованием квазизеркальной антисимметрии (α), которые расположены на структуре додекаэдра симметрично относительно плоскости I (см. рис. 5.2, *а* и б).

Близкие по размеру, но разные по свойствам боковые цепи должны воссоздавать векторы, расположенные симметрично по отноше-

118

нию к плоскости II. Как видно из рис. 6.1, расстояния от C_i^{α} -атома до вершин, в которые направлены векторы, находящиеся за плоскостью II (направлены от нас) и перед плоскостью II (направлены к нам), не очень сильно отличаются. По этой причине боковые цепи, воссоздающие эти векторы, не должны существенно отличаться и в размере. Их отличия могут состоять лишь в замене терминальных групп. И действительно, пары аминокислот, расположенные перед плоскостью II и за ней и связанные преобразованием незеркальной антисимметрии (β), близки по размеру, но отличаются функциональными свойствами (см. 5.3).

Векторы, расположенные над плоскостью III, воссоздаются более короткими боковыми цепями аминокислот. Как видно из рис. 6.1, вершины векторов, направленных в сторону группы HN_{i} -C=O_{i-1}, расположенные над плоскостью III, находятся значительно ближе к C_i^{α} -атому, чем вершины векторов, направленных в противоположную сторону – к атому С_{і-4} (он находится под плоскостью III). По этой причине все векторы, находящиеся над плоскостью III, должны воссоздаваться более короткими боковыми цепями аминокислот (физическими операторами), чем векторы, находящиеся под плоскостью III. Поскольку эти векторы связаны операцией вращения вокруг оси С₂(γ), то аминокислоты с самыми короткими боковыми цепями должны противостоять аминокислотам с самыми длинными боковыми цепями. В то же время боковые цепи, воссоздающие векторы, расположенные в экваториальной области додекаэдра, должны отличаться в меньшей степени (например, на одно звено в цепи). Именно это мы наблюдаем на примере антисимметрии антиподов в модели структуры канонического набора аминокислот на додеказдре (см. 5.3).

Свойства боковых цепей, воссоздающих векторы противоположного направления. Боковые цепи, воссоздающие векторы противоположного направления, должны обладать свойствами дополнительности. Это связано с тем, что наклон оси этих векторов один и тот же и вместе они составляют его диаметр. При этом, поскольку диаметры соединяют вершины додекаэдра, расположенные по разные стороны от плоскости I, то боковые цепи, находящиеся на одной оси справа от плоскости I, будут максимально приближены к C_i^{α} -атому, а находящиеся слева от этой плоскости будут максимально удалены от C_i^{α} -атома. Это позволяет понять, почему в противоположных вершинах додекаэдра в модели структуры канонического набора аминокислот, если соединить их диаметрами (см. рис. 5.2), находятся боковые цепи, свойства которых дополнительны [37]. Такой тип антисимметрии боковых цепей можно назвать антисимметрией коплементарности. Самую короткую полярную цепь (Ser) можно связать диаметром с самой массивной (Trp), а меньшую, по сравнению с Trp, цепь (His) – с более массивной, по сравнению с Ser, цепью (Thr). Минимальная по размеру отрицательно заряженная боковая цепь Asp находится на одном диаметре с максимальной по длине положительно заряженной цепью Lys, а более короткая цепь Arg связана с более длинной цепью Glu и т. д. Можно ожидать, что направление связи $C_i^{\alpha} - C_{i+1}^{\alpha}$, задаваемое каждой из боковых цепей, воссоздающих векторы одинакового наклона, должно быть одним и тем же.

Таким образом, анализ свойств боковых цепей аминокислот в структуре MBM_6 в рамках теоретико-группового подхода позволяет в значительной степени объяснить природу канонического набора аминокислот. Например, становится понятным, почему в группе боковых цепей аминокислот присутствуют такие аномальные, на первый взгляд, боковые цепи, как Gly и Pro. Кроме того, этот анализ позволил выявить природу антисимметрии боковых цепей аминокислот.

Свойства тетраэдрического С_i^a-атома

Общие свойства тетраэдрического α -атома, описанные в 3.4, полностью могут быть отнесены и к свойствам C_i^{α} -атомов белковой цепи. По этой причине не будем повторять содержание 3.4, а сделаем лишь три дополнения.

Первое дополнение касается структуры MBM_б, которая возникает при формировании водородных связей $N_{i+1}H$ группы белка с основной цепью. Очевидно, что при образовании связи $N_{i+1}H...O_{i-3}$ будет воспроизводиться структура α -спирали и структура MBM останется неизменной. Однако уже при образовании связи $N_{i+1}H...O_{i-2}$ в этой области будет происходить редукция. В самом деле, анализ более сотни

белков показал, что в этой области практически не обнаруживаются такие аминокислоты, как Gln, Lys, Tyr, Trp, т. е. именно те, которые занимают левый нижний край додекаэдра в MBM, как было предсказано в 3.4. Что касается образования связи $N_{i+1}H...O_{i-1}$, то крутизна этого участка не позволяет MBM вписаться в эту область, за исключением фрагмента самой цепи. И действительно, в этой области встречается практически только одна аминокислота – Gly [38].

Второе дополнение касается формирования групп боковых цепей, задающих одинаковое направление связи $C_i^{\alpha} - C_{i+1}$. Вполне вероятно, что такое направление должны задавать боковые цепи, воссоздающие симметричные векторы. Например, такую группу образуют аминокислоты Asp, Glu, Lys и Arg. Действительно, эти аминокислоты часто встречаются в α -спиральных участках [21], и можно ожидать, что они сходным образом детерминируют направление $C_i^{\alpha} - C_{i+1}^{\alpha}$, способствующее образованию связи $N_{i+1}H...O_{i-3}$.

Третье дополнение касается учета вариантов, связанных с переходом не только от связи $N_iH...O_{i-4}$ к связи $N_{i+1}H...O_{i-2}$ или $N_{i+1}H...O_{i-1}$, но и в противоположном направлении, т. е. от связи $N_{i+1}H...O_{i-1}$ к $N_{i+1}H...O_{i-2}$ или от связи $N_{i+1}H...O_{i-2}$ к связи $N_iH...O_{i-4}$. Эти варианты нами пока не рассматривались. Можно ожидать, что в таком случае возникнет проблема сложения векторов, поэтому необходимость в анализе этих вопросов существует.

Вырожденность свойств боковых цепей аминокислот и вырожденность триплетов генетического кода. Двухъярусная модель MBM

В нашей модели (см. 3.1) было выделено лишь 20 векторов и 20 физических операторов для их воссоздания. В то же время, в топологическом коде могут существовать порядка 62 значащих триплетов, кодирующих 62 конформации цепного полимера. Им должно соответствовать 62 вектора, для воссоздания которых необходимо 62 физических оператора. Для решения этого несоответствия было предположено, что боковые цепи физических операторов должны обладать вырожденными состояниями, которые и обеспечивают воссоздание всех конформаций топологического кода. В современном генетическом коде лишь три триплета являются стоп-кодонами (кодируют окончание синтеза), а остальным триплетам соответствуют определенные аминокислоты [23]–[25]. При этом код является вырожденным, и одной аминокислоте в коде часто соответствует несколько значащих триплетов. Так, большинство полярных аминокислот, содержащих резонансные группы в боковой цепи, кодируется двумя триплетами. Перечислим их (см. рис. 4.7): His, Gln, Asp, Glu, Tyr, Asn, Lys. В 3.3 предпологалось, что поскольку большинство операторов связности может иметь два варианта водородных связей, то этим вариантам должны соответствовать две конформации, закодированные триплетами. Эта идея вполне оправдывается для блока А.

В то же время в блоке G боковая цепь Arg кодируется шестью триплетами. Если наше предположение справедливо, то эта аминокислота должна иметь не два, а шесть вариантов водородных связей. Для проверки этого предложения был выполнен анализ ПФ, образованных боковой цепью Arg с помощью программы Protein 3D. Результаты этого анализа для вариантов водородных связей, образованных боковой цепью Arg с O_{i-4} -атомом, представлены на рис. 6.2. Для сопоставления вверху приведены все триплеты, кодирующие Arg.

Четыре триплета, кодирующих Arg, содержат дуплет CG, а два триплета – дуплет AG. Последние два триплета, очевидно, имеют наибольшую связность (111010 и 111011). Проведенный анализ водородных связей, образованных боковой цепью Arg, позволил обнаружить в белках следующие 6 вариантов этих связей: NE; NH1; NH2; NE, NH1 + NE, NH2; NH1, NH2; NE, NH1, NH2 (рис. 6.2). Варианты NE, NH1 и NE, NH2 были практически идентичны и потому их объединили. Степени распространения различных типов циклов достаточно сильно отличаются.

Как видно из рис. 6.2, все эти варианты, если смотреть на них в профиль, имеют наклон водородной связи вправо, как и предсказывает модель MBM. Если же взглянуть на водородные связи Arg сверху (рисунки даны в гл. 9 [1]), то все они, как и ожидалось в рамках модели MBM, расположены справа от плоскости I. Наибольшей связностью обладают, вероятно, варианты ∂ и *е*. В целом, как мы и предполагали, 6 триплетам, кодирующим аминокислоту Arg, соответствует 6 возможных сочетаний водородных связей, образуемых ее боковой цепью. Однако вопрос о том, существует ли корреляция между тем или иным трип-



Рис. 6.2. Триплеты, кодирующие Arg, и варианты водородных связей, образованные атомами боковой цепи Arg с O_{i-4} -атомом: $a - NE; \delta - NH1; e - NH2; e - NE, NH1 + NE, NH2; <math>\partial - NH1, NH2; e - NE, NH1, NH2$

летом и тем или иным сочетанием водородных связей боковой цепи (в частности, для Arg), требует дальнейшего изучения. Такой связи может и не быть, поскольку, в отличие от соответствия «триплет – аминокислота», которое задается в последовательности триплетов, однозначной связи между вырожденностью триплетов и типом водородной связи боковой цепи нет.

Обсудим вопрос о вырожденности неполярных боковых цепей. В генетическом коде степень вырожденности триплетов, кодирующих эти цепи, составляет от 6 для Leu до 3 для Ile. В отличие от водородных связей вопрос о различных состояниях этих боковых цепей не так ясен. Можно предполагать, что из всех неполярных цепей боковая цепь Leu, благодаря присутствию β -CH₂-группы, может иметь наибольшее число конформаций. При соударении с О_{*i*-4}-атомом может возникнуть пучок векторов, соответствующий шести триплетам кода. В то же время, остальные неполярные цепи могут создавать меньшее количество конформаций и воссоздавать меньшее количество векторов.

Высказанные предположения приводят к модели двухъярусной структуры MBM (см. 3.3), в которой нижним ярусом может служить ромбоикосододекаэдр (см. 3.6). Однако дальнейшее обсуждение этого вопроса предполагает построение MBM на основе реальных данных. Принципиальных препятствий к разработке этого вопроса нет. Отметим также, что сделанные на основе анализа MBM выводы можно проверить с помощью данных, полученных при РСА белков. Это касается в первую очередь свойств боковых цепей, образующих водородные связи с O_{i-4} -атомом. Эти данные приведены в [1, 9.2].

Роль пентафрагментов как элементов МВМ_б

Хотя большая часть событий, связанных с работой физических операторов, разыгрывается в области связи $N_iH...O_{i-4}$, нельзя не отметить, что структура и состав пентафрагментов, на которых, собственно, все и держится, также имеет большое значение. В частности, наличие в (i - 4)-м положении боковых цепей, способных образовывать H-связи, при появлении в *i*-м положении боковой цепи, обладающей подобными свойствами (например, пары C–N⁺H...⁻O–C=O), может обеспечить образование циклической конформации, несмотря на отсутствие склонности к формированию спирали у предшествующих боковых цепей. Рабо-

та MBM_6 оказывается связанной с логикой появления тех или иных боковых цепей в процессе их биосинтеза. Так, если в (i - 4)-м положении находится боковая цепь A, а в *i*-м положении – боковая цепь B, то возникает структура S. В то же время, если в (i - 4)-м положении находится боковая цепь C, а в *i*-м положении – боковая цепь B, то возникает структура AS. Частично этот вопрос обсуждался нами в [1, 9.2.3.4], однако он требует более внимательного и углубленного анализа.

6.3. Перспективы развития модели молекулярной векторной машины белков

Существует два пути дальнейшей разработки модели МВМб.

Первый путь – это компьютерное моделирование MBM_б. Какихлибо принципиальных препятствий на этом пути не видно. Использование компьютерной модели позволит, в частности, провести детальный анализ логики формирования циклических и ациклических структур основной цепи с учетом положения боковых цепей на основной цепи. В принципе, этот путь может привести к разработке прямого способа прогнозирования как вторичной, так и третичной структур белка.

Второй путь, технически менее сложный, но более трудоемкий, связан с выделением пентафрагментов вместе с Н-связями и их систематизацией. По сути это значит анализировать не механизм работы MBM, а результаты ее работы и строить модель механизма по этим результатам. Такой путь оказался для нас более приемлемым. В результате была создана база данных пентафрагментов, сортированных по количеству и положению в них водородных связей. На ее основе были разработаны метод прогнозирования вторичной структуры белка по заданной первичной структуре и метод проектирования первичной структуры белка по заданной вторичной структуре. Изложению этих вопросов посвящена ч. 3 настоящего издания.

Вместе с тем, развитие этого направления не ограничено только работой с основной цепью. Последующий анализ других типов взаимодействий (например, взаимодействий при формировании спиралей 3₁₀, взаимодействий боковых цепей с основной цепью и боковых цепей между собой) может привести к решению проблем формирования третичной структуры белка. Это направление представляется наиболее перспективным и практически значимым.

Часть 3. ПРАКТИЧЕСКИЕ ПРИМЕНЕНИЯ ТЕОРИИ

Глава 7. БАЗА ДАННЫХ ПЕНТАФРАГМЕНТОВ БЕЛКА

7.1. Обоснование перехода к практическому использованию теории

Фрагменты из пяти аминокислот – новый уровень организации белков

В первой части была сформулирована общая теория топологического кодирования цепных полимеров. Во второй части на примере генетического кода и модели молекулярной векторной машины белков была показана возможность применения этой теории к биоструктурам. При этом как в исходной теории, так и в ее применениях отправной точкой построений являлся пентафрагмент. Перечислим основные причины использования в этом качестве именно ПФ:

1. Анализ конформаций, проведенный на основе аналогов ПФ – четырехзвенных цепных графов, показал, что количество возможных конформаций составляет 64 и они описываются с помощью 64 матриц из шести переменных, образуя суперматрицу. Кодированием трех пар переменных с помощью четырехбуквенного алфавита она трансформируется в 64 триплета топологического кода.

2. Внутри ПФ имеется одна водородная связь, на которую может воздействовать боковая цепь полимера. Это приводит к появлению физических операторов двух типов: операторов связности и операторов антисвязности, способствующих формированию соответственно циклических и ациклических ПФ. Показано, что соответствие «триплет – боковая цепь» полимера (белка) можно объяснить в рамках этих представлений.

3. Детальный анализ водородной связи $Q_iH...X_{i-4}=R$ (в белках – $N_iH...O_{i-4}=C$), возникающей при образовании циклического ПФ, привел к модели MBM. В рамках такой модели в этой области были выделены 20 векторов действия, которые могут быть воссозданы с помощью 20 боковых цепей, рассматриваемых в качестве неприводимых представлений этих векторов.

Таким образом, во всех перечисленных узловых моментах наших теоретических представлений фигурируют ПФ. В применении к бел-

кам можно сказать, что ПФ выполняют функцию промежуточного уровня между его первичной и вторичной структурами. Именно анализ этого переходного уровня и позволил объяснить многие особенности биоструктур. По этой причине следует полагать, что фрагменты из пяти аминокислот (пентафрагменты) можно выделить в качестве нового уровня структурной организации белков [30], [36].

Возможные пути анализа состава ПФ

В конце гл. 6 были рассмотрены пути развития модели MBM_{6} . Один из них – это компьютерное моделирование MBM_{6} . К сожалению, первые попытки компьютерного моделирования [109] пока не дали ощутимых результатов. Однако это, скорее, связано не с «порочностью» самой идеи, а с достигнутым техническим уровнем моделирования.

Другой путь, которым пошли мы, связан с выделением ПФ вместе с описанием их H-связей и последующей систематизацией. В результате была создана база данных ПФ, сортированных по числу и положению в них водородных связей. С ее использованием был разработан способ прогнозирования вторичной структуры белка по заданной первичной структуре и метод проектирования первичной структуры белка по заданной вторичной структуре. Полученные в процессе развития этого направления свидетельства о регистрации базы данных и компьютерных программ, а также патенты [110]–[114] свидетельствуют о перспективности выбранного пути. В данной главе изложен процесс создания базы данных ПФ. Отдельные этапы этой процедуры, проведенной для ПФ, выделенных из 540 файлов белков Protein Data Bank, были опубликованы [115]–[117]. В расширенном виде (для выборки в 2500 белков) эти результаты изложены в [118].

Краткий обзор подходов к созданию баз данных белковых фрагментов

Для предсказания вторичной структуры белка и конструирования первичных структур белка с заданной вторичной и третичной структурами часто создают библиотеки белковых фрагментов [119]–[121]. Такие библиотеки основаны на различных принципах и содержат фрагменты с длиной от 3 до 14 аминокислотных остатков. Одним из способов получения фрагментов является нарезание фрагментов, перекрывающихся со сдвигом на одну аминокислоту [120], [121]. Для описания конформаций таких фрагментов обычно используют двугранные углы между связями основной цепи (φ, φ, ω) [38]. Длина фрагментов, используемых при создании библиотек, как правило, диктуется задачами исследования и обычно не имеет какого-либо фундаментального теоретического обоснования.

Перечисленные ранее причины использования ПФ в качестве элементарных единиц белковой структуры показывают, что выбор фрагментов именно из 5 остатков аминокислот имеет для построения базы данных фундаментальное значение. Однако, как оказалось, не менее важным является и способ описания конформации этих фрагментов. Косвенно такое описание связано с работой MBM и обусловлено связностью белковой молекулы, аналогом которой может служить *n*-звенный линейный цепной граф. Избранный нами способ описания на основе бинарного кодирования связей позволил по-новому систематизировать ПФ, а построенная система создала условия для разработки последующих методов прогнозирования и проектирования структуры белков.

7.2. Получение и сортировка пентафрагментов

Исходные файлы для получения ПФ и этапы их последующей обработки

Для нарезки ПФ использовались текстовые файлы (табл. 7.1), описывающие H-связи вторичной структуры белков. Эти файлы получали на основе pdb-файлов 2500 белков Protein Data Bank с помощью визуализатора белковых структур PROTEIN 3D [122]–[124]. Как видно из приведенного примера, последовательность аминокислот в текстовых файлах записана снизу вверх, т. е. от N-конца к C-концу (именно так происходит рост белковой цепи при биосинтезе), и содержит участки с различным числом H-связей. Так, в данном примере снизу вверх последовательно представлены: фрагмент с β -структурой (нет H-связей ближнего порядка); переходный участок « β -структура – α -спираль», в котором возникают H-связи типа C = O_{i-4} ...HN_i; фрагмент с α -спиралью; переходный участок « α -спираль – β -структура» и снова фрагмент с β -структурой. В табл. 7.1 (третий столбец) приведены общие обозначения типов H-связей, использованных в txt-файлах. В чет-

Таблица 7.1

	100
Пример текстового файла,	получаемого с помощью программы
PROTEIN 3D (и	ісходный файл 1CSN.pdb)

Фрагмент	Тип	Общее обозначение	Описание
текстового файла	структуры	Н-связей в txt-файле	Н-связей
68 Ile	1,5 ,51	1	
67 Gly			
66 Thr	В-структура	X_1X_2 Abc	00
65 Cvs	p cipjnijpu	1 2	00
64 Gly			
63 Ala N – 59 Tyr O			
63 Ala			
62 Leu N – 58 Thr O			
62 Leu	Переход	X_1X_2 Abc N – Y_3Y_4 Ehf O	
61 Leu N – 57 Arg O	«α-спираль –	X_1X_2 Abc	10
61 Leu	β-структура»		
60 Lys N – 56 Tyr O			
60 Lys			
59 Tyr O – 63 Ala N			
59 Tyr N – 55 Glu O			
59 Tyr			
58 Thr O – 62 Leu N			
58 Thr N – 54 Asp O			
58 Thr		X_1X_2 Abc O – Y_1Y_2 Deh N	
57 Arg O – 61 Leu N	а-спираль	X_1X_2 Abc $N - Y_2Y_4$ Ehf O	11
57 Arg N – 53 Arg O	a empand	$X_1 X_2$ Abc	
57 Arg		$A_1 A_2 A b c$	
56 Tyr O – 60 Lys N			
56 Tyr N – 52 Leu O			
56 Tyr			
55 Glu O – 59 Tyr N			
55 Glu N – 51 Gln O			
55 Glu			
54 Asp O – 58 Thr N	Переход		
54 Asp	ив отруктура	X_1X_2 Abc O – Y_1Y_2 Deh N	01
53 Arg O – 57 Arg N	«р-структура –	$X_1 X_2$ Abc	01
53 Arg	а-спираль»		
52 Leu O – 56 Tyr N			
52 Leu			
51 Gln O – 55 Glu N			
51 Gln			
50 Pro		X X A1	
49 Ala	β-структура	$X_1 X_2$ Abc	00
48 Asp			
47 Ser			
46 Arg			

вертом столбце таблицы дано их закодированное описание в виде пар переменных. Отсутствие H-связей описывается парой переменных 00, связь О...HN – парой 01, связь NH...O – парой 10, а связи О...HN и NH...O – парой 11.

Поскольку в ПФ имеется пять пар возможных связей, принадлежащих α-углеродным атомам групп, и каждая пара может принимать одно из четырех булевых значений, то в результате ПФ получают описания в виде десятизначных булевых чисел.

Последующую обработку полученных текстовых файлов (нарезку, сортировку и упрощение их структуры) проводили с помощью специальных мини-программ, разработанных А. И. Беляевым. Список этих программ и их свойства представлены в табл. 7.2. В результате обработки каждый ПФ содержал лишь название файла, из которого он получен, номера аминокислот (снизу вверх) в исходном белке и последовательность аминокислот в выделенном фрагменте (табл. 7.3).

Таблица 7.2

Программы, использованные при обработке текстовых файлов для получения пентафрагментов

N⁰	Программы	Результат работы программ
1	PROTEIN 3D	Текстовые файлы на основе PDB-файлов
2	cutter.exe	Нарезка файлов на пентафрагменты
3	selectorexe	Сортировка ПФ на папки по числу строк
4	separator.exe	Сортировка ПФ с одинаковым числом строк на файлы с индивидуальными видами ПФ
5	simplificator. exe	Упрощение пентафрагментов внутри файлов, удаление описания связей

Таблица 7.3

Исходный и окончательный вид пентафрагментов после обработки с помощью мини-программ

Исходный вид	Окончательный вид
lcsn 56 Tyr O – 60 Lys N 56 Tyr N – 52 Leu O 56 Tyr 55 Glu O – 59 Tyr N 55 Glu N – 51 Gln O 55 Glu 54 Asp O – 58 Thr N 54 Asp 53 Arg O – 57 Arg N 53 Arg 52 Leu O – 56 Tyr N 52 Leu	1csn 56 Tyr 55 Glu 54 Asp 53 Arg 52 Leu

Оставшейся информации достаточно, чтобы найти и оценить конформацию ПФ в исходной структуре белка. Рассмотрим более детально описанные выше процедуры.

Нарезка и обозначение пентафрагментов

Нарезку текстовых файлов на ПФ проводили с помощью программы cutter.exe со сдвигом в одну аминокислоту, начиная с первого нижнего ПФ (табл. 7.4). При этом их обозначение в виде 10-значных булевых чисел проводили сверху вниз, начиная с *i*-го атома, и записывали в одной строке слева направо (показаны в нижней части табл. 7.4). В процессе нарезки ПФ на каждом этапе в их описании к десятизначному

Таблица 7.4

Фрагмент текстового файла 1CSN	Обозна- чения	Э	тапы нарезки пентафр	рагментов
61 Leu N – 57 Arg O	← 10			61 Leu N – 57 Arg O
61 Leu 60 Lys N – 56 Tyr O 60 Lys	← 10			61 Leu 60 Lys N – 56 Tyr O 60 Lys
59 Tyr O – 63 Ala N 59 Tyr N – 55 Glu O 59 Tyr	← 11		→	59 Tyr O – 63 Ala N 59 Tyr N – 55 Glu O 59 Tyr
58 Thr O – 62 Leu N 58 Thr N – 54 Asp O 58 Thr	← 11		2	58 Thr O – 62 Leu N 58 Thr N – 54 Asp O 58 Thr
57 Arg O – 61 Leu N 57 Arg N – 53 Arg O 57 Arg	← 11			57 Arg O – 61 Leu N 57 Arg N – 53 Arg O 57 Arg
51 Gln O – 55 Glu N 51 Gln	← 01	→	51 Gln O – 55 Glu N 51 Gln	
50 Pro	← 00	50 Pro	50 Pro	
49 Ala	← 00	49 Ala	49 Ala	
48 Asp	← 00	48 Asp	48 Asp	
47 Ser	← 00	47 Ser	47 Ser	
46 Arg	← 00	46 Arg		
Десятизначное обозначание		0000000000	01 00000000	10 10111111
Описание			1010111111 61	
конформации			1011111111 60)
нарезаемых подряд			111111111 59)
пентафрагментов				
			0101010100 54	L
			0101010000 53	
			0101000000 52	
			010000000 51	
			0000000000 50	

Пример процедуры нарезки на пентафрагменты (белок 1SCN)

числу добавлялась одна пара переменных слева и удалялась одна пара переменных справа (табл. 7.4, столбец внизу). Таким образом, каждый ПФ получал свое обозначение в виде 10-значного числа. Общее число 10-значных обозначений на основе четырех вариантов пар булевых переменных – 1024.

Принципы сортировки пентафрагментов

В процессе нарезки ПФ было замечено, что сортировку ПФ удобно проводить на основе иерархической организации.

Сортировка в папки. Самый верхний уровень такой организации – это класс пентафрагмента, обозначенный на основе первой и пятой пар переменных, описывающих связи ПФ. Обозначения классов присваивали общим папкам. Всего было получено 16 классов ПФ, которые подразделились на два типа: основные и минорные (табл. 7.5).

Таблица 7.5

Цва типа папон	к, в кото	рые помещен	ны пентафрагменть
----------------	-----------	-------------	-------------------

Основные ПФ	Минорные ПФ
00–00	00–10
00–01	00–11
01–00	01–01
01–10	01–11
10-01	10-00
10–11	10-10
11-01	11-00
11–11	11-10

Поскольку в описание папок входят четыре переменные, то нетрудно построить пространственную структуру папок в виде булева гиперкуба В⁴ (рис. 7.1). На этом рисунке кружки, содержащие обозначения папок с основными ПФ, закрашены серым цветом, а с минорными – не закрашены. Из рисунка видно, что названия папок с основными ПФ локализованы в гиперкубе вверху и внизу, а с минорными – справа и слева.

При этом названия папок из верхней группы преобразуются в названия из нижней группы на основе правил антисимметрии (0 $\leftarrow \rightarrow$ 1), например: 00–00 $\leftarrow \rightarrow$ 11–11. Аналогичные преобразования можно произвести для папок правой и левой групп.



Рис. 7.1. Пространственная структура папок с пентафрагментами. Закрашенные кружки – названия папок с основными ПФ, незакрашенные – с минорными ПФ

Чтобы понять особенности первой и пятой пар переменных для основных и минорных ПФ, в табл. 7.6 и 7.7 показан их общий вид. Сравнение таблиц показывает, что особенностью основных ПФ является наличие в этой группе циклических ПФ (на рис. 7.1 названия папок с такими фрагментами локализованы в верхнем квартете), о чем можно заключить на основе повторения связей. Например, для папки 10-01 связь Abc N_i–Ehj O_{i–4} идентична связи Ehj O_{i–4}–Abc N_i (в табл. 7.6 они выделены жирным шрифтом). То же можно сказать о связях ПФ в папках 11 – 01, 10 – 11 и 11 – 11. В то же время, особенностью минорных ПФ является наличие в них связей типа N_i – O_{i+4}, O_i – N_{i+4}, а также типов O_{i–4}–N_{i–8} и N_{i–4}–O_{i–8}, которые не обнаруживаются в папках с основными ПФ.

Сортировка в подпапки. Как следует из табл. 7.6 и 7.7, три центральные пары переменных, описывающие Н-связи групп основной цепи при (i - 1)-м, (i - 2)-м и (i - 3)-м α -атомах, обозначают в ПФ сход-

ные связи. По этой причине ПФ далее сортировали в подпапки. В качестве основы для такой систематизации нами была использована блочная матрица, в которой (i - 2)-я пара переменных была общей для каждого блока (табл. 7.8), причем (i - 1)-е пары переменных располагались в строках, а (i - 3)-и пары – в столбцах.

Таблица 7.6

At.	00-00 00-10				10-01 11-01			
No		Acy	/clic			Сус	elic	
i	Abc	00	Abc	00	Abc N _i – Ehj O _{i–4}	10	Abc $O_i - Jkl N_{i+4}$	11
					Abc		Abc N _i –EhjO _{i–4}	
							Abc	
<i>i</i> –1	Bcd O_{i-1} – Klm N_{i-1}	+3	Bcd $O_{i-1} - Klm N_{i+1}$	-3	Bcd $O_{i-1} - Klm N_{i+1}$	+3	Bcd O_{i-1} – Klm N_i	+3
	Bcd N_{i-1} – Hfg O_{i-1}	-5	Bcd N_{i-1} – Hfg O_{i-3}	5	Bcd N_{i-1} – Hfg O_{i-1}	5	Bcd N_{i-1} – Hfg O_i	-5
<i>i</i> –2	Bcd	-	Bcd		Bcd	-	Bcd	-
; 3	Cde O_{i-2} – Lmn N_i	+2	Cde O_{i-2} – Lmn N_{i+1}	-2	$Cde O_{i-2} - Lmn N_i$	+2	Cde O _{i-2} – Lmn N	i+2
1-5	Cde N _{i-2} – Fgj O _{i-6}	5	Cde N _{i-2} – Fgj O _{i-6}		Cde N _{i-2} – Fgj O _{i-e}	5	Cde N _{i-2} – Fgj O _{i-}	6
	Cde		Cde		Cde		Cde	
	Deh O _{i-3} – Mno N _i	i+1	Deh O _{i-3} – Mno N _i +	+1	Deh O _{i-3} – Mno N _i	+1	Deh O _{i-3} – Mno N	i+1
	Deh N _{i-3} – Gjk O _{i-}	7	Deh N _{<math>i=3 – Gjk O$i=7$</math>}	7	Deh N _{i-3} – Gjk O _{i-}	7	Deh N _{<math>i=3 – Gjk O$i=1$</math>}	-7
	Deh	,	Deh		Deh		Deh	
<i>i</i> –4	Ehj	00	Ehj N _{i-4} – Jkm O _{i-8}	10	Ehj O _{i-4} – Abc N _i	01	Ehj O _{i-4} – Abc N _i	01
			Ehj		Ehj		Ehj	
			-		5		5	
	01-00		01-10		10-11		11-11	
i	$\frac{01-00}{\text{Abc O}_i - \text{Jkl N}_{i+4}}$	01	01-10 Abc O _i – Jkl N _{i+4}	01	$\frac{10-11}{\text{Abc N}_i - \text{Ehj O}_{i-4}}$	11	$\frac{11-11}{\text{Abc O}_i - \text{Jkl N}_{i+4}}$	11
i	$\frac{01-00}{\text{Abc O}_i - \text{Jkl N}_{i+4}}$ Abc	01	$\frac{01-10}{\text{Abc O}_i - \text{Jkl N}_{i+4}}$ Abc	01	$\frac{10-11}{\text{Abc } N_i - \text{Ehj } O_{i-4}}$ Abc	11	$\frac{11-11}{\text{Abc } O_i - \text{Jkl } N_{i+4}}$ $\text{Abc } N_i - \text{Ehj}O_{i-4}$	11
i	$\frac{01-00}{\text{Abc O}_i - \text{Jkl N}_{i+4}}$ Abc	01	$\frac{01-10}{\text{Abc O}_i - \text{Jkl N}_{i+4}}$ Abc	01	10-11 Abc N _i – Ehj O _{i–4} Abc	11	$\frac{11-11}{\text{Abc O}_i - \text{Jkl N}_{i+4}}$ $\frac{\text{Abc N}_i - \text{EhjO}_{i-4}}{\text{Abc}}$	11
<i>i</i> <i>i</i> -1	$\frac{01-00}{\text{Abc O}_i - \text{Jkl N}_{i+4}}$ Abc Bcd O _{i-1} - Klm N _i .	01	01-10 Abc O _i – Jkl N _{i+4} Abc Bcd O _{i-1} – Klm N _{i+4}	-3	10-11 Abc N _i – Ehj O _{i-4} Abc Bcd O _{i-1} – Klm N _i	+3	$\frac{11-11}{\text{Abc } O_i - \text{Jkl } N_{i+4}}$ $\frac{\text{Abc } N_i - \text{EhjO}_{i-4}}{\text{Abc}}$ $\frac{\text{Bcd } O_{i-1} - \text{Klm } N_i}{\text{Bcd } O_{i-1} - \text{Klm } N_i}$	+3
<i>i</i> <i>i</i> -1	$\begin{array}{c} 01\text{-}00\\ \mbox{Abc }O_i - \mbox{Jkl }N_{i+4}\\ \mbox{Abc}\\ \end{array}$	01 +3 5	01-10 Abc O _i – Jkl N _{i+4} Abc Bcd O _{i-1} – Klm N _{i+4} Bcd N _{i-1} – Hfg O _{i-1}	01 -3 5	10-11 Abc N _i – Ehj O _{i-4} Abc Bcd O _{i-1} – Klm N _i Bcd N _{i-1} – Hfg O _{i-}	+3	$\frac{11-11}{\text{Abc } O_i - \text{Jkl } N_{i+4}}$ $\frac{\text{Abc } N_i - \text{EhjO}_{i-4}}{\text{Abc}}$ $\frac{\text{Bcd } O_{i-1} - \text{Klm } N_i}{\text{Bcd } N_{i-1} - \text{Hfg } O_i}$	+3
<i>i</i> <i>i</i> -1 <i>i</i> -2	$\begin{array}{c} 01\text{-}00\\ \mbox{Abc } {\rm O}_i - \mbox{Jkl } {\rm N}_{i+4}\\ \mbox{Abc}\\ \hline \\ \mbox{Bcd } {\rm O}_{i-1} - \mbox{Klm } {\rm N}_{i}\\ \mbox{Bcd } {\rm N}_{i-1} - \mbox{Hfg } {\rm O}_{i-1}\\ \mbox{Bcd } \end{array}$	01 +3 5	01-10 Abc O _i – Jkl N _{i+4} Abc Bcd O _{i-1} – Klm N _{i+4} Bcd N _{i-1} – Hfg O _{i-4} Bcd	01 -3 5	$10-11$ Abc $N_i - Ehj O_{i-4}$ Abc Bcd $O_{i-1} - Klm N_{i}$ Bcd $N_{i-1} - Hfg O_{i-1}$ Bcd	11 +3 5	$11-11$ Abc $O_i - Jkl N_{i+4}$ Abc $N_i - EhjO_{i-4}$ Abc Bcd $O_{i-1} - Klm N_i$ Bcd $N_{i-1} - Hfg O_i$ Bcd	+3
<i>i</i> <i>i</i> -1 <i>i</i> -2 <i>i</i> -3	$\begin{array}{c} 01\text{-}00\\ \mbox{Abc } {\rm O}_i - \mbox{Jkl } {\rm N}_{i+4}\\ \mbox{Abc}\\ \end{array}$	01 +3 -5 +2	01-10 Abc O _i – Jkl N _{i+4} Abc Bcd O _{i-1} – Klm N _{i+} Bcd N _{i-1} – Hfg O _{i-1} Bcd Cde O _{i-2} – Lmn N _i .	01 -3 5 +2	10-11 Abc N _i – Ehj O _{i-4} Abc Bcd O _{i-1} – Klm N _i Bcd N _{i-1} – Hfg O _{i-} Bcd Cde O _{i-2} – Lmn N _i	11 +3 -5 +2	$\frac{11-11}{\text{Abc } O_i - \text{Jkl } N_{i+4}}$ $\frac{\text{Abc } N_i - \text{EhjO}_{i-4}}{\text{Abc }}$ $\frac{\text{Bcd } O_{i-1} - \text{Klm } N_i}{\text{Bcd } N_{i-1} - \text{Hfg } O_{i}}$ $\frac{\text{Bcd } O_{i-2} - \text{Lmn } N_i}{\text{Cde } O_{i-2} - \text{Lmn } N_i}$	11 +3 -5 <i>i</i> +2
<i>i</i> <i>i</i> -1 <i>i</i> -2 <i>i</i> -3	$\begin{array}{c} 01\text{-}00\\ \mbox{Abc }O_i - \mbox{Jkl }N_{i+4}\\ \mbox{Abc}\\ \hline \\ \mbox{Bcd }O_{i-1} - \mbox{Klm }N_{i}\\ \mbox{Bcd }N_{i-1} - \mbox{Hfg }O_{i-1}\\ \mbox{Bcd }\\ \mbox{Cde }O_{i-2} - \mbox{Lmn }N_i\\ \mbox{Cde }N_{i-2} - \mbox{Fgj }O_{i-4}\\ \hline \end{array}$	01 +3 .5 +2 5	01-10 Abc O _i – Jkl N _{i+4} Abc Bcd O _{i-1} – Klm N _{i+} Bcd N _{i-1} – Hfg O _{i-1} Bcd Cde O _{i-2} – Lmn N _{i-} Cde N _{i-2} – Fgj O _{i-6}	-3 5 +2	10-11 Abc N _i – Ehj O _{i-4} Abc Bcd O _{i-1} – Klm N _i Bcd N _{i-1} – Hfg O _{i-} Bcd Cde O _{i-2} – Lmn N _i Cde N _{i-2} – Fgj O _{i-4}	+3 5 +2	$\frac{11-11}{\text{Abc } O_i - \text{Jkl } N_{i+4}}$ $\frac{\text{Abc } N_i - \text{EhjO}_{i-4}}{\text{Abc}}$ $\frac{\text{Bcd } O_{i-1} - \text{Klm } N_i}{\text{Bcd } N_{i-1} - \text{Hfg } O_i}$ $\frac{\text{Bcd } O_{i-2} - \text{Lmn } N_i}{\text{Cde } N_{i-2} - \text{Fgj } O_{i-2}}$	111 +3 -5 <i>i</i> +2 6
<i>i</i> <i>i</i> -1 <i>i</i> -2 <i>i</i> -3	$\begin{array}{c} 01\text{-}00\\ \mbox{Abc }O_i - \mbox{Jkl }N_{i+4}\\ \mbox{Abc}\\ \hline\\ \mbox{Bcd }O_{i-1} - \mbox{Klm }N_{i}\\ \mbox{Bcd }N_{i-1} - \mbox{Hfg }O_{i-1}\\ \mbox{Bcd }\\ \mbox{Cde }O_{i-2} - \mbox{Lmn }N_i\\ \mbox{Cde }N_{i-2} - \mbox{Fgj }O_{i-6}\\ \mbox{Cde }\end{array}$	01 +3 -5 +2	01-10 Abc O _i – Jkl N _{i+4} Abc Bcd O _{i-1} – Klm N _{i+} Bcd N _{i-1} – Hfg O _{i-1} Bcd Cde O _{i-2} – Lmn N _i Cde N _{i-2} – Fgj O _{i-6} Cde	01 -3 5 +2	10-11 Abc N _i – Ehj O _{i-4} Abc Bcd O _{i-1} – Klm N _i Bcd N _{i-1} – Hfg O _{i-} Bcd Cde O _{i-2} – Lmn N _i Cde N _{i-2} – Fgj O _{i-0} Cde	+3 -5 +2 -5	$\frac{11-11}{\text{Abc } O_i - \text{Jkl } N_{i+4}}$ $\frac{\text{Abc } N_i - \text{EhjO}_{i-4}}{\text{Abc }}$ $\frac{\text{Bcd } O_{i-1} - \text{Klm } N_i}{\text{Bcd } N_{i-1} - \text{Hfg } O_{i}}$ $\frac{\text{Bcd } O_{i-2} - \text{Lmn } N_i}{\text{Cde } O_{i-2} - \text{Fgj } O_{i-1}}$ $\frac{\text{Cde } N_{i-2} - \text{Fgj } O_{i-1}}{\text{Cde } N_i}$	11 +3 -5 <i>i</i> +2 6
<i>i</i> <i>i</i> -1 <i>i</i> -2 <i>i</i> -3	$\begin{array}{c} 01\text{-}00\\ \mbox{Abc }O_i - \mbox{Jkl }N_{i+4}\\ \mbox{Abc}\\ \hline\\ \mbox{Bcd }O_{i-1} - \mbox{Klm }N_{i}\\ \mbox{Bcd }N_{i-1} - \mbox{Hfg }O_{i-1}\\ \mbox{Bcd }O_{i-2} - \mbox{Lmn }N_i\\ \mbox{Cde }O_{i-2} - \mbox{Lmn }N_i\\ \mbox{Cde }N_{i-2} - \mbox{Fg }O_{i-6}\\ \mbox{Cde }\\ \mbox{Deh }O_{i-3} - \mbox{Mno }N_i\\ \end{array}$	01 +3 5 +2 5	$\begin{array}{c} 01-10\\ \mbox{Abc } O_i - \mbox{Jkl } N_{i+4}\\ \mbox{Abc}\\ \hline \mbox{Bcd } O_{i-1} - \mbox{Klm } N_{i+4}\\ \mbox{Bcd } N_{i-1} - \mbox{Hfg } O_{i-4}\\ \mbox{Bcd } Cde \ O_{i-2} - \mbox{Lmn } N_{i-4}\\ \mbox{Cde } N_{i-2} - \mbox{Fgj } O_{i-6}\\ \mbox{Cde } Deh \ O_{i-3} - \mbox{Mno } N_{i}. \end{array}$	01 -3 5 +2 -3 -3 -3 -3 -3 -3 -3 -3 -3 	10-11 Abc N _i – Ehj O _{i-4} Abc Bcd O _{i-1} – Klm N _i Bcd N _{i-1} – Hfg O _{i-4} Bcd Cde O _{i-2} – Lmn N _i Cde N _{i-2} – Fgj O _{i-4} Cde Deh O _{i-3} – Mno N _i	11 +3 5 +2 5 +1	$\frac{11-11}{\text{Abc } O_i - \text{Jkl } N_{i+4}}$ $\frac{\text{Abc } N_i - \text{EhjO}_{i-4}}{\text{Abc }}$ $\frac{\text{Bcd } O_{i-1} - \text{Klm } N_i}{\text{Bcd } N_{i-1} - \text{Hfg } O_i}$ $\frac{\text{Bcd } O_{i-2} - \text{Lmn } N_i}{\text{Cde } O_{i-2} - \text{Enj } O_i}$ $\frac{\text{Cde } O_{i-2} - \text{Fgj } O_i}{\text{Cde } Deh O_{i-3}} - \text{Mno } N_i$	11 +3 -5 <i>i</i> +2 6 <i>i</i> +1
<i>i</i> <i>i</i> -1 <i>i</i> -2 <i>i</i> -3	$\begin{array}{c} 01\text{-}00\\ \mbox{Abc }O_i - \mbox{Jkl }N_{i+4}\\ \mbox{Abc}\\ \end{array}$	01 +3 55 +2 55 	$\begin{array}{c} 01-10\\ \mbox{Abc } O_i - Jkl \ N_{i+4}\\ \mbox{Abc}\\ \hline \\ \mbox{Bcd } O_{i-1} - Klm \ N_{i+4}\\ \mbox{Bcd } N_{i-1} - Hfg \ O_{i-4}\\ \mbox{Bcd } O_{i-2} - Lmn \ N_{i-6}\\ \mbox{Cde } O_{i-2} - Fgj \ O_{i-6}\\ \mbox{Cde } O_{i-3} - Mno \ N_{i-6}\\ \mbox{Deh } O_{i-3} - Gjk \ O_{i-7}\\ \end{array}$	01 -3 5 +2 +1 7	10-11 Abc N _i – Ehj O _{i-4} Abc Bcd O _{i-1} – Klm N _i Bcd N _{i-1} – Hfg O _{i-} Bcd Cde O _{i-2} – Lmn N _i Cde N _{i-2} – Fgj O _{i-0} Cde Deh O _{i-3} – Mno N _i Deh N _{i-3} – Gjk O _{i-}	11 +3 55 +2 55 +1 7	$\frac{11-11}{\text{Abc } O_i - \text{Jkl } N_{i+4}}$ $\frac{\text{Abc } N_i - \text{EhjO}_{i-4}}{\text{Abc }}$ $\frac{\text{Bcd } O_{i-1} - \text{Klm } N_i}{\text{Bcd } N_{i-1} - \text{Hfg } O_{i-1}}$ $\frac{\text{Bcd } O_{i-2} - \text{Lmn } N_i}{\text{Cde } N_{i-2} - \text{Fgj } O_{i-1}}$ $\frac{\text{Cde } O_{i-3} - \text{Mno } N_i}{\text{Deh } N_{i-3} - \text{Gjk } O_{i-3}}$	111 +3 -5 <i>i</i> +2 6 <i>i</i> +1 -7
<i>i</i> <i>i</i> -1 <i>i</i> -2 <i>i</i> -3	$\begin{array}{c} 01\text{-}00\\ \mbox{Abc }O_i - \mbox{Jkl }N_{i+4}\\ \mbox{Abc}\\ \end{array}$	01 +3 5 +2 5 	$\begin{array}{c} 01-10\\ \mbox{Abc } O_i - Jkl \ N_{i+4}\\ \mbox{Abc}\\ \hline \\ \mbox{Bcd } O_{i-1} - Klm \ N_{i+4}\\ \mbox{Bcd } N_{i-1} - Hfg \ O_{i-2}\\ \mbox{Bcd } O_{i-2} - Lmn \ N_{i-1}\\ \mbox{Cde } O_{i-2} - Fgj \ O_{i-6}\\ \mbox{Cde } O_{i-3} - Mno \ N_{i-1}\\ \mbox{Deh } O_{i-3} - Gjk \ O_{i-1}\\ \mbox{Deh } N_{i-3} - Gjk \ O_{i-1}\\ \mbox{Deh } \end{array}$	01 3 5 +-2 	10-11 Abc N _i – Ehj O _{i-4} Abc Bcd O _{i-1} – Klm N _i Bcd N _{i-1} – Hfg O _{i-} Bcd Cde O _{i-2} – Lmn N _i Cde N _{i-2} – Fgj O _{i-0} Cde Deh O _{i-3} – Mno N _i Deh N _{i-3} – Gjk O _{i-} Deh	11 +3 5 +2 5 +2 5 5 +1 7	$\frac{11-11}{\text{Abc } O_i - \text{Jkl } N_{i+4}}$ $\frac{\text{Abc } N_i - \text{EhjO}_{i-4}}{\text{Abc }}$ $\frac{\text{Bcd } O_{i-1} - \text{Klm } N_i}{\text{Bcd } N_{i-1} - \text{Hfg } O_i}$ $\frac{\text{Bcd } O_{i-2} - \text{Lmn } N_i}{\text{Cde } O_{i-2} - \text{Enj } O_i}$ $\frac{\text{Cde } O_{i-2} - \text{Fgj } O_i}{\text{Cde } O_{i-3} - \text{Mno } N_i}$ $\frac{\text{Deh } O_{i-3} - \text{Gjk } O_i}{\text{Deh } N_{i-3} - \text{Gjk } O_i}$	+3 -5 <i>i</i> +2 6 <i>i</i> +1 -7
<i>i</i> <i>i</i> -1 <i>i</i> -2 <i>i</i> -3 <i>i</i> -4	$\begin{array}{c} 01{\text{-}}00 \\ \mbox{Abc } {\rm O}_i - {\rm Jkl } {\rm N}_{i+4} \\ \mbox{Abc} \\ \hline \mbox{Bcd } {\rm O}_{i-1} - {\rm Klm } {\rm N}_{i} \\ \mbox{Bcd } {\rm N}_{i-1} - {\rm Hfg } {\rm O}_{i-1} \\ \mbox{Bcd } {\rm O}_{i-2} - {\rm Lmn } {\rm N}_i \\ \mbox{Cde } {\rm O}_{i-2} - {\rm Fgj } {\rm O}_{i-6} \\ \mbox{Cde } {\rm N}_{i-2} - {\rm Fgj } {\rm O}_{i-6} \\ \mbox{Cde } \\ \mbox{Deh } {\rm O}_{i-3} - {\rm Mno } {\rm N}_i \\ \mbox{Deh } {\rm O}_{i-3} - {\rm Gjk } {\rm O}_{i-1} \\ \mbox{Deh } \\ \mbox{Ehj} \end{array}$	01 +3 55 ++2 5 7 00	$\begin{array}{c} 01-10\\ \mbox{Abc } O_i - \mbox{Jkl } N_{i+4}\\ \mbox{Abc}\\ \hline \mbox{Bcd } O_{i-1} - \mbox{Klm } N_{i+4}\\ \mbox{Bcd } N_{i-1} - \mbox{Hfg } O_{i-4}\\ \mbox{Bcd } O_{i-2} - \mbox{Lmn } N_{i-4}\\ \mbox{Cde } O_{i-2} - \mbox{Lmn } N_{i-4}\\ \mbox{Cde } O_{i-3} - \mbox{Mn } N_{i-4}\\ \mbox{Deh } N_{i-4} - \mbox{Jkm } O_{i-8}\\ \hline \end{array}$	01 -3 5 +2 7 10	10-11 Abc N _i – Ehj O _{i-4} Abc Bcd O _{i-1} – Klm N _i Bcd N _{i-1} – Hfg O _{i-4} Bcd Cde O _{i-2} – Lmn N _i Cde N _{i-2} – Fgj O _{i-4} Cde Deh O _{i-3} – Mno N _i Deh Ehj O _{i-4} – Abc N _i	11 +3 5 +2 5 +1 7 10	$\frac{11-11}{\text{Abc } O_i - \text{Jkl } N_{i+4}}$ $\frac{\text{Abc } N_i - \text{EhjO}_{i-4}}{\text{Abc }}$ $\frac{\text{Bcd } O_{i-1} - \text{Klm } N_i}{\text{Bcd } N_{i-1} - \text{Hfg } O_i}$ $\frac{\text{Bcd } O_{i-2} - \text{Lmn } N_i}{\text{Cde } N_{i-2} - \text{Fgj } O_{i-1}}$ $\frac{\text{Cde } O_{i-3} - \text{Mno } N_i}{\text{Deh } O_{i-3} - \text{Gjk } O_i}$ $\frac{\text{Deh } O_{i-4} - \text{AbcN}_i}{\text{Ehj } O_{i-4} - \text{AbcN}_i}$	111 +3 -5 <i>i</i> +2 6 <i>i</i> +1 -7
<i>i</i> <i>i</i> -1 <i>i</i> -2 <i>i</i> -3 <i>i</i> -4	$\begin{array}{c} 01\text{-}00\\ \mbox{Abc }O_i - \mbox{Jkl }N_{i+4}\\ \mbox{Abc}\\ \hline \mbox{Bcd }O_{i-1} - \mbox{Klm }N_{i},\\ \mbox{Bcd }N_{i-1} - \mbox{Hfg }O_{i-1}\\ \mbox{Bcd }O_{i-2} - \mbox{Lmn }N_i\\ \mbox{Cde }O_{i-2} - \mbox{Lmn }N_i\\ \mbox{Cde }N_{i-2} - \mbox{Fg }O_{i-6}\\ \mbox{Cde }Deh O_{i-3} - \mbox{Mno }N_i\\ \mbox{Deh }N_{i-3} - \mbox{Gjk }O_{i-1}\\ \mbox{Deh }Deh\\ \mbox{Ehj} \end{array}$	01 +3 5 +2 5 5 +1 7 00	$\begin{array}{c} 01-10\\ \mbox{Abc } O_i - {\rm Jkl } {\rm N}_{i+4}\\ \mbox{Abc}\\ \hline \mbox{Bcd } {\rm O}_{i-1} - {\rm Klm } {\rm N}_{i+4}\\ \mbox{Bcd } {\rm N}_{i-1} - {\rm Hfg } {\rm O}_{i-4}\\ \mbox{Bcd } {\rm Ode } {\rm O}_{i-2} - {\rm Lmn } {\rm N}_{i-4}\\ \mbox{Cde } {\rm N}_{i-2} - {\rm Fgj } {\rm O}_{i-6}\\ \mbox{Cde } {\rm N}_{i-2} - {\rm Fgj } {\rm O}_{i-6}\\ \mbox{Cde } {\rm Deh } {\rm O}_{i-3} - {\rm Mno } {\rm N}_{i}\\ \mbox{Deh } {\rm N}_{i-3} - {\rm Gjk } {\rm O}_{i-4}\\ \mbox{Deh } {\rm N}_{i-4} - {\rm Jkm } {\rm O}_{i-8}\\ \mbox{Ehj } \\ \mbox{Ehj } \end{array}$	01 -3 5 +2 7 10	10-11 Abc N _i – Ehj O _{i-4} Abc Bcd O _{i-1} – Klm N _i . Bcd N _{i-1} – Hfg O _{i-4} Bcd Cde O _{i-2} – Lmn N _i Cde N _{i-2} – Fgj O _{i-4} Cde Deh O _{i-3} – Mno N _i Deh N _{i-3} – Gjk O _{i-} Deh Ehj O_{i-4} – Abc N_i Ehj N _{i-4} – Jkm O _{i-8}	11 +3 55 +2 56 +1 7 10	$\frac{11-11}{\text{Abc } O_i - \text{Jkl } N_{i+4}}$ $\frac{\text{Abc } N_i - \text{EhjO}_{i-4}}{\text{Abc } N_i - \text{EhjO}_{i-4}}$ $\frac{\text{Bcd } O_{i-1} - \text{Klm } N_i}{\text{Bcd } N_{i-1} - \text{Hfg } O_i}$ $\frac{\text{Bcd } O_{i-2} - \text{Lmn } N_i}{\text{Cde } O_{i-2} - \text{Egj } O_i}$ $\frac{\text{Cde } O_{i-2} - \text{Fgj } O_i}{\text{Cde } O_{i-3} - \text{Mno } N_i}$ $\frac{\text{Deh } O_{i-3} - \text{Gjk } O_i}{\text{Deh } O_{i-4} - \text{AbcN}_i}$ $\frac{\text{Ehj } O_{i-4} - \text{AbcN}_{i-8}}{\text{Ehj } N_{i-4} - \text{JkmO}_{i-8}}$	111 ++3 -5 <i>i</i> +2 6 <i>i</i> +1 -7

Основные пентафрагменты: общий вид и положение Н-связей

Следует отметить, что пространственной структурой такой матрицы является булев гиперкуб В⁶ [32]. Эта структура будет использована нами при обсуждении полученных результатов.

Таблица	7.	7
---------	----	---

	onus in nouradnorwouror		HUY U ADDAA
Оошии вид мино	лоных пентафратментов	з и положение в	5 них п-связеи
	F STREES		

At. No	00-01		00-11		10-00		10-10	
i	Abc	00	Abc	00	Abc N_i – Jkl O_{i+4}	10	Abc $O_i - Jkl N_{i+4}$	10
					Abc		Abc	
<i>i</i> –1	Bcd		Bcd		Bcd		Bcd	
<i>i</i> –2	Cde		Cde		Cde		Cde	
<i>i</i> –3	Deh		Deh		Deh		Deh	
<i>i</i> –4	Ehj O _{i-4} – Jkm N _{i-8}	01	Ehj O _{i-4} – Jkm N _{i-8}	11	Ehj	00	Ehj N _{i-4} – JkmO _{i-8}	10
			Ehj N _{i-4} – Jkm O _{i-8}				Ehj	
			Ehj					
	01-01		01-11		11-00		11-10	
i	Abc $O_i - Jkl N_{i+4}$	01	Abc $O_i - Jkl N_{i+4}$	01	Abc N_{1} – Ikl O_{2}	11	Abc $N_1 - Jkl O_{114}$	11
	$l = l^{+4}$		1 1 1 1		i = 100 $i = 100$ $i = 4$		10011_l 1101_{l+4}	
	Abc		Abc		Ehj $O_i - JknN_{i+4}$		Ehj N _{<math>i=4 – JknO$i+4$</math>}	
	Abc		Abc		Ehj $O_i - JknN_{i+4}$ Abc		Ehj N _{$i-4$} – JknO _{$i+4$} Abc	
<i>i</i> –1	Abc Bcd		Abc Bcd		Ehj $O_i - JknN_{i+4}$ Abc Bcd		Ehj N _{i-4} – JknO _{i+4} Abc Bcd	
<i>i</i> -1 <i>i</i> -2	Abc Bcd Cde		Abc Bcd Cde		$\begin{array}{c} \text{Ehj } O_i - \text{JknN}_{i+4} \\ \text{Abc} \\ \text{Bcd} \\ \text{Cde} \end{array}$		$\begin{array}{c} \text{Ehj } N_{i} \rightarrow \text{JknO}_{i+4} \\ \text{Abc} \\ \hline \text{Bcd} \\ \text{Cde} \end{array}$	
<i>i</i> -1 <i>i</i> -2 <i>i</i> -3	Abc Bcd Cde Deh		Abc Bcd Cde Deh		$\begin{array}{c} \text{Ehj } O_i - \text{JknN}_{i+4} \\ \text{Abc} \\ \hline \text{Bcd} \\ \hline \text{Cde} \\ \hline \text{Deh} \end{array}$		$ Ehj N_{i-4} - JknO_{i+4} Abc Bcd Cde Deh $	
$ \begin{array}{c} i-1\\ i-2\\ i-3\\ i-4 \end{array} $	Abc Bcd Cde Deh Ehj O _{i-4} – Jkm N _{i-8}	01	Abc Bcd Cde Deh Ehj O _{i-4} – Jkm N _{i-8}	11	$\begin{array}{c} \text{Ehj } O_i - \text{JknN}_{i+4} \\ \text{Abc} \\ \text{Bcd} \\ \text{Cde} \\ \text{Deh} \\ \hline \end{array}$	00	Ehj $N_{i-4} - JknO_{i+4}$ Abc Bcd Cde Deh Ehj $N_{i-4} - JkmO_{i-8}$	10
$ \begin{array}{c} i-1\\ i-2\\ i-3\\ i-4 \end{array} $	Abc Bcd Cde Deh Ehj O_{i-4} – Jkm N_{i-8} Ehj	01	Abc Bcd Cde Deh Ehj O_{i-4} – Jkm N_{i-8} Ehj N_{i-4} – Jkm O_{i-8}	11	Ehj $O_i - JknN_{i+4}$ Abc Bcd Cde Deh Ehj	00	Ehj $N_{i-4} - JknO_{i+4}$ Bcd Cde Deh Ehj $N_{i-4} - JkmO_{i-8}$	10

Таблица 7.8

Блочная суперматрица подпапок в базе данных пентафрагментов белков

α- атом	Описание центральной части пентафрагментов							
<i>i</i> –2	00			01				
<i>i</i> -3 <i>i</i> -1	00	01	10	11	00	01	10	11
00	000000	000001	000010	000011	000100	000101	000110	000111
10	100000	100001	100010	100011	100100	100101	100110	100111
01	010000	010001	010010	010011	010100	010101	010110	010111
11	110000	110001	110010	110011	110100	110101	110110	110111
<i>i</i> –2	10			11				
<i>i</i> -3 <i>i</i> -1	00	01	10	11	00	01	10	11
00	001000	001001	001010	001011	001100	001101	001110	001111
10	101000	101001	101010	101011	101100	101101	101110	101111
01	011000	011001	011010	011011	011100	011101	011110	011111
11	111000	111001	111010	111011	111100	111101	111110	111111

Создаваемые подпапки обозначали в виде десятизначных чисел, в которых первая и пятая пары ПФ соответствовали номерам общей папки, а три центральные пары – описанию связей центральных α-атомов.

Сортировка в файлы. Для целей поиска нужных ПФ внутри подпапок их (ПФ) сортировали в отдельные файлы. Для этого использовали разделение аминокислот на группы антисимметрии согласно табл. 5.4. Детали такой систематизации описаны в нашем патенте [113] и в монографии [1]. В соответствии с этой систематизацией каждая из 20 аминокислот, которые присутствуют в белках, входит в свою подгруппу преобразований:

1-я подгруппа – Gly, Pro;

2-я подгруппа – Ala, Leu;

3-я подгруппа – Ser, Thr, Cys, Met, His, Trp, Phe, Tyr;

4-я подгруппа – Asp, Glu, Asn, Gln, Arg, Lys, Val, Ile.

С помощью программы separator.exe ПФ внутри подпапок кодировались на основе отнесения аминокислот ПФ (начиная сверху) к той или иной подгруппе преобразований и сортировались в файлы, в названия которых входили: пятизначный код ПФ и десятизначное название подпапки. Примеры названий файлов приведены в табл. 7.9.

Таблица 7.9

Примеры названий файлов в базе данных пентафрагментов белков

Кодирование	Название файла	Кодирование	Название файла
пентафрагмента	в подпапке	пентафрагмента	в подпапке
Icmb 5 Gly – 1 4 Ser – 3 3 Trp – 3 2 Glu – 4 1 Ala – 2	13342_0000000000.txt	2c2c 5 Ala - 2 4 Ala - 2 3 Asp - 4 2 Gly - 1 1 Glu - 4	22414_0101010000.txt

В результате этой процедуры внутри файлов оказывались ПФ, имеющие идентичную кодировку аминокислот. Общее число файлов, на которые можно подразделить ПФ в соответствии с заложенными принципами кодировки, – 1024.

7.3. Анализ особенностей пентафрагментов, содержащихся в базе данных

Общая характеристика пентафрагментов

Процедуру обработки PDB-файлов проводили дважды. В первой выборке было использовано порядка 500 PDB-файлов [115]–[117], а во второй – 2500 [118]. Полученные данные в обоих случаях выявили сходные закономерности. В данном издании будут обсуждаться наи-

более интересные результаты, полученные во второй выборке. Желающим ознакомиться с полными данными рекомендуем [115]–[118].

В результате обработки текстовых файлов, полученных на основе 2500 PDB-файлов белков, было выделено 603 569 ПФ. В соответствии с описанными ранее процедурами они были рассортированы в 16 папок, внутри которых содержались 64 подпапки. Внутри подпапок содержалось различное количество файлов с ПФ, которые и составили полученную нами базу данных. Ее описание (в том виде, как она зарегистрирована) дано в материалах регистрации базы данных в РОСПАТЕНТе [110].

Основные пентафрагменты: относительное содержание и симметрия названий подпапок

Общее количество ПФ, отнесенных к группе основных, составило 598 664 (99,2 %). Было найдено, что в результате сортировки во всех восьми папках, содержащих основные ПФ (см. табл. 7.5) почти все 64 подпапки оказались заполненными. При этом было обнаружено, что имеется ряд подпапок, в которых количество ПФ превышает во много раз их среднее содержание. Названия этих подпапок в виде десятизначных булевых чисел, относительное содержание в них ПФ и найденная для них симметрия представлены в табл. 7.10.

Таблица 7.10

№	Названия подпапок с пентафрагментами	Количество пентафраг- ментов (%)	Симметрия названий подпапок с пентафрагментами		
18	000000010	14 397 (2,385)	Зеркальная	симметрия	(отражение)
17	0000001010	11 016 (1,825)	010000000	$\leftarrow \rightarrow$	000000010
16	0000101010	10 878 (1,802)	0101000000	$\leftarrow \rightarrow$	0000001010
15	0010101010	11 571 (1,917)	0101010000	$\leftarrow \rightarrow$	0000101010
14	1010101011	12 550 (2,079)	0101010100	$\leftrightarrow \rightarrow$	0010101010
13	1010101111	12 003 (1,989)	1101010101	$\leftarrow \rightarrow$	1010101011
12	1010111111	11 721 (1,942)	1111010101	$\leftrightarrow \rightarrow$	1010101111
11	1011111111	11 155 (1,848)	1111110101	$\leftarrow \rightarrow$	1010111111
10	111111111	49 888 (8,266)	1111111101	$\leftrightarrow \rightarrow$	1011111111
9	111111101	11 163 (1,849)	Поворотная си	мметрия и а	нтисимметрия
8	1111110101	12 005 (1,989)	1111111101	Поворот	000000010
7	1111010101	12 170 (2,016)	1111110101	блоков	0000001010
6	1101010101	12 722 (2,108)	1111010101	на 180°	0000101010
5	0101010100	12 327 (2,042)	1101010101		0010101010
4	0101010000	11 669 (1,933)	0101010100	Переход	1010101011
3	0101000000	12 088 (2,003)	0101010000	$0 \leftarrow \rightarrow 1$	1010101111
2	010000000	15 213 (2,521)	0101000000		1010111111
1	0000000000	222 426 (36,852)	0100000000		1011111111

Названия подпапок с наиболее распространенными основными пентафрагментами и их симметрия

Как следует из табл. 7.10, количество ПФ в представленных подпапках весьма существенно изменяется. Больше всего ПФ присутствует в подпапках 000000000 (222 426 – 36,85 % от общего числа ПФ) и 1111111111 (49 888 – 8,266 %), чего и следовало ожидать, поскольку они содержат ПФ, выделенные из участков с конформацией β-структуры (десятизначное описание 000000000) и α -спирали (десятизначное описание 11111111), являющихся основными вторичными структурами белков.

Содержание ПФ в остальных подпапках, представленных в табл. 7.10, составляет ~2 %. Эти подпапки образуют последовательный ряд переходов с однобитовыми изменениями на участках от β-структуры к α-спирали (этапы 2–9) и от α-спирали к β-структуре (этапы 11–18).

Симметрия пентафрагментов, отраженная в названиях содержащих их подпапок, показана в правой половине табл. 7.10. Так, если названия подпапок с переходными участками 2–9 записать сверху вниз, а рядом выписать аналогичные подпапки 11–18, то становится очевидным, что эти два ряда обладают зеркальной симметрией. В то же время, запись рядом этих двух групп подпапок без изменений показывает, что между ними имеется поворотная симметрия, а если их просматривать в одной строке, то эти названия преобразуются друг в друга по принципу антисимметрии (0 $\leftarrow \rightarrow$ 1), аналогично рассмотренному в 2.2.

Следует также обратить внимание, что ПФ, участвующие в стадиях переходов « β -структура $\leftarrow \rightarrow \alpha$ -спираль», расположены в папках 01– 00 (стадии 2–5) и 11–01 (стадии 6–9), а переходов « α -спираль $\leftarrow \rightarrow$ β -структура» – в папках 10–11 (стадии 11–14) и 00–10 (стадии 15–18). При этом в каждой папке, описывающей часть переходного участка, содержится по четыре стадии. Причина этих особенностей, как оказалось, может быть вскрыта, если использовать пространственные структуры, описывающие содержащиеся в папках подпапки.

Пространственные структуры папок с основными пентафрагментами

Как отмечалось в 7.2, наличие в папках 64 подпапок с тремя варьирующими центральными парами переменных (табл. 7.8) позволяет построить их пространственную структуру в виде булева гиперкуба В⁶. В качестве примера на рис. 7.2 показана такая структура для папки 01–00. При этом первая и пятая пары переменных, как видно из рис. 7.2, вынесены за пределы гиперкуба. Они расположены на рисунке вверху и внизу. Каждая вершина гиперкуба B^6 связана с шестью соседними вершинами однобитовыми переходами. В папке 01–00 находятся подпапки, описываемые десятизначными числами 0100000000, 0101000000, 0101010000 и 0101010100. Соответственно три центральные пары переменных будут записаны так: 000000, 0100000, 010100, 101010. Мы нанесли эти шестизначные значения на гиперкуб и соединили линиями со стрелками (рис. 7.2). При этом все четыре названия подпапок оказались на гиперкубе связаными единичными переходами, причем на одном гиперкубе можно расположить только эти четыре значения, описывающих стадии 2–5.



Рис. 7.2. Пространственная структура подпапок пентафрагментов папки 01–00 и расположение на ней номеров подпапок стадий 2–4

Для дальнейшего анализа подпапок, входящих в остальные стадии переходов « α -спираль $\leftarrow \rightarrow \beta$ -структура», была построена модель структуры названий папок с основными ПФ (рис. 7.3).



Рис. 7.3. Модель структуры папок с основными ПФ

Предложенная модель представляет собой трехмерный куб (В³), в вершинах которого расположены гиперкубы В⁶, обозначенные в соответствии с числовыми наименованиями папок. Эти названия папок связаны единичными переходами. Кажущееся нарушение этого принципа для папок 01–00 $\leftarrow \rightarrow$ 11–01 и 10–11 $\leftarrow \rightarrow$ 00-10 на самом деле не является таковым, поскольку в папках 11–01 и 10–11 расположены циклические ПФ и образующаяся единичная связь повторяется два раза (10–01). Из этого рисунка видно, что последовательные стадии переходов « α -спираль $\leftarrow \rightarrow \beta$ -структура» образуют на этой структуре траекторию, в которой имеется четыре папки (гиперкубы 01–00, 11–01, 10–11, 00–10), содержащие по четыре стадии переходов, и две папки (00–00 и 11–11), содержащие *n* раз повторяющиеся стадии для β -структурных или α -спиральных участков.

Приведенный пример траектории для переходов « α -спираль $\leftarrow \rightarrow \beta$ -структура» является частным случаем. Используя эту структуру из восьми гиперкубов можно прокладывать траектории различной длины и в различных сочетаниях переменных, т. е. она может служить генератором описаний вторичных структур белка, содержащих основные ПФ.

Особенности минорных пентафрагментов

Минорные ПФ поначалу вообще не были замечены. Однако подробный анализ оставшихся ПФ позволил выявить их реальное присутствие в выборке в количестве 4905, что составляет 0,8 % от их общего числа. Последующая их сортировка показала, что минорные ПФ также можно подразделить на 8 папок: 00–01, 00–11, 01–01, 01–11, 10–00, 10–10, 11–00, 11–10.

При создании базы данных внутри каждой папки было выделено также по 64 подпапки. Однако, в отличие от основных ПФ, степень заполнения подпапок для минорных ПФ оказалась невысокой (не более 10...15 % общего числа подпапок). Было установлено [118], что имеется лишь две подпапки с максимальным числом ПФ: 0000000001 (1892 ПФ) и 1000000000 (1489 ПФ). Как правило, наибольшее число ПФ обнаруживается в подпапках, которые описываются десятизначными числами, расположенными в начальных областях булевых гиперкубов В⁶, в блоках 00. Это наблюдается практически во всех папках, за исключением 11–10.

Было обнаружено, что чаще всего минорные ПФ выделяются из участков, содержащих β -изгибы белков. Как известно, в β -изгибах основной цепи (реверсивных поворотах) возникают водородные связи между атомами группы N_iH...O_{i-4} или N_iH...O_{i-3} [38]. При этом могут иметь место право- и левозакрученные варианты. Рассмотрим их для

случая Н-связей между атомами группы $N_iH...O_{i-4}$ (рис. 7.4, *a* и *б*). Видно (рис. 7.4, *б*), что в случае правозакрученного изгиба водородная связь возникает между атомами группы N_iH и O_{i-4} , а в случае левозакрученного – между атомами группы $N_{i-4}H$ и O_i (рис. 7.4, *a*).



a – левозакрученный изгиб; δ – правозакрученный изгиб

В табл. 7.11 представлено описание формирующихся ПФ в виде десятизначных булевых чисел для обоих типов изгибов. Сопоставление этих чисел показывает, что в состав правозакрученных β -изгибов входят только числа, описывающие основные ПФ, а в состав левозакрученных – числа, описывающие как минорные ПФ (выделены жирным шрифтом), так и основные [118]. Как отмечалось ранее, именно фрагменты видов 1000000000 и 000000001 представлены в таблице минорных ПФ в наибольшем количестве [118].

Таблица 7.11

	Левозакрученные β-изгибы	Правозакрученные β-изгибы
<i>i</i> + 4	000000001	000000010
<i>i</i> + 3	000000100	0000001000
<i>i</i> + 2	0000010000	0000100000
<i>i</i> + 1	0001000000	001000000
i	010000010	100000001
<i>i</i> – 1	0000001000	000000100
i-2	0000100000	0000010000
<i>i</i> – 3	001000000	0001000000
<i>i</i> – 4	100000000	010000000

Описание β-изгибов с помощью десятизначных булевых чисел

Из этого примера следует, что минорные ПФ в качестве одиночных элементов включаются в редко встречающиеся левозакрученные изгибы β -структур и другие элементы вторичных структур белков. Кроме того, как показал анализ, в отличие от основных ПФ они не склонны к формированию протяженных участков вторичной структуры. Это объясняет, почему существует группа минорных ПФ. Отсутствие этих редких ПФ в структурах будет существенно снижать степень разнообразия вариантов вторичных структур белков.

Полная пространственная структура базы данных пентафрагментов

В 7.2 была предложена пространственная структура папок, содержащих основные и минорные ПФ (см. рис. 7.1). Она основана на принципе единичных переходов между описывающими их четырехзначными булевыми числами. Аналогичная структура была построена для двух уровней – уровня папок и уровня подпапок [117], [118]. Эта структура имеет те же закономерности, что и приведенная на рис. 7.1. В частности, на основе правил антисимметрии (0 $\leftarrow \rightarrow$ 1) названия папок верхней группы преобразуются в названия папок нижней группы, а названия папок левой группы – в названия папок правой группы. В 7.3 на структуре для папок с основными ПФ (см. рис. 7.2) была показана возможность использования данной структуры для генерации описаний вторичных структур, содержащих основные ПФ. Предложенная в [117], [118] структура, являющаяся полной для всех папок с ПФ, также, вероятно, может быть использована для таких целей, хотя этот вопрос требует дальнейшей детальной проработки.

7.4. Перспективы развития базы данных пентафрагментов

Получение пентафрагментов с различными вариантами связей боковых цепей с основной цепью и между собой

Полученная и зарегистрированная нами база данных [110] содержит лишь описание ПФ основной цепи. Такое описание, конечно, является определяющим для вторичных структур белка. Однако оно является лишь первым этапом в системной «разборке» белков. Следующими (дополнительными) этапами могут служить получение текстового файла и нарезка ПФ на основе образования связей N_iH...O_{i-3}, имеющих в основном отношение к начальным и конечным участкам α -спиралей, а также образование H-связей боковых цепей с основной цепью ($R_i...R_{i-4}$) и боковых цепей между собой ($R_i...R_{i-4}$). Такие возможности у программы Protein 3D и других мини-программ имеются. Использование полученных при этом закономерностей открывает при прогнозировании и проектировании белков перспективу перехода от вторичной структуры к третичной.

Проблема построения искусственной базы данных пентафрагментов

Созданная нами база данных ПФ хотя и достаточно велика, но отнюдь не исчерпывает всех возможных вариантов этих фрагментов в белках. В процессе разработки методов прогнозирования и проектирования структуры белков с использованием данной базы мы столкнулись с ограничениями ее практических возможностей. Так, применение метода прогнозирования вторичной структуры белка, которое осуществляется для белков, имеющихся в базе данных, или их близких аналогов с точностью порядка 99 %, оказывается малоэффективным при переходе к неизвестным белкам.

Можно ли существенно расширить базу данных? Конечно, возможно постепенное пополнение этой базы посредством обработки вновь исследованных белков и в этом смысле нет никаких ограничений. Но можно ли, анализируя закономерности построения экспериментальной базы, создать искусственную базу данных, содержащую все возможные варианты ПФ? По-видимому, можно. В качестве аргументов в пользу этого приведем результаты предварительного анализа нашей базы данных.

С целью выяснения возможности построения искусственной базы данных ПФ было проанализировано относительное содержание аминокислот в ряде подпапок, содержащих большое количество ПФ. Так, в табл. 7.12 приведено относительное содержание всех аминокислот в i-м - (i - 4)-м положениях для подпапки 0000000000 (β-структура). Из таблицы следует, что относительное содержание аминокислот в каждом положении отличается от среднего (оно должно быть ~5 %). Например, для Gly и Leu оно выше среднего, а для His и Trp – ниже среднего. Однако в каждом положении оно примерно одинаково, что свидетельствует об отсутствии какой-либо тенденции к изменениям в

¹⁴⁴
Таблица 7.12

	i	%	<i>i</i> – 1	%	<i>i</i> – 2	%	<i>i</i> – 3	%	<i>i</i> – 4	%
Gly	20 395	9,2	20 313	9,1	20 124	9,0	19 635	8,8	19 479	8,8
Pro	12 491	5,6	12 485	5,6	12 473	5,6	12 774	5,7	13 115	5,9
Ala	14 812	6,7	14 753	6,6	14 663	6,6	14 666	6,6	14 752	6,6
Leu	17 624	7,9	17 494	7,9	17 710	8,0	17 674	7,9	17 580	7,9
Ser	14 014	6,3	13 932	6,3	13 719	6,2	13 588	6,1	13 495	6,1
Thr	14 568	6,5	14 456	6,5	14 440	6,5	14 383	6,5	14 137	6,4
Cys	2853	1,3	2816	1,3	2793	1,3	2739	1,2	2691	1,2
Met	3500	1,6	3384	1,5	3374	1,5	3357	1,5	3565	1,6
His	5052	2,3	4987	2,2	4875	2,2	4891	2,2	4903	2,2
Trp	3365	1,5	3289	1,5	3288	1,5	3239	1,5	3192	1,4
Phe	9722	4,4	9802	4,4	9882	4,4	9636	4,3	9340	4,2
Tyr	8865	4,0	8846	4,0	8785	3,9	8727	3,9	8545	3,8
Asp	12 720	5,7	12 566	5,6	12 537	5,6	12 681	5,7	12 942	5,8
Glu	11 603	5,2	11 606	5,2	11 546	5,2	11 657	5,2	11 832	5,3
Asn	9696	4,4	9563	4,3	9461	4,3	9513	4,3	9541	4,3
Gln	6506	2,9	6520	2,9	6485	2,9	6657	3,0	6783	3,0
Arg	9520	4,3	9551	4,3	9494	4,3	9674	4,3	9853	4,4
Lys	10 881	4,9	11 051	5,0	11 191	5,0	11 596	5,2	11 991	5,4
Val	19 502	8,8	20 049	9,0	20 367	9,2	20 243	9,1	19 767	8,9
Ile	14 737	6,6	14 962	6,7	15 218	6,8	15 094	6,8	14 898	6,7
Ν	222 426		222 42	25	222 4	25	222 4	24	222 4	01

Относительное содержание аминокислот в подпапке 000000000

 β -структуре. В табл. 7.13 аналогичные данные приведены для подпапки 111111111 (ПФ из α -спирали). Из таблицы видно, что выборка ПФ для α -спирали оказалась в четыре раза меньше, чем для β -структуры. Это означает, что уменьшилась возможность перестановок под влиянием ограничений на состав аминокислот. И действительно, характерной особенностью этой подпапки является полное отсутствие Pro, что и должно иметь место, поскольку эта аминокислота является оператором антисвязности и разрушает α -спирали (см. 6.2). Повышено содержание неполярных аминокислот (Ala, Leu), а также полярных (Glu, Arg) и понижено содержание циклических (His, Trp) и серосодержащих (Cys, Met). Однако распределение относительного содержания аминокислот в зависимости от положения в ПФ, как и в случае с β структурой, примерно одинаково.

Анализ подпапок, расположенных в области перехода « β -структура – α -спираль» (табл. 7.14, стадии 6–9) показывает, что в них последовательно происходит исчезновение Рго в положениях *i* (6-я ста-

Таблица 7.13

			· · 1							
	i	%	<i>i</i> – 1	%	<i>i</i> – 2	%	<i>i</i> – 3	%	<i>i</i> – 4	%
Gly	1712	3,4	1750	3,5	1831	3,7	1917	3,8	1955	3,9
Pro	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
Ala	6625	13,3	6465	13,0	6662	13,4	6786	13,6	6731	13,5
Leu	7018	14,1	6802	13,6	6269	12,6	6451	2,9	6507	13,0
Ser	1991	4,0	1970	3,9	2059	4,1	2066	4,1	2002	4,0
Thr	2069	4,1	2138	4,3	2172	4,4	2245	4,5	2237	4,5
Cys	672	1,3	574	1,2	555	1,1	558	1,1	575	1,2
Met	1591	3,2	1520	3,0	1394	2,8	1399	2,8	1412	2,8
His	1046	2,1	1032	2,1	1029	2,1	973	2,0	978	2,0
Trp	742	1,5	737	1,5	627	1,3	617	1,2	698	1,4
Phe	2182	4,4	2256	4,5	2012	4,0	2020	4,0	2121	4,3
Tyr	1706	3,4	1709	3,4	1598	3,2	1500	3,0	1528	3,1
Asp	1927	3,9	2007	4,0	2176	4,4	2063	4,1	1911	3,8
Glu	3618	7,3	3668	7,4	3949	7,9	3697	7,4	3371	6,8
Asn	1595	3,2	1590	3,2	1763	3,5	1800	3,6	1669	3,3
Gln	2177	4,4	2170	4,3	2367	4,7	2386	4,8	2407	4,8
Arg	3167	6,3	3122	6,3	3478	7,0	3354	6,7	3423	6,9
Lys	3026	6,1	2915	5,8	3152	6,3	3068	6,1	2929	5,9
Val	3526	7,1	3762	7,5	3516	7,0	3637	7,3	3814	7,6
Ile	3480	7,0	3684	7,4	3262	6,5	3338	6,7	3607	7,2
Ν	49 888		49 8	88	49 888		49 888		49 888	

Относительное содержание аминокислот в подпапке 111111111

Таблица 7.14

Относительное содержание Pro в папках цикла «β-структура – α-спираль»

№	Названия	i		<i>i</i> –	1	<i>i</i> –	2	<i>i</i> –	3	<i>i</i> –	4
	подпапок	Ν	%	Ν	%	Ν	%	Ν	%	N	%
18	000000010	899	6,2	1056	7,3	1240	8,6	1100	7,6	604	4,2
17	0000001010	878	8,0	991	9,0	862	7,8	486	4,4	0	0,0
16	0000101010	949	8,7	840	7,7	433	4,0	0	0,0	0	0,0
15	0010101010	853	7,4	443	3,8	0	0,0	0	0,0	0	0,0
14	1010101011	467	3,7	0	0,0	0	0,0	0	0,0	0	0,0
13	1010101111	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
12	1010111111	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
11	1011111111	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
10	11111111111	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
9	1111111101	0	0,0	0	0,0	0	0,0	0	0,0	294	2,6
8	1111110101	0	0,0	0	0,0	0	0,0	262	2,2	789	6,6
7	1111010101	0	0,0	0	0,0	185	1,5	598	4,9	1545	12,7
6	1101010101	0	0,0	170	1,3	590	4,6	1532	12,0	787	6,2
5	0101010100	171	1,4	566	4,6	1462	1,9	749	6,1	790	6,4
4	0101010000	513	4,4	1370	11,7	710	6,1	775	6,6	950	8,1
3	0101000000	1410	11,7	762	6,3	798	6,6	972	8,0	837	6,9
2	010000000	972	6,4	938	6,2	1161	7,6	997	6,6	842	5,5
1	0000000000	12 491	5,6	12 485	5,6	12 473	5,6	12774	5,7	13 115	5,9

дия); *i*, *i* – 1 (7-я стадия); *i*, *i* – 1, *i* – 2 (8-я стадия); *i*, *i* – 1, *i* – 2, *i* – 3 (9-я стадия). В области перехода « α -спираль – β -структура» (стадии 14–17) происходит обратный процесс: последовательное появление Рго в положениях *i*, *i* – 1, *i* – 2, *i* – 3. Совершенно иная динамика изменений наблюдается на этих стадиях для Gly (табл. 7.15).

Таблица 7.15

№	Название	i		<i>i</i> – 1		<i>i</i> – 2		<i>i</i> – 3		<i>i</i> – 4	
	подпапки	Ν	%	Ν	%	Ν	%	Ν	%	Ν	%
18	000000010	1076	7,5	1148	8,0	1024	7,1	1246	8,7	4248	29,5
17	000001010	888	8,1	793	7,2	953	8,7	3796	34,5	385	3,5
16	0000101010	780	7,2	914	8,4	3933	36,2	352	3,2	229	2,1
15	0010101010	896	7,7	4301	37,2	360	3,1	248	2,1	199	1,7
14	1010101011	4693	37,4	396	3,2	276	2,2	224	1,8	330	2,6
13	1010101111	412	3,4	263	2,2	226	1,9	317	2,6	290	2,4
12	1010111111	268	2,3	241	2,1	313	2,7	292	2,5	328	2,8
11	1011111111	255	2,3	301	2,7	296	2,7	308	2,8	299	2,7
10	11111111111	1712	3,4	1750	3,5	1831	3,7	1917	3,8	1955	3,9
9	1111111101	340	3,0	378	3,4	394	3,5	337	3,0	546	4,9
8	1111110101	399	3,3	403	3,4	358	3,0	575	4,8	683	5,7
7	1111010101	380	3,1	351	2,9	565	4,6	675	5,5	612	5,0
6	1101010101	352	2,8	577	4,5	714	5,6	633	5,0	1175	9,2
5	0101010100	564	4,6	701	5,7	625	5,1	1187	9,6	1035	8,4
4	0101010000	677	5,8	610	5,2	1176	10,1	989	8,5	1223	10,5
3	0101000000	618	5,1	1261	10,4	1066	8,8	1197	9,9	1176	9,7
2	0100000000	1503	9,9	1277	8,4	1447	9,5	1478	9,7	1405	9,2
1	00000000000	20 395	9,2	20 313	9,1	20 124	9,0	19 635	8,8	19 479	8,8

Относительное содержание Gly в папках цикла «β-структура – α-спираль»

Начиная с 6-й стадии, количество Gly в ПФ в 3 раза снижается по сравнению с 1-й стадией, и столь низкое содержание наблюдается на всех стадиях – с 6-й по 13-ю включительно. При этом на 14-й стадии содержание Gly в ПФ резко увеличивается (почти до 40 %), а затем восстанавливается до исходного (7...9 %).

При создании искусственной базы данных можно ограничивать или увеличивать содержание тех или иных аминокислот в определенных положениях ПФ и моделировать процессы, аналогичные естественным изменениям содержания аминокислот на различных стадиях формирования вторичной структуры белка.

В процессе анализа особенностей содержания ПФ в файлах внутри подпапок также были найдены определенные закономерности. Так, файлы 11111_0000000000 содержат практически все варианты сочетаний Pro и Gly – аминокислот, принадлежащих к первой группе согласно нашей классификации (см. табл. 5.4). В то же время, для других папок и подпапок файлы и варианты, содержащие Pro и Gly, не встречаются. Это означает, что можно задавать определенные правила отбора вариантов сочетаний и приблизиться к реальным закономерностям формирования ПФ. В целом, перспективы создания искусственной базы данных представляются вполне реальными. Создание такой базы позволит с точностью почти 100 % предсказывать вторичную структуру любого известного белка, что обеспечит существенный прогресс в исследовании белковых структур без больших экономических затрат.

Глава 8. МЕТОД ПРОГНОЗИРОВАНИЯ ВТОРИЧНОЙ СТРУКТУРЫ БЕЛКА

8.1. Общие вопросы

Идеологическое обоснование подхода

Предлагаемый материал базируется на полученном нами патенте [113]. Количество исследований кристаллических белков, проводимое для установления их полной структуры, растет в арифметической прогрессии, а количество известных первичных структур с неизвестной функцией (в связи с применением методов секвенирования) – в геометрической. По этой причине прогнозирование вторичной структуры белков (т. е. положения α -спиральных или β -структурных фрагментов на первичной структуре белков) как этап в предсказании полной структуры имеет важное теоретическое и практическое значение [125].

Хотя в последние годы наметилась тенденция к более активной разработке методов прогнозирования третичной структуры [126], значимость методов прогнозирования вторичной структуры белка не исчезла. Идеологической основой для этого являются современные данные о процессах биосинтеза белка, которые показывают, что вторичная структура белка формируется ко-трансляционно, т. е. практически сразу после синтеза полипептидной цепи [1], [13]. Образование структуры белка при этом подразделяется на два этапа: формирование вторичной структуры в канале рибосомы, расположенном в большой субъединице и изолированном от внешней среды, и принятие белком правильной третичной структуры после выхода белка из рибосомы (при участии специальных белков – шаперонов). Надежность образования правильной надмолекулярной структуры белка обеспечивает именно данная последовательность этапов. Перенос акцентов на третичную структуру без знания точных механизмов формирования вторичной структуры, по-видимому, является преждевременным. Выражаясь образно, это напоминает строительство здания на песке.

Введение в проблему предсказания вторичной структуры белка

Стимулом к разработке методов предсказания вторичных структур белка явились работы К. Анфинсена [127], в которых на примере белка рибонуклеазы была установлена связь между последовательностью аминокислот и конформацией биологически активной молекулы. В ранний период наибольшее признание получил метод прогнозирования вторичной структуры белка, предложенный в [128]. В дальнейшем среди подходов и методов предсказания вторичной структуры белка выделились статистический [128]–[134], структурный [135]–[140] и квантово-химический [141] подходы, а также методы, основанные на использовании принципа нейронных сетей [142]–[144] и на аналогии с известными белковыми фрагментами [145]–[146]. Необходимо отметить, что ни один из предложенных методов не обладает точностью предсказания выше 80 %.

Метод [128], развитый в [147], процедурно схож с нашим методом. Сходство состоит в том, что в процессе прогнозирования используется группа аминокислот (в методе [128] их шесть). В процессе прогнозирования такая группа последовательно выделяется и продвигается по первичной структуре белка с трансляционным сдвигом в одну аминокислоту. Однако в методе [128] расчитывается потенциал спирализации для выделяемых фрагментов на основе предварительной статистической обработки вероятности нахождения той или иной аминокислоты в различных типах вторичной структур. В нашем методе никаких расчетов нет. Программа PREDICTOR на основе заданного алгоритма отыскивает в базе данных тот или иной ПФ, выделяемый в первичной структуре белка. Описание конформации этого ПФ записано в виде десятизначного числа в названии папки, где находится найденный ПФ. На каждом этапе это название записывается в результирующий файл, и по записанным десятизначным числам судят о конформации вторичной структуры белка. Точность предсказания вторичной структуры белков, пентафрагменты которых имеются в базе данных, близка к 100 %.

8.2. Суть метода

Компоненты, необходимые для начала работы

Для прогнозирования необходимо иметь на компьютере следующие компоненты:

– базу данных ПФ, полученную согласно описанию, приведенному в 7.2, и обладающую структурой, описанной в 7.3;

 – файлы, содержащие описание первичной структуры белка, представленные в различных форматах (последовательность аминокислот, последовательность триплетов нуклеотидов, переводимая в белковую последовательность и т. д.);

– программу, переводящую различные форматы в формат, используемый программой для прогнозирования (формат .dbk);

- программу для прогнозирования.

В качестве одного из примеров (см. патент [113]) мы провели процедуру предсказания вторичной структуры белка репрессора метионина, для которого известна его первичная структура (файл из GenBank). Известна также третичная структура этого белка, исследованного методом PCA [148] (индекс в Protein Data Bank – 1cmb), которая использована для сравнения с результатами прогнозирования.

Подготовка файлов

Для представления информации о первичной структуре исследуемого белка в виде рабочего файла нами была написана вспомогательная программа FILEMAKER, которая переводит различные форматы записи первичной структуры белка в файлах банков данных (например, в Genbank) в формат, читаемый программой для прогнозирования вторичной структуры.

Для исследуемого белка репрессора метионина информация о последовательности нуклеотидов, кодирующих последовательность аминокислот (GenBank Ген CP001665.1), имеет следующий вид (табл. 8.1).

Таблица 8.1

Последовательность нуклеотидов, кодирующих последовательность
аминокислот репрессора метионина (GenBank Ген CP001665.1)
atggctgaat ggagcggcga atatatcagc ccatacgctg agcacggcaa gaagagtgaa

1	
61	caagtcaaaa agattacggt ttccattcct cttaaggtgt taaaaatcct caccgatgaa
121	cgcacgcgtc gtcaggtgaa caacctgcgt cacgctacca acagcgagct gctgtgcgaa
181	gcgtttctgc atgcctttac cgggcaacct ttgccggatg atgccgatct gcgtaaagag
241	cgcagcgacg aaatcccgga agcggcaaaa gagatcatgc gtgagatggg gattaacccg
301	gagacgtggg aatactaa

С помощью программы FILEMAKER данный текстовый файл был представлен в виде рабочего файла, содержащего последовательность аминокислот исследуемого белка репрессора метионина (табл. 8.2). В этой таблице приведены начальный и конечный участки исходного рабочего файла. Поясним особенности рабочего файла.

Таблица 8.2

1	2	3	4	5
105	TAA	Z	STP	bbbbbbbbb
104	TAC	Y	Tyr	bbbbbbbbbb
103	GAA	Е	Glu	bbbbbbbbb
102	TGG	W	Trp	bbbbbbbbbb
101	ACG	Т	Thr	bbbbbbbbb
100	GAG	Е	Glu	bbbbbbbbbb
6	GAA	Е	Glu	bbbbbbbbbb
5	GGC	G	Gly	bbbbbbbbbb
4	AGC	S	Ser	bbbbbbbbbb
3	TGG	W	Trp	bbbbbbbbbb
2	GAA	Е	Glu	bbbbbbbbbb
1	GCT	A	Ala	bbbbbbbbbb
0	ATG	М	MET	bbbbbbbbbb

Начальный и конечный участки исходного рабочего файла репрессора метионина

Нумерация аминокислот белка записана в первом столбце и имеет направление снизу вверх, что связано с моделированием процесса биосинтеза белка, который осуществляется от N-конца к C-концу. Во втором столбце находятся триплеты, кодирующие аминокислотную последовательность. В третьем столбце записано однобуквенное обозначение аминокислот, а в четвертом – трехбуквенное. Аминокислота метионин (Met), которая является сигналом инициации синтеза белка, записана под номером 0, а сигнал окончания синтеза обозначается в этом столбце трехбуквенным символом (Stp). Результаты прогнозирования в виде десятизначных булевых чисел записываются в пятом столбце.

В том случае, когда для форматирования используются аминокислотные последовательности, второй столбец в файле отсутствует.

Алгоритм прогнозирования вторичной структуры белка

Согласно предложенному методу [113] процедура прогнозирования любого белка может быть разделена на несколько этапов, которые включены в алгоритм написанной для ускорения прогнозирования программы PREDICTOR [111]. Эти этапы таковы:

– выделение и кодирование ПФ для его поиска в базе данных;

- поиск ПФ в файлах заданного списка подпапок;
- запись названий подпапок с найденным ПФ в рабочий файл;
- задание подпапок для последующего поиска ПФ;
- интерпретация результатов прогнозирования.

Придерживаясь данного перечня, последовательно рассмотрим эти этапы на приведенном в [113] примере.

Выделение и кодирование $\Pi \Phi$ для его поиска в базе данных. В рабочем файле белка репрессора метионина программа выделяет $\Pi \Phi$ в самом начале полипептидной цепи (табл. 8.3). При этом Met, имеющий нулевой номер, в состав $\Pi \Phi$ не включается.

Таблица 8.3

	-			
1	2	3	4	5
5	GGC	G	Gly	bbbbbbbbbb
4	AGC	S	Ser	bbbbbbbbbb
3	TGG	W	Trp	bbbbbbbbbb
2	GAA	Е	Glu	bbbbbbbbbb
1	GCT	A	Ala	bbbbbbbbbb
0	ATG	М	MET	bbbbbbbbbb

Выделение первого пентафрагмента в рабочем файле белка репрессора метионина

Для целей поиска ПФ в файлах подпапок базы данных программа запоминает названия аминокислот и кодирует их в соответствии с разбиением аминокислот на подгруппы (см. 7.2, табл. 7.9). Программа присваивает каждой аминокислоте ПФ номер той группы преобразований, в которую она входит.

Для первого ПФ отнесение к группам преобразований таково:

5. Gly – 1; 4. Ser – 3; 3. Trp – 3; 2. Glu – 4; 1. Ala – 2.

Считывание аминокислот производится сверху вниз, а запись числа – слева направо: 13342. Полученное кодировочное число записывается в память компьютера.

 0111111111, 0011111111. Эти варианты отражают все изменения в парах переменных в процессе их нарезки и сортировки в папки и подпапки (см. 7.2).

В данном примере программа ведет поиск запомненного кодового номера пентафрагмента в подпапках данного списка и находит его в первой группе подпапок, в подпапке 0000000000. В этой подпапке имеется ряд файлов, среди которых находится файл с запомненным кодовым номером 13342 (выделен жирным шрифтом):

> 11111_000000000.txt 11112_000000000.txt 11113_0000000000.txt 13341_0000000000.txt 13342_0000000000.txt 13343_0000000000.txt 13344_0000000000.txt 44442_0000000000.txt 44443_000000000.txt 44444_0000000000.txt

В этом файле обнаруживается следующая последовательность аминокислот, читаемая сверху вниз:

Gly Ser Trp Glu Ala

Данная последовательность совпадает с запомненной последовательностью выделенного пентафрагмента: Gly, Ser, Trp, Glu, Ala. Это означает, что искомый первый пентафрагмент найден в папке 0000000000.

Запись названий подпапок с найденным ПФ в рабочий файл. При нахождении первого ПФ в базе данных на основе исходного списка папок этот ПФ считают начальным, фиксируют номер папки базы дан-

ных, содержащей начальный ПФ, и вносят номер подпапки базы данных, содержащей начальный ПФ, в рабочий файл исследуемого белка.

В табл. 8.4 приведен пример записи начального ПФ (номера папки с найденным ПФ **000000000**) в рабочем файле репрессора метионина (см. пятую строку, выделенную жирным шрифтом, пятый столбец).

Таблица 8.4

1	2	3	4	5
6	GAA	E	Glu	bbbbbbbbbb
5	GGC	G	Gly	000000000
4	AGC	S	Ser	bbbbbbbbbb
3	TGG	W	Trp	bbbbbbbbb
2	GAA	Е	Glu	bbbbbbbbb
1	GCT	A	Ala	bbbbbbbbbb
0	ATG	М	MET	bbbbbbbbbb

Запись номера папки начального пентафрагмента и выделение следующего пентафрагмента в рабочем файле белка репрессора метионина

Задание подпапок для последующего поиска $\Pi \Phi$. После нахождения начального П Φ , в случае, если этот П Φ совпадает с первым П Φ в последовательности аминокислот исследуемого белка, производят сдвиг вперед вдоль последовательности аминокислот в рабочем файле исследуемого белка на одну аминокислоту и выделяют новый П Φ . Его также запоминают, кодируют и на основе вновь созданного списка подпапок для поиска ищут новый П Φ в этих подпапках.

Составление списка подпапок основано на наблюдении (см. 7.2), что в процессе нарезки ПФ на каждом этапе в их описании к десятизначному числу добавлялась одна пара переменных слева и удалялась одна пара переменных справа (см. табл. 7.4, столбец внизу). Всего имеется четыре варианта пар переменных, которые были использованы при кодировании Н-связей ПФ с помощью десятизначных булевых чисел: 00, 01, 10, 11 (см. табл. 7.1). В результате нарезки каждый ПФ получал свое обозначение в виде десятизначного числа. Именно эта особенность формирования базы данных и была использована при разработке алгоритма прогнозирования.

Подпапка, содержащая ранее найденный начальный П Φ (он является в данном случае и первым в прогнозируемом белке), имеет номер 0000000000. На основе этого номера программа создает список из четырех папок приписыванием к этому номеру с левой стороны

Далее проводятся поиск нового ПФ в базе данных на основе созданного списка папок, фиксирование номера папки базы данных, содержащей найденный пентафрагмент, и внесение номера папки базы данных, содержащей найденный ПФ, в рабочий файл исследуемого белка.

Дальнейшая процедура прогнозирования происходит на основе повторного составления нового списка подпапок на основе предыдущего номера папки с найденным ПФ и поиска последующих ПФ до конца исследуемого белка. В качестве примера в табл. 8.5 приведен фрагмент рабочего файла, отражающий результаты прогнозирования вторичной структуры белка репрессора метионина с помощью программы PREDICTOR.

Интерпретация результатов прогнозирования. Из табл. 8.5 видно, что прогнозирование начинается с пятой аминокислоты, что соответствует первому выделенному ПФ, который и был найден в подпапке 000000000. При этом обнаружение ПФ в подпапке 0000000000 продолжается до 28-й аминокислоты. Это означает, что весь участок белка с 5-й по 28-ю аминокислоты соответствует β-структуре. Начиная с 29-й аминокислоты, ПФ обнаруживаются уже в других подпапках. В табл. 8.5 этот фрагмент белка с 29-й по 36-ю аминокислоты предсказывается как переходный от β-структуры к α -спирали. Такие переходы даны в табл. 7.1 и 7.10. Далее фрагмент белка с 37-й по 43-ю аминокислоты прогнозируется десятизначными числами 1111111111, что соответствует номеру подпапки, содержащей α -спиральные ПФ. Наконец, начиная с 44-й аминокислоты, прогнозируется участок перехода от α -спирали к β-структуре.

Таким образом, зная, к какому типу структуры относится то или иное десятизначное его описание, легко «прочитать» положение и тип вторичной структуры исследуемого белка. Приведенный пример является простейшим и не исчерпывает все возможные варианты десяти-

Таблица 8.5

Фрагмент записи результатов прогнозирования структуры	
белка репрессора метионина в рабочем файле программы PREDICTOR	Ľ

45	GTG	V	Val	1010111111	
44	CAG	Q	Gln	1011111111	
43	CGT	R	Arg	1111111111	
42	CGT	R	Arg	1111111111	
41	ACG	Т	Thr	1111111111	Vuortor
40	CGC	R	Arg	1111111111	y daelok
39	GAA	Е	Glu	1111111111	α-спирали
38	GAT	D	Asp	1111111111	
37	ACC	Т	Thr	1111111111	
36	CTC	L	Leu	1111111101	
35	ATC	Ι	Ile	1111110101	
34	AAA	K	Lys	1111010101	Переходный
33	TTA	L	Leu	1101010101	участок
32	GTG	V	Val	01010101 00	«β-структура –
31	AAG	K	Lys	010101 0000	α-спираль»
30	CTT	L	Leu	0101 000000	
29	CCT	Р	Pro	01 00000000	
28	ATT	Ι	Ile	000000000	
27	TCC	S	Ser	000000000	
26	GTT	V	Val	0000000000	
25	ACG	Т	Thr	000000000	
24	ATT	Ι	Ile	0000000000	
23	AAG	K	Lys	000000000	
22	AAA	K	Lys	000000000	
21	GTC	V	Val	0000000000	
20	CAA	Q	Gln	0000000000	
19	GAA	Е	Glu	0000000000	
18	AGT	S	Ser	000000000	
17	AAG	K	Lys	0000000000	
16	AAG	K	Lys	0000000000	
15	GGC	G	Gly	000000000	Участок
14	CAC	Н	His	0000000000	β-структуры
13	GAG	Е	Glu	0000000000	
12	GCT	А	Ala	000000000	
11	TAC	Y	Tyr	0000000000	
10	CCA	Р	Pro	0000000000	
9	AGC	S	Ser	000000000	
8	ATC	Ι	Ile	0000000000	
7	TAT	Y	Tyr	0000000000	
6	GAA	E	Glu	0000000000	
5	GGC	G	Gly	0000000000	
4	AGC	S	Ser	bbbbbbbbbb	
3	TGG	W	Trp	bbbbbbbbbb	
2	GAA	Е	Glu	bbbbbbbbbb	
1	GCT	А	Ala	bbbbbbbbbb	
0	ATG	М	MET	bbbbbbbbbb	

значных описаний. Чтобы представить эти варианты, можно создать специальную программу по их визуализации или использовать каталог вторичных структур (см. гл. 9).

8.3. Проблемы алгоритма прогнозирования и предложенные пути их решения

Несовпадение начального ПФ с первым выделяемым ПФ

В процессе прогнозирования мы столкнулись с рядом моментов, усложняющих алгоритм прогнозирования. Однако последующая их проработка позволила решить все проблемы введением в алгоритм дополнительных вариантов возможных ситуаций. Рассмотрим некоторые из них.

Вариант несовпадения начального ПФ с первым выделяемым ПФ является более общим, по сравнению с рассмотренным ранее примером. В [113] он рассмотрен на примере прогнозирования вторичной структуры белка цитохрома С2. Для детального ознакомления с этим вопросом мы отсылаем читателя к [113]. Отличие от приведенного примера состоит в том, что программа запоминает найденный начальный ПФ и его положение в цепи белка. Прогнозирование начинается с найденного начального ПФ и проводится до конца белковой цепи (до обнаружения STPсигнала). После этого программа возвращается к запомненному начальному ПФ и ищет ПФ в списке подпапок, задаваемых от исходной подпапки, но на основе добавления пар переменных 00, 01, 10, 11 не с левой, а с правой стороны. Например, если исходная подпапка была 000000000, то список будет таким: 000000000, 000000001, 000000010, 000000011. Все остальные процедуры такие же, как и в приведенном ранее примере. Прогнозирование ведется до обнаружения нулевой аминокислоты (Met), означающей начало синтеза белка.

Обнаружение ложного начального ПФ

В процессе поиска начального ПФ могут обнаруживаться ложные начальные ПФ. Суть их состоит в том, что при последующем трансляционном сдвиге на одну аминокислоту новый ПФ не обнаруживается и программа останавливается. Чтобы этого не происходило, в алгоритм введена запись каждого из проверяемых ПФ на четыре шага. Этим обеспечивается страховка программы от останова, если ложный ПФ проявляется не сразу, а спустя 3–4 шага. При появлении тупика программа возвращается к исходному записанному ПФ и осуществляет следующий трансляционный сдвиг по белковой цепи без записи в рабочий файл. Только когда пройдено четыре шага и останова программы не произошло, обнаруженный начальный ПФ записывается в файл и производятся дальнейшие действия согласно алгоритму.

Наличие одинаковых ПФ в нескольких альтернативных подпапках

Согласно алгоритму после нахождения очередного ПФ программа PREDICTOR на основе номера подпапки, в которой обнаружен $\Pi \Phi$, составляет новый список подпапок и просматривает их. Иногда возникает ситуация, когда одинаковые ПФ обнаруживаются одновременно в двух и более подпапках. На случай возникновения такой ситуации в алгоритм введена запись названия подпапки, на основе которой был сформирован список. Программа просматривает поочередно каждую из альтернативных подпапок на несколько этапов вперед с записью в свою память просматриваемой папки и всего пути от нее. Было установлено, что просмотр вариантов должен осуществляться максимально на 3-4 стадии. Если просмотр четырех стадий приводит к тупику, то программа возвращается на этап формирования исходного списка подпапок и просматривает следующую подпапку на 4 стадии вперед и т. д. Если после четырех стадий тупика не обнаруживается, программа записывает в память этот номер подпапки и возвращается в исходную подпапку, на основе которой был сформирован список подпапок. При этом программа записывает номер этой подпапки в рабочий файл, переходит в запомненную подпапку, где нет тупика, и записывает ее номер в рабочий файл. Далее программа обнаруживает все последующие ПФ и записывает все последующие номера подпапок согласно алгоритму. При продвижении от начального ПФ к началу белковой цепи эта же проблема решается аналогично.

8.4. Сопоставление результатов прогнозирования с экспериментальными данными

Критерием эффективности предлагаемого способа прогнозирования вторичной структуры белка является сопоставление положений α-спиральных и β-структурных фрагментов, а также изгибов β-структуры в первичной структуре белка, прогнозируемое на основе предлагаемого метода, с данными экспериментального исследования тех же белков методом РСА. Для сравнения мы использовали структуры репрессора метионина [148], цитохрома C₂ [149] и фрагмента рецептора гормона эстрогена [150], приведенные в качестве примеров в [113].

Сопоставление результатов прогнозирования положения α-спиральных и β-структурных фрагментов, а также изгибов β-структуры, полученных с помощью предлагаемого метода, с экспериментальными данными иллюстрирует табл. 8.6. Видно, что они практически полностью совпадают (с точностью ± 1 аминокислота). Существующие мето-

Таблица 8.6

	TT ~	Тип вторичной структуры								
Nº	Название белка, индекс в Protein	α-спираль		β-стру	ктура	Центры изгибов β-структуры				
	алинокислот (AK)	Прогноз	Экспер.	Прогноз	Экспер.	Прогноз	Экспер.			
		Положение на белке								
	Репрессор	33 - 47,	32 - 47,	1 - 32,	3 - 31,	76 и 77,	76 и 77,			
1	метионина, 1cmb	56 – 67,	55 - 67,	80 - 88	80 - 87	102	102			
	[148] 107 AK	89 - 96	88 - 97							
	Цитохром C ₂ ,	7 – 12,	7 – 12,	20 - 52,	19 – 51,	18,	18,			
2	2c2c [149]	53 – 59,	52 – 59,	84 - 100	84 - 100	63	63			
2	111 AK	67 – 83,	66 – 83,							
		101 - 109	101 - 111							
	Фрагмент рецептора	28 - 37,	27 - 36,	1 – 10,	3 - 10,	11,	11,			
3	гормона эстрогена,	63 - 71	63 – 71	12 - 27,	12 – 26,	56	56			
	1hcq [150] 74 AK			38 - 55	38 - 54					

Сопоставление результатов прогнозирования вторичной структуры белков с помощью предлагаемого метода с экспериментальными данными

ды прогнозирования вторичной структуры белка на основе его первичной структуры обладают такими недостатками, как предсказание ложных фрагментов вторичной структуры или неполное предсказание всех фрагментов вторичной структуры, что связано с методологическими недостатками этих методов (например, с вероятностным характером проводимых вычислений [128], [147]). При прогнозировании вторичной структуры белка нашим методом не обнаруживается ни ложных фрагментов вторичной структуры, ни неполного выявления фрагментов вторичной структуры. Близкая к 100 % точность прогнозирования вторичной структуры белка связана с особенностями данного метода, основанного на выделении ПФ, идущих со сдвигом в одну аминокислоту, и их поиске в базе данных ПФ белков. В названиях подпапок найденных ПФ содержится информация о последовательности аминокислот и H-связях ПФ во вторичной структуре белка.

В заключение отметим, что пополнение имеющейся базы данных будет постепенно расширять предсказательные возможности предлагаемого метода. Тем не менее, проблему его универсальности таким способом вряд ли удастся решить. Единственным кардинальным решением является разработка искусственной базы данных ПФ белков.

Глава 9. КАТАЛОГ ДЕСЯТИЗНАЧНЫХ ОПИСАНИЙ ВТОРИЧНЫХ СТРУКТУР БЕЛКА

9.1. Постановка задачи

Причины создания каталога

База данных ПФ белка [110], описанная в гл. 7, была использована нами при разработке метода прогнозирования вторичных структур белка на основе заданной первичной структуры (последовательности аминокислот) [113] (см. гл. 8). Однако возможности ее применения оказались значительно шире, и в дальнейшем нами был предложен метод проектирования первичной структуры белка с заданной вторичной структурой [114]. По сути оба метода решают задачи, имеющие разную постановку, но одну принципиальную основу, – базу данных ПФ.

Уже при разработке первого метода, когда получали последовательность десятизначных булевых чисел, описывающих прогнозируемую вторичную структуру, возникла необходимость их структурной интерпретации (см. 8.2). В простейших случаях провести такую интерпретацию не составляет большого труда. Однако вариабельность вторичных структур такова, что во многих случаях необходимо «иметь под рукой» каталог десятизначных описаний вторичных структур и желательно – вид структуры, чтобы осуществить правильное ее прочтение. С другой стороны, разработанный нами метод проектирования первичных структур белка по заданному описанию вторичной структуры [114] также требует наличия каталога вторичных структур, поскольку выбор последовательности вводимых пар переменных определяется десятизначным описанием задаваемой структуры.

Идеальным решением является создание визуализатора, который мог бы использовать десятизначное описание вторичной структуры для перевода его в трехмерное изображение на экране монитора. Принципиальных ограничений по его созданию нет. Однако пока такой визуализатор не разработан, можно на основе имеющихся возможностей создать каталог десятизначных описаний вторичной структуры.

Есть два подхода к построению такого каталога.

В рамках первого подхода в 7.3 была предложена модель структуры папок ПФ. Используя структуру из восьми гиперкубов можно прокладывать траектории различной длины и в различных сочетаниях переменных, т. е. такая структура может служить генератором описаний вторичных структур белка, содержащих основные ПФ. Этот подход продуктивен, но есть опасность сгенерировать структуры, не встречающиеся в реальных белках.

В рамках второго подхода поиск вторичных структур можно вести непосредственно в текстовых файлах. Этот подход трудоемок, но опирается на реальные структуры и потому более надежен.

Мы избрали второй подход, чему способствовало наличие большой выборки текстовых файлов. Хотя нельзя исключить, что приведенные таблицы описывают не все возможные варианты и спустя некоторое время созданный каталог можно будет дополнить и откорректировать.

Принципы составления каталога

В процессе нарезания и систематизации ПФ на основе текстовых файлов (см. 7.2) были выявлены две особенности десятизначных описаний вторичных структур.

Первая особенность состоит в том, что эти описания формируются последовательно, в виде дискретных стадий. На каждой стадии происходит добавление пары переменных с левой стороны и удаление пары переменных с правой стороны. В результате этого описания вторичных структур формируются в виде столбцов, в которых десятизначные цифры идут снизу вверх, в порядке выделения ПФ на первичной структуре белка. Это означает, что описания фрагментов вторичной структуры должны состоять из столбцов с последовательно идущими снизу вверх десятизначными числами.

Вторая особенность связана с антисимметрией вторичных структур. Если десятизначное описание имеет отношение к α -спиральному участку, то при замене булевых чисел 1 $\leftarrow \rightarrow 0$ точно такое же описание будет существовать для β -структурного участка. Внешне такие структуры совершенно не похожи друг на друга, но в этом и состоит суть антисимметрии. В качестве примера в табл. 7.10 отмечена антисимметрия переходных участков « β -структура – α -спираль» и « α -спираль – β -структура». Наличие антисимметрии позволило сделать вывод, что систематизацию вариантов описаний необходимо проводить парами: слева расположить варианты на β -структуре, а рядом (справа) – варианты на α -спирали. Отмеченные особенности были использованы в качестве принципиальной основы при составлении каталога.

В составе вторичных структур было выделено 4 класса:

- β-структуры и α-спирали;

– переходные участки « β -структура $\leftarrow \rightarrow \alpha$ -спираль»;

 – спиральные вздутия на фоне β-структур и впадины на фоне α-спиралей;

- изгибы β-структур и изломы α-спиралей.

Такое разделение весьма условно и, как будет показано далее, провести четкую границу между разными классами трудно. Далее будет показано также, что классы структур естественным образом переходят друг в друга.

Количество представителей в классах значительно отличается. Наименее интересен основной класс вторичных структур – « β -структуры и α -спирали», в котором каждая структура имеет лишь один вариант конформации. Много структур имеется в классе переходных участков « β -структура $\leftarrow \rightarrow \alpha$ -спираль». Интересен класс, «спиральные вздутия на фоне β -структур и впадины на фоне α -спиралей», в котором образуются серии вторичных структур. Важен и четвертый класс («Изгибы β -структур и изломы α -спиралей»), имеющий немного представителей.

В каталоге также были выделены побочные подклассы. Они возникли из основных классов в результате изменений в некоторых структурах. Так, были выделены трансляционные сдвиги и модификации изгибов β-структуры и изломов α-спирали, а также модификации переходных участков вторичных структур.

Для удобства использования таблицы каталога вынесены в приложение. Поскольку все таблицы приложения относятся к гл. 9, то их нумерация начинается с цифры 9. Приведенные в каталоге иллюстративные материалы состоят из трех элементов:

– фрагментов текстовых файлов, полученных с помощью программы PROTEIN 3D на основе файлов PROTEIN DATA BANK;

- десятизначных описаний вторичной структуры этих фрагментов;

– пространственных изображений этих вторичных структур, визуализированных с помощью программы PROTEIN 3D. Для сопоставления визуализированных структур начальную аминокислоту в них располагали справа (она выделена крупным шрифтом в прямоугольнике). При этом нумерация аминокислот в цепи направлена справа налево.

9.2. Участки β-структуры и α-спирали: примеры описания

В процессе изложения неоднократно упоминалось описание β-структурных и α-спиральных участков белка (см., например, табл. 7.4 и 7.10) с помощью десятизначных чисел 000000000 и 1111111111. В табл. 9.1 приведены ПФ этих двух типов структур и их пространственная структура (исходный PDB-файл – 3adk). Поскольку далее таблицы будут содержать различные примеры структур и их описаний, то для идентификации им будут присваиваться порядковый номер (в порядке их появления в таблицах) и буквенный индекс. Структуры, расположенные в таблицах слева, будут записываться с индексом A, а структуры, расположенные справа, – с индексом Б. В табл. 9.1 (см. приложение) они обозначены соответственно как 1-А и 1-Б. Если несколько структур будут относиться к одному классу, то они будут записываться через дефис, например: 2-1-А; 2-2-А; 2-3-А; 2-1-Б; 2-2-Б; 2-3-Б и т. д.

Как видно из табл. 9.1, аминокислоты ПФ β -структуры не имеют водородных связей в ближайшем окружении (i...i - 4, i...i + 4), и ПФ описывается десятизначным числом 000000000. В ПФ α -спирали эти связи имеются:

132 Arg O – 136 Ser N,	132 Arg N – 128 Arg O – 11 ;
131 Lys O – 135 Thr N,	131 Lys N – 127 Lys O – 11 ;
130 Leu O – 134 Glu N,	130 Leu N – 126 Thr O – 11 ;
129 Leu O – 133 Gly N,	129 Leu N – 125 Met O – 11 ;
128 Arg O – 132 Arg N,	128 Arg N – 124 Thr O – 11 ,

и потому они могут быть описаны десятизначным числом 1111111111.

Обратите внимание, что связь 132 Arg N – 128 Arg O повторяется еще раз: 128 Arg O – 132 Arg N, что обусловлено циклическим характером этого фрагмента (см. табл. 7.6). Вид пространственной структуры ПФ с β-структурой и α-спиралью показан в нижней части табл. 9.1.

9.3. Канонические переходные участки «β-структура – α-спираль» и «α-спираль – β-структура»

Нумерация связей и этапов

Следующим типом структур, которые будут рассмотрены, являются канонические переходные участки « β -структура – α -спираль» и « α -спираль – β -структура». В силу того, что данные последовательности десятизначных чисел (табл. 7.10) представляют собой постепенное изменение вторичной структуры (α -спираль $\leftarrow \rightarrow \beta$ -структура) без каких-либо дефектов, эти участки были названы каноническими, в отличие от других вариантов, которые могут содержать отличия от исходных структур.

Для более упорядоченного рассмотрения вариантов, отличающихся от канонических переходов, в табл. 9.2 (см. приложение) слева от текстовых файлов приведены номера связей, которыми в этих файлах будут обозначаться переходные участки. На этапе перехода от β -структуры к α -спирали эти связи нумеруются снизу вверх с индексом «*a*» (левые структуры), а на этапе перехода от α -спирали к β -структуре – сверху вниз с индексом «*б*», что связано с симметрией этих связей. Стадии, приведенные в табл. 9.2, нумеруются точно так же, как в табл. 7.10.

Пример канонических переходных участков

Названная пара переходных участков в виде фрагментов текстовых файлов, десятизначных описаний и пространственной структуры приведена в табл. 9.2 (текстовый файл белка 3adk). Из этой таблицы видно, что этапу переходов от β -структуры к α -спирали (структура 2-А) предшествует участок β -структуры, состоящий из пяти аминокислот, который описывается десятизначным числом 0000000000. Протяженность β -структуры может изменяться в широких пределах и будет иметь вид повторяющихся десятизначных чисел 0000000000. При переходе от β -структуры к α -спирали описание участка состоит из двух этапов. Первый этап содержит четыре стадии (2...5). Он связан с добавлением к десятизначному числу пар переменных 01 справа (они выделены жирным шрифтом и соответствуют связям 1а–4а в текстовом файле) и удалением пар переменных 00 слева. На втором этапе, также состоящем из четырех стадий (6...9), происходит добавление пар переменных 11 справа и удаление пар переменных 01 слева.

166

На стадии 10 этого перехода приведена полная α-спираль (в табл. 9.2 ее составляют только пять аминокислот). Повторяемость аминокислот, имеющих пару Н-связей (переменные 11), может изменяться в широких пределах (до нескольких десятков), что зависит от свойств конкретного белка.

Переходу от α -спирали к β -структуре (структура 2-Б) также предшествует участок α -спирали (десятизначное число 1111111111). В самом переходе наблюдается также два этапа. *На первом этапе* при прохождении стадий 11...14 происходит добавление справа пар переменных 10 (они соответствуют связям 4б–1б) и удаление слева пар 11. *На втором этапе* при прохождении стадий 15...18 происходит добавление пар переменных 00 и удаление пар переменных 10. Заканчивается переход β -структурой (стадия 19). Отдельные стадии этих двух переходов можно преобразовать друг в друга по принципу антисимметрии (0 $\leftarrow \rightarrow$ 1). Так стадия 2 (**01**0000000) преобразуется в стадию 11 (**10**1111111), стадия 3 (**0101**000000) – в стадию 12 (**1010**11111) и т. д. Вид пространственной структуры рассмотренных переходных участков приведен в нижней части табл. 9.2.

9.4. Спиральные вздутия на фоне β-структур и впадины на фоне α-спиралей

Эти два класса вторичных структур образуют серии, в которые входит по пять структур и их десятизначных описаний, приведенных парами в табл. 9.3 (см. приложение).

Как видно из этой таблицы, первая структура 3-1-А содержит 5 пар Н-связей (спиральное вздутие), а в парной к ней структуре 3-1-Б пропущены 5 пар Н-связей. Мы использовали их в качестве исходных структур, так как в структуре 3-1-А содержится полный α-спиральный участок из пяти аминокислот на фоне β-структуры (текстовый файл 1ald). На нем представлены все связи переходных участков и десятизначное описание стадий со 2-й по 18-ю. Аналогично в структуре 3-1-Б представлена неспиральная впадина – фрагмент β-структуры из пяти аминокислот на фоне α-спирали (текстовый файл 3n5о). Она также содержит все связи, но они расположены в иной последовательности. Сначала снизу вверх идут связи 4б, 3б, 2б, 1б, а затем – связи 1а, 2а, 3а и 4а. Цифровое описание содержит сначала описание стадий с 11-й по 18-ю, а затем – с 1-й по 9-ю. Сравнение двух структур позволяет увидеть, что десятизначные описания стадий, включая стадии 10 и 1 (полная α -спираль и полная β -структура), превращаются друг в друга на основе преобразования антисимметрии (0 $\leftarrow \rightarrow$ 1), например: **01**00000000 $\leftarrow \rightarrow$ **10**11111111... ...**1111111111** $\leftarrow \rightarrow$ 0000000000 и т. д. Внизу показан вид структур 3-1-А и 3-1-Б, на которых хорошо просматривается их антисимметрия.

Следующая в этой серии пара структур – 3-2-А и 3-2-Б. Они содержат фрагменты текстовых файлов, в которых уже имеется только четыре аминокислоты, описываемые парами переменных 11 (структура 3-2-А) и 00 (структура 3-2-Б). Связи в переходных участках остаются неизменными, а в десятизначных описаниях исчезают соответственно стадии 10 и 19. При этом β-структуры в структуре 3-2-А более сближены, по сравнению со структурой 3-1-А, протяженность α-спирального участка сокращается, относительное расположение β-структурных участков меняется. Сходные изменения наблюдаются для спиралей структуры 3-2-Б при одновременном сокращении протяженности β-структуры. Однако изменения в относительном расположении α-спиралей менее заметны.

В третьей паре структур также исчезает по две стадии: 9-я и 11-я в структуре 3-3-А и 2-я и 18-я в структуре 3-3-Б, но появляются соответственно десятизначные числа **1011111101** и **0100000010**. Относительное расположение участков β-структуры в структуре 3-3-А, по сравнению со структурой 3-2-А, меняется, а участков α-спирали в структуре 3-3-Б почти не меняется. Количество связей в переходных участках не меняется.

В четвертой паре структур 3-4-А и 3-4-Б исчезают еще по две стадии (8-я, 12-я и 17-я, 3-я) и в каждой из них появляется по две новые стадии, описываемые двумя десятизначными числами. Для структуры 3-4-Б заметно сокращается расстояние между спиральными участками. В паре структур 3-5-А и 3-5-Б исчезают стадии 7-я, 13-я и 16-я, 4-я и появляется по три новые стадии, описываемые тремя новыми десятизначными числами.

Наконец, в шестой паре структур (3-6-А и 3-6-Б) исчезают соответственно стадии 6-я, 14-я и 15-я, 5-я и появляется четыре новые стадии. В парах структур 4-6 в переходных участках количество связей также не меняется. Это означает, что все модификации, которые происходят в канонических переходных участках (см. 9.6) могут проявиться и в рассмотренных структурах.

9.5. Изгибы β-структур и изломы α-спиралей Исходные структуры

Дальнейшие варианты, возможные на основе приведенных в табл. 9.3 примеров, связаны изменениями, которые могут происходить в другом направлении: вместо удаления аминокислот с двумя водородными связями (варианты группы А) – внедрение аминокислот без водородных связей и, наоборот, вместо удаления аминокислот без водородных связей – внедрение аминокислот с двумя водородных связей – внедрение аминокислот с двумя водородными связями (варианты группы Б). Все эти варианты последовательно представлены в табл. 9.4 (см. приложение), они выглядят как изгибы β-структур и изломы α-спиралей.

В качестве исходных структур для получения новых вариантов были использованы структуры 3-6-А и 3-6-Б. Внедрение в структуру 3-6-А одной аминокислоты, не содержащей Н-связи, приводит к структуре 4-1-А, а введение в структуру 3-6-Б одной аминокислоты с двумя Н-связями – к структуре 4-1-Б (табл. 9.4). В результате в структуре 4-1-А, по сравнению со структурой 3-6-А, исчезают канонические стадии 5-я и 15-я и появляется пять новых стадий, описываемых десятизначными числами 0001010100, 1000010101, 1010000101, 1010100001, 0010101000. С правой стороны внедрение аминокислоты с парой Н-связей в структуру 3-6-Б приводит к исчезновению в структуре 4-1-Б канонических стадий 6-й и 14-й и появлению пяти стадий: 1110101011, 0111101010, 0101111010, 0101011110, 1101010111. Парное сопоставление этих стадий показывает, что по своему описанию они антисимметричны (0 $\leftarrow \rightarrow$ 1), например: 0001010100 ← → 1110101011, 1000010101 ← → 0111101010 и т. д. В изгибе можно увидеть наличие трех идущих подряд водородных связей, а в изломе – разрыв трех систем HN–С=О-групп (ССИВС).

Последующее внедрение двух и трех аминокислот, не образующих Н-связи слева (варианты 4-2-А и 4-3-А в табл. 9.4), и аналогичного количества аминокислот, образующих по две Н-связи (структуры 4-2-Б и 4-3-Б), приводит в этих вариантах к соответствующему исчезновению канонических стадий и появлению новых промежуточных стадий. Так, в варианте 4-2-А исчезают стадии 4-я и 16-я и появляется 6 стадий, описываемых числами 0001010000, 0000010100, 1000000101, 1010000001, 0010100000, 0000101000, а в варианте 4-2-Б исчезают канонические стадии 7-я и 13-я и появляется 6 стадий, описываемых числами 1110101111, 1111101011, 0111111010, 0101111110, 1101011111, 1111010111. В структуре изгиба 4-2-А видны две водородные связи, а в структуре излома 4-2-Б – два разрыва ССИВС.

Аналогично в варианте 4-3-А дополнительно исчезают канонические стадии 3-я, 17-я и появляется 7 новых стадий, описываемых числами 000100000, 0000010000, 0000000100, 100000001, 0010000000, 0000100000, 0000001000. Можно отметить также, что в структуре 4-3-Б исчезают стадии 8-я, 12-я и появляется 7 стадий с числами 111011111, 1111101111, 111111011, 011111110, 110111111, 1111011111, 1111101111. В изгибе β-структуры 4-3-А остается лишь одна H-связь, а в изломе αспирали 4-3-Б сохраняется разрыв одной ССИВС. Во всех приведенных примерах новые десятизначные числа слева преобразуются в десятизначные числа справа посредством переходов 0 $\leftarrow \rightarrow 1$.

Можно также отметить, что количество связей, относимых к переходным участкам, постепенно снижается. Так, в структурах 4-1-А и 4-1-Б их остается три, в структурах 4-2-А и 4-2-Б – две, а в структурах 4-3-А и 4-3-Б – только одна.

Таким образом, последовательное внедрение аминокислот в структуры 3-6-А и 3-6-Б привело к появлению трех вариантов изгибов β-структуры и трех вариантов изломов α-спиралей, в которых последовательно исчезают канонические стадии переходов от β-структуры к α-спирали.

Трансляционные сдвиги

Следует отметить, что структуры 4-2-А и 4-2-Б можно рассматривать также как трансляционный сдвиг двух повторяющихся структур 4-3-А и 4-3-Б. Это означает, что на их основе можно получить новые варианты, отличающиеся от исходных вариантов значением трансляционного сдвига. Две пары таких вариантов были найдены (табл. 9.5, приложение).

Если в исходных структурах 4-2-А и 4-2-Б наблюдался трансляционный сдвиг циклических ПФ на одну стадию, то в первой паре структур табл. 9.5 (5-1-А и 5-1-Б) этот сдвиг происходит на две стадии. Каноническим стадиям в десятизначном описании соответствуют лишь две стадии: 2-я, 18-я для структуры 5-1-А и 9-я, 11-я для структуры 5-1-Б. В изгибе 5-1-А присутствуют две Н-связи, а в изломе 5-1-Б – два разрыва ССИВС, что хорошо видно на приведенных пространственных структурах.

Во второй паре структур (5-2-А и 5-2-Б) происходит трансляционный сдвиг в три стадии. В их описании также присутствуют лишь по два канонических десятизначных числа (2, 18 и 9, 11). В структуре 5-2-А также наблюдается две Н-связи, а в структуре 5-2-Б – разрыв двух ССИВС.

Легко видеть, что десятизначные числа приведенных двух пар структур преобразуются друг в друга на основе антисимметрии (0 $\leftarrow \rightarrow$ 1), например в парах 5-1-А и 5-1-Б: **0001**000000 $\leftarrow \rightarrow$ **1110**111111, **000001**0000 $\leftarrow \rightarrow$ **111110**1111, **01000001**00 $\leftarrow \rightarrow$ **10111110**11, **1001000001** $\leftarrow \rightarrow$ **0110111110** и т. д.

Модификации исходных структур

Если взять за основу для последующего построения каталога структуры 4-3-А и 4-3-Б, то кроме трансляционных сдвигов в них возможны еще и модификации. В частности, связи, описываемые парой переменных 01, при однобитовых переходах превращаются в связи, описываемые парами переменных 00 и 11, т. е. $00 \leftarrow 01 \rightarrow 11$. Аналогично связи, описываемые парой переменных 10, превращаются в связи, описываемые парами переменных 00 и 11, т. е. $00 \leftarrow 10 \rightarrow 11$. Кроме того, между структурами можно обнаружить антисимметричные преобразования типа $0 \leftarrow \rightarrow 1$. В текстовых файлах были найдены лишь две структуры, удовлетворяющие этим модификациям, которые приведены в табл. 9.6 (см. приложение).

Как видно из табл. 9.6, десятизначные описания этих вариантов не содержат стадий канонических переходов. Структура 6-1-А представляет собой обратный изгиб β -структур. Если сопоставить десятизначное описание этой структуры с описанием исходной структуры 4-3-А, то можно сказать, что пара переменных 01 в прямом изгибе преобразуется посредством перехода 0 $\leftarrow \rightarrow$ 1 в пару 10, при этом изменений в фоновой β -структуре не происходит, например: **01**00000000 $\leftarrow \rightarrow$ **10**0000000, 00**01**000000 $\leftarrow \rightarrow$ 001**0**000000, 0000**1**00000 $\leftarrow \rightarrow$ 00001**0**0000 и т. д. Структуры, полностью антисимметричной структуре 6-1-А, нами в текстовых файлах пока не найдено.

Другой структурой, соответствующей переходам $01 \rightarrow 11$ и $10 \rightarrow 11$, является вариант 6-2-А, также показанный в табл. 9.6. Как хорошо видно на пространственной структуре, этот изгиб содержит две Н-связи. Антисимметричного варианта для этой структуры также пока не найдено.

9.6. Структурные модификации канонических переходов Предварительные замечания

Пронумерованность отдельных стадий канонических переходов и их связей позволила рассмотреть модификации канонических переходов системно. Принципиально такие модификации возможны в двух направлениях:

– связи, описываемые парами переменных 01 и 10, посредством однобитовых переходов превращаются в пары 00 и 11, т. е. $00 \leftarrow 01 \rightarrow 11$ и $00 \leftarrow 10 \rightarrow 11$;

– связи, описываемые парой переменных 01 и 10, посредством двубитовых переходов превращаются в пары переменных 01 и 10, т. е. $01 \rightarrow 10$ и $10 \rightarrow 01$.

Однако на практике мы обнаружили лишь один тип однобитовых переходов: $01 \rightarrow 11$ и $10 \rightarrow 11$.

Естественно, что такие изменения должны сопровождаться появлением дополнительной водородной связи, что и подтвердилось в процессе обнаружения таких переходов. При этом изменения происходили в канонических переходах в трех положениях связей $O \rightarrow N$ (1a, 2a, 3a) и трех положениях связей $N \rightarrow O$ (1б, 2б, 3б), а их количество менялось от одного до трех. Далее приведены результаты систематизации для модификаций одной, двух и трех H-связей.

Модификации одной Н-связи в канонических переходах

Модификации типа 01 \rightarrow 11 и 10 \rightarrow 11 для одной Н-связи обнаружены в положениях 1а, 2а, 3а для перехода « β -структура – α -спираль» и в положениях 1б, 2б, 3б для перехода « α -спираль – β -структура». Положения 4a и 4б не могут быть использованы, так как сливаются со спиральными участками белка. Примеры единичных модификаций и их описания с помощью десятизначных чисел, трехмерная структура, а также нумерация чисел, совпадающих с каноническими переходами, приведены в табл. 9.7 (см. приложение). Рассмотрим более подробно вариант модификации в положениях 1а и 1б. Как следует из табл. 9.7, в варианте 1а появляется дополнительная H-связь типа $N_iH...O_{i-4}$, предшествующая каноническому переходу. Аналогичная связь появляется в положении 1б, но после канонического перехода. Перемещение модификаций из положения 1а в положения 2а и 3а, а также из положения 1б в положения 2б и 3б сопровождается синхронным перемещением этой дополнительной связи, что хорошо видно в приведенных примерах. Характер остальных связей не изменяется. При этом, как видно из табл. 9.7, перед α -спиралью и после α спирали появляется характерная петля структуры с водородной связью.

Что касается стадии десятизначных описаний, то они совпадают с каноническими описаниями в следующих положениях: 1а – 2 и 7, 8, 9; 1б –11, 12, 13 и 18; 2а – 2 и 8, 9, 2б –11, 12 и 18; За – 2 и 9, 3б – 11 и 18. Иными словами: по мере перемещения модификации степень совпадения с каноническими десятизначными числами уменьшается.

По сравнению с описанием канонических переходов, где наблюдалась антисимметрия (0 $\leftarrow \rightarrow$ 1) на всех стадиях (табл. 9.2), в данных модификациях между десятизначными описаниями антисимметрия наблюдается только для стадий 2-й и 11-й, а также 9-й и 18-й. В остальных случаях также присутствуют элементы симметрии, но они носят комбинированный характер. Суть этой симметрии состоит в том, что между стадиями переходов левых и правых структур имеется преобразование только в отношении заменяемой пары переменных, тогда как замены «фона» не происходит, например: 0010000000 и 000000100, 1100000001 и 1000000011 и т. д.

Модификации двух Н-связей в канонических переходах

Поскольку в обоих канонических переходах имеется всего три из четырех положений, которые могут подвергаться модификациям (1a, 2a, 3a и 1б, 2б, 3б), то при модификации двух Н-связей теоретически возможны три пары вариантов: 1a, 2a и 1б, 2б; 1a, 3a и 1б, 3б; 2a, 3a и 2б, 3б. Все эти варианты были реально найдены в текстовых файлах и приведены в табл. 9.8 (см. приложение).

Из табл. 9.8 следует, что, как и в случае с модификацией одной Н-связи, в результате модификации наблюдается появление двух дополнительных Н-связей типа N_iH....O_{i-4}: для левых вариантов – предшествующих каноническому переходу, а для правых вариантов – следующих за каноническим переходом. При разном положении двух модифицированных связей смещается и положение дополнительных H-связей, что хорошо видно в приведенных примерах, причем характер остальных связей не изменяется. Наличие характерных структурных петель до и после α-спирали, как хорошо видно на рисунках в табл. 9.8, сохраняется.

Стадии десятизначных описаний совпадают с каноническими описаниями в следующих положениях: 1a, 2a – 2, 3 и 8, 9; 16, 26 – 11, 12 и 17, 18; 2a, 3a – 2, 3 и 9, 26, 36 – 11 и 17, 18; 1a, 3a – 2 и 9, 16, 36 – 11 и 18. В случае двух модификаций, как и в случае одной модификации, по мере их перемещения вдоль связей степень совпадения с каноническими десятизначными числами уменьшается. Антисимметрия $(0 \leftarrow 1)$, как и для случая модификации одной Н-связи, наблюдается лишь для стадий 2-й и 11-й, а также 9-й и 18-й. В остальных случаях симметрия носит комбинированный характер.

Модификации трех Н-связей в канонических переходах

Максимальное количество возможных модификаций – три Н-связи. Как уже отмечалось, модификация положений 4a и 4б сокращает количество связей в переходах до трех. По этой причине возможна лишь одна модификация трех положений в канонических переходах – это положения 1a, 2a, 3a и 1б, 2б, 3б. Эта модификация была обнаружена для обоих вариантов переходов (табл. 9.9, см. приложение).

Как и в предыдущих случаях, в результате модификации наблюдается появление дополнительных H-связей типа $N_iH...O_{i-4}$, (в данном случае – трех), причем для левых вариантов в положениях, предшествующих каноническому переходу, а для правых вариантов – в положениях, следующих за каноническим переходом. Все три модификации идут последовательно, блоком. Таким же блоком расположены и дополнительные H-связи, что хорошо видно из табл. 9.9. Характер остальных связей не изменяется. При этом видна характерная петля перед α -спиралью и после α -спирали, стабилизированная тремя H-связями.

Стадии десятизначных описаний совпадают с каноническими описаниями в следующих положениях: 1a, 2a, 3a – 2, 3, 4 и 9 – слева; 1 δ , 2 δ , 3 δ – 11, 16, 17 и 18 – справа. Антисимметрия (0 $\leftarrow \rightarrow$ 1), как и

для других модификаций Н-связей, наблюдается лишь для стадий 2-й, 11-й, а также 9-й, 18-й. В остальных случаях симметрия элементов носит комбинированный характер.

Значение систематизации модификаций канонических переходов

Нами была рассмотрена и систематизирована большая группа модификаций канонических переходов. Однако полезность проведенной систематизации оказалось значительно шире, чем только для применения к каноническим переходам. Можно отметить следующие аспекты.

Во-первых, сразу же прояснилось, в каком направлении подобная модификация происходит, и отпали нереальные варианты.

Во-вторых, просмотр текстовых файлов с использованием результатов анализа позволил увидеть, каким образом разные типы вторичных структур сочетаются между собой. До систематизации модификаций картина казалась весьма запутанной. Сейчас же стало ясно, что различные типы вторичных структур без всяких переходов присоединяются друг к другу в зависимости от функциональных и структурных требований структуры.

В-третьих, поскольку аналогичные каноническим последовательностям типы связей встречаются и в других вариантах структур (см. 9.4 и 9.5), то в них также возможны приведенные модификации, причем в тех же положениях и с теми же дополнительными H-связями. Фрагменты десятизначных описаний, содержащих описание модификаций, приведенные в табл. 9.7–9.9, можно с успехом использовать для интерпретации различных вариантов структур, получаемых в результате прогнозирования, а также для конструирования первичных структур на основе разнообразных модификаций вторичных структур.

Таким образом, круг структур, к которым можно применить полученные в этом разделе результаты, существенно расширяется.

9.7. Перспективы использования каталога

В данной главе не ставилось целью привести все варианты вторичных структур. Однако в процессе анализа было показано, что количество вариантов вторичных структур не столь уж велико. Есть основания полагать, что представленный каталог, с учетом возможных модификаций, охватывает не менее 90 % вариантов вторичных структур. Приведенный в данной главе материал может быть с успехом использован для интерпретации прогнозируемых вторичных структур (см. гл. 7).

В гл. 10 описан метод проектирования первичных структур по заданной вторичной структуре, который целиком основан на использовании данного каталога. Отдельные элементы вторичных структур (например, каноническая последовательность связей и ее модификации) являются общими для многих типов структур и имеют сходное описание в виде десятизначных чисел. Это существенно расширяет возможности использования каталога для конструирования первичных структур.

В целом, заложенные в данной главе принципы систематизации вторичных структур (в частности, рассмотрение их в виде отдельных стадий и параллельный анализ антисимметричных структур) полностью себя оправдали.

Глава 10. МЕТОД ПРОЕКТИРОВАНИЯ ПЕРВИЧНОЙ СТРУКТУРЫ БЕЛКА С ЗАДАННОЙ ВТОРИЧНОЙ СТРУКТУРОЙ

10.1. Современное состояние проблемы дизайна белков

Дизайн белков: история появления подхода

Успехи в исследовании пространственной структуры белков, инициированные классическими работами Полинга, Перутца и других кристаллографов, с течением времени были обобщены [38] и позволили поставить вопрос о конструировании первичных структур белков, принимающих заданные вторичную и третичную структуры.

Первыми попытками в области дизайна белков явились работы группы Де Градо [151], [152]. Авторы исходили из простой идеи: гидрофобные взаимодействия белковых структур должны быть минимизированы и упрятаны в гидрофобное ядро, а гидрофильные взаимодействия должны обеспечивать контакт с растворителем. Исходя из этих соображений был спроектирован и синтезирован искусственный белок, содержащий лишь несколько типов аминокислот (Leu, Glu, Lys) и состоящий из четырех α-спиралей. Однако такой упрощенный подход не позволяет проектировать сложные белки, состоящие из 20 различных типов аминокислот и обладающие заданными структурными и функциональными свойствами.

В основу искусственного белка – альбебетина – была положена не существовавшая в природе структура, которая состояла из двух повторов типа: $\alpha - \beta - \beta$ [153]. Структура альбебетина была разработана на основе физической теории формирования вторичной структуры белков, развиваемой авторами [139]. Структурное исследование альбебетина показало, что он обладает заданной вторичной структурой и находится в состоянии расплавленной глобулы. Следует отметить, что точность подхода, использованного авторами, не превышает 80 %, что не позволяет с полной уверенностью проектировать белки с заданной структурой. Авторы практически спроектировали лишь один белок, и далее исследования были прекращены.

С целью улучшения предсказательных свойств известного метода, использующего физические потенциалы, было предложено ввести ряд параметров, учитывающих свойства последовательностей аминокислот [154]. На основе этого метода и с учетом введенных параметров авторы спроектировали *de novo* ряд белков [155]. Однако такой подход имеет компилятивный характер и обеспечивает лишь некоторое улучшение взятых за основу методов, но не меняет сущность исходного физического метода, который носит вероятностный характер.

Для проектирования новых белков в качестве прототипа структуры часто используют специально выбранный образцовый белок. Так, в [156] на основе домена цинк-фингер-белка, имевшего структуру с мотивом ββα, была сконструирована последовательность аминокислот, которая хотя и имела малое сходство с исходным доменом, но образовывала структуру, идентичную предложенному для нее дизайну (белок FSD-1).

С помощью методологии Розетта [157], основанной на оптимизации выбранных структур, был спроектирован и синтезирован неизвестный в природе искусственный белок Тор 7, структура которого была подтверждена экспериментально. Ядро Розетта – это физическая модель макромолекулярных взаимодействий и алгоритмов поиска аминокислотной последовательности с наименьшей энергией для заданной белковой структуры. Авторы [158] применили свой метод к разработке конструкций ряда других белков. Однако этот метод требует довольно сложных расчетов и не всегда приводит к успешным результатам. Для его использования также необходимо наличие образцов.

Методология Розетта, которая является наиболее известной и разработанной, продолжает активно развиваться и в настоящее время. К последним работам в этом направлении относится публикация [159], авторы которой обращают внимание на принципиальную важность поиска системы правил, касающихся шаблонов вторичных структур для белковых третичных мотивов, особенно таких шаблонов, которые существенно влияют на формирование третичных структур. Экспериментально полученные структуры, основанные на системе этих правил, обладают высокой стабильностью и близки расчетным моделям по структуре. По сути авторы [159] вплотную подошли к положительной оценке важности вторичных структур в их детерминации третичной структуры. Тем не менее, в этом подходе не сложилось той идеологии, на которой основан предлагаемый нами способ проектирования первичных структур.

178

Выбор идеологии

Чтобы начать разрабатывать принципиально новый метод проектирования белковых структур, необходимо заложить в него и принципиально новую идеологию. Элементы этой идеологии в той или иной степени уже затрагивались в гл. 7–9. Проектирование белка является наиболее сложным аспектом применения того или иного подхода, в котором сходятся все элементы идеологии. Поэтому попытаемся еще раз переосмыслить приведенные ранее соображения, чтобы глубже продвинуться в понимании нашей идеологии.

Создание базы данных $\Pi \Phi$. Теоретической основой созданной базы данных ПФ явилось представление о MBM, работающей на каждой стадии биосинтеза белковой цепи. Какой бы сложной ни была белковая молекула, синтез ее осуществляется последовательно, и в любой ее части конформация выделяемых ПФ определяется условиями их формирования на момент биосинтеза (конформацией предшествующих и последующих стадий), т. е. траекторией, по которой идет развитие белковой структуры. Именно эта простая мысль лежит в основе создания базы данных ПФ. Если бы это было не так, то мы бы, во-первых, не получили системы ПФ (в силу нелинейных изменений механизмов формирования вторичной структуры ПФ), а во-вторых, не смогли бы применить эту систему ПФ для прогнозирования вторичной структуры тех белков, которые были использованы для создания базы данных ПФ.

Прогнозирование вторичной структуры белка. Предсказания вторичной структуры белка с использованием базы данных ПФ оказались правильными. При этом 10 булевых знаков, использованных для описания конформации ПФ, оказались необходимыми и достаточными для сохранения информации о контексте, в рамках которого этот ПФ был сформирован и использован. По этой причине программа безошибочно находит выделяемый в первичной структуре ПФ в соответствующем файле и подпапке с десятизначным описанием его конформации, считывает описание его конформации в соответствующем названии подпапки и записывает данные в рабочий файл.

Каталог десятизначных описаний вторичных структур белка. Создание такого каталога основано на той же идеологии: формирование различных типов вторичных структур, описанных в гл. 9, как и ПФ, связано с работой MBM и не зависит от того, на каком именно участке белковой цепи – в начале, в середине или в конце – они обнаружены. Есть данные [160], что процесс формирования третичной структуры напоминает наматывание нити на клочок бумаги и формирование клубка. Именно эта идея и заложена в скрытом виде в проводимые нами процедуры систематизации вторичных структур и ПФ. Если бы это было иначе, то систематизация вторичных структур, как и систематизация ПФ, была бы лишена смысла. При наличии нелинейности процессов тот или иной тип вторичной структуры формировался бы всякий раз под действием различных условий, например: в начале цепи – одни условия, в середине (в глобуле) – другие, в конце синтеза (на поверхности глобулы) – третьи. Очевидно, что практическое использование десятизначных описаний вторичных структур также было бы невозможно.

Таким образом, в основе принятой нами идеологии лежит следующая биологически важная идея: синтез белка и формирование его надмолекулярной структуры осуществляются линейно и под контролем клеточных механизмов по принципу конвейерной сборки. Каждый последующий этап сборки может быть начат лишь после того, как будет полностью закончен предыдущий. При этом каждый этап относительно независим и не влияет на реализацию последующих этапов.

Эта идея, однако, не означает, что между этапами формирования структуры белка нет никакой взаимосвязи. Последовательность аминокислот в первичной структуре белка, рассматриваемая на этапе синтеза, определяет возможность формирования того или иного типа вторичной структуры белка. В то же время, сформированные фрагменты вторичной структуры будут определять возможности укладки этих фрагментов в третичную структуру. Естественно, что если проектировать третичную структуру, не предполагая формирования в ней определенных функциональных связей (например, непрерывных ССИВС), то такая третичная структура будет бессмысленной и нефункциональной.

К процедуре конструирования белковой структуры применим следующий алгоритм:

1) задается идея функциональной третичной структуры;

2) на вторичных структурах задается место расположения контактов боковых цепей аминокислот при формировании из них третичной структуры;
3) выбираются варианты вторичных структур, обеспечивающие возможность формирования заданных контактов;

4) проектируется первичная структура белка, которая должна формировать заданные варианты вторичных структур и заданное положение боковых цепей, в результате чего формируется задуманная функциональная третичная структура.

Описанная идея принципиально отличается от «идеи образцов» хотя бы уже тем, что исходит не от заданной структуры, а от заданной функции. Это означает, что заданную функцию можно сконструировать различными способами с помощью различных исходных вторичных структур и, естественно, на основе разных белковых последовательностей. Тем самым снимаются ограничения, связанные с использованием прототипов структур. И еще: реализация данной схемы должна осуществляться биосинтезом, а не химическим синтезом. Процесс формирования белковой молекулы в рибосоме направлен от N-конца к С-концу и обеспечивает возможность правильной работы боковых цепей в качестве физических операторов MBM_б как элемента белкового конвейера. Тем самым организация белковой молекулы находится под жестким контролем клеточных структур, что обеспечивает ее правильную и однозначную укладку. Химический синтез отличается от биосинтеза тем, что осуществляется от С-конца к N-концу и самоорганизация белковой молекулы оставлена «на волю божью».

10.2. Сущность метода проектирования первичной структуры белка

Компоненты, необходимые для начала работы

Разработанный нами метод проектирования первичной структуры белка имеет много общего с методом прогнозирования вторичных структур. Это и неудивительно, поскольку оба метода имеют общее основание – базу данных ПФ белков. Однако, несмотря на сходство, есть и кардинальные различия, связанные с различиями в алгоритмах. Для проектирования первичной структуры белка необходимо иметь на компьютере следующие компоненты:

 – базу данных ПФ, полученную согласно описанию, приведенному в 7.2, и обладающую структурой, описанной в 7.3; – каталог вторичных структур, описанных с помощью десятизначных булевых чисел;

– программу (PROTCOM), с помощью которой осуществляется ввод пар переменных для поиска новых ПФ в базе данных, выбор аминокислот для конструируемой полипептидной цепи и запись сконструированной последовательности в рабочий файл (формат .dbk).

Ограничительные условия применения данного метода

Проектирование первичной структуры белка предлагаемым методом нужно рассматривать как важный, но не единственный и не последний этап конструирования белка. Созданная нами база данных ПФ белков включает в рассмотрение далеко не все типы Н-связей, возможные на уровне вторичной структуры. Так, в процессе ее создания были полностью исключены водородные связи типа N_iH...O_{i-3}. Хотя они имеют локальное положение (главным образом, в начале и в конце α-спиралей, а также в изгибах β-структуры), их исключение влияет на вид конформации белка. Кроме того, в базе данных отсутствует информация о водородных связях боковых цепей с основной цепью (типа R_iH...O_{i-4}, R_iH...O_{i-3}), а также водородных связях между боковыми цепями ($R_iH...R_{i-4}, R_iH...R_{i-3}, R_iH...R_{i-2}$). Такая информация также имеет значение, особенно при переходе от конструирования вторичной структуры к разработке элементов и узлов третичной структуры. Это, конечно, было сделано намеренно, поскольку вначале принципиально важно создать «скелет» заявляемого метода, а уж затем покрывать его «мясом». Постепенно информацию об описанных выше типах Н-связей предполагается вводить в базу данных и учитывать в процессе конструирования. Однако важно отметить, что введение дополнительной информации принципиально не будет влиять на ход описанного далее алгоритма проектирования, а послужит лишь уточнению параметров конструируемой структуры.

Алгоритм проектирования

Далее будут рассмотрены основные этапы проектирования, которые подробно описаны в [114].

Выбор вторичных структур из каталога. Наличие каталога вторичных структур является важной предпосылкой конструирования первичной

структуры белка. Естественно, что если проводится формальное конструирование, не ориентированное на получение функциональной структуры, то получение набора вторичных структур не составляет большого труда.

В 9.6 на основе анализа текстовых файлов был сделан важный вывод о взаимной независимости различных типов структур, что позволяет использовать их в любой последовательности в соответствии с целями конструирования. По этой причине исследователь (будем далее называть его «оператор») сам определяет набор вторичных структур, необходимых ему для конструирования первичной структуры белка.

Выбор и введение начального ПФ. Процедура проектирования производится оператором с самого начала и начинается с задания и ввода в компьютерную программу начальной последовательности из пяти аминокислот, принадлежащих к группе из 20 канонических аминокислот белков. Такую последовательность мы называем заданным начальным ПФ.

Задуманная начальная последовательность из пяти аминокислот представляется в виде столбца из трехбуквенных (или однобуквенных, что не принципиально) обозначений аминокислот с их нумерацией, записанной снизу вверх:

5	Efg
5	Eig
4	Def
3	Cde
2	Bcd
1	Abc

Порядок введения этой последовательности в программу не имеет значения. В данном случае эта последовательность вводится сверху вниз и записывается в память программы. Процедура выбора заданного начального ПФ является очень ответственной. Ее успех в значительной степени зависит и от задаваемого описания вторичной структуры начального ПФ. Если это описание известно уже заранее, то можно просто выбрать в базе данных несколько ПФ (выбор определяется опытом оператора) и использовать их в процессе конструирования белка.

Введение описания вторичной структуры. Согласно алгоритму [114] задают и вводят в память компьютера описание вторичной структуры заданного начального ПФ в виде десятизначного числа в двоичной системе. Взгляд на приведенные в гл. 9 вторичные структуры показывает, что большинство вариантов начинается с фрагмента β-структуры или α-спирали. По этой причине можно посоветовать выбрать ПФ в папке 0000000000 или 1111111111. После введения ПФ программа предложит ввести описание вторичной структуры, и оператор введет его в соответствии с происхождением заданного начального ПФ.

Поиск заданного начального $\Pi \Phi$ в папке с заданным описанием его структуры. Независимо от метода задания начального $\Pi \Phi$ и десятизначного описания его структуры дальнейшая процедура состоит в поиске заданного начального $\Pi \Phi$ проектируемого белка в базе данных с помощью программы PROTCOM. При этом алгоритм поиска включает в себя следующие этапы:

а) кодирование заданного начального пентафрагмента*;

б) поиск заданного начального ПФ в базе данных в папке с заданной вторичной структурой пентафрагмента.

Для введенного десятизначного числа 000000000 заданный начальный ПФ ищут в папке базы данных с номером 000000000, в файле с кодовым номером 12344, т. е. 12344 0000000000.

При нахождении в папке заданного начального П Φ считают этот П Φ первым из возможного числа N пентафрагментов проектируемой первичной структуры белка и производят:

– фиксирование номера папки базы данных, содержащей первый ПФ;

 – запись последовательности аминокислот первого ПФ в рабочий файл программы;

– запись десятизначного номера папки, описывающего вторичную структуру найденного первого ПФ, в рабочий файл.

Формат рабочего файла программы PROTCOM показан в табл. 10.1.

Запись последовательности аминокислот исследуемого белка в рабочий файл производится снизу вверх, что отражает порядок синтеза белка на рибосоме (удлинение белка происходит добавлением аминокислот к верхней аминокислоте). Столбцы файла имеют следующее назначение:

^{*} Программа считывает аминокислоты ПФ сверху вниз, кодирует их в соответствии с принадлежностью к той или иной группе антисимметрии (см. 7.2, табл. 7.9) и записывает кодовый номер слева направо, аналогично сформированным индексам файлов, например: Efg – 1, Def – 2, Cde – 3, Bcd- 4, Abc – 4, кодовый номер – 12344.

1 – номера аминокислот в проектируемом белке, записанные снизу вверх;

2 – последовательность аминокислот в проектируемом белке, записанная снизу вверх с помощью трехбуквенных обозначений;

3 – десятизначные номера папок (bbbbbbbbbb) базы данных, описывающих вторичную структуру проектируемых ПФ, записанные снизу вверх.

Таблица 10.1 Формат рабочего файла, созданного программой PROTCOM

1	2	3
N	STP	bbbbbbbbbb
5	Efg	bbbbbbbbbb
4	Def	
3	Cde	
2	Bcd	
1	Abc	

В строке N размещается сигнальное значение конца белковой последовательности (STP).

Жирным шрифтом выделен первый ПФ и десятизначный номер папки, в которой он найден.

При ненахождении заданного начального П Φ в папке с заданным описанием его вторичной структуры производят необходимые изменения в последовательности аминокислот заданного начального П Φ и продолжают поиск в папке с заданным описанием его структуры. Поиск повторяют до тех пор, пока П Φ с заданной последовательностью аминокислот не будет найден.

Задание десятизначных описаний вторичной структуры и поиск новых аминокислот для проектирования. После нахождения начального ПФ процедура проектирования входит в нормальную колею. Программа предлагает оператору ввести любую из четырех возможных пар переменных (00, 01, 10 или 11). Выбор той или иной пары определяется десятизначным описанием планируемой на данном этапе вторичной структуры.

Ввод соответствующей пары переменных переводит программу в режим поиска ПФ в базе данных, содержащих четыре аминокислоты каждой из (N-1) стадий проектирования, записанных в рабочем файле ПФ, и одну новую аминокислоту. Кодирование четырех аминокислот осуществляется в соответствии с ранее изложенным пунктом б), а сам поиск ПФ производится в файлах четырех типов. Например, если поиск проводится в папке 0000000000, то обозначения типов файлов будут такими:

> 1XXXX_0000000000: 2XXXX_0000000000; 3XXXX_0000000000; 4XXXX_00000000000.

Найденные новые аминокислоты, которые можно использовать для последующего проектирования первичной структуры белка, записываются в интерфейсе для каждого типа файлов по отдельности. Оператор выбирает новую аминокислоту и вводит ее в память программы. Программа присоединяет новую аминокислоту к четырем предыдущим аминокислотам, записывает ее в рабочий файл и снова выделяет четыре предыдущие аминокислоты, но уже с учетом введенной. Далее программа предлагает снова ввести четыре варианта пар переменных для последующего задания десятизначного описания вторичной структуры. Оператор, следуя выбранному из каталога вторичных структур десятизначному описанию, выбирает и вводит в программу следующую пару переменных и снова переводит программу в режим поиска ПФ в базе данных, содержащих четыре аминокислоты из предыдущей (N – 1)стадии, и т. д. В качестве примера в табл. 10.2 жирным шрифтом выделены четыре последние аминокислоты предыдущего ПФ и введенное описание вторичной структуры для поиска нового ПФ.

Таблица 10.2

, 1		
1	2	3
6		000000000
5	Efg	0000000000
4	Def	
3	Cde	
2	Bcd	
1	Abc	

Выделение аминокислот и описание их вторичной структуры для поиска пентафрагментов в базе данных

В результате действий программы PROMCOM и работы оператора, проектирующего белок, в рабочем файле оказывается полностью заполненным второй столбец, содержащий первичную структуру белка, и третий столбец, содержащий десятизначное описание задаваемой вторичной структуры этого белка. Наличие в третьем столбце идущих подряд папок 000000000 характеризует фрагмент как β -структурный. Несколько идущих подряд папок с нумерацией 1111111111 показывает, что сконструированный фрагмент первичной структуры должен принимать α -спиральную конформацию (см. приложение, структуры 1-А и 1-Б). Переходные участки между α -спиральной и β -структурной конформациями, а также изгибы β -структуры проектируются на основе соответствующих десятизначных описаний этих структур, представленных в каталоге (см. приложение, структуры 2-А, 2-Б, 4-3-А). Для лучшего понимания далее приведен конкретный пример проектирования фрагмента первичной структуры белка.

10.3. Пример проектирования первичной структуры белка с заданной вторичной структурой

Выбор задаваемого описания вторичной структуры

Для иллюстрации предложенного нами способа проектирования первичной структуры белка на основе заданной вторичной структуры непринципиально, какую первичную структуру проектировать, поскольку алгоритм проектирования универсален. При проектировании элементов третичной структуры (например, взаимодействия двух α-спиральных фрагментов) принципиально важным является учет предполагаемых взаимодействий между контактирующими участками α-спиралей. Это будет сказываться на выборе боковых цепей аминокислот, предоставляемых программой для проектирования вторичной структуры, но никак не повлияет на ход алгоритма. По этой причине представим в качестве примера проектирование первичной структуры белкового фрагмента для относительно простого заданного типа вторичной структуры – обратного изгиба с одной H-связью. Несмотря на простоту, все особенности использованного нами алгоритма проектирования будут показаны вполне наглядно.

Перед началом проектирования оператор находит задаваемую структуру в каталоге десятизначных описаний вторичных структур (она соответствует структуре 6-1-А (см. приложение, табл. 9.6)).

Процедура проектирования аминокислотной последовательности

Рассмотрим основные особенности проектирования аминокислотной последовательности на основе заданного десятизначного описания вторичной структуры. Дополнительные подробности можно найти в нашем патенте [114].

Выбор и введение начального ПФ. Как отмечалось ранее, важным моментом проектирования является выбор начальной последовательности из пяти аминокислот. После нескольких попыток, детали которых опустим, была выбрана следующая последовательность:

5	Val
4	Asp
3	Cys
2	Gly
1	Tyr

Она записана в порядке номеров снизу вверх: Туг, Gly, Cys, Asp, Val. Оператор вводит ее в программу, которая записывает ее в память компьютера.

Введение описания вторичной структуры. Согласно алгоритму, описанному в [114], в память компьютера вводят в виде десятизначного числа в двоичной системе описание вторичной структуры заданного начального ПФ. Как видно из табл. 9.6 (см. приложение), начальным описанием обратного изгиба с одной H-связью является десятизначное число 0100000000. Однако последовательность из 5 аминокислот была выбрана нами для фрагмента β -структуры (папка 000000000), которая предшествует началу обратного изгиба. Поэтому после введения ПФ программа предлагает ввести описание вторичной структуры, и оператор вводит значение 000000000 в соответствии с той папкой, для которой будет проводиться поиск начального ПФ.

Поиск заданного начального $\Pi \Phi$ в папке с заданным описанием его структуры. Независимо от метода задания начального $\Pi \Phi$ и десятизначного описания его структуры, дальнейшая процедура алгоритма состоит в поиске заданного начального $\Pi \Phi$ проектируемого белка в базе данных с помощью программы PROTCOM. При этом этапы алгоритма поиска включают в себя: а) кодирование заданного начального ΠΦ, которое осуществляется программой путем отнесения каждой аминокислоты заданного начального ПΦ к той или иной группе антисимметрии (см. 7.2, табл. 7.9)*;

б) поиск заданного начального ПФ в базе данных в папке с заданной вторичной структурой ПФ.

Для введенного десятизначного числа 000000000 заданный начальный ПФ ищут в папке базы данных с номером 0000000000, в файле с кодовым номером 44313 0000000000.

Дальнейшая процедура поиска описана в 10.2. Программа обнаружила заданный начальный ПФ в файле 44313 0000000000:

Этот ПФ был выделен из текстового файла, полученного программой PROTEIN 3D на основе обработки координат атомов белка из Protein Data Bank и структуры, описываемой как 0000000000, не содержащей H-связей в ближайшем окружении ПФ.

Табл. 10.3–10.5 иллюстрируют процедуру проектирования первичной структуры белка с помощью программы PROTCOM. В левой части таблицы, озаглавленной «Ввод», размещаются: в первом столбце – вводимые в программу PROTCOM порядковые номера проектируемых аминокислот, во втором столбце – аминокислоты или пары переменных, выбираемые оператором на основе заданной вторичной структуры (см. приложение, табл. 9.6). В третьем столбце записывается введенное в программу описание вторичной структуры в виде десятизначного числа. В центральной части «Поиск пентафрагмента в базе данных» помещаются: в первом столбце – наименования файлов с номерами кодировки и заданной папки, а во втором – наименования найденных в ПФ аминокислот. В правой части таблицы находится запись,

^{*} В данном примере: Val – 4, Asp – 4, Cys – 3, Gly – 1, Tyr – 3. Эта числовая последовательность записывается в память программы слева направо 44313 и используется для поиска заданного начального пентафрагмента в папке 000000000 базы данных в файле 44313 0000000000.

осуществляемая программой PROTCOM в рабочем файле после обнаружения заданного начального ПФ. В дальнейшем аналогичная запись производится после выбора аминокислоты в файле с номером кодировки и номером заданной папки.

Как следует из табл. 10.3, программа обнаружила введенный начальный ПФ в файле с соответствующей кодировкой и номером папки и делает запись в рабочем файле. Действия оператора согласно алгоритму поиска, относящиеся к случаю ненахождения в папке заданного начального ПФ, опускаем.

Таблица 10.3

Ввод		од	Поиск пентафрагмента в базе данных		Запись в рабочем файле		
№	Амино- кислоты или пары переме- нных	Введенное описание вторичной структуры	Наименование файла с номером кодировки и номером заданной папки	Наимено- вания амино- кислот	Nº	Наимено- вания амино- кислот	Описание вторичной структуры
5	Val	000000000	44313_0000000000		5	Val	000000000
4	Asp				4	Asp	
3	Cys				3	Cys	
2	Gly				2	Gly	
1	Tyr				1	Tyr	

Поиск заданного начального пентафрагмента в базе данных

Задание десятизначных описаний вторичной структуры и поиск новых аминокислот для проектирования. Далее производится проектирование первичной структуры аминокислот в соответствии с заданной вторичной структурой. Программа PROTCOM предлагает ввести пары переменных 00, 01, 10 или 11. Из табл. 9.6 (см. приложение) видно, что первым десятизначным числом, описывающим вторичную структуру (обратный изгиб с одной Н-связью) является 1000000000. По этой причине из предложенных программой пар переменных оператор выбирает 10 и вводит эту пару переменных в программу (столбец «Аминокислоты или пары переменных» в табл. 10.4). Программа добавляет пару переменных 10 слева и удаляет пару цифр (00) справа. Это приводит к изменению десятизначного числа, описывающего вторичную структуру предыдущего ПФ, что отражено в столбце «Описание вторичной структуры» табл. 10.4.

Таблица 10.4

Ввод			Поиск пентафрагмента в базе данных			Запись в рабочем файле		
N⁰	Амино- кислоты или пары переме- нных	Введенное описание вторичной структуры	Наименование файла с номером кодировки и номером заданной папки		№	Наимено- вания амино- кислот	Описание вторичной структуры	
6	10	100000000	14431_100000000	Gly	6	Gly	100000000	
5	Val	000000000	44313_0000000000		5	Val	000000000	
4	Asp				4	Asp		
3	Cys				3	Cys		
2	Gly				2	Gly		
1	Tyr				1	Tyr		

Поиск второго пентафрагмента в базе данных

Далее согласно алгоритму программа ищет в базе данных П Φ четыре аминокислоты каждого из (N – 1) П Φ , записанных в рабочем файле, и одну новую. Для этого программа выделяет в П Φ , записанном в рабочем файле (табл. 10.3), четыре аминокислоты, записанные сверху вниз: Val, Asp, Cys, Gly.

Затем программа кодирует их в соответствии с принадлежностью к той или иной группе антисимметрии и записывает кодовый номер слева направо, аналогично сформированным индексам файлов, но без первой аминокислоты: 4431. Далее в базе данных проводится поиск ПФ, содержащих четыре выделенные аминокислоты, в папке с заданной структурой следующего ПФ (100000000), т. е. в файлах X4431_1000000000, где X может принимать значения 1, 2, 3, 4, соответствующие номерам групп антисимметрии – 14431_100000000, 24431_100000000, 34431_100000000, 44431_100000000.

В результате поиска были найдены П Φ , содержащие четыре последние аминокислоты (Val, Asp, Cys, Gly) и следующие пятые аминокислоты, записанные вместе с кодами белков, из которых они были получены:

– в файле группы 1 (14431_100000000): Gly;

в файле группы 2 (24431_100000000) – пентафрагменты не найдены;

в файле группы 3 (34431_100000000): – пентафрагменты не найдены;

в файле группы 4 (44431_100000000): – пентафрагменты не найдены.

Обращаем внимание на то, что в файлах групп 2, 3, 4 пентафрагменты не найдены. В связи с этим для проектирования используем единственную аминокислоту Gly, которую оператор вводит в программу. После ввода этой аминокислоты программа производит:

– запись новой аминокислоты в рабочий файл («Запись в рабочем файле» табл. 10.4), отражающий проектируемую первичную структуру белка (Gly);

- запись десятичного номера папки, описывающего вторичную структуру каждого найденного ПФ (100000000).

Поскольку ПФ был найден, то далее повторяются действия согласно пп. а) и б) до окончания процесса проектирования. Как видно из табл. 10.5, в процессе проектирования последовательности аминокислот на всех этапах появлялась возможность выбора лишь из одной аминокислоты.

Таблица 10.5

	Вв	од	Поиск пентафра в базе данни	гмента ых	Запись в рабочем фай		очем файле
N⁰	Амино- кислоты или пары переме- нных	Введенное описание вторичной структуры	Наименование файла с номером кодировки и номером заданной папки	Наимено- вания амино- кислот	N⁰	Наимено- вания амино- кислот	Описание вторичной структуры
14	00	000000001	14224_000000001	Gly	14	Gly	000000001
13	00	000000100	42241_0000000100	Arg	13	Arg	000000100
12	00	0000010000	22414_0000010000	Leu	12	Leu	0000010000
11	00	0001000000	24141_0001000000	Leu	11	Leu	0001000000
10	01	010000010	41411_010000010	Arg	10	Arg	010000010
9	00	0000001000	14114_0000001000	Gly	9	Gly	000001000
8	00	0000100000	41144_0000100000	Asp	8	Asp	0000100000
7	00	0010000000	11443_0010000000	Pro	7	Pro	001000000
6	10	100000000	14431_100000000	Gly	6	Gly	100000000
5	Val	000000000	44313_000000000		5	Val	000000000
4	Asp				4	Asp	
3	Cys				3	Cys	
2	Gly				2	Gly	
1	Tyr				1	Tyr	

Поиск последующих пентафрагментов в базе данных

Спроектированной первичной структурой белка считают полученную в рабочем файле последовательность аминокислот с соответствующим описанием ее вторичной структуры, записанной в рабочем файле программы PROTCOM и представленной в правой части табл. 10.5. Для экспериментального подтверждения существования спроектированной в данном примере первичной структуры белка с заданной вторичной структурой в Protein Data Bank было найдено несколько фрагментов белков, имеющих последовательность аминокислот, полностью совпадающую либо со всей спроектированной последовательностью аминокислот нашего примера, либо с участком этой последовательности (табл. 10.6).

Таблица 10.6

№	Спроектированная		Фрагменты белков						
	первичная	номер	1AGD	номер	2R37	номер	3B02	номер	1BAS
	структура	амино-		амино-		амино-		амино-	
	примера 2	кислот		кислот		кислот		кислот	
14	Gly	112	Gly						
13	Arg	111	Arg						
12	Leu	110	Leu					84	Leu
11	Leu	109	Leu			39	Leu	83	Leu
10	Arg	108	Arg			38	Arg	82	Arg
9	Gly	107	Gly	192	Gly	37	Gly	81	Gly
8	Asp	106	Asp	191	Asp	36	Asp	80	Asp
7	Pro	105	Pro	190	Pro	35	Pro		
6	Gly	104	Gly	189	Gly				
5	Val	103	Val	188	Val				
4	Asp	102	Asp						
3	Cys	101	Cys						
2	Gly	100	Gly				_		
1	Tyr	99	Tyr						

Сопоставление первичных структур фрагментов белков

Как следует из табл. 10.6, спроектированная нами первичная структура аминокислот оказалась идентичной последовательности аминокислот фрагмента белка 1AGD, что видно из сравнения ее с фрагментом текстового файла этого белка в табл. 10.7. Очевидно, что и десятизначное описание этого фрагмента оказалось также полностью идентичным заданному описанию вторичной структуры белка нашего примера (ср. табл. 10.5 и 10.7). Кроме того, было установлено (табл. 10.6), что отдельные участки этой последовательности могут быть составлены на основе ПФ ряда других белков, а именно 2R37 (№ 9), 3B02 (№ 11) и 1BAS (№ 12), которые не имеют никакого родства с белком 1AGD. В табл. 10.7 приведено описание их вторичной структуры в виде десятизначных булевых чисел, которое полностью совпадает с заданным описанием вторичной структуры спроектированной первичной структуры нашего примера.

Таблица 10.7

Фрагменты	Фрагменты вторичной структуры белков из Protein Data Bank						
				В	торичной		
14.00	2R37	3B02	1BAS	С	труктуры		
IAGD	0000001000	00 01 000000	0000010000	d	ррагмента		
				бе	елка 1AGD		
112 GLY				14	00000000001		
111 ARG				13	0000000100		
110 LEU				12	0000010000		
109 LEU			84 LEU	11	0001000000		
108 ARG O – 104 GLY N	39 LEU 83 LEU	10	0100000010				
108 ARG		38 ARG O – 34 LEU N 82 AF	82 ARG O – 78 LYS N	9	0000001000		
107 GLY	37 GLY 81 GLY	8	0000100000				
106 ASP		36 ASP	SP 80 ASP RO	7	0010000000		
105 PRO	192 GLY	35 PRO		6	1000000000		
104 GLY N – 108 ARG U	191 ASP			5	0000000000		
104 GL Y	190 PKU 190 CLVN - 102 ILE O			4			
103 VAL	189 GLY N – 195 ILE U 199 VAI			3			
101 CVS	100 VAL			2			
100 GLV				1			
99 TYR				1			

Вторичная структура фрагментов белков

В принципе могут быть приведены и пространственные структуры представленных фрагментов [114]. Однако никакой дополнительной информации они не добавят.

Таблица 10.8

Перечень белков, в которых методом РСА была исследована структура, совпадающая с заданной вторичной структурой

Номер	Код	Название белка	Степень	Ссылки
этапа	белка	разреше		
5-8, 10, 13, 14	1AGD	Histocompatipility complex	2,05 A	[162]
9	2R37	Human glutathione peroxidase 3	1,85 A	[163]
11	3B02	Transcriptional regulator, CRP family	1,92 A	[164]
12	1BAS	Fibroblast growth factor	1,9 A	[165]

Таким образом, в приведенном примере спроектированная первичная структура белка подтверждается с помощью двух вариантов вторичных структур.

Первый вариант: первичная структура белка спроектирована на основе десятизначного описания вторичной структуры и состоит из ПФ, происходящих только из одного белка – 1AGD.

Второй вариант: первичная структура белка спроектирована также на основе десятизначного описания вторичной структуры, но состоит из ПФ, полученных из четырех разных белков: 1AGD, 2R37, 3B02 и 1BAS. Описание вторичных структур ПФ белков 1AGD, 2R37, 3B02 и 1BAS также полностью совпадает в нашем примере. Сведения о белках 1AGD, 2R37, 3B02 и 1BAS и ссылки на публикации представлены в табл. 10.8.

Обсуждение результатов проектирования

Тот факт, что одна и та же последовательность аминокислот нами получена как на основе одного белка, так и из «лоскутков», т. е. из ПФ разных белков, имеет принципиальное значение.

Попытаемся сформулировать и обсудить выводы, которые следуют из полученных результатов.

Вывод первый. Все ПФ, найденные в разных белках, имеют одно и то же описание конформации, которое не зависит от положения этого ПФ в структуре белка (как следует из табл. 10.7, они все имеют разные номера в первичной структуре этих белков). Это описание связано, прежде всего, контекстом той конформации белка, в которой находятся приведенные ПФ. Как показывает анализ этого контекста, они все расположены на участках изгибов β -структуры.

В связи с этим уместно обсудить роль контекста в формировании той или иной структуры. Имеются сведения, что одни и те же последовательности аминокислот (например, ПФ) можно обнаружить как в α-спирали, так и в β-структуре. И действительно, при анализе ПФ созданной нами базы данных мы неоднократно сталкивались с такими фактами.

Если к этим последовательностям применить представление о том, что формирование вторичной структуры белка определяется взаимодействием белка с окружающей водной средой, то получается «нонсенс»: одна и та же последовательность аминокислот принимает при одних и тех же условиях разную вторичную структуру. «Нонсенс» исчезает, если предположить, что формирование вторичной структуры той или иной последовательности определяется тем контекстом, в котором она расположена. Это означает, что если ПФ, предшествующие данному ПФ, уже имели определенный тип вторичной структуры, то велика вероятность того, что и данный ПФ будет иметь ту же структуру. Таким образом, факт существования одной и той последовательности аминокислот в ПФ, находящихся в разных папках, в рамках контекстного подхода находит вполне рациональное объяснение.

Вывод второй. Если посмотреть на состав аминокислот, которые входят в приведенные П Φ , то легко увидеть, что все они содержат такие аминокислоты, как Pro и Gly, а также одну полярную аминокислоту, например Arg (оператор связности). Кроме того, они идут в данном фрагменте в определенной последовательности: Pro, Gly, Arg. Это означает, что формирование заданной вторичной структуры зависит как от определенного состава аминокислот, так и от порядка их расположения на сконструированной первичной структуре. С позиции модели MBM именно так и должно быть: если поменять местами хотя бы даже две рядом расположенные аминокислоты, то это может привести к принципиально разным результатам, поскольку меняются условия работы MBM. С позиции иных представлений (например, представлений об определяющей роли водного окружения на формирование вторичной структуры белка) такое разительное влияние вряд ли можно было бы предположить.

В [1, 9.2] были выдвинуты предположения о различных вариантах последовательностей аминокислот и их влиянии на формирование вторичной структуры белка с позиции работы MBM. Из этих предположений следует, что процесс формирования вторичной структуры на основе исходной последовательности аминокислот в значительной мере подчиняется логике, похожей на ту логику, которую мы наблюдаем во фразах, например: если А, то Б, или А – или Б, ни А, ни Б и т. д. В свое время эти идеи развивались в работах В. А. Ратнера [166], однако в отношении проектирования первичных структур по заданному описанию вторичной структуры они не применялись.

10.4. Перспективы развития предложенного подхода

Объем монографии не позволяет подробно рассмотреть разнообразные варианты проектирования первичных структур, которые возможны на основе нашего подхода. По этой причине лишь кратко остановимся на возможностях дальнейшего его развития. Многие из этих соображений, в применении к другим аспектам подхода, уже были высказаны, однако по мере необходимости будем их повторять.

Во-первых, созданная нами база данных ПФ имеет ограничения, поскольку использует лишь 2500 вариантов белков. Их количество постепенно прирастает, однако предела пока не видно. В то же время анализ закономерностей формирования папок ПФ вполне доступен и при наличии определенного опыта позволит в ближайшее время приступить к созданию базы данных, основанной на моделировании наблюдаемых закономерностей. Всестороннее рассмотрение этих закономерностей должно привести к раскрытию принципов работы MBM. По сути, в результате этих исследований база данных будет заменена на генератор ПФ, который и лежит в основе всех существующих третичных структур. Это откроет практически безграничные возможности для конструирования новых белков с заданной структурой и новыми свойствами.

Во-вторых, при проектировании первичной структуры используется база данных, в которой описаны лишь основные (хотя и наиболее важные) типы взаимодействий, а именно: образование циклических ПФ с Н-связями между $N_iH...O_{i-4}$ атомами. Как отмечалось в 10.2, во вторичных структурах возможно формирование еще целого ряда H-связей, например боковых цепей с основной цепью типа $R_iH...O_{i-4}$, $R_iH...O_{i-3}$, а также водородных связей между боковыми цепями ($R_iH...R_{i-4}$, $R_iH...R_{i-3}$, $R_iH...R_{i-2}$). Учет этих взаимодействий – ближайшая задача проводимых исследований. Мы представляем это таким образом, что проанализированные закономерности будут служить в качестве дополнительных баз данных, которые программа PROTCOM будет просматривать синхронно с основной базой данных.

В-третьих, дальнейшая разработка созданного нами каталога описаний вторичных структур в перспективе позволит вывести общие закономерности формирования такого каталога, т. е. создать закономерную систему вторичных структур и их десятизначных описаний.

В-четвертых, учет вышеупомянутых взаимодействий боковых цепей аминокислот позволит включить в сферу проектирования логику появления тех или иных аминокислот в первичной структуре и тем самым перебросить мостик к формированию элементов (логических узлов) третичной структуры. Разработка этих вопросов позволит ввести ограничения на выбор тех или иных боковых цепей для проектирования первичной структуры, что может существенно способствовать успешности получения функциональных структур.

Наконец, в-пятых, процесс проектирования, конечно, можно и нужно визуализировать. Это сделает процедуру проектирования белков визуально наглядной и ускорит процесс проектирования, что также будет способствовать большей успешности получения новых функциональных структур.

Таким образом, предложенный нами подход к проектированию первичной структуры белков по заданному десятизначному описанию вторичной структуры имеет весьма широкие перспективы.

ЗАКЛЮЧЕНИЕ

Подведем краткий итог. Настоящая монография направлена на построение теории кодирования топологии цепных полимеров. Всякая теория базируется на каких-то элементарных понятиях. В данной теории таким элементарным понятием стал ПФ цепного полимера, математическим аналогом которого является четырехзвенный цепной граф. Читатель мог проследить, как на основе этого понятия или его аналога происходит развитие теории «самой из себя».

В первой части, посвященной развитию общей теории топологического кодирования цепных полимеров, описано построение топологического кода на основе аналога $\Pi \Phi$ – четырехзвенного графа, введено понятие физического оператора, действующего на образование циклического $\Pi \Phi$, проведен анализ соответствия элементов кода физическим операторам, воссоздающим закодированную структуру $\Pi \Phi$ (гл. 2).

Анализ работы физических операторов в области формирования связи $Q_iH...X_{i-4}$ =R привел к модели MBM. С использованием теоретико-группового подхода в модели выделены группа из 20 векторов действия и группа из 20 сменяемых физических операторов (канонический набор), рассматриваемых в качестве неприводимых представлений группы векторов. Проведен также анализ роли тетраэдрического R_i^{α} -го атома (к нему прикреплены физические операторы), задающего направление роста цепи полимера (гл. 3).

Вторая часть посвящена конкретизации идей общей теории топологического кодирования на примере биополимеров (в основном, белков), рассматриваемых в качестве частных случаев линейных цепных полимеров. С позиции топологического кодирования проанализирована структура генетического кода, а боковые цепи аминокислот рассмотрены как физические операторы, воссоздающие закодированную триплетами структуру ПФ. Тем самым найдено принципиальное решение проблемы соответствия «триплет – аминокислота» в генетическом коде (гл. 4).

Дальнейшая конкретизация представлений общей теории связана с разработкой модели MBM_б. Модель пространственной структуры канонического набора аминокислот на додекаэдре является одним из элементов MBM_б. Предпосылками к построению модели послужили совпадение количества аминокислот и количества вершин в додекаэд-

ре, а также наличие в каноническом наборе пар аминокислот, близких по свойствам. На додекаэдре проведены три плоскости симметрии, введены обозначения вершин, построена таблица преобразований симметрии элементов канонического набора, состоящая из четырех подгрупп, а также выделено четыре типа антисимметрии, в соответствии с которыми распределились боковые цепи аминокислот (гл. 5).

С позиции теоретико-группового подхода система векторов, выделенная в MBM_б, рассматривается как математическая группа, в которой имеются нейтральный и обратный элементы, а боковые цепи аминокислот представляют собой группу неприводимых представлений векторов. В целом, с позиции этого подхода оказалось возможным понять природу канонического набора аминокислот, антисимметрии аминокислот внутри этого набора и порядок их расположения на додекаэдре (гл. 6).

В третьей части представлены практические результаты разработки подхода. Модель MBM_{6} , в основе которой лежит ПФ белка, может иметь два возможных этапа использования. Первый этап – это анализ результатов работы MBM_{6} путем выделения и систематизации белковых ПФ. Второй этап – построение модели пространственной структуры MBM_{6} , использующей закономерности структуры системы ПФ. В монографии приведены результаты разработки первого этапа.

Основой всех последующих результатов явилась созданная база данных ПФ белков. Она имеет иерархическую структуру: 16 папок (8 с основными ПФ и 8 – с минорными), внутри каждой папки по 64 подпапки, в каждой из которых, в свою очередь, может содержаться до 1024 файлов с ПФ. Проведен теоретический анализ особенностей ПФ, содержащихся в базе данных (гл. 7).

База данных послужила основой для разработки метода прогнозирования вторичной структуры белка на основе первичной структуры (гл. 8). Подробно описаны компоненты, необходимые для прогнозирования, а также алгоритм прогнозирования. Результатами прогнозирования являлись текстовые файлы, содержащие описание вторичной структуры белка в виде столбцов десятизначных булевых чисел. В отличие от существующих методов прогнозирования, точность которых не превышает 80 %, предлагаемый метод позволяет на структурах, использованных для создания базы данных, предсказывать вторичные структуры с точностью, близкой к 100 %. Для интерпретации прогнозируемых структур был создан каталог описаний вторичных структур белка и их десятизначных описаний (гл. 9). В основу систематизации структур положено их описание в виде отдельных стадий, а также парное представление описываемых структур, основанное на принципе антисимметрии. Описано более 20 вариантов пар вторичных структур, что охватывает большую их часть.

С использованием созданной базы данных ПФ и каталога десятизначных описаний вторичных структур белка разработан способ проектирования первичной структуры белка по заданному описанию его вторичной структуры (гл. 10). В основе принятой нами идеологии дизайна лежит идея конвейерной сборки надмолекулярных структур, осуществляемой последовательно под контролем клеточных механизмов, которая была изложена в гл. 1. Рассмотрены условия применения и алгоритм данного метода проектирования. Приведен пример проектирования первичной структуры белкового фрагмента по заданному описанию вторичной структуры.

Следует отметить, что сфера применения предложенных методов прогнозирования и проектирования белков может существенно расшириться благодаря созданию искусственной базы данных ПФ, полученной на основе анализа закономерностей экспериментальной базы. В конечном счете, это должно привести к следующему этапу – появлению MBM_б, работающей на основе найденных теоретических принципов.

Сформулированные в издании идеи и принципы, практическая значимость которых продемонстрирована на примерах из молекулярной биологии, могут иметь не меньшее значение и для развития органической наноэлектроники. Для этих целей необходимы широкое использование цепных полимеров, разработка для них вариантов топологического кода, систем физических операторов и молекулярных устройств, обеспечивающих конвейерное тиражирование и воссоздание пространственной структуры сконструированных последовательностей.

В целом, можно надеяться, что предложенный в настоящей монографии подход, основанный на принципах топологического кодирования цепных полимеров, в процессе дальнейшей разработки будет иметь широкие перспективы для использования полученных на его основе результатов – как в биологии, так и в наноэлектронике.

СПИСОК ЛИТЕРАТУРЫ

1. Карасев В. А., Лучинин В. В. Введение в конструирование бионических наносистем. М.: Физматлит, 2009.

2. Дорфман В. Ф. О топологической разрешимости интегральных структур без совмещений // Микроэлектроника. 1975. Т. 4, вып. 3. С. 213–219.

3. Anfinsen C. B. On the possibility of predicting tertiary structure from primary sequence / Ed. Sela M. // New perspectives in biology. Amsterdam Elsevier. 1964. P. 42–50.

4. Chen H., Gu F., Huang Z. Improved Chou-Fasman method for protein secondary structure prediction // BMC Bioinformatics. 2006. Vol. 7, N_{O} 4. P. 14.

5. Advances, interactions, and future developments in the CNS, Phenix, and Rosetta structural biology software systems / P. D. Adams, D. Baker, A. T. Brunger et al. // Annu. Rev Biophys. 2013. Vol. 42. P. 265–287.

6. Kister A. E., Potapov V. Amino acid distribution rules predict protein fold // Biochem. Soc Trans. 2013. Vol. 41. P. 616–619.

7. Меклер Л. Б., Идлис Р. Г. Общий стереохимический генетический код – путь к биологии и универсальной медицине XXI века // Природа. 1993. № 5. С. 29–63.

8. Blalock J. E., Smith E. M. Hydropathic anticomplementarity of amino acids based on the genetic code // Biochem. Biophys .Res. Commun. 1984. Vol. 121. P. 203–207.

9. Чипенс Г. И. Скрытая симметрия генетического кода и законы взаимодействия аминокислот // Биоорганич. химия. 1991. Т. 17, № 10. С. 1335–1346.

10. The polypeptide tunnel system in the ribosome and its gating in erythromycin resistance mutants of L4 and L22 / I. S. Gabashvili, S. T. Gregory, M. Valle et al. // Mol. Cell. 2001. Vol. 8. P. 181–188.

11. Structure of the 80S ribosome from Saccharomyces cerevisiae– tRNA–ribosome and subunit-subunit interactions / C. M. Spahn, R. Beckmann, N. Eswar et al. // Mol. Cell. 2001. Vol. 107. P. 373–386.

12. Three-dimensional structures of translating ribosomes by Cryo-EM / R. J. Gilbert, P. Fucini, S. Connell et al. // Mol. Cell. 2004. Vol. 14. P. 57–66.

13. Zhang G., Ignatova Z. Folding at the birth of the nascent chain: coordinating translation with co-translational folding // Curr. Opin. Struct. Biol. 2011. Vol. 21. P. 25–31.

14. Карасев В. А. О роли систем сопряженных ионно-водородных связей в надмолекулярных структурах // Вестн. Ленингр. ун-та. 1974. № 9. С. 74–86.

15. Карасев В. А., Стефанов В. Е., Курганов Б. И. Надмолекулярные биоструктуры: организация, функционирование, происхождение // Итоги науки и техники. Сер. «Биол. химия». М.: ВИНИТИ, 1989. Т. 31.

16. Карасев В. А., Стефанов В. Е. Эволюционный структурнофункциональный подход к надмолекулярным структурам // Успехи биол. хим. 1991. Т. 32. С. 114–145.

17. Карасев В. А., Лучинин В. В., Стефанов В. Е. Как построить биочип // Биотехнология. 1993. № 2. С. 3–15.

18. Karasev V. A., Luchinin V. V., Stefanov V. E. A Model of molecular electronics based on the concept of conjugated ionic-hydrogen bond systems // Adv. Mater. Opt. Electron. 1994. Vol. 4. P. 203–218.

19. Карасев В. А., Стефанов В. Е. Об элементной базе молекулярной биоэлектроники // Биомолекулярная электроника и проблема самосборки надмолекулярных структур / под ред. П. И. Лазарева; Науч. центр биол. исслед. Пущино, 1987. С. 45–53.

20. Карасев В. А., Лучинин В. В. Молекулярная архитектура органических сенсорных наносистем // Петербург. журн. электроники. 2001. № 4. С. 12–32.

21. Карасев В. А. Архитектура, принципы организации и функционирования биоорганических наноструктур // Нанотехнология. Физика. Процессы. Диагностика. Приборы / под ред. В. В. Лучинина и Ю. М. Таирова. М.: Физматлит, 2006. С. 65–97.

22. Карасев В. А., Лучинин В. В., Соколов А. И. Био- и квантовоинформационные технологии в наноэлектронике / СПбГЭТУ «ЛЭТИ». СПб., 2013.

23. Pauling L. The Nature of the Chemical Bond. 3rd ed. Ithaca: New York: Cornell Univ. Press, 1960.

24. Карасев В. А. Как закодировать топологию биочипа? // Биотехнология. 1998. № 3. С. 62–75.

25. Карасев В. А., Лучинин В. В. Проблемы создания искусственных бионических микро- и наносистем // Изв. вузов. Электроника. 1998. № 6. С. 5–15.

26. Karasev V. A., Demchenko E. L., Stefanov V. E. Topological coding of polymers and protein structure prediction // Chemical topology: applications and techniques. Ser. «Math. Chem.» / Ed. D. Bonchev, D. Rouvray. NewYork; London; Paris: Gordon&Breach, 2000. Vol. 6. P. 295–345.

27. Karasev V. A., Stefanov V. E. Topological nature of the genetic code // J. Theor. Biol. 2001. Vol. 209. P. 303–317.

28. Karasev V. A., Luchinin V. V., Stefanov V. E. Topological coding: Towards new materials for molecular electronics // Adv. Funct. Mater. 2002. № 12. P. 461–469.

29. Карасев В. А. Генетический код: новые горизонты. СПб.: Тэсса, 2003.

30. Карасев В. А., Лучинин В. В. Модель топологического кодирования цепных полимеров для бионической наноэлектроники. І. Топологический код и соответствие физических операторов триплетам кода // Биотехносфера. 2009. № 1. С. 2–10.

31. Bochmann D., Posthoff Ch. Binare Dynamische Systeme. Berlin: Akademie-Verlag, 1981.

32. Яблонский С. В. Введение в дискретную математику. М.: Наука, 1986.

33. Ling G. N. The role of inductive effect in the determination of protein structure // Physiolog. Chem. and Phys. and Med. NMR. 1986. Vol. 18. P. 3–16.

34. Карасев В. А. Аминокислоты канонического набора как неприводимые представления группы векторов – диаметров додекаэдра. Деп. в ВИНИТИ 25.04.2007, № 461-В2007.

35. Karasev V. A., Luchinin V. V., Stefanov V. E. A model of the «molecular vector machine» for protein folding // Proc. of the 3-rd Moscow conf. on computational mol. biology, Moscow, 2007, July 27–31. P. 134–136.

36. Карасев В. А., Лучинин В. В. Модель топологического кодирования цепных полимеров для бионической наноэлектроники. II. Молекулярная векторная машина и структура канонического набора физических операторов // Биотехносфера. 2009. № 2. С. 6–12.

37. Karasev V. A. Theory of topological coding of proteins and nature of antisymmetry of the amino acids canonical set // Symmetry: Culture and Science. 2012. Vol. 23. P. 427–447.

38. Шульц Г., Ширмер Р. Принципы структурной организации белков / пер. с англ. М.: Мир, 1982.

39. Петрашень М. И., Трифонов Е. Д. Применение теории групп в квантовой механике. М.: Наука, 1967.

40. Ичас М. Биологический код. М.: Мир, 1971.

41. Pelc S. R. Correlation between coding-triplets and amino-acids // Nature. 1965. Vol. 207. P. 597–599.

42. Ратнер В. А. Структура и эволюция генетического кода // Итоги науки и техники. Сер. «Мол. биол.». М.: ВИНИТИ, 1985. Т. 21. С. 158–197.

43. Волькенштейн М. В. Физика ферментов. М.: Наука, 1967.

44. Swanson R. A unifying concept for the amino acid code // Bull. Math. Biol. 1984. Vol. 46. P. 187–204.

45. Shcherbak V. I. Rumer's rule and transformation in the context of the co-operative symmetry of the genetic code // J. Theor. Biol. 1989. Vol. 139. P. 271–276.

46. Румер Ю. Б. О систематизации кодонов в генетическом коде // Докл. АН СССР. 1967. Т. 167. С. 1393–1394.

47. Румер Ю. Б. Систематизация кодонов в генетическом коде // Докл. АН СССР. 1968. Т. 183. С. 225–226.

48. Walker G. W. R. Genetics and the origin of the genetic code // Orig. Life. 1974. Vol. 5. P. 351–356.

49. Dankwerts H. J., Neubert D. Symmetryes of the genetic code-doublets // J. Mol. Evol. 1975. Vol. 5. P. 327–332.

50. Карасев В. А. Ромбический вариант генетического словаря на основе комплементарности кодирующих нуклеотидов // Вестн. Ленингр. ун-та. 1976. № 3, вып. 1. С. 93–97. Сер. «Биол.».

51. Карасев В. А., Сорокин С. Г. О топологической структуре генетического кода // Генетика. 1997. Т. 33. С. 744–751.

52. Hasegava M., Miyata T. On the antisymmetry of the amino acid code table // Orig. Life. 1980. Vol. 10. P. 265–270.

53. Bertman M. O., Jungck J. R. Group graph of the genetic code // J. Heredity. 1979. Vol. 70. P. 379–384.

54. Jimenez-Montaño M. A., de la Mora-Basañez C. R., Pöschel Th. The hypercube structure of the genetic code explains conservative and non-conservative aminoacid substitutions in vivo and in vitro // Biosystems. 1996. Vol. 39. P. 117–125.

55. Bourbaki N. Elements de Mathematique. 3-rd. Ed. Livre III. Topologie generale. Paris: Hermann, 1961.

56. Волохонский А. Г. О формальной структуре генетического кода // Цитология и генетика. 1972. № 6. С. 487–492.

57. Trainor L. E. H., Rowe G. W., Szabo V. L. A tetrahedral representation of polycodon sequenses and a possible origin of codon degeneracy // J. Theor. Biol. 1984. Vol. 104. P. 459–468.

58. Jimenez-Montaño M. A., de la Mora-Basañez R. The Genetic code as a six-dimensional boolean hypercube // Proc. Soc. for Math. Biol. Annual Meeting, University of California at Berkeley, July $23^{rd} - 26^{th}$, 1992.

59. Klump H. H. The physical basis of the genetic code: the choice between speed and precision // Arch. Biochem. Biophys. 1993. Vol. 301. P. 207–209.

60. Петухов С. В. Матричная генетика, алгебра генетического кода, помехоустойчивость. М.; Ижевск: НИЦ «Регулярная и хаотическая динамика», 2008.

61. Gamow G. Possible relation between DNA and protein structure // Nature. 1954. Vol. 173. P. 318.

62. Crick C. H. F. The Origin of the Genetic Code // J. Mol. Biol. 1968. Vol. 38. P. 367–379.

63. Shcherbak V. I. Twenty canonical amino acids of the genetic code: the arithmetical regularities. Pt. I // J. Theor. Biol. 1993. Vol. 162. P. 399–401.

64. Shcherbak V. I. Sixty-four triplets and 20 canonical amino acids of the genetic code: the arithmetical regularities. Pt. II // J. Theor. Biol. 1994. Vol. 166. P. 475–477.

65. Damjanović Z. M., Rakočević M. M. Genetic code: a new understanding of codon – amino acid assignment // Posted or updated: 2006-11-09 to q-bio.OT, q-bio.BM.

66. Recent evidence for evolution of the genetic code / S. Osawa, Th. H. Jukes, K. Watanabe, A. Muto // Microbiol. Rev. 1992. Vol. 56, № 1. P. 229–264.

67. Di Giulio M. The origin of the genetic code: theories and their relationships (review) // Biosystems. 2005. Vol. 80. P. 175–184.

68. Sengupta S., Higgs P. G. A Unified model of codon reassignment in alternative genetic codes // Genetics. 2005. Vol. 170, № 2. P. 831–40.

69. Elzanowski A., Ostell J. The Genetic Codes // National Center for Biotech. Information. 2010. Retrieved 6 May 2013.

70. Wilhelm Th., Nikolajewa S. A New Classification Scheme of the Genetic Code // J. Mol. Evol. 2004. Vol. 59. P. 598–605.

71. Castro-Chavez F. The rules of variation: Amino acid exchange according to the rotating circular genetic code // J. Theor. Biol. 2010. Vol. 264. P. 711–721.

72. Shcherbak V. I. Arithmetic inside the universal genetic code // Biosystems. 2003. Vol. 70. P. 187–209.

73. Bashfold Y. D., Tsohantjis I., Yarvis P. D. Codon and nucleotide assignments in a supersymmetric model of the genetic code // Phys. Lett. A. 1997. Vol. 233. P. 481.

74. Фраппат Л., Сорба П., Сциаррино А. Квантовые группы и генетический код // Теор. и матем. физ. 2001. Т. 128, № 1. С. 27–42.

75. Negadi T. Rumer's transformation in biology as the negation in classical logic // Int. J. Quant. Chem. 2003. Vol. 94. P. 65–74.

76. Negadi T. Symmetry Groups for the Rumer–Konopel'chenko– Shcherbak «Bisections» of the Genetic Code and Applications // Internet Electron. J. Mol. Des. 2004. Vol. 3. P. 247–270.

77. Duplij D., Duplij S. DNA sequence representation by trianders and determinative degree of nucleotides // J Zhejiang Univ. Sci. B. 2005. Vol. 6. P. 743–755.

78. Gonzalez D. L. Can the genetic code be mathematically described? // Med. Sci. Monit. 2004. Vol. 10. HY 11-17.

79. Gonzalez D. L., Giannerini S., Rosa R. Detecting structure in parity binary sequences // IEEE Eng. Med. Biol. Magazine. 2006. № 1/2. P. 69–81.

80. Петухов С. В. Бипериодическая таблица генетического кода и число протонов. М.: Изд-во МКЦ, 2001.

81. Petoukhov S. V., He M. Symmetrical Analysis Techniques for Genetic Systems and Bioinformatics: Advanced Patterns and Applications. Hershey, USA: IGI Global, 2010.

82. Yang C. M. On the 28-gon symmetry inherent in the genetic code intertwined with aminoacyl-tRNA synthetases the Lucas series // Bull. Math. Biol. 2004. Vol. 66. P. 1241–1257.

83. Yang C. M. Molecular versus atomic information logic behind the genetic coding content constrained by two evolutionary axes and Fibonacci-Lucas sequence // Ser. Math. Biol. and Med. Vol. 8. Proc. Intern. Conf. «Advances in Bioinformatics and its applications» / Eds.: M. He, G. Narasimhan, S. Petoukhov. New Jersey et al.: World Scientific Publishing Co. Ltd. 2005. P. 554–562.

84. Кабалдин Ю. Г., Дементьев В. И., Рыжова Т. Ю. Пространственная структура триплетного генетического кода и некоторые ее свойства в плоских изображениях // Тр. Нижегородского государственного технического университета им. Р. Е. Алексеева. 2001. № 1 (86). С. 89–97.

85. Sánchez R., Morgado E., Grau R. The Genetic Code Boolean Lattice // MATCH Commun. Math. Comput. Chem. 2004. Vol. 52. P. 29–46.

86. Jimenez-Montaño M. A. Applications of huper genetic code to bioinformatics // Ser. Math. Biol. And Med. Vol. 8. Proc. Intern. Conf. «Advances In Bioinformatics And Its Applications» / Eds.: M. He, G. Narasimhan, S. Petoukhov. New Jersey et al.: World Scientific Publishing Co. Ltd. 2005. P. 473–481.

87. The 24 Possible Algebraic Representations of the Standard Genetic Code in Six or in Three Dimensions / M. V. José, E. Morgado, R. Sánchez, T. Govezensky // Advanced Studies in Biol. 2012. Vol. 4, № 3. P. 119–152.

88. Stambuk N. Universal properties of the genetic code // Croatica Chemica ACTA. 2000. Vol. 73. P. 1123–1139.

89. José M. V., Morgado E. R., Govezensky T. An extended RNA code and its relationship to the standard genetic code: an algebraic and geometrical approach // Bull. Math. Biol. 2007. Vol. 69. P. 215–243.

90. He M. X., Petoukhov S. V., Ricci P. E. Genetic code, hamming distance and stochastic matrices // Bull. Math. Biol. 2004. Vol. 66. P. 1405–1421.

91. Sánchez R., Morgado E., Grau R. A. Gene algebra from a genetic code algebraic structure // J. Math. Biol. 2005. Vol. 51. P. 431–457.

92. Sánchez R., Grau R., Morgado E. A novel Lie algebra of the genetic code over the Galois field of four DNA bases // Math. Biosci. 2006. Vol. 202. P. 156–174.

93. Sánchez R., Morgado E., Grau R. A. Genetic code Boolean structure. I. The meaning of Boolean deductions // Bull. Math. Biol. 2005. Vol. 67. P. 1–14.

94. Jiménez-Montaño M. A., de la Mora-Basañez C. R., Pöschel T. T. On the hypercube structure of the genetic code / Ed. A. Hwa, C. A. Cantor. World Scientific, 1994. P. 445.

95. Campbell P. N., Smith A. D. Biochemistry Illustrated. Edinburgh–London–Madrid–Tokio: Curchill Livingstone, 1994. P. 8–9.

96. Siemion I. Z., Cebrat M., Kluczyk A. The problem of amino acid complementarity and antisense peptides // Curr. Protein Pept. Sci. 2004. Vol. 5. P. 507–527.

97. A common periodic table of codons and amino acids /J. C. Biro, B. Benyo, C. Sansom et al. // Biochem. Biophys. Res. Commun. 2003. Vol. 306. P. 408-415.

98. Balakrishnan J. Symmetry scheme for amino acid codons // Phys. Rev. E. 2002. Vol. 65 (2 Pt. 1). 021912.

99. Sachse S., Roeder Ch. A classification scheme of amino acids in the genetic code by group theory. Leading EDGE forum CSC papers. 2011. pdf–arXiv 1202.0448.pdf.

100. Taylor W. R. The classification of amino acid conservation // J. Theor. Biol. 1986. Vol. 119. P. 205–218.

101. Esteve J. G., Falceto F. Classification of amino acids induced by their associated matrices // Biophys. Chem. 2005. Vol. 115. P. 177–180.

102. May A. C. W. Towards more meaningful hierarchical classification of amino acid scoring matrices // Protein Engineering. 1999. Vol. 12. P. 707–712.

103. Kosiol C., Goldman N., Buttimore N. H. A new criterion and method for amino acid classification // J. Theor. Biol. 2004. Vol. 228. P. 97–106.

104. Лукашенко Н. Е. Расширяющийся генетический код: 21-я и 22-я аминокислоты – селеноцистеин и пирролизин // Генетика. 2010. Т. 46. С. 1013–1032.

105. Karasev V. A., Luchinin V. V., Stefanov V. E. A dodecahedronbased model of spatial representation of the canonical set of amino acids // Ser. «Math. Biol. And Med.». Vol. 8. Proc. Intern. Conf. «Advances In Bioinformatics And Its Applications» / Eds.: M. He, G. Narasimhan, S. Petoukhov. New Jersey et al.: World Scientific Publishing Co. Ltd. 2005. P. 482–493. 106. Карасев В. А. Об антисимметриях канонического набора аминокислот. Деп. в ВИНИТИ 23.03.2004, №470–В2004.

107. Jackman J. E., Alfonzo J. D. Transfer RNA modifications: nature's combinatorial chemistry playground // Advanced Rev. 2013. Vol. 4. P. 35–48.

108. Jalan A. A., Hartgerink J. D. Pairwise interactions in collagen and the design of heterotrimeric helices // Curr. Opin. Chem. Biol. 2013. Vol. 17. P. 960–967.

109. Калишенко Е. Л., Кринкин К. В. Система топологического моделирования структуры белковых молекул // Прикладная информатика. 2009. № 4. С. 114–124.

110. Карасев В. А. Беляев А. И., Лучинин В. В. База данных пентафрагментов белков // Зарег. в фед. агентстве по интеллектуальной собственности «РОСПАТЕНТ». 2010. № 2010620364.

111. Карасев В. А. Беляев А. И., Лучинин В. В. Компьютерная программа для прогнозирования вторичной структуры белка (PREDICTOR) // Зарег. в фед. агентстве по интеллектуальной собственности «РОСПАТЕНТ». 2010. № 2010613374.

112. Карасев В. А. Беляев А. И., Лучинин В. В. Компьютерная программа для конструирования первичной структуры с заданной вторичной структурой (PROTCOM) // Зарег. в фед. агентстве по интеллектуальной собственности «РОСПАТЕНТ». 2011. № 2011611105.

113. Пат. РФ № 2425837 / В. А. Карасев, В. В. Лучинин. Способ прогнозирования вторичной структуры белка. 2011.

114. Пат. РФ № 2511002 / В. А. Карасев, В. В. Лучинин. Способ проектирования первичной структуры белка с заданной вторичной структурой. 2014.

115. Карасев В. А. О принципах построения цепных полимеров. І. Система пентафрагментов // Биотехносфера. 2009. № 5. С. 7–17.

116. Карасев В. А. О принципах построения цепных полимеров. II. Пространственная структура системы пентафрагментов и алгоритмы формирования вторичных структур // Биотехносфера. 2009. № 6. С. 2–13.

117. Карасев В. А. О принципах построения цепных полимеров. III. Минорные элементы и полная пространственная структура пентафрагментов на гиперкубе В⁴ // Биотехносфера. 2011. № 1–2. С. 66–74. 118. Karasev V. A., Stefanov V. E. 10-digits boolean system in description of protein pentafragments // Symmetry: Culture and Science. 2013. Vol. 24. P. 275–293.

119. Reconstruction of protein backbones from the BriX collection of canonical protein fragments / L. Baeten, J. Reumers, V. Tur et al. // PLoS Comput. Biol. 2008. Vol. 4: e1000083.

120. Budowski-Tal I., Nov Y., Kolodny R. FragBag, an accurate representation of protein structure, retrieves structural neighbors from the entire PDB quickly and accurately // Proc. Natl Acad. Sci. USA. 2010. Vol. 107. P. 3481–3486.

121. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules / A. Leaver-Fay, M. Tyka, S. M. Lewis et al. // Methods Enzymol. 2011. Vol. 487. P. 545–574.

122. Демченко Е. Л., Карасев В. А. Компьютерная программа «Декодер надмолекулярной структуры белка – Протеин 3Д» / СПбГЭТУ «ЛЭТИ». РАПО РФ. 05.03.1998. № 980143.

123. Карасев В. А., Демченко Е. Л. Компьютерная программа PROTEIN 3D (Decoder of Protein Supramolecular Structure) и ее применение для декодирования надмолекулярной структуры белков. СПбГЭТУ «ЛЭТИ». Деп. в ВИНИТИ 24.04.1997, № 1379–В1997.

124. Карасев В. А., Демченко Е. Л. Компьютерный анализ систем сопряженных ионно-водородных связей в белках // Тез. докл. II съезда биофизиков России, М., 23–27 авг. 1999. Т. III. С. 946–947.

125. Финкельштейн А. В., Птицын О. Б. Физика белка. М.: Книжный дом «Университет», 2002.

126. Shenoy S. R., Jayaram B. Proteins: sequence to structure and function – current status // Curr. Protein. Pept. Sci. 2010. Vol. 11. P. 498–514.

127. Anfinsen C. B. On the possibility of predicting tertiary structure from primary sequence // New perspectives in biology / Ed. M. Sela. American Elsevier. 1964. P. 42–50.

128. Chou P. Y., Fasman G. D. Prediction of protein conformation // Biochemistry. 1974. Vol. 13. P. 222–245.

129. Beghin F., Dirks J. Une methode statistique simple de prediction des conformations proteiques // Arch. Int. Physiol. Biochem. 1975. Vol. 83. P. 167–168.

130. Periti P. F. Bayesian approach to the recognition of discrete patterns with an applications to a problem of protein molecular structure // Bull. Chim. Pharm. 1974. Vol. 113. P. 187–218.

131. Garnier J., Osguthorpe D. J., Robson B. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins // J. Mol. Biol. 1978. Vol. 120. P. 97–120.

132. Gibrat J.-F., Garnier J., Robson B. Further developments of protein secondary structure prediction using information theory. New parameters and consideration of residue pairs // J. Mol. Biol. 1987. Vol. 198. P. 425–443.

133. Wu T. T., Kabat E. A. An attempt to locate the non-helical and permissively helical sequences of proteins: Application to the variable regions of immunoglobulin light and heavy chains // Proc. Natl. Acad. Sci. USA. 1971. Vol. 68. P. 1501–1506.

134. Nagano K. Triplet information in helix prediction applied to the analysis of supersecondary structures // J. Mol. Biol. 1977. Vol. 109. P. 251–274.

135. Kuntz I. D. An approach to the tertiary structure of globular proteins // J. Amer. Chem. Soc. 1972. Vol. 94. P. 4009–4012.

136. Palau J., Puigdomenech P. The structural code for proteins: Zonal distribution of amino acids residues and stabilisation of helices by hydrophobic triplets // J. Mol. Biol. 1974. Vol. 88. P. 457–469.

137. Tanaka S., Scheraga H. A. Statistical mechanical treatment of protein conformation. III. Prediction of protein conformation based on a three-state model // Macromolecules. 1976. Vol. 9. P. 168–182.

138. Scheraga H. A. Recent progress in the theoretical treatment of protein folding // Biopolymers. 1983. Vol. 22. P. 1–14.

139. Ptitsyn O. B., Finkelstein A. V. Theory of protein secondary structure and algorithm of its prediction // Biopolymers. 1983. Vol. 22. P. 15–25.

140. Lim V. I. Structural principles of globular organisation of protein chains. A stereochemical theory of globular secondary structure // J. Mol. Biol. 1974. Vol. 88. P. 857–872.

141. Functional inferences from blind ab initio protein structure predictions / R. Bonneau, J. Tsai, I. Ruczinski, D. Baker // J. Struct. Biol. 2001. Vol. 134. P. 186–190. 142. Qian N., Sejnowski T. J. Predicting the secondary structure of globular proteins using neural network models // J. Mol. Biol. 1988. Vol. 202. P. 865–884.

143. Stolorz P., Lapedes A., Xia Y. Predicting protein secondary structure using neural net and statistical methods // J. Mol. Biol. 1992. Vol. 225. P. 363–377.

144. Frishman D, Argos P. Recognition of distantly related protein sequences using conserved motifs and neural networks // J. Mol. Biol. 1992. Vol. 228. P. 951–962.

145. Nishikava K., Ooi T. Amino acid sequence homology applied to the prediction of protein secondary structures, and joint prediction with existing methods // Biochim. Biophys. Acta. 1986. Vol. 871. P. 45–54.

146. Levin J. M., Robson B., Garnier J. An algorithm for secondary structure determination in proteins based on sequence similarity // FEBS Lett. 1986. Vol. 205. P. 303–308.

147. Chen H., Gu F., Huang Z. Improved Chou-Fasman method for protein secondary structure prediction // BMC Bioinformatics. 2006. Vol. 7, N_{2} 4. P. 14.

148. Rafferty J. B., Somers W. S., Saint-Girons I., Phillips S. E. V. Three dimensional crystal structures of Escherichia coli met repressor with and without corepressor // Nature. 1989. Vol. 341. P. 705–710.

149. The structure of oxidized cytochrome c 2 of Rhodospirillum rubrum / F. R. Salemme, S. T. Freer, N. H. Xuong et al. // J. Biol. Chem. 1971. Vol. 248. P. 3910–3921.

150. The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements / J. W. R. Schwabe, L. Chapman, J. T. Finch, D. Rhodes // Cell (Cambridge, Mass.). 1993. Vol. 75. P. 567–578.

151. The design, synthesis, and crystallization of an alpha-helical peptide / D. Eisenberg, W. Wilcox, S. M. Eshita et al. // Proteins: Structure, Function, and Bioinformatics. 1986. Vol. 1. P. 16–22

152. DeGrado W. F., Regan L., Ho S. P. The Design of a Four-helix Bundle Protein // Cold Spring Harb. Symp. Quant. Biol. 1987. Vol. 52. P. 521–526.

153. A new approach to artificial and modified proteins: theory-based design, synthesis in a cell-free system and fast testing of structural proper-

ties by radiolabels / V. V. Chemeris, D. A. Dolgikh, A. N. Fedorov et al. // Protein Eng. 1994. Vol. 7. P. 1041–1052.

154. Poole A. M., Ranganathan R. Knowledge-based potentials in protein design // Current Opinion in Structural Biology. 2006. Vol. 16. P. 508–513.

155. Пат. WO2007030594. МПК G06F19/22; G06F19/18. Methods of using and analyzing biological sequence data. Опубл.15.03.2007. Бюл. № 18.

156. Dahiyat B. I., Mayo S. L. De Novo Protein Design: Fully Automated Sequence Selection // Science. 1997. Vol. 278. P. 82–87.

157. Design of a novel globular protein fold with atomic-level accuracy / B. Kuhlman, G. Dantas, G. C. Ireton et al. // Science. 2003. Vol. 302. P. 1364–1368.

158. Пат. US7574306. MПК G06F19/00. Method and system for optimization of polymer sequences with stable, 3-dimensional conformations. Опубл. 11.08.2009.

159. Koga N., Tatsumi-Koga R., Liu G. Principles for designing ideal protein structures // Nature. 2012. Vol. 1491. P. 222–227.

160. Kolb V. A., Makeyev E. V., Spirin A. S. Co-translational folding of an eukaryotic multidomain protein in a prokaryotic translation system // J. Biol. Chem. 2000. Vol. 275. P. 16597–16601.

161. Antagonist Hiv-1 Gag Peptides Induce Structural Changes In Hla B8 / S. W. Reid, S. McAdam, K. J. Smith et al. // J. Exp. Med. 1996. Vol. 184. P. 2279–2286.

162. Crystal structure of human glutathione peroxidase 3 (selenocysteine to glycine mutant) / E. S. Pilka, K. Guo, O. Gileadi et al. // Данные Protein Date Bank.

163. Agari Y., Kuramitsu S., Shinkai A. X-ray crystal structure of tthb099, a crp/fnr superfamily transcriptional regulator from thermus thermophilus hb8, reveals a DNA-binding protein with no required allosteric effector molecule // Proteins. 2012. Vol. 80. P. 1490–1494.

164. Three-dimensional structures of acidic and basic fibroblast growth factors / X. Zhu, H. Komiya, A. Chirino et al. // Science. 1991. Vol. 251. P. 90–93.

165. Ратнер В. А. Молекулярно-генетические системы управления. Новосибирск: Наука, 1975.

СОКРАЩЕНИЯ И ОБОЗНАЧЕНИЯ

Азотистые основания:

- А аденин;
- Т тимин;
- G гуанин;
- С цитозин.

Аминокислоты:

- Ala аланин (alanine);
- Arg аргинин (arginine);
- Asp аспарагиновая кислота (asparagic acid);
- Asn аспарагин (asparagine);

Cys-цистеин (cysteine);

Glu – глутаминовая кислота (glutamic acid);

Gln – глутамин (glunamine);

Gly – глицин (glycine);

His – гистидин (histidine);

Ile – изолейцин (isoleucine);

Leu – лейцин (leucine);

Lys – лизин (lysine);

Met – метионин (methionine);

Pro – пролин (proline);

Phe-фенилаланин (phenylalanine);

Ser – серин (serine);

Thr – треонин (threonine);

Trp – триптофан (tryptophane);

Tyr – тирозин (tyrosine);

Val – валин (valine).

Буквы четырехбуквенного алфавита для кодирования пар переменных:

K, L, N, P.

Обозначения вершин додекаэдра в модели МВМ:

$$A, -A, {}^{1}A, A_{1}, {}^{2}A, A_{2}, -{}^{1}A, -A_{1}, -{}^{2}A, -A_{2},$$

 $B, -B, {}^{1}B, B_{1}, {}^{2}B, B_{2}, -{}^{1}B, -B_{1}, -{}^{2}B, -B_{2}.$

РНК – рибонуклеиновая кислота;

т-РНК – транспортная рибонуклеиновая кислота;

ДНК – дезоксирибонуклеиновая кислота;

МВМ – молекулярная векторная машина;

MBM_б – молекулярная векторная машина белков.

ПФ – пентафрагмент цепного полимера или белка;

ССИВС – системы сопряженных ионно-водородных связей;

Н-связи – водородные связи;

N-конец – обозначение аминогруппы в начале белковой цепи;

С-конец – обозначение карбоксильной группы в конце белковой цепи;

R-Z – простая группа;

R-Z₁, R-Z_m – варианты обозначений простых групп;

Q-R=X – резонансная группа;

 $\mathbf{HQ}-\mathbf{R}=\mathbf{X}, \mathbf{HQ}_{1}-\mathbf{R}=\mathbf{X}_{1}, \mathbf{HQ}_{n}-\mathbf{R}=\mathbf{X}_{n}, \mathbf{HQ}_{i}-\mathbf{R}_{i}=\mathbf{X}_{i}, \mathbf{HQ}_{i-1}-\mathbf{R}=\mathbf{X}_{i},$

HQ_{*i*-3}-R=X_{*i*-4} – варианты обозначений резонансных групп;

HN-С=О-группа – пептидная группа, пептидная связь в белках;

Q_{*i*}H…X_{*i*−4}=R-связь – водородная связь в ПФ цепного полимера;

N_{*i*}H…O_{*i*−4}=С-связь – водородная связь в ПФ белка;

*sp*² – обозначения орбиталей, принятые в квантовой химии.

ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

Антисимметрия вторичных структур – такое рассмотрение вторичных структур, которое предполагает, что если десятизначное описание имеет отношение к α -спиральному участку, то при замене булевых чисел 0 $\leftarrow \rightarrow 1$ точно такое же описание будет существовать для β -структурного участка.

Ациклические $\Pi \Phi$ – пентафрагменты, в которых нет H-связи между атомами Q_i H...X_{*i*-4}=R.

Биоструктуры – биологические цепные полимеры (белки и нуклеиновые кислоты), а также структуры, обеспечивающие хранение информации о них, их воспроизведение (генетический код, канонический набор аминокислот, система транспортных РНК) и их пространственную организацию (рибосомы, шапероны и т. д.).

Боковые цепи аминокислот как неприводимые представления группы векторов – в рамках модели MBM_6 группа векторов, действующих в области связи $N_iH...O_{i-4}=C_{i-4}$, которые воссоздаются группой канонических аминокислот, являющихся неприводимыми (минимальными) представлениями этих векторов.

Вектор действия – одно из направлений действия физического оператора в области связи Q_iH...X_{i-4}=R.

Верхние треугольные матрицы из шести булевых переменных $(x_i = 0,1)$ – матрицы, описывающие расположение ребер связности в ПФ или четырехзвенном цепном графе (иначе – конформацию ПФ или четырехзвенного цепного графа)

Гиперкуб В⁶ – шестимерный булев гиперкуб, в котором соседние вершины связаны однобитовыми переходами (является пространственным представлением суперматрицы).

Канонический набор аминокислот – полный набор боковых цепей аминокислот, встречающихся в структуре белков^{*}.

Канонический набор физических операторов – полный набор боковых цепей полимера, обеспечивающих воссоздание всех векторов действия в молекулярной векторной машине.

^{*} В составе большинства белков этот набор содержит 20 аминокислот. В специализированных клеточных структурах встречается еще две аминокислоты, аналоги цистеина (селеноцитеин) и лизина (пирролизин).
Линейные цепные полимеры – неразветвленные полимеры, состоящие из звеньев, связанных в последовательную цепь.

Минорные $\Pi \Phi$ – название группы редко встречающихся пентафрагментов.

Молекулярная векторная машина – теоретическая абстракция, позволяющая рассматривать механизм действия физических операторов в области образования водородной связи Q_iH...X_{i-4}=R в ПФ цепного полимера.

Операторы антисвязности – боковые цепи полимера, которые препятствуют формированию циклического ПФ, т. е. способствуют образованию ациклических ПФ.

Операторы связности – такие боковые цепи полимера, которые способствуют образованию циклического ПФ и его дополнительной фиксации (например, за счет водородных связей).

Описание пентафрагментов в виде десятизначного булева числа – способ описания водородных связей (конформации) ПФ белков парами переменных^{*}.

Основные $\Pi \Phi$ – название группы наиболее распространенных пентафрагментов.

Пентафрагмент цепного полимера – минимальный объект для теоретического анализа, содержащий пять альфа-атомов полимера.

Преобразование Румера – взаимный переход двух множеств дуплетов или триплетов генетического кода, подчиняющийся правилу $C \leftrightarrow A, G \leftrightarrow U.$

Проблема соответствия «триплет – аминокислота» – проблема, связанная с объяснением причин соответствия боковых цепей аминокислот определенным триплетам генетического кода.

Простая группа (R–Z) – сочетание из двух атомов элементоворганогенов, содержащее одинарную σ-связь.

Пространственная структура дуплетного генетического кода – структура, использующая для своего построения (вследствие вырожденности триплетного генетического кода) первые два основания триплетов и изоморфная булеву гиперкубу В⁴.

^{*} Отсутствие Н-связей описывается парой переменных 00, связь О...HN – парой 01, связь NH...O – парой 10, а связи О...HN и NH...O – парой 11.

Пространственная структура триплетного генетического коda – структура, использующая для своего построения триплеты генетического кода и изоморфная булеву гиперкубу В⁶.

Пространственная структура триплетного топологического кода – структура, изоморфная булеву гиперкубу В⁶, в которой соседние вершины связаны между собой однобуквенными единичными переходами.

Ребра связности – ребра четырехзвенного цепного графа, соединяющие пары несмежных вершин.

Резонансная группа (Q–R=X) – сочетание из трех атомов элементов-органогенов, содержащее две σ -связи и одну π -связь, способную к перемещению, резонансу (Q–R=X $\leftarrow \rightarrow$ Q=R–X).

Системы сопряженных ионно-водородных связей (ССИВС) – системы, состоящие из простых и резонансных групп, связанных водородными связями.

Структурные ребра – ребра четырехзвенного цепного графа, соединяющие пары смежных вершин.

Суперматрица – блочная матрица, содержащая 64 верхние треугольные матрицы, описывающие конформации ПФ или четырехзвенного цепного графа.

Триплетный топологический код – суперматрица, в которой матрицы, описывающие конформации ПФ или четырехзвенного цепного графа, трансформированы в триплеты с помощью четырехбуквенного алфавита.

Физический оператор – боковая цепь цепного полимера, выступающая в роли регулятора самоорганизации этого полимера.

Циклические и ациклические графы – математические аналоги пентафрагментов.

Циклические $\Pi \Phi$ – пентафрагменты, в которых есть H-связь между атомами Q_iH...X_{i-4}=R.

Четырехзвенный цепной граф – математический аналог ПФ, состоящий из пяти вершин и четырех ребер, связующих эти вершины.

Элементы-органогены – химические элементы, являющиеся основой биоорганических веществ (С, N, O, P, S).

приложение

Таблицы каталога десятизначных описаний вторичных структур белка (табличный материал к гл. 9)

Таблица 9.1

Фрагменты текстовых файлов с участками β-структуры и α-спирали, их описание с помощью 10-значных чисел и пространственное представление

Участок	10-значное	Участок	10-значное
β-структуры	описание	α-спирали	описание
1.	-A	1-Б	L
3adk 120 Ala 119 Asp 118 Val 117 Tyr 116 Leu 115 Leu	000000000	3adk 132 Arg O – 136 Ser N 132 Arg N – 128 Arg O 132 Arg 131 Lys O – 135 Thr N 131 Lys N – 127 Lys O 131 Lys 130 Leu O – 134 Glu N 130 Leu N – 126 Thr O 130 Leu 129 Leu O – 133 Gly N 129 Leu N – 125 Met O 129 Leu 128 Arg O – 132 Arg N 128 Arg N – 124 Thr O 128 Arg	1111111111
LIS AGY DITE OF	12 TR LEU Leu 115	Clist CLU	Arg 128
Вид струг	ктуры 1-А	Вид структурь	11-Б

Таблица 9.2

Канонические переходные участки «β-структура – α-спираль» (слева) и «α-спираль – β-структура»

№ свя-	Фрагмент	№ ста-	10-значное	№ свя-	Фрагмент	№ ста-	10-значное
ЗИ	тхт-фаила	дии	описание	ЗИ	іхі-фаила	дии	описание
	2-A	1			2-Б		n
4a 3a 2a 1a	3adk 129 Leu O – 133 Gly N 129 Leu N – 125 Met O 129 Leu 128 Arg O – 132 Arg N 128 Arg N – 124 Thr O 128 Arg 127 Lys O – 131 Lys N 127 Lys N – 123 Glu O 127 Lys 126 Thr O – 130 Leu N 126 Thr N – 122 Pro O 126 Thr 125 Met O – 129 Leu N 125 Met N – 121 Gly O 125 Met 124 Thr O – 128 Arg N 124 Thr 123 Glu O – 127 Lys N 123 Glu 122 Pro O – 126 Thr N 122 Pro 121 Gly O – 125 Met N 121 Gly 120 Ala 119 Asp 118 Val 117 Tyr 116 Leu	10 9 8 7 6 5 4 3 2 1	1111111111 111111101 111110101 1111010101 1101010100 0101010000 01000000	16 26 36 46	 Jadk 142 Asn 141 Asp 140 Asp 139 Val 138 Arg 137 Gly N – 133 Gly O 137 Gly 136 Ser N – 132 Arg O 136 Ser 135 Thr N – 131 Lys O 135 Thr 134 Glu N – 130 Leu O 134 Glu 133 Gly O – 137 Gly N 133 Gly N – 129 Leu O 133 Gly 132 Arg O – 136 Ser N 132 Arg N – 128 Arg O 132 Arg 131 Lys O – 135 Thr N 131 Lys N – 127 Lys O 131 Lys 130 Leu O – 134 Glu N 130 Leu N – 126 Thr O 130 Leu 129 Leu O – 133 Gly N 129 Leu N – 125 Met O 129 Leu 	19 18 17 16 15 14 13 12 11 10	0000000000 000000010 000001010 000101010 101010111 101011111 10111111
••			Leu 116	241 ASP	SS 4.40 ASP 5.22 VAL 5.22 VAL 5.		131 US 132 LOS 135 TR 5 GER

Таблица 9.3

Спиральные вздутия на фоне β-структур и впадины на фоне α-спиралей

№	Фрагмент	№	10-значное	№	Фрагмент	№	10-значное
СВЯ-	txt-файла	ста-	описание	свя-	txt-файла	ста-	описание
311	3-1-A	дии		311	3-1-Б	дии	
	Пять связей ОHN	NHC)	Пропуск пяти пар Н-связей			
16 26 36 46 46 4a 3a 2a 1a	Find Si Thr 50 Asn 49 Glu 48 Thr 47 Gly N – 43 Leu O 47 Gly 46 Ile N – 42 Arg O 46 Ile 45 Ser N – 41 Lys O 45 Ser 44 Gln N – 40 Ala O 44 Gln 43 Leu O – 47 Gly N 43 Leu N – 39 Ile O 42 Arg O – 46 Ile N 42 Arg N – 38 Ser O 42 Arg 41 Lys O – 45 Ser N 41 Lys O – 45 Ser N 41 Lys N – 37 Gly O 41 Lys 40 Ala O – 44 Gln N 40 Ala N – 36 Thr O 40 Ala 39 Ile O – 43 Leu N 39 Ile O – 43 Leu N 39 Ile O – 42 Arg N 38 Ser 37 Gly O – 41 Lys N 37 Gly 36 Thr O – 40 Ala N 36 Thr 35 Ser O – 39 Ile N 35 Ser 34 Glu 33 Asp 32 Ala 31 Ala	18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2	0000000010 000001010 00101010 001010101	4a 3a 1a 16 26 36 46	205 Phe O - 209 Leu N 205 Phe O - 201 Thr O 205 Phe 204 Val O - 208 Met N 204 Val N - 200 Ile O 204 Val N - 200 Ile O 204 Val 203 Arg O - 207 Glu N 203 Arg O - 207 Glu N 203 Arg O - 207 Glu N 203 Arg O - 206 Glu N 202 Lys O - 206 Glu N 202 Lys O - 206 Glu N 202 Lys O - 205 Phe N 201 Thr O - 205 Phe N 201 Thr O - 205 Phe N 201 Ihr O - 203 Arg N 199 Pro O - 202 Lys N 198 Phe 197 Arg 196 Ala 195 Leu 194 Asp 193 Met 192 Gly N - 188 Ala O 192 Gly N - 187 Ala O 191 Val 190 Arg N - 186 Trp O 190 Arg 189 Glu N - 185 Val O 189 Glu 188 Ala O - 192 Gly N 188 Ala O - 192 Gly N 188 Ala O - 191 Val N 187 Ala O - 191 Val N 187 Ala N - 183 Pro O 187 Ala N - 183 Pro O 187 Ala N - 183 Pro O 187 Ala N - 180 Pro O 187 Ala N - 180 Pro O 186 Trp O - 190 Arg N 186 Trp N - 182 Val O 186 Trp N - 182 Val O 186 Trp N - 182 Val O 185 Val O - 189 Glu N 185 Val O - 18	9 8 7 6 5 4 3 2 1 8 8 17 16 15 14 13 12 11	111111101 111110101 1111010101 10101010
to the set of the set				15 300 IIII 201 AR	9 F90 - 130 FHE - 202 LVS - 205 FHE - 205 FHE	193 mr	192 @LT 199 GLD 199 GLD 199 ALA 6 THE
	Вид структуры	3-1-A			Вид структуры	3-1-Б	

Продолжение табл. 9.3

№ свя-	Фрагмент	№ ста-	10-значное	№ свя-	Фрагмент	№ ста-	10-значное
3И	txt-файла	дии	описание	3И	txt-файла	дии	описание
	3-2-A				3-2-Б		
	Четыре связи О HN	NH(C		Пропуск четырех пар	Н-связе	ей
16 26 36 46 46 4a 3a 2a 1a	49 Gin 48 Ala 47 Asn 46 Ser 45 Arg N - 41 Leu O 45 Arg N - 40 Thr O 44 Ser N - 40 Thr O 43 Thr N - 39 Leu O 44 Ser 41 Leu N - 38 Ile O 42 Leu 41 Leu N - 37 Ser O 41 Leu N - 37 Ser O 41 Leu O - 44 Ser N 40 Thr O - 44 Ser N 40 Thr N - 36 Glu O 40 Thr 39 Leu O - 43 Thr N 39 Leu O - 43 Thr N 39 Leu O - 43 Thr N 39 Leu O - 42 Leu N 38 Ile 37 Ser O - 41 Leu N 36 Glu O - 40 Thr N 36 Glu O - 39 Leu N 35 Glu O - 39 Leu N 35 Glu 34 Asp O - 38 Ile N 34 Asp 33 Thr 32 Gly 31 Leu 30 Gly	18 17 16 15 14 13 12 11 9 8 7 6 5 4 3 2	0000000010 000001010 0000101010 001010101 101010111 1010111111	4a 3a 2a 1a 16 26 36 46	89 Val O - 93 Asn N 89 Val 81 Leu O - 92 Val N 88 Leu N - 84 Gln O 88 Leu N - 84 Gln O 88 Leu N - 84 Gln O 87 Ala O - 91 Met N 87 Ala O - 91 Met N 87 Ala O - 91 Met N 88 Ala O - 90 Asp N 86 Ala O - 90 Asp N 86 Ala O - 90 Asp N 86 Ala O - 80 Ala N 85 Glu O - 89 Val N 85 Glu O - 89 Val N 84 Gln O - 88 Leu N 84 Gln O - 87 Ala N 83 Gln O - 87 Ala N 83 Gln B 82 Asp O - 86 Ala N 82 Asp O - 86 Ala N 82 Asp O - 86 Ala N 82 Asp O - 73 Gly O 77 Gly N - 73 Gly O 76 Leu N - 72 Leu O 76 Leu N - 71 His O 75 Thr 74 Arg 73 Gly O - 77 Gly N 73 Gly O - 77 Gly N 73 Gly O - 75 Thr N 71 His O - 74 Arg N 70 Arg O - 74 Arg N 70 Arg O - 74 Arg N	9 8 7 6 5 4 3 2 18 17 16 15 14 13 12 11	111111101 111110101 111101010 10101010
вил структуры 3-2-А					и из каке	3-2-5	74 ASC 74 ASC 72 LE9 ATE 70
	Вид структуры	3-2-A			Вид структуры	3-2-Б	

Продолжение табл. 9.3

N⁰	Фрагмент	N⁰	10-значное	N⁰	Фрагмент	N⁰	10-значное
свя-	txt-файла	ста-	описание	свя-	txt-файла	ста-	описание
311	3-3-A	дии		311	3-3-Б	дии	
	Три связи О HN N	НО			Пропуск трех пар Н	-связей	
16 26 36 46 46 4a 3a 2a 1a	$\begin{array}{c} 1a4p\\ 63 \ Asp\\ 62 \ Arg\\ 61 \ Cys\\ 60 \ Gln\\ 59 \ Asp \ N-55 \ Met \ O\\ 59 \ Asp\\ 58 \ Leu \ N-55 \ Met \ O\\ 58 \ Leu \ S^{-} \ 57 \ Asp\\ 56 \ Lys\\ 57 \ Asp\ N-52 \ Asp\ O\\ 56 \ Lys\\ 55 \ Met\ O-59 \ Asp \ N\\ 55 \ Met\ N-51 \ Val \ O\\ 55 \ Met\ N-51 \ Val \ O\\ 54 \ Ile \ N-50 \ Ala \ O\\ 54 \ Ile \ N-50 \ Ala \ O\\ 51 \ Val \ O-55 \ Met \ N\ O-55 \ Met$	18 17 16 15 14 13 12 8 7 6 5 4 3 2	0000000010 000001010 000010101 001010101 101010111 101011111 10111111	4a 3a 2a 1a 16 26 36 46	$\begin{array}{c} 3n5l\\ 307 Lys O - 311 Leu N\\ 307 Lys N - 303 Glu O\\ 307 Lys N - 303 Glu O\\ 307 Lys \\ 306 Ala O - 310 Ala N\\ 306 Ala N - 302 Glu O\\ 306 Ala \\ 305 Ala N - 302 Glu O\\ 306 Ala \\ 305 Ala N - 301 Ala O\\ 305 Ala \\ 305 Ala \\ 305 Ala \\ 304 Lys O - 308 Leu N\\ 304 Lys O - 308 Leu N\\ 304 Lys O - 307 Lys N\\ 303 Glu O - 306 Ala N\\ 302 Glu O - 306 Ala N\\ 302 Glu \\ 301 Ala O - 305 Ala N\\ 301 Ala O - 304 Lys N\\ 300 Gly \\ 299 Leu \\ 298 Asn \\ 297 Ala \\ 296 Asn N - 292 Asp O\\ 296 Asn \\ 295 Asn N - 291 Thr O\\ 295 Asn \\ 294 Ala N - 290 Arg O\\ 294 Ala N - 290 Arg O\\ 293 Val N - 289 Gln O\\ 292 Asp O - 296 Asn N\\ 292 Asp N - 288 Lys O\\ 292 Asp N - 288 Lys O\\ 292 Asp N - 288 Lys O\\ 292 Asp N - 286 Leu O\\ 290 Arg O - 294 Ala N\\ 290 Arg O - 294 Ala N\\ 290 Arg N - 285 Glu O\\ 289 Gln N - 285 Glu O\\ 289 Gln N - 285 Glu O\\ 289 Gln \\ \end{array}$	9 8 7 6 5 4 3 17 16 15 14 13 12 11	111111101 111110101 111010101 10101010
Вид структуры 3-3-А						297 294 294 294	ALA 295 ASH 4 ATA 295 ASH 291 THR 291 THR Gln 289
	энд өгруктуры	551	•		End cibling bi		

Продолжение табл. 9.3

N⁰	Фрагмант	N₂	10 202000	No	Фрагмант	N⁰	10 202000
свя-	Фрагмент txt-файла	ста-	описание	свя-	Фрагмент txt-файла	ста-	описание
3И	txt-panna	дии	onneanne	ЗИ	txt-yanna	дии	описание
	<u>3-4-A</u>				3-4-Б	x	
	Две связи ОНN	NH()		Пропуск двух пар Н	1-связе	еи
16 26 36 46 4a 3a 2a 1a	area Trp 376 Trp 375 His 374 Met 373 Lys 372 Ala 371 Gly N – 367 Phe O 371 Gly 370 Met N – 366 Leu O 370 Met 369 Leu N – 365 Gly O 369 Leu 368 Leu N – 364 Lys O 368 Leu 367 Phe O – 371 Gly N 367 Phe N – 363 Pro O 367 Phe 366 Leu O – 370 Met N 366 Leu N – 362 Asp O 366 Leu 365 Gly O – 369 Leu N 365 Gly 364 Lys O – 368 Leu N 363 Pro 362 Asp O – 366 Leu N 362 Asp 361 Asp 360 Glu 359 Lys 358 Val	18 17 16 15 14 13 7 6 5 4 3 2	0000000010 000001010 000101010 001010101 101010111 1010111101 10111101 111010101 111010101 0101010000 010100000 01000000	4a 3a 2a 1a 16 26 36 46	731 Asp $O - 735$ Val N 731 Asp $O - 735$ Val N 731 Asp $N - 727$ Lys O 731 Asp 730 Arg $O - 734$ Ala N 730 Arg $O - 734$ Ala N 730 Arg $- 726$ Glu O 730 Arg 729 Val $O - 733$ Val N 729 Val $O - 733$ Val N 729 Val $N - 725$ Gln O 729 Val 728 Phe $O - 732$ Phe N 728 Phe $N - 724$ Ala O 728 Phe $N - 724$ Ala O 728 Phe $- 724$ Ala O 728 Phe $- 724$ Ala O 728 Old $O - 730$ Arg N 726 Glu $O - 730$ Arg N 725 Gln $O - 729$ Val N 724 Ala $O - 728$ Phe N 724 Ala $O - 715$ Leu O 721 Ser 720 Gly $N - 716$ Ala O 720 Gly $N - 716$ Ala O 720 Gly $N - 715$ Leu O 719 Tyr 718 Val $N - 714$ Ala O 718 Val $N - 714$ Ala O 717 Glu $O - 721$ Ser N 717 Glu $O - 721$ Ser N 717 Glu $O - 720$ Gly N 716 Ala $O - 710$ Gly N 716 Ala $O - 710$ Tyr N 715 Leu $O - 719$ Tyr N 715 Leu $O - 711$ Gln O 714 Ala $O - 718$ Val N 714 Ala $O - 718$ Val N 714 Ala $O - 718$ Val N 714 Ala $N - 710$ Ser O 714 Ala	9 8 7 6 5 4 16 15 14 13 12 11	111111101 11110101 111010101 10101010
The second secon				, ²⁴	125 GER 725 GE	715 LL	au Ala 714
	Вид структуры	3-4-A			Вид структуры	3-4-Б	

Продолжение табл. 9.3

N⁰	Фрагмент	N⁰	10-значное	№	Фрагмент	N⁰	10-значное
свя-	txt-файла	ста-	описание	свя-	txt-файла	ста-	описание
3И		дии		ЗИ	255	дии	
	3-5-A	NU	0		3-5-Б Прописи отной нари	U opg	ъой
	Одна связь Опіх,	<u>INП</u>	0		пропуск одной пары	п-свя	зеи
16 26 36 46 4a 3a 2a 1 a	1aio 208 Leu 207 Thr 206 Gly 205 Ala 204 Pro 203 Met N – 199 Leu O 203 Met 202 Lys N – 198 Asp O 202 Lys 201 Leu N – 197 Ala O 200 Gly N – 196 Gly O 200 Gly N – 196 Gly O 200 Gly 199 Leu O – 203 Met N 199 Leu N – 195 Tyr O 199 Leu 198 Asp O – 202 Lys N 198 Asp O – 202 Lys N 198 Asp O – 201 Leu N 197 Ala O – 201 Leu N 197 Ala O – 201 Leu N 197 Ala O – 200 Gly N 196 Gly O – 200 Gly N 196 Gly 195 Tyr O – 199 Leu N 195 Tyr 194 Asp 193 Leu 192 Thr	18 17 16 15 14 6 5 4 3 2	0000000010 000001010 0000101010 0010101011 101010101 101010101 101010101 010101000 010100000 01000000	4a 3a 2a 1a 16 26 36 4 6	3148 27 Glu O = 31 Arg N 27 Glu N = 23 Asp O 27 Glu 26 Gly O = 30 Asp N 26 Gly N = 22 Gly O 26 Gly 25 Gly O = 29 Leu N 25 Gly N = 21 Ala O 25 Gly 24 Tyr O = 28 Ala N 24 Tyr N = 20 His O 24 Tyr 20 His O 24 Tyr 23 Asp O = 27 Glu N 23 Asp O = 27 Glu N 23 Asp 22 Gly O = 26 Gly N 21 Ala O = 25 Gly N 21 Ala O = 25 Gly N 21 Ala O = 25 Gly N 21 Ala O = 24 Tyr N 20 His 19 Gly 18 Gly N = 14 Trp O 18 Gly 17 Ile N = 13 Thr O 17 Ile 16 Lys N = 12 Ser O 16 Lys 15 Asp = 11 Lys O 15 Asp 14 Trp O = 18 Gly N 14 Trp N = 10 Ile O 14 Trp 13 Thr O = 17 Ile N 13 Thr N = 9 Asn O 13 Thr 12 Ser O = 16 Lys N 12 Ser N = 8 Thr O 12 Ser I 11 Lys N = 7 Lys O 11 Lys N = 7 Lys O 11 Lys N = 7 Lys O 11 Lys N = 7 Lys O	9 8 7 6 5 15 14 13 12 11	111111101 11110101 111010101 110101010 0101010010
2144 128 117 128 118 128						TA AN	11 ANY
	Вид структуры	3-5-A		•	Вид структуры	3-5-Б	

Окончание табл. 9.3

N₂	Фрагмент	N⁰	10-2020000	N₂	Фрагмент	N⁰	10-211211100
свя-	txt-файла	ста-	описание	свя-	txt-файла	ста-	описание
3И	2 ()	дии		3И	2 (Г	дии	
	Э-0-А Нет связей О. НМ	INH	0		Э-О-Б Нет пропуска Н-	оразеі	ž
	1alo	1111.			5tnc	связет	1
16 26 36 46 4a 3a 2a 1a	195 Tyr 194 Asp 193 Leu 192 Thr 191 Gly N – 187 Ala O 191 Gly 190 Thr N – 186 Val O 190 Thr 189 Val N – 185 Ala O 189 Val 188 Lys N – 184 Thr O 188 Lys 187 Ala O – 191 Gly N 187 Ala 186 Val O – 190 Thr N 186 Val 185 Ala O – 189 Val N 185 Ala 184 Thr O – 188 Lys N 184 Thr 183 Pro 182 Arg 181 Pro 180 Tyr	18 17 16 15 5 4 3 2	0000000010 000001010 001010100 10101001 1010010	4a 3a 1a 16 26 36 46	22 Phe O – 26 Phe N 22 Phe O – 26 Phe N 22 Phe N – 18 Met O 22 Phe 21 Glu O – 25 Ala N 21 Glu N – 17 Glu O 21 Glu 20 Ala O – 24 Ala N 20 Ala N – 16 Glu O 20 Ala 19 Ile O – 23 Lys N 19 Ile N – 15 Ser O 19 Ile 18 Met O – 22 Phe N 18 Met O – 22 Phe N 18 Met 17 Glu O – 21 Glu N 17 Glu 16 Glu O – 20 Ala N 16 Glu 15 Ser O – 19 Ile N 15 Ser 14 Leu N – 10 Ala O 14 Leu 13 Phe N – 9 Glu O 13 Phe 12 Ala N – 8 Ala O 12 Ala 11 Arg N – 7 Gln O 11 Arg 10 Ala O – 14 Leu N 10 Ala O – 13 Phe N 9 Glu N – 5 Asp O 9 Glu 8 Ala O – 12 Ala N 8 Ala N – 4 Thr O 8 Ala 7 Gln O – 11 Arg N 7 Gln N – 3 Met O	9 8 7 6 14 13 12 11	111111101 111110101 1111010101 10101010
	189 YAL 194 GEV 194 GEV 195 GEV 195 GEV 195 JEP 195 JEP 195 JEP 195 JEP 195 JEP 195 JEP 195 JEP 195 JEP	181 PRO Tyr 180		7, 5111 1, 0,3 1, 0,		to all	
	Вид структуры	3-6-A			Вид структуры	3-6-Б	

Таблица 9.4

Изгибы β-структур и изломы α-спиралей

N⁰	Фрагмент	№	10-значное	N⁰	Фрагмент	N⁰	10-значное
свя-	txt-файла –	ста-	описание	свя-	txt-файла – изгио	ста-	описание
311	излом а-спирали 4-1-А	дии		311	р-структуры 4-1-Б	дии	
	Внелрение олной ами	ноки	споты		Внелрение олной ами	нокис	поты
	не образующей Н	-связ	ей		образующей пару 1	Н-связ	вей
16 26 36 3a 2a 1a	1alo 237 His 236 Val 235 Gly 234 Pro 233 Met N - 229 Glu O 233 Met N - 228 Ser O 232 Thr 231 Leu N - 227 ThrO 231 Leu 230 Ala 229 Glu O - 233 Met N 229 Glu 228 Ser O - 232 Thr N 228 Ser O - 232 Thr N 228 Ser 227 Thr O - 231 Leu N 225 Ile 224 Gly 223 Lys	18 17 16 4 3 2	0000000010 000001010 000101010 0010100001 1010000101 000101010 0101010000 0101000000	3a 2a 1a 16 26 36	1bhm 26 Tyr O - 30 Lys N 26 Tyr N - 22 Ile O 26 Tyr 25 Ala O - 29 Val N 25 Ala O - 29 Val N 25 Ala - 21 Leu O 25 Ala 24 Gln O - 28 Glu N 24 Gln O - 28 Glu N 24 Gln O - 27 Asn N 23 Gln O - 27 Asn N 23 Gln O - 27 Asn N 23 Gln O - 26 Tyr N 22 Ile 21 Leu O - 26 Tyr N 22 Ile 21 Leu O - 25 Ala N 21 Leu 20 Lys O - 24 Gln N 20 Lys 19 Asp O - 23 Gln N 19 Asp N - 15 Leu O 19 Asp 18 Lys N - 14 Glu O 18 Lys 17 Ser N - 13 Lys O 17 Ser 16 Leu N - 12 Ala O 16 Leu 15 Leu O - 19 Asp N 15 Leu O - 18 Lys N 14 Glu O - 18 Lys N 14 Glu N - 10 Asp O 14 Glu 13 Lys O - 17 Ser N 13 Lys N - 9 Thr O 13 Lys	9 8 7 13 12 11	1111111101 111110101 111010101 1010101111 01010111100 011110100 1110101011 1010101111 1010111111
224 P		Lys 223	•••	Р Р	тако	V 98 4 cm Lys 13	
	Вид структуры	4-1-A			Вид структуры	4-1-Б	

Продолжение табл. 9.4

Mo	Фрагмент	Mo		Mo	Фрагмент	Mo		
CB4-	тхt-файла –	л≌ ста-	10-значное	CB4-	тхт-файла – изгиб	л≞	10-значное	
зи	изпом а-спираци	лии	описание	зи	В-структуры	лии	описание	
JI	4-2-A	дпп		JI	<u> р-структуры</u> 4-2-Б	дии		
	Виелоение прух эми	IIOVIA	спот	Внедрение двух аминокислот				
	не образующих Н	-CB93	ей		образующих две пар	ы Н-сі	лот, зязей	
16 26 2a 1a	1b8v 85 Asp 84 Ala 83 Asp 82 Lys 81 Phe N – 77 Pro O 81 Phe 80 Ala N – 76 Asp O 80 Ala 79 Thr 78 Met 77 Pro O – 81 Phe N 77 Pro 76 Asp O – 80 Ala N 76 Asp 75 Ala 74 His 73 Ala 72 Thr	18 17 3 2	0000000010 000001010 0000101000 0010100000 1000000	2a 1a 16 26	1avh 253 Glu O – 257 Tyr N 253 Glu N – 249 Ala O 253 Glu 252 Ala N – 248 Pro O 252 Ala 251 Leu O – 255 Leu N 251 Leu N – 247 Ile O 251 Leu N – 244 Thr N 250 Tyr O – 254 Thr N 250 Tyr N – 246 Ser O 250 Tyr 249 Ala O – 253 Glu N 249 Ala 248 Pro O – 252 Ala N 248 Pro 247 Ile O – 251 Leu N 247 Ile O – 251 Leu N 247 Ile N – 243 Ser O 247 Ile 246 Ser N – 242 Lys O 246 Ser 245 Arg 244 Ile N – 240 Val O 243 Ser 242 Lys O – 246 Ser N 242 Lys O – 246 Ser N 242 Lys O – 246 Ser N 242 Lys N – 238 Leu O 241 Val O – 245 Arg N 241 Val O – 245 Arg N 241 Val N – 237 Leu O <t< td=""><td>9 8 12 11</td><td>1111111101 111110101 1111010111 101011111 010111111</td></t<>	9 8 12 11	1111111101 111110101 1111010111 101011111 010111111	
93 A27 93 A27 94 A1A 95 A27 95 A1A 95 A27 96 A1A 97 A1A 97 A1A					ти но пула пула ти те та Двух Н-связей	243 5	HI ANT ALL THE VAL	
	Вид структуры	4-2-A	12		Вид структуры	4-2-Б		

Окончание табл. 9.4

№ CB9-	Фрагмент txt-файла –	№ ста-	10-значное	Nº CB9-	Фрагмент txt-файла – изгиб	№ ста-	10-значное		
ЗИ	излом α-спирали	дии	описание	зи	β-структуры	дии	описание		
	4-3-A				4-3-Б				
	Внедрение трех ами	ноки	слот,	Внедрение трех аминокислот,					
	не ооразующих н	I-СВЯЗ	еи		ооразующих три пар	ы н-сі	вязеи		
16 1a	1gzm 30 Tyr 29 Tyr 28 Gln 27 Pro 26 Ala N – 22 Ser O 26 Ala 25 Glu 24 Phe 23 Pro 22 Ser O – 26 Ala N 22 Ser 21 Arg 20 Val 19 Val 18 Gly	18	0000000010 000001000 000100000 100000001 000000	1a 16	$\begin{array}{l} 332 \ Leu \ O &= 336 \ Asp \ N \\ 332 \ Leu \ N &= 328 \ Leu \ O \\ 332 \ Leu \ N &= 328 \ Leu \ O \\ 331 \ Ala \ O &= 335 \ Trp \ N \\ 331 \ Ala \ N &= 327 \ Gln \ O \\ 331 \ Ala \\ 330 \ Asp \ O &= 334 \ Gln \ N \\ 330 \ Asp \ O &= 334 \ Gln \ N \\ 330 \ Asp \ N &= 326 \ Glu \ O \\ 330 \ Asp \\ 329 \ Tyr \ O &= 333 \ Tyr \ N \\ 329 \ Tyr \ N &= 325 \ Ala \ O \\ 329 \ Tyr \ N &= 325 \ Ala \ O \\ 327 \ Gln \ O &= 331 \ Ala \ N \\ 327 \ Gln \ O &= 331 \ Ala \ N \\ 327 \ Gln \ O &= 331 \ Ala \ N \\ 327 \ Gln \ O &= 331 \ Ala \ N \\ 327 \ Gln \ O &= 331 \ Ala \ N \\ 327 \ Gln \ N &= 323 \ Ala \ O \\ 327 \ Gln \ S26 \ Glu \ O &= 330 \ Asp \ N \\ 326 \ Glu \ O &= 330 \ Asp \ N \\ 326 \ Glu \ O &= 330 \ Asp \ N \\ 326 \ Glu \ O &= 320 \ Tyr \ N \\ 326 \ Glu \ O &= 320 \ Tyr \ N \\ 325 \ Ala \ O &= 329 \ Tyr \ N \\ 325 \ Ala \ N &= 321 \ Thr \ O \\ 323 \ Ala \ O &= 327 \ Gln \ N \\ 323 \ Ala \ O &= 327 \ Gln \ N \\ 323 \ Ala \ O &= 327 \ Gln \ N \\ 323 \ Ala \ O &= 327 \ Gln \ N \\ 323 \ Ala \ O &= 327 \ Gln \ N \\ 322 \ Leu \ O &= 326 \ Glu \ N \\ 322 \ Leu \ O &= 326 \ Glu \ N \\ 322 \ Leu \ O &= 326 \ Glu \ N \\ 322 \ Leu \ O &= 326 \ Glu \ N \\ 322 \ Leu \ O &= 326 \ Glu \ N \\ 322 \ Leu \ O \ S26 \ Glu \ N \\ 321 \ Thr \ O \ S25 \ Ala \ N \\ S21 \ Thr \ O \ S24 \ Ala \ N \\ S20 \ Cys \ O \ S24 \ Ala \ N \\ S20 \ Cys \ O \ S24 \ Ala \ N \\ S20 \ Cys \ O \ S24 \ Ala \ N \ S20 \ Cys \ O \ S24 \ Ala \ N \ S20 \ Cys \ O \ S24 \ Ala \ N \ S20 \ Cys \ O \ S24 \ Ala \ N \ S20 \ Cys \ O \ S24 \ Ala \ N \ S20 \ Cys \ O \ S24 \ Ala \ N \ S20 \ Cys \ O \ S24 \ Ala \ N \ S20 \ Cys \ O \ S24 \ Ala \ N \ S20 \ Cys \ O \ S24 \ Ala \ N \ S20 \ Cys \ O \ S24 \ Ala \ N \ S20 \ Cys \ O \ S24 \ Ala \ N \ S20 \ Cys \ O \ S24 \ Ala \ N \ S20 \ Cys \ O \ S24 \ Ala \ N \ S20 \ Cys \ O \ S24 \ Ala \ N \ S20 \ Cys \ O \ S24 \ Ala \ N \ S20 \ Cys \ O \ S24 \ Ala \ S20 \ Cys \ O \ S24 \ Ala \ S20 \ Cys \ O \ S24 \ Ala \ S20 \ Cys \ O \ S24 \ Ala \ S20 \ Cys \ O \ S24 \ Ala \ S20 \ Cys \ O \ S24 \ Ala \ S20 \ Cys \ O \ S24 \ Ala \ S20 \ Cys \ O \ S24 \ Ala \ S20 \ Cys \ O \ S$	9	1111111101 1111110111 1111011111 110111111		
	20 TTZ	0 VAL	Gly 18		7322 U (223 A		Cys 320 21 THR		
28 GGK 27 990 46 AZA 25/900 22 65K					инан инания инан инания инан инания инан инания инан инания инан инания инан инания инан инания инан инания инан инан инания инан	— 324 АГА В И			
	Вид структуры	4-3-A			Вид структуры	4-3-Б			

Таблица 9.5

Фрагмент txt-файла	№ ста-	10-значное	№	Фрагмент	N⁰	10-значное
txt-файла	ста-					10 Juli moc
· · · · · · · · · · · · · · · · · · ·		описание	свя-	txt-файла	ста-	описание
5-1-A	дии		3и	5-1-Б	дии	
Сдвиг на две аминокис	лоты вве	px		Сдвиг на две аминокисл	оты ввер)X
2oz1 32 Arg 31 Ile 30 Glu 29 Asp 28 Phe N - 24 Leu O 28 Phe 27 Arg 26 Asp 25 Ala N - 21 Pro O 25 Ala 24 Leu O - 28 Phe N 24 Leu 23 His 22 Ala 21 Pro O - 25 Ala N 21 Pro 20 Thr 19 Pro 18 Ala 17 Arg	2	0000000010 000001000 0001000010 1000001001	la 16	2v7q 406 Phe O - 410 Leu N 406 Phe N - 402 Ala O 406 Phe N - 402 Ala O 406 Phe N - 402 Ala O 405 Ghn O - 409 Asp N 405 Gh N - 401 Ala O 405 Gh 404 Ala O - 408 Ser N 404 Ala N - 400 Val O 404 Ala 403 Phe O - 407 Gly N 403 Phe O - 407 Gly N 403 Phe O - 407 Gly N 403 Phe O - 407 Gly N 404 Ala 404 Ala O - 406 Phe 402 Ala O - 407 Gln N 401 Ala O - 405 He N 396 Gln O 400 Val O - 404 Ala N 400 Val N - 396 Gln O 397 Tyr O - 401 Ala N 397 Tyr O - 400 Val N 396 Gln O - 400 Val N 396 Gln O - 392 Leu O 396 Gln O - 392 Hey O 395 Ala N - 391 Lys O 394 Leu N - 390 Met O 394 Leu N - 390 Met O 394 Leu N - 397 Tyr N 393 Glu O - 397 Tyr N 393 Glu O - 397 Tyr N 393 Glu N - 389 Thr O 393 Glu S - 386 Gly 392 Leu N - 386 Gly 392 Leu N - 387 Ala 391 Lys O - 387 Ala 391 Lys O - 387 Ala 391 Lys O	9	11111101 11111011 11011110 01111010 011110101 110101111 010111101 10111101 111101111 111011111 111011111
Агд 17 Агд 17 19 АГА 19 АГА 19 АГА 19 АГА 19 АГА 19 БО 19 БО 10 БО 10 10 БО 10 10 БО 10 10 БО 10 10 БО 10 10 10 10 10 10 10 10 10 10 10 10 10				исе ака начини по разрыв не <u>Н-связи</u> Вид структуры	399 AR NH 399 GLU Pe H- 5-1-6	ээ анд ээ анд ээ анд ээ анд оо анд ээ анд оо анд ээ анд оо анд ээ анд оо анд ээ анд оо анд ээ анд ээ анд эн эн анд эн эн эн анд эн ан
	Сдвиг на две аминокис 2 ог.1 3 2 Arg 3 1 Пе 30 Glu 29 Asp 28 Phe N – 24 Leu O 28 Phe 27 Arg 25 Ala N – 21 Pro O 25 Ala N – 21 Pro O 25 Ala N – 21 Pro O 25 Ala N – 21 Pro O 23 His 22 Ala 21 Pro O – 25 Ala N 21 Pro O – 25 Ala N 22 ABC – 25 Ala N 23 His 24 Leu O – 28 Phe N 24 Leu O – 28 Phe N 25 Ala N – 21 Pro O – 25 Ala N 25 Ala N – 21 Pro O – 25 Ala N 26 Jaco – 27	Сдвит на две аминокислоты вве 2021 32 Arg 31 Ile 30 Glu 29 Asp 28 Phe N – 24 Leu O 28 Phe 27 Arg 26 Asp 25 Ala N – 21 Pro O 25 Ala N – 21 Pro O 25 Ala N – 21 Pro O 20 Thr 19 Pro 20 Thr 19 Pro 18 Ala 17 Arg 18 Ala 17 Arg 18 Ala 17 Arg 19 Ro 19 Gas 19 Gas 10 Ga	Сдвиг на две аминокислоты вверх 2 ozl 32 Arg 31 Ile 30 Glu 29 Asp 29 Asp 26 Asp 26 Asp 26 Asp 26 Asp 27 Arg 31 Le 0000010000 000100000 001000001 001000001 001000001 001000001 001000001 001000000	Сдвиг на две аминокислоты вверх 2021 32 Arg 31 Ile 30 Glu 29 Asp 28 Phe N - 24 Leu O 28 Phe 7 Arg 26 Asp 25 Ala N - 21 Pro O 25 Ala N - 21 Pro O 25 Ala N - 21 Pro O 25 Ala 21 Pro O - 25 Ala N 21 Pro O - 25 Ala N 21 Pro O 20 Thr 19 Pro 16 16 16 16 16 16 16 16 16 16	Сдвит на две аминовисати верх Сдвит на две аминовисати верх 2074 32 Аге 32 Аге 31 Пе 31 Пе 31 Пе 32 Аге 31 Пе 31 Пе 32 Аге 31 Пе 32 Аге 31 Пе 32 Аге 31 Пе 32 Аге 33 Пе 32 Аге 33 Пе 32 Аге 33 Пе 44 Ала О - 408 Ser N 440 Ала О - 407 Gity N 440 Ала О - 406 Phe 440 Ала О - 406 Cin N 440 Ала О - 406 Cin N 440 Ала О - 406 Phe 440 Ала О - 406 Phe 440 Ала О - 406 Ala N 400 Val N - 306 Cin O 310 Fro 22 Ala O - 28 Phe N 20 Thr 16 398 Arg 397 Tyr N - 393 Cin O 396 Cin O - 403 Phe N 397 Tyr N - 393 Cin O 396 Cin O - 403 Phe N 397 Tyr N - 393 Cin O 398 Arg 397 Tyr N - 393 Cin O 396 Cin N - 392 Leu O 397 Tyr N - 393 Cin N 392 Leu D - 398 Cin N 393 Cin N - 393 Dir O 393 Cin N - 393 Dir O 393 Cin N - 393 Dir O 393 Cin N - 393 Cin N 392 Leu D - 398 Cin N 392 Leu D - 398 Cin N 392 Leu D - 398 Cin N 393 Liys D - 395 Ala N 391 Lys D - 3	Сдвит на дле аминокислоты вверх Сдвит на дле аминокислоты вверх 2001 32 Агд 33 Агд 33 Агд 30 СВ 30

Варианты трансляционных сдвигов повторяющихся структур

Окончание табл. 9.5

Фрагмент	N⁰	10-2020000	№	Фрагмент	№	10-211211100
txt-файла —	ста-	описание	свя-	txt-файла – изгиб	ста-	описание
излом α-спирали	дии	onneunne	ЗИ	β-структуры	дии	onneume
5-2-А Слвиг на три аминокисло	оты ввег	x		5-2-Б Слвиг на три аминокисл	оты вверу	x
Сдвиг на три аминокисл 9 Asn 8 Arg 7 Gly 6 Leu 5 Tyr N – 61 Thr O 5 Tyr 3 Pro 1 Thr O – 65 Tyr N 1 Thr N – 57 Ser O 1 Thr O – 65 Tyr N 1 Thr N – 57 Ser O 1 Thr O 9 Lys 8 Pro 7 Ser O – 61 Thr N 7 Ser 6 Arg 5 Lys 4 Asp 3 Phe	18 2	x 0000000010 000001000 000100000 001000000	la 16	Сдвиг на три аминокисл 2W02 433 Leu O – 437 His N 433 Leu N – 429 Pro O 433 Leu N – 429 Pro O 433 Leu N – 429 Pro O 432 Ser O – 436 Asn N 432 Ser O – 436 Asn N 432 Ser O – 436 Asn N 432 Ser N – 428 Arg O 431 Leu O – 435 Phe N 431 Leu O – 435 Phe N 431 Leu O – 435 Phe N 430 Val O – 434 Phe N 430 Val O – 434 Phe N 430 Val O – 434 Leu N 429 Pro O – 433 Leu N 429 Pro O – 432 Ser N 428 Arg O – 432 Ser N 428 Arg O – 432 Ser N 428 Arg O – 432 Gin O 427 Leu O – 431 Leu N 427 Leu N – 423 Gin O 427 Leu N – 423 Gin O 426 Leu N – 422 Tyr O 426 Leu N – 422 Tyr O 426 Gin O – 427 Leu N 423 Gin O – 427 Leu N 423 Gin N – 419 Phe O 423 Gin N – 419 Phe O 422 Tyr N – 418 Trp O 422 Tyr M – 418 Trp O 420 Asp O – 424 Thr N 420 Asp N – 416 Leu O 420 Asp N – 416 Leu O 419 Phe O – 423 Gin N 419 Phe O – 423 Gin N 419 Phe O – 421 Asp N 417 Tyr O – 421 Asp N 417 Tyr N – 413 Asn O 417 Tyr	9	<pre></pre>
69 ASN 668 ARC 67 GLP 50 FTP 66 LEU 65 TTP 64 ASP 63 FRO BLAIL CODVECTUDIN	53	1		кате кале Разрыв Н-связя	Тут 417 Разрыв Н-связи	
	69 ASN 669 ASN 668 ANG 57 GLÝ 56 LEU 65 TVE 65 TVE 61 65 TVE 63 PRO ВИД СТРУКТУРЫ	69 ASN 69 ASN 668 ARG 67 GLY 56 ARG 57 SER 56 ARG 57 SER 56 ARG 61 THR 63 PRO Вид структуры 5-2-А	69 АБИ 69 АБИ 69 АБИ 69 АБИ 657 БЕК 67 БЕК 63 ГРЮ Бы, АБР 65 ТГР 65 ТГР 65 ТГР 65 ТГР 65 ТГР 65 ТГР 65 ТГР 65 ТГР 65 ТГР 65 ТРК 65 ТЕК 65 ТЕС	16 16 16 16 16 16 16 16 16 16	16 421 Аур Алар А-417 Туг О 421 Аур 420 Аур 0 - 424 Тыг N 420 Аур 0 - 424 Тыг N 420 Аур 0 - 423 Gin N 419 Phe 0 - 422 Tyr N 418 Trp 0 - 423 Asp N 417 Tyr 0 - 421 Asp N	16 422 1yr 421 Asp N-417 Туг О 420 Asp N-416 Leu O 420 Asp N-416 Leu O 420 Asp N-416 Leu O 420 Asp N-416 Leu O 419 Phe O-423 Gln N 419 Phe N-413 Leu O 418 Trp O-421 Asp N 417 Tyr N-413 Asn O 417 Tyr N-413 Asn O 417 Tyr N-413 Asn O 417 Tyr 10 16 110 17 111 18 112 17 113 17 114 17 115 18 116 18 117 17 117 17 118 17 117 17 117 17 117 17 118 17 117 17 117 17 117 17 117 17 117 17 117 17 118 17 119 18 110 17 117 17 117 17 117 17 117 17 117 18 118 19 119 10

Таблица 9.6

Модификации исходных структур



Таблица 9.7

Единичные модификации Н-связей ($01 \rightarrow 11$ и $10 \rightarrow 11$	l) в канонических переходах
----------------------------------	---	-----------------------------

№	Фрагмент	N⁰	10-2020000	N⁰	Фрагмент	N⁰	10-2020000
свя-	tyt doğu	ста-	10-значное	свя-		ста-	10-значное
ЗИ	іхі-файла	дии	описание	ЗИ	іхі-файла	дии	описание
	7-1-A				7-1-Б		
Ν	Иодификация Н-связи в	поло	жении 1а	N	Лодификация Н-связи в	полож	кении 1б
	1bag				1aso		
	339 Asn O – 343 Asn N				264 Ile		
	339 Asn N – 335 Ile O				263 Ser		
	339 Asn				262 Thr		
	338 Val O – 342 His N				261 Asp		
	338 Val N – 334 Ala O				260 Thr N – 256 Asn O		
	338 Val				260 Thr		
	337 Ala O – 341 Phe N				259 Phe		
	337 Ala N – 333 Gln O				258 Trd		
	337 Ala				257 Lvs		
	336 Thr O – 340 Arg N	9	1111111101	16	256 Asn O – 260 Thr N	18	0000000010
	336 Thr N – 332 Asp O	8	1111110101		256 Asn N – 252 Ser O		0000001000
	336 Thr	7	1111010101		256 Asn		0000100000
	335 Ile O - 339 Asn N	,	1101010111		255 Asn N = 251 Asn O		001000000
	335 Ile		0101011100		255 Asn		1000000011
	$334 \text{ Ala } \Omega = 338 \text{ Val N}$		0101110000		254 Cys N = 250 Phe O		0000001110
	334 Ala		011100000		254 Cvs		0000111010
	333 Gln O = 337 Ala N		110000000		253 Ile N $= 249$ Leu O		0011101010
	333 Gln		0000000100		253 Ile		11 10101011
	$332 \operatorname{Asn} \Omega = 336 \operatorname{Thr} N$		0000000100		252 Ser $\Omega = 256 \text{ Asn N}$	13	101010101111
19	332 Asp N = 328 Alg O		000100000		252 Ser 0 = 230 Ash 1	12	10101011111
14	332 Asp 1 = 520 Ala O	2	010000000		252 Ser 240 Lys C	11	1010111111
	331 Glu	2	0100000000		$251 \text{ Asn } \Omega = 255 \text{ Asn } N$	11	1011111111
	330 Phe				251 Asp V = 247 Met O		
	329 Leu				251 Asp = 247 Wet 0		
	$328 \text{ Alg } \Omega = 332 \text{ Asn N}$				$250 \text{ Phe } \Omega = 254 \text{ Cys } N$		
	$\frac{320 \text{ Ala O} - 352 \text{ Asp } 1}{328 \text{ Ala}}$				250 Phe N $= 246$ Ser O		
	327 Ser				250 Phe		
	326 Gly				$240 \text{ Let } \Omega = 253 \text{ He N}$		
	325 Arg				249 Leu N $= 245$ Glu O		
	323 Alg				249 Leu N = 245 Olu O		
	239 ASE 239 ASE 239 ASE 239 TEX 239 TEX 230 TEX 230 TEX 230 TEX 230 TEX 230 TEX 231 TEX 232 ASE 232 ASE				257 139 257 139 258 TBP 258 TBP 259 TBP 259 TBP 259 TBP	251 J	sp 250 Priz
	337 ALA			•			
	Вил структуры						

Продолжение табл. 9.7

Nº	Фрагмент	Nº cT2	10-значное	N⁰	Фрагмент	Nº c⊤a	10-значное
свя- 3И	txt-файла	ста- лии	описание	свя- 3И	txt-файла	лии	описание
-	7-2-A				7-2-Б	, ,	
Ν	Модификация Н-связи в	полох	кении 2а	Ν	Иодификация Н-связи в	полож	кении 2б
2a	lbcc 337 Val O – 341 Gln N 337 Val N – 333 Asp O 337 Val 336 Phe O – 340 Gly N 336 Phe N – 332 Asp O 336 Phe 335 Met O – 339 Gln N 335 Met N – 331 Ile O 335 Met N – 331 Ile O 334 Met O – 338 Leu N 334 Met N – 330 Ser O 334 Met 333 Asp O – 337 Val N 333 Asp 332 Asp O – 336 Phe N 332 Asp 331 Ile O – 335 Met N 331 Ile N – 327 Asp O 331 Ile 330 Ser O – 334 Met N 330 Ser 329 Met 328 Arg 327 Asp O – 331 Ile N 327 Asp 326 Cys 325 Val 324 Phe 323 Tyr	9 8 2	1111111101 1111110101 11110101110 01011101000 01110100001 0100000100 000010000 0000100000 01000000	26	laok 117 Tyr 116 Asn 115 Lys 113 Tyr N - 109 Asn O 113 Tyr 112 Thr 111 Asn 110 Val N - 106 Leu O 100 Asn O - 113 Tyr N 109 Asn O - 113 Tyr N 109 Asn N - 105 Cys O 109 Asn 108 Glu 107 Gly N - 103 Ala O 106 Leu O - 110 Val N 106 Leu O - 102 Ala O 105 Cys O - 109 Asn N 105 Cys N - 101 Ala O 105 Cys 104 Ile O - 108 Glu N 104 Ile N - 100 Arg O 104 Ile 103 Ala O - 107 Gly N 103 Ala N - 99 Asp O 103 Ala	18 12 11	0000000010 0000010000 0010000010 1000001011 0000101110 0010111010 10111010111 1100111111
<u>323 Тут</u> <u>Тут 323</u> <u>324 Рие</u> <u>325 VAL</u> <u>326 CV5</u> <u>332 ASP</u> <u>332 ASP</u> <u>332 ASP</u> <u>329 ККТ</u> <u>329 ККТ</u>						7-2-Б	4 ILE 100 CLU 9 AGN 10 VAL 111 ASN

Окончание табл. 9.7

N⁰	Фрагмент	N⁰	10-значное	N⁰	Фрагмент	N⁰	10-значное
свя- зи	txt-файла	ста-	описание	свя- зи	txt-файла	ста-	описание
m	7-3-A	дии		<u>зи</u> дии 7-3-Б			
N	Иодификация Н-связи в	полох	кении За	Ν	Иодификация Н-связи в	полож	кении 3б
3a	$\begin{array}{c} 3 \mbox{dzc} \\ 138 \ Leu \ O = 142 \ Leu \ N \\ 138 \ Leu \ N = 134 \ Gly \ O \\ 138 \ Leu \ N = 134 \ Gly \ O \\ 138 \ Leu \ N = 134 \ Gly \ O \\ 137 \ Lys \ O = 141 \ Ala \ N \\ 137 \ Lys \ N = 133 \ Glu \ O \\ 137 \ Lys \ N = 133 \ Glu \ O \\ 136 \ Arg \ N = 132 \ Glu \ O \\ 136 \ Arg \ N = 132 \ Glu \ O \\ 135 \ Asn \ N = 131 \ Pro \ O \\ 133 \ Glu \ O = 137 \ Lys \ N \\ 133 \ Glu \ O = 137 \ Lys \ N \\ 133 \ Glu \ O = 137 \ Lys \ N \\ 133 \ Glu \ O = 136 \ Arg \ N \\ 132 \ Glu \ N \\ 132 \ Glu \ N \\ 132 \ Glu \ N \ D = 136 \ Sc \ N \ N \ N \ N \ N \ N \ N \ N \ N \ $	9	1111111101 1111110111 1111101101 011101000 1101000100 010001000 0001000000	36	1a3q 189 189 189 187 188 187 188 187 188 187 188 187 188 187 188 185 184 183 Lys 182 Lys 181 Leu 181 Leu 181 Leu 181 Leu 180 Glu 179 Lys 181 Leu 180 Glu 179 Lys 179 Lys 179 Lys 179 Lys 179 Lys 178 Ala 177 Glu 176	18	0000000010 0000010001 0010001010 1000101011 0010101110 1010111011 111011111 10111111
125 ТКМ 125 ТКМ 138 10 137 175 133 GLV 133 GLV 133 GLV 133 GLV 134 FRO 139 FRO 139 FRO 139 SER ВИД СТРУКТУРЫ 7-3-А						170 ALA 177 101 1 103 LYS 104 VA	INS 180 GLY

Таблица 9.8

Модификации двух Н-связей в канонических переходах

txt-файла	ста-	OTHODINA	свя-	tut horre	ста-				
	пии	описание	314	іхі-файла	лии	описание			
8-1-A	8-1-A				<u>зи</u> дии 8-1-Б				
одификация Н-связей в по	ложе	ниях 1а и 2а	Mo	дификация H-связей в по.	ложен	иях 1б и 2б			
lamp 126 Thr O $-$ 130 Arg N 126 Thr N $-$ 122 Ile O 126 Thr 125 Val O $-$ 129 Ile N 125 Val N $-$ 121 Gly O 125 Val 124 Ala O $-$ 128 Val N 124 Ala O $-$ 128 Val N 124 Ala N $-$ 120 Ser O 124 Ala 123 Ala O $-$ 127 Glu N 123 Ala O $-$ 126 Thr N 122 Ile O $-$ 126 Thr N 122 Ile O $-$ 126 Thr N 121 Gly 120 Ser O $-$ 124 Ala N 120 Ser 119 Ala O $-$ 123 Ala N 119 Ala O $-$ 123 Ala N 119 Ala O $-$ 122 Ile N 118 Asp O $-$ 122 Ile N 118 Asp 117 Asp 116 Asp 115 Ala O $-$ 119 Ala N 115 Ala 114 Gly O $-$ 118 Asp N 114 Gly 113 Pro 112 Ala 111 Val 110 Ser	9 8 3 2	1111111101 1111110101 1111010111 1010111110 01111100001 11100000101 0000010100 0001010000 0101000000	16 26	2epo 27 Ile 26 Glu 25 Gly 24 Leu 23 Asn N – 19 Asp O 23 Asn 22 Leu N – 18 Leu O 22 Leu 21 Tyr 20 Lys 19 Asp O – 23 Asn N 19 Asp N – 15 Leu O 19 Asp 18 Leu O – 22 Leu N 18 Leu N – 14 Ala O 18 Leu 17 Arg N – 13 Lys O 17 Arg 16 Val N – 12 Glu O 16 Val 15 Leu O – 19 Asp N 15 Leu N – 11 Gln O 15 Leu 14 Ala O – 18 Leu N 14 Ala O – 18 Leu N 14 Ala N – 10 Lys O 14 Ala 13 Lys O – 17 Arg N 13 Lys N – 9 Ser O 13 Lys 12 Glu O – 16 Val N 12 Glu N – 8 Ser O	18 17 12 11	0000000010 0000001010 00001010000 10100000111 10000011110 0011111010 11110101111 101011111 10111111			
126 THR 126 THR 127 VA 128 VA 129 VA 121 GEY 121 GE	Ser 110 III VAL	Ý			175 17 ALA 17 ANO 18 LUU 29 LV5 21 LV6				
	Одификация Н-связей в по Патр 126 Thr O – 130 Arg N 126 Thr N – 122 Ile O 126 Thr N – 122 Ile O 126 Thr N – 122 Ile O 125 Val O – 129 Ile N 125 Val N – 121 Gly O 125 Val 124 Ala O – 128 Val N 124 Ala O – 128 Val N 124 Ala O – 128 Val N 123 Ala 123 Ala 122 Ile O – 126 Thr N 122 Ile O – 126 Thr N 122 Ile O – 126 Thr N 121 Gly O – 125 Val N 121 Gly O – 125 Val N 121 Gly O – 125 Val N 120 Ser O – 124 Ala N 120 Ser O – 122 Ile N 118 Asp O – 122 Ile N 118 Asp N – 114 Gly O 118 Asp N – 114 Gly O 115 Ala O – 119 Ala N 115 Ala 114 Gly O – 118 Asp N 114 Gly 113 Pro 112 Ala 111 Val 110 Ser	Папр Положен 1 атр 126 Thr O – 130 Arg N 126 Thr N – 122 Ile O 126 Thr N – 122 Ile O 126 Thr N – 122 Ile O 126 Thr 125 Val O – 129 Ile N 125 Val N – 121 Gly O 125 Val N – 121 Gly O 125 Val N – 121 Gly O 125 Val N – 121 Gly O 125 Val N – 120 Ser O 124 Ala O – 128 Val N 8 124 Ala N – 120 Ser O 124 Ala 123 Ala N – 119 Ala O 123 Ala 122 Ile O – 126 Thr N 122 Ile O 121 Gly O – 125 Val N 3 121 Gly O – 125 Val N 3 121 Gly O – 123 Ala N 120 Ser 119 Ala O – 123 Ala N 120 Ser 119 Ala O – 122 Ile N 118 Asp O – 122 Ile N 118 Asp N – 114 Gly O 118 Asp 115 Ala O – 119 Ala N 115 Ala 115 Ala O – 119 Ala N 115 Ala 114 Gly O – 118 Asp N 114 Gly 113 Pro 112 Ala 111 Val 110 Ser	Элификация Н-связей в положениях 1а и 2а Гатр 126 Thr O – 130 Arg N 126 Thr N – 122 Ile O 126 Thr 125 Val O – 129 Ile N 125 Val N – 121 Gly O 125 Val 124 Ala O – 128 Val N 124 Ala O – 128 Val N 124 Ala O – 127 Glu N 123 Ala O – 127 Glu N 123 Ala N – 119 Ala O 123 Ala 122 Ile O – 126 Thr N 122 Ile O – 126 Thr N 120 Ser O – 124 Ala N 120 Ser 121 Gly 120 Ser O – 124 Ala N 120 Ser 119 Ala O – 123 Ala N 119 Ala O – 123 Ala N 119 Ala O – 123 Ala N 119 Ala N – 115 Ala O 119 Ala 118 Asp O – 122 Ile N 118 Asp N – 114 Gly O 118 Asp 117 Asp 116 Asp 117 Asp 116 Asp 117 Asp 118 Aug O – 118 Asp N 114 Gly O – 118 Asp N 114 Gly O – 118 Asp N 115 Ala 110 Ser	Олификация H-связей в положениях la и 2а Мо, lamp 126 Thr O - 130 Arg N 126 Thr N - 122 Ile O 126 Thr N - 122 Ile O 126 Thr N - 122 Ile O 125 Val N - 121 Gly O 9 111111101 124 Ala O - 128 Val N 8 111110101 11110001 124 Ala O - 128 Val N 8 111110101 14 123 Ala O - 127 Glu N 01011110000 16 1123 Ala O - 127 Glu N 01011110000 16 1123 Ala O - 127 Glu N 01011110000 16 123 Ala N - 119 Ala O 1110000010 26 0001010000 2 0100000000 26 121 Gly O - 125 Val N 3 01010000000 26 120 Ser O - 124 Ala N 2 0100000000 26 119 Ala O - 123 Ala N 2 0100000000 26 119 Ala O - 123 Ala N 2 0100000000 26 119 Ala 114 Gly O - 118 Asp N 115 Ala 114 Gly O - 118 Asp N 113 Ala 114 Gly O - 118 Asp N 116 Asp 111 Val 110 Ser 111 Val	Одификация H-связей в положениях la и 2a Модификация H-связей в положениях la и 2a Iamp 126 Thr N - 130 Arg N 2epo 126 Thr N - 122 Ile O 27 Ile 126 Thr N - 122 Ile O 25 Val N - 121 Gly O 125 Val N - 121 Gly O 9 125 Val N - 121 Gly O 9 124 Ala N - 120 Ser O 1111111100 123 Ala N - 19 Ala O 0101111100 122 Ile O - 126 Thr N 1100000101 122 Ile O - 126 Thr N 00000010100 121 Gly O - 125 Val N 3 122 Ile O - 126 Thr N 100000101 122 Ile O - 126 Thr N 1000001010 122 Ile O - 126 Thr N 010100000 121 Gly O - 125 Val N 3 120 Ser 3 010100000 121 Gly O - 123 Ala N 1 120 Ser 113 Asp O - 122 Ile N 118 Asp N - 114 Gly O 1 119 Ala O - 112 Ala N 1 119 Ala O - 112 Ala N 1 115 Ala O 119 Ala O 115 Ala O 119 Ala N 115 Ala O 119 Ala N <t< td=""><td>Задификация H-связей в положениях I а и 2а Модификация H-связей в положениях I а и 2а Iamp 26 Thr O - 130 Arg N 26 Thr N - 122 IIe O 126 Thr N - 122 IIe O 26 Glu 25 Gly 125 Val O - 129 IIe N 25 Val N - 121 Gly O 23 Asn N - 19 Asp O 124 Ala O - 128 Val N 8 1111110101 22 Leu N - 18 Leu O 17 124 Ala O - 128 Val N 8 1111110101 22 Leu N - 18 Leu O 17 123 Ala O - 127 Glu N 0101111000 16 19 Asp O - 23 Asn N 18 122 IIe O - 126 Thr N 1100000101 19 Asp O - 23 Asn N 112 Leu O 12 122 IIe O - 126 Thr N 110000000 18 Leu O - 22 Leu N 18 111 122 IIe O - 126 Thr N 100000000 18 Leu O - 22 Leu N 110 121 Gly O - 125 Val N 3 0101000000 18 Leu O - 12 Leu N 12 120 Ser O - 124 Ala N 2 010000000 18 Leu N - 14 Ala O 15 Leu N - 11 Gln O 123 Ala O - 122 IIe N 115 Ala O - 119 Ala N 15 Leu N - 11 Gln O 15 Leu N - 11 Gln O 119 Ala 115 Ala O - 119 Ala N 13 Lys N - 9 Ser O 13 Lys N - 9 Ser O</td></t<>	Задификация H-связей в положениях I а и 2а Модификация H-связей в положениях I а и 2а Iamp 26 Thr O - 130 Arg N 26 Thr N - 122 IIe O 126 Thr N - 122 IIe O 26 Glu 25 Gly 125 Val O - 129 IIe N 25 Val N - 121 Gly O 23 Asn N - 19 Asp O 124 Ala O - 128 Val N 8 1111110101 22 Leu N - 18 Leu O 17 124 Ala O - 128 Val N 8 1111110101 22 Leu N - 18 Leu O 17 123 Ala O - 127 Glu N 0101111000 16 19 Asp O - 23 Asn N 18 122 IIe O - 126 Thr N 1100000101 19 Asp O - 23 Asn N 112 Leu O 12 122 IIe O - 126 Thr N 110000000 18 Leu O - 22 Leu N 18 111 122 IIe O - 126 Thr N 100000000 18 Leu O - 22 Leu N 110 121 Gly O - 125 Val N 3 0101000000 18 Leu O - 12 Leu N 12 120 Ser O - 124 Ala N 2 010000000 18 Leu N - 14 Ala O 15 Leu N - 11 Gln O 123 Ala O - 122 IIe N 115 Ala O - 119 Ala N 15 Leu N - 11 Gln O 15 Leu N - 11 Gln O 119 Ala 115 Ala O - 119 Ala N 13 Lys N - 9 Ser O 13 Lys N - 9 Ser O			

Продолжение табл. 9.8

N⁰	Фрагмент	№	10-значное	№	Фрагмент	N⁰	10-значное
свя-	txt-файла	ста-	описание	свя-	txt-файла	ста-	описание
31	- 8-7-A	дии		31		дии	
Mo	дификация Н-связей в по.	ложен	ниях 2а и 3а	Mo	цификация Н-связей в по.	ложен	иях 2б и 3б
3a 2a	laul 128 Arg O - 132 Tyr N 128 Arg N - 124 Arg O 127 Gly O - 131 His N 127 Gly N - 123 Lys O 127 Gly N - 123 Lys O 126 Tyr O - 130 Leu N 126 Tyr N - 122 Leu O 125 Tyr O - 129 Ile N 125 Tyr N - 121 His O 125 Tyr N - 127 Gly N 124 Arg O - 128 Arg N 123 Lys O - 127 Gly N 123 Lys N - 119 Ser O 123 Lys 122 Leu O - 126 Tyr N 122 Leu O - 125 Tyr N 121 His 120 Leu 119 Ser O - 123 Lys N 119 Ser 118 Ser O - 122 Leu N 118 Ser 117 Met 116 Leu 115 Lys 114 Gly	9 3 2	1111111101 11111101111 1111011111 1101111100 1101000101 01000101000 0001010000 0101000000	26 36	$\begin{array}{l} \begin{array}{l} \begin{array}{l} \begin{array}{l} \begin{array}{l} \begin{array}{l} \begin{array}{l} \begin{array}{l} $	18 17 11	00000000010 0000001010 00000101000 00101000101 10000101111 0010111110 10111110111 1111011111 10111111
Вид структуры 8-2-А					36 THR 35 VAL 34 LYS 33 LIE 33 LIE 32 THY 31	4 UY 25 U 27 CI	Glu 22 V3 VAL 26 LYS PRE 29 LYS 30 ASP
	Бид структуры а	5-2-A		l	бид структуры	0-∠-D	

Окончание табл. 9.8

№ свя-	Фрагмент	№ ста-	10-значное	№ свя-	Фрагмент	№ ста-	10-значное		
ЗИ	txt-фаила	дии	описание	ЗИ	txt-фаила	дии	описание		
8-3-A					8-3-Б				
Mo	дификация Н-связей в по	ложен	ниях 1а и За	Mo,	дификация Н-связей в по.	ложен	иях 1б и 3б		
	2bu2				319v				
	108 Leu O - 112 Arg N				100 Ser				
	108 Leu N - 104 Pne O				99 HIS 08 L au				
	100 Leu 107 Ala O 111 Ile N				98 Leu 97 Arg				
	107 Ala V = 103 Gln O				97 Alg 96 Lys N $=$ 92 Cly O				
	107 Ala N = 105 Gm O				96 Lys				
	106 Asp O - 110 Thr N				95 Glu				
	106 Asp N – 102 Ser O				94 Val N – 90 Thr O				
	106 Asp				94 Val				
	105 Thr O – 109 Val N				93				
	105 Thr N – 101 Leu O	9	111111110 1	16	92 Gly O – 96 Lys N	18	000000010		
	105 Thr		1111111 0111		92 Gly N – 88 Phe O		000000 1000		
	104 Phe O – 108 Leu N		1111 011101		92 Gly		0000100010		
	104 Phe		1101110111		91 Ala N – 87 Ala O		0010001000		
•	103 Gln O – 107 Ala N		0111011100		91 Ala		1000100011		
3 a	103 Gln N – 99 Arg O			20	90 Thr O – 94 Val N		0010001110		
	103 Gm 102 Ser O = 106 Asp N				90 Thr N $-$ 80 Lys U 90 Thr				
	102 Ser O = 100 Asp N		00010001		89 Phe N – 85 Glu O		1110111011		
	102.501 101 Leu O – 105 Thr N		010001000100		89 Phe		1011101111		
1a	101 Leu N – 97 Asp O		0001 000000		88 Phe O – 92 Glv N		1110 111111		
	101 Leu	2	01 00000000		88 Phe N – 84 Ile O –	11	10 11111111		
	100 Thr				88 Phe		-		
	99 Arg O – 103 Gln N				87 Ala O – 91 Ala N				
	99 Arg				87 Ala N – 83 Val O				
	98 His				87 Ala				
	97 Asp O – 101 Leu N				86 Lys O – 90 Thr N				
	97 Asp				6 Lys N – 82 Asp O				
	96 Glu				86 Lys 85 Cha O - 80 Dha N				
	95 Pro				85 Glu O = 89 Phe N				
	93 I vs				85 Glu				
	, , , , , , , , , , , , , , , , , , ,	I		-		1			
	TEN TEN			100 S	ER				
	107 ALA		۰						
•	A A			•99 HIS					
	104 PHE		Lys 93		•				
•	IDE EP O105 THR		•		98 LEU				
	103 GLN	94 ASP		•					
		95 PRO		92 GLY	8 PHE	Sci. aci			
	102 SER		-		РН	• <u>1Glu 85</u>			
		Z		Y in the	é,	>			
THERE STARP						Port	R . B6 LYS		
	•-••••••••••••••••••••••••••••••••••••	HIS	•		94 VAL 90 THR 95 GLU				
Вид структуры 8-3-А					Вид структуры	8-3-Б			

Таблица 9.9

Модификации трех Н-связей в канонических перехода	X
---	---

No	_	No		No	_	No	
свя-	Фрагмент	ста-	10-значное	свя-	Фрагмент	ста-	10-значное
3И	txt-файла	дии	описание	3И	txt-файла	дии	описание
511	9_1_A	дш		511	9-1-6	дпп	
Mo	ификация Н_срязей в пол	oven	agy 19 29 39	Mo	ификация Н_срязей в поц	overu	av 16 26 36
10102	ификация п-связеи в пол	Ожені	лях та, 2а, 3а	IVIOL	ификация п-связеи в пол	Ожени	ях 10, 20, 30
3a 2a 1a	$\begin{array}{l} 1 \text{XJO} \\ 105 \ \text{Val} \ \text{O} = 109 \ \text{Ala} \ \text{N} \\ 105 \ \text{Val} \ \text{N} = 101 \ \text{Gly} \ \text{O} \\ 105 \ \text{Val} \ \text{N} = 101 \ \text{Gly} \ \text{O} \\ 105 \ \text{Val} \ \text{N} = 100 \ \text{Ser} \ \text{O} \\ 104 \ \text{Ala} \ \text{N} = 100 \ \text{Ser} \ \text{O} \\ 104 \ \text{Ala} \ \text{N} = 100 \ \text{Ser} \ \text{O} \\ 104 \ \text{Ala} \ \text{N} = 100 \ \text{Ser} \ \text{O} \\ 103 \ \text{Ala} \ \text{N} = 99 \ \text{Gly} \ \text{O} \\ 103 \ \text{Ala} \ \text{N} = 99 \ \text{Gly} \ \text{O} \\ 103 \ \text{Ala} \ \text{N} = 99 \ \text{Gly} \ \text{O} \\ 102 \ \text{Ser} \ \text{O} = 106 \ \text{Leu} \ \text{N} \\ 102 \ \text{Ser} \ \text{N} = 98 \ \text{Asn} \ \text{O} \\ 102 \ \text{Ser} \ \text{N} = 98 \ \text{Asn} \ \text{O} \\ 100 \ \text{Ser} \ \text{N} = 96 \ \text{Asn} \ \text{O} \\ 100 \ \text{Ser} \ \text{N} = 96 \ \text{Asn} \ \text{O} \\ 100 \ \text{Ser} \ \text{N} = 96 \ \text{Asn} \ \text{O} \\ 100 \ \text{Ser} \ \text{N} = 96 \ \text{Asn} \ \text{O} \\ 100 \ \text{Ser} \ \text{N} = 96 \ \text{Asn} \ \text{O} \\ 100 \ \text{Ser} \ \text{N} = 96 \ \text{Asn} \ \text{O} \\ 100 \ \text{Ser} \ \text{N} = 96 \ \text{Asn} \ \text{O} \\ 100 \ \text{Ser} \ \text{N} = 96 \ \text{Asn} \ \text{O} \\ 100 \ \text{Ser} \ \text{N} = 96 \ \text{Asn} \ \text{O} \\ 100 \ \text{Ser} \ \text{N} = 96 \ \text{Asn} \ \text{O} \\ 100 \ \text{Ser} \ \text{N} \\ 99 \ \text{Gly} \ \text{N} = 95 \ \text{Ile} \ \text{O} \\ 99 \ \text{Gly} \ \text{O} = 102 \ \text{Ser} \ \text{N} \\ 98 \ \text{Asn} \ \text{N} = 94 \ \text{Gly} \ \text{O} \\ 98 \ \text{Asn} \ \text{N} = 94 \ \text{Gly} \ \text{O} \\ 95 \ \text{Ile} \ \text{O} = 99 \ \text{Gly} \ \text{N} \\ 95 \ \text{Ile} \ \text{O} = 98 \ \text{Asn} \ \text{N} \\ 94 \ \text{Gly} \ \text{O} = 98 \ \text{Asn} \ \text{N} \\ 94 \ \text{Gly} \ \text{O} = 98 \ \text{Asn} \ \text{N} \\ 94 \ \text{Gly} \ \text{O} = 98 \ \text{Ser} \ \text{N} \ 90 \ \text{Ser} \ 100 \ \text{Ser} \ \text{N} \ 100 \ \text{Ser} \ 100 \ 100 \ \text{Ser} \ 100 \ 100 \ 100 \ 100 \ 100 \ 100 \ 100 \ 100 \ 100$	9 4 3 2	1111111101 11111101 111101111 1111011111 111111001 1111100101 1111000101 1000101010 001010000 010100000 0100000000	16 26 36	2 VSZ 563 Arg 562 Phe 561 Cys 560 Thr 559 Gly N - 555 Arg O 559 Gly 558 Glu N - 554 Asn O 558 Glu 557 Val N - 553 Leu O 557 Val 556 Leu 555 Arg N - 551 Gln O 555 Arg N - 551 Gln O 555 Arg 554 Asn N - 550 Gln O 554 Asn 553 Leu O - 557 Val N 553 Leu N - 549 Lys O 553 Leu 552 Arg 551 Gln O - 555 Arg N 551 Gln N - 547 Leu O 551 Gln 550 Gln O - 554 Asn N 550 Gln O - 554 Asn N 550 Gln O - 554 Asn N 550 Gln N - 546 Glu O 550 Gln 549 Lys O - 553 Leu N 549 Lys N - 545 Leu O 548 Ile O - 552 Arg N 548 Ile O - 552 Arg N	18 17 16	0000000010 0000010100 001010000 1010100011 10100011111 000111111
THE CONVERTIONAL 9-1-A							11e 348
L	вид структуры	9-1-A			вид структуры	7-1-D	

Научное издание

Карасев Владимир Александрович

Принципы топологического кодирования цепных полимеров и структура белков

Литературный редактор В. В. Малиновский Технический редактор О. В. Афанасьева

Оригинал-макет подготовлен коллективом кафедры микро- и наноэлектроники СПбГЭТУ «ЛЭТИ»

Подписано в печать 21.07.2014. Формат 70×100 1/16. Бумага офсетная. Печать офсетная. Печ. л. 15,0. Гарнитура «Times New Roman». Тираж 200 экз. Заказ

Отпечатано с готовых диапозитивов в типографии ЗАО «Электронстандарт–Принт» 196158, Санкт-Петербург, Московское ш., д. 23, корп. 1, лит. А