

УДК 547.963.3, 577.322.4

СПОСОБ МОДЕЛИРОВАНИЯ ПРОСТРАНСТВЕННОЙ СТРУКТУРЫ БЕЛКА ПО ДЕТЕРМИНИРУЮЩЕЙ ЕГО НУКЛЕОТИДНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

В.В. Соколик

ГУ «Институт неврологии, психиатрии и наркологии АМН Украины», 61068, Харьков, ул. Академика Павлова, 46, Украина; e-mail: sokolik67@rambler.ru

Поступила в редакцию 10 ноября 2009 г.

Принята 5 июня 2010 г.

Разработан новый способ моделирования пространственной структуры белка по детерминирующей его нуклеотидной последовательности. Вводится понятие конфигурации пептидной связи в качестве основного, кодируемого в геноме наряду с аминокислотами, элемента пространственной структуры белка. Представлена таблица генетического кода пространственной структуры белка, используя которую можно построить структурный шаблон белка, а также декодировать наличие и положение фрагментов его вторичной структуры. Важным достоинством метода является то, что можно построить структурный шаблон индивидуально для любого неизвестного белка лишь «прочитав» детерминирующую его нуклеотидную последовательность. Проведенный сравнительный анализ схем вторичной структуры, декодированных по нуклеотидным последовательностям, и схем из Protein Data Bank, построенных на основе экспериментальных данных спектроскопии ядерного магнитного резонанса и рентгено-структурного анализа, для α -субъединицы гемоглобина, основного актина и α -актина 1 свидетельствовал о большей разрешимости (до 1 аминокислотного остатка), точности (100%), однозначности (индивидуальный структурный шаблон) и простоты предложенного в данной работе способа моделирования.

КЛЮЧЕВЫЕ СЛОВА: персональный структурный шаблон белка, генетический код пространственной структуры белка, конфигурация пептидной связи, α -субъединица гемоглобина, основной актин, α -актин 1.

СПОСІБ МОДЕЛЮВАННЯ ПРОСТОРОВОЇ СТРУКТУРИ БІЛКУ ЗА НУКЛЕОТИДНОЮ ПОСЛІДОВНІСТЮ, ЩО ЙОГО ДЕТЕРМІНУЄ

В.В. Соколік

ДУ «Інститут неврології, психіатрії і наркології АМН України», 61068, Харків, вул. Академіка Павлова, 46, Україна; e-mail: sokolik67@rambler.ru

Розроблено новий спосіб моделювання просторової структури білку за нуклеотидною послідовністю, що його детермінує. Вводиться поняття конфігурації пептидного зв'язку в якості основного елементу просторової структури білку, що закодований у геномі разом з амінокислотами. Представлена таблиця генетичного коду просторової структури білку, використання якої надає можливість побудувати структурний шаблон білку, а також декодувати наявність і розташування фрагментів його вторинної структури. Суттєвою перевагою метода є та обставина, що можна побудувати структурний шаблон індивідуально для будь-якого невідомого білку лише «прочитавши» нуклеотидну послідовність, яка його детермінує. Порівняльний аналіз схем вторинної структури, що були декодовані за нуклеотидними послідовностями, та схем з Protein Data Bank, які були побудовані на основі експериментальних даних спектроскопії ядерного магнітного резонансу і рентгено-структурного аналізу, для α -суб'єдиниці гемоглобіну, основного актину і α -актину 1 засвідчив більшу чутливість (до 1 амінокислотного залишка), точність (100%), однозначність (індивідуальний структурний шаблон) і простоту способу моделювання, який запропоновано у навіній роботі.

КЛЮЧОВІ СЛОВА: персональний структурний шаблон білку, генетичний код просторової структури білку, конфігурація пептидного зв'язку, α -суб'єдиница гемоглобіну, основний актин, α -актин 1.

THE WAY OF MODELING OF SPATIALS STRUCTURE OF PROTEIN ON DETERMINING IT NUCLEOTIDE SEQUENCES

V.V. Sokolik

SI « Institute of neurology, psychiatry and narcology of the AMS Ukraine», Academician Pavlov St.,
46, Kharkov, 61068, Ukraine, e-mail: sokolik67@rambler.ru

The novel way of modeling of spatialstructure of protein on determining it nucleotide sequences is elaborated. The concept of a configuration of peptide bond as the major coded in genome alongside with amino acids by the rests, an element of spatial structure of protein. The table of a genetic code of spatial structure of protein is presented, using which it is possible to construct a structural template of protein, and also to decode presence and position of fragments of secondary structure of it. The important advantage of a method is that it is possible to construct a structural pattern individually for any unknown protein only "having read through" determining it nucleotide sequence. The lead comparative analysis of schemes of secondary structure, decoded on nucleotide sequences, and schemes from Protein Data Bank, constructed on the basis of experimental data of spectroscopy of a nuclear magnetic resonance and the roentgen-structural analysis, for α -subunit of hemoglobin, major actin and α -actin 1 testified to greater resolvability (up to 1 amino acids rest), accuracy (100 %), unambiguity (an individual structural pattern) and simplicity of the way of modeling offered in given work.

KEY WORDS: personal structural template of protein, genetic code of three-dimensional structure of protein, configuration of peptide bond, α -subunit of hemoglobin, major actin, α -actin 1.

В настоящее время алгоритмы моделирования пространственной структуры белка исходят из его аминокислотной последовательности [1]. В этом заслуга Anfinsen С.В., который первым указал на наличие взаимосвязи между последовательностью аминокислотных остатков и конформацией биологически активной молекулы белка: «...необходимая [для сворачивания белка] информация заключена в линейной последовательности аминокислот пептидной цепочки и никакой дополнительной генетической информации, больше той, что содержится в ДНК, не требуется» [2]. Поэтому дальнейшее развитие этого направления исследований вылилось в классификацию аминокислот по способности входить в состав определённых вторичных структур в зависимости от свойств боковых цепей. В последнее время на основе анализа экспериментального материала и результатов физического моделирования были предприняты попытки найти генетический код пространственной структуры белка, по аналогии с генетическим кодом его аминокислотной последовательности [3]. Однако вновь всё свелось к характеристике аминокислотных остатков в зависимости от дуплета первых двух нуклеотидов кодона, при этом определяющая роль отводилась второму нуклеотиду триплета [4,5].

Хотелось бы обратить внимание на то, что при обсуждении аминокислотной последовательности белка из понятия его первичной структуры невольно ускользает рассмотрение пептидных связей между аминокислотными остатками. Подразумевают, что пептидные связи качественно абсолютно одинаковые, а это не так. В данной работе вводится понятие **конфигурации пептидной связи (КПС)** в качестве основного кодируемого в геноме элемента пространственной структуры белка, а точнее его структурного шаблона. Под конфигурацией пептидной связи следует понимать ротамерные варианты ориентации двух смежных аминокислотных остатков относительно оси пептидной связи.

Накопилось огромное количество противоречий вследствие сведения понятия первичной структуры белка до «линейной последовательности аминокислотных остатков». Рассмотрим два из них.

Wobble гипотеза неоднозначности спаривания кодон-антикодон для третьего нуклеотида триплета при рибомальном синтезе белка. Crick F. предположил, что для кодирования и однозначной трансляции аминокислотного остатка достаточно дуплета первых двух нуклеотидов кодона. Третий нуклеотид может быть любым: по

гипотезе Crick F. он не участвует в кодировании аминокислотного остатка, а «воблирует» [6].

Мы полагаем, что третий нуклеотид кодона действительно не детерминирует аминокислотный остаток в полипептидной цепи, он кодирует конфигурацию пептидной связи в композиции смежных аминокислотных остатков. Данное обстоятельство только подчеркивает значимость третьего нуклеотида: он, в первую очередь, взаимодействует с первым нуклеотидом антикодона тРНК, образуя криковскую пару, и в конечном итоге определяет скелет пространственной структуры белка. В пользу ведущей кодирующей роли третьего нуклеотида свидетельствует анализ распределения точечных мутаций между нуклеотидами кодонов в нуклеотидных последовательностях, детерминирующих белки. Однонуклеотидные замены в триплете нуклеотидов выявляются преимущественно в первом нуклеотиде кодонов, в два раза реже во втором и практически никогда в третьем [7]. Это также свидетельствует о том, что основной вклад в формирование пространственной структуры белка вносит конфигурация пептидных связей, в то время как боковые цепи аминокислотных остатков детерминированы менее строго и некоторые их замены не приводят к катастрофическому нарушению укладки полипептидной цепи и его функциональной конфигурации.

Пространственная структура «идентичных» белков может быть различна. Путём введения замен в аминокислотные последовательности негомологичных белков G_A и G_{B1} (IgG-связывающие домены streptococcal protein) с различной нативной пространственной структурой – 3α и α/β , соответственно, He Y. и соавт. достигли 95% идентичности их аминокислотных последовательностей. Тем не менее, мутанты G_A и G_{B1} сохранили свои исходные пространственные структуры. Ответственность за это исследователи возложили на три не совпадающих аминокислотных остатка [8].

Мы полагаем, что He Y. и соавт. в процессе мутагенеза не тронули третий нуклеотид. Поэтому индивидуальные структурные шаблоны из последовательностей пептидных связей в мутантах G_A и G_{B1} не изменились, несмотря на многочисленные замены в боковых цепях аминокислот. Это ещё раз подчёркивает ведущую кодирующую роль третьего нуклеотида в детерминировании последовательности конфигурации пептидных связей (скелета пространственной структуры белков).

Привлекать понятие конфигурации химической связи, вообще, и пептидной, в частности, целесообразно для химической связи между двумя несимметричными группами атомов в сложной молекуле или полимере, как, например, в аминокислотах или белках (рис. 1). Известно, что различные пространственные формы молекулы, переходящие друг в друга путем вращения вокруг σ -связей C–C, называют ротамерами или *поворотными изомерами* [9]. Три аминокислоты (серин, лейцин и аргинин) имеют ротамеры, для которых предусмотрены разные кодоны [10]. В молекулах глицина или аланина варианты конфигураций химических связей неотличимы друг от друга вследствие отсутствия или симметричности их боковых цепей (рис. 1). Для остальных аминокислот теоретически возможно не менее трех поворотных изомеров по C–C связи, присоединяющей их боковые цепи к остову молекул. Однако в геноме кодируется только один из них, обладающий наибольшей стабильностью вследствие заторможенной конфигурации, в отличие от менее стабильных заслонённых. Показано, что энергетические барьеры, разделяющие ротамеры, не превышают ~ 100 кДж/моль, поэтому времена их жизни в индивидуальном состоянии $\sim 10^{-5}—10^{-13}$ с [11]. Это делает бессмысленным детерминирование в геноме определённого варианта ротамера аминокислотного остатка, если не принять во внимание, что в составе полипептида он дополнительно фиксируется внутримолекулярными водородными связями.

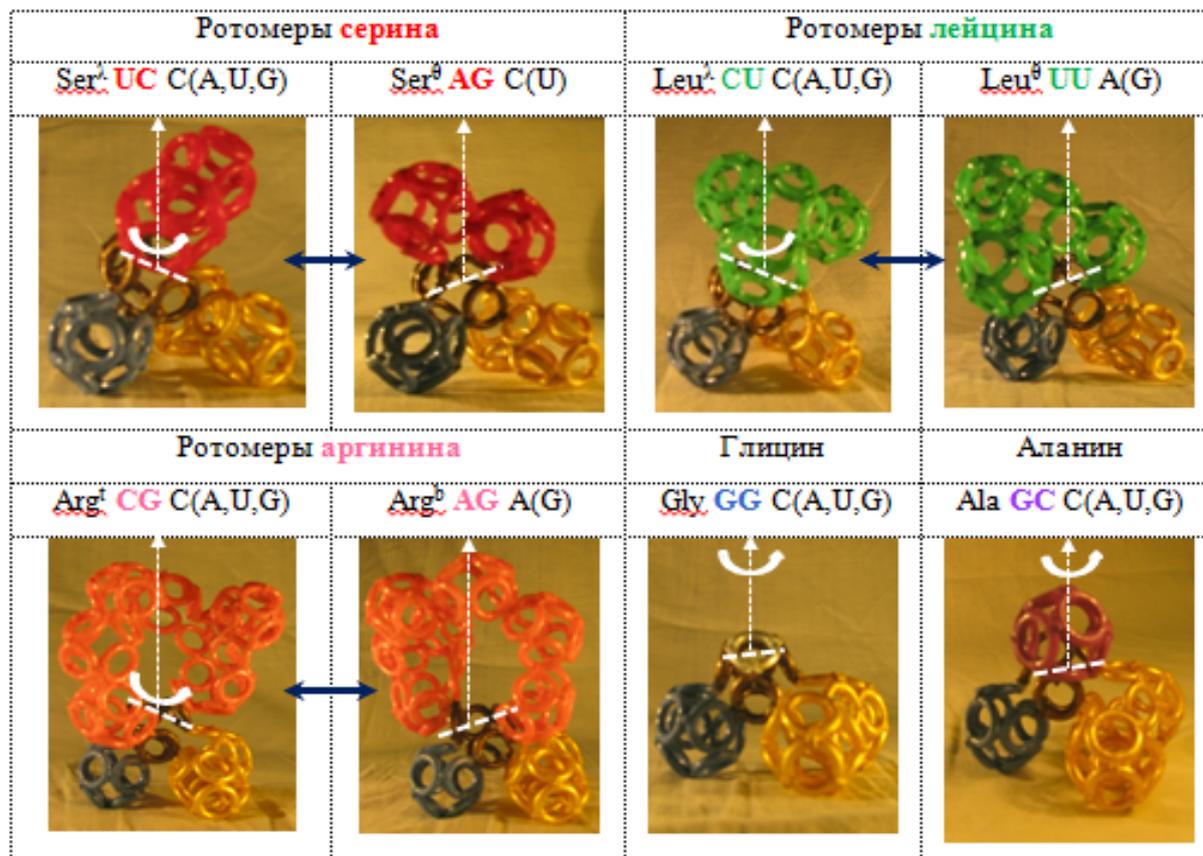


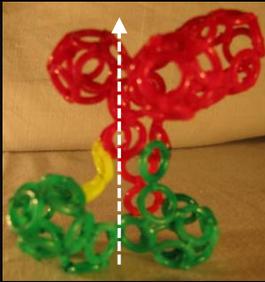
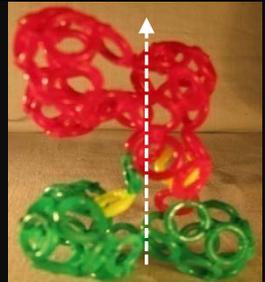
Рис. 1. Кольцеобразные модели поворотных изомеров серина, лейцина, аргинина и их кодоны. Глицин и аланин не имеют поворотных изомеров. Маркировка ротомеров аминокислот соответствует первоисточнику [10].

Аналогичным образом кольцеобразные модели молекул дали возможность выделить три варианта конфигурации пептидной связи R, 0 и L на примере димера аланина (табл. 1). Присоединяемая к фиксированной нижней модели аланина верхняя модель этой аминокислоты имеет три варианта ориентации в пространстве относительно оси симметрии, проходящей через пептидную связь (табл. 1). Далее, на примере физической кольцеобразной модели полиаланина установили, что повторение в полипептиде **R-конфигурации пептидной связи** заставляет его аминокислотную цепочку сворачиваться в правую спираль без какого-либо дополнительного участия боковых цепей аминокислотных остатков. Повторение **0-конфигурации** – в β -тяж, а повторение **L-конфигурации** – в левую спираль (табл. 1). Чередование R, 0 и L-конфигураций пептидных связей (рис. 2) обуславливает неупорядоченный участок полипептидной цепи или выпетливание между фрагментами с вторичной структурой.

Способ моделирования пространственной структуры белка по детерминирующей ...

Таблица 1

Варианты конфигураций пептидной связи в трёх базовых разновидностях вторичных структур белка на примере физической кольцеванной модели полиаланина

Конфигурация пептидной связи	R	0	L
Фото поворотных изомеров по пептидной связи на примере димера аланина			
Кодон	XYC или XYG	XYA	XYU
Вариант вторичной структуры	правая спираль	β -тяж	левая спираль
Фото вторичной структуры полиаланина			
Нуклеотидная последовательность	$(XYC/G)_n$	$(XYA)_n$	$(XYU)_n$

Примечание: Варианты конфигурации пептидной связи: R – правая, 0 – нулевая и L - левая, X и Y – первый и второй нуклеотиды кодирующей аминокислоты триплета. Стрелкой обозначена виртуальная ось симметрии поворотных изомеров по пептидной связи. Цветными кольцами маркированы электроны атомов пептидной группы, принимающие участие в образовании пептидной связи.



Рис. 2. Неупорядоченный фрагмент полипептидной цепи.

Как же кодируется структурный шаблон из пептидных связей белка и как этот код реализуется? На основе эмпирической таблицы композиционного генетического кода (табл. 2) автором данного исследования была составлена **таблица генетического кода пространственной структуры белка** (табл. 3). Под композиционным кодом А.Ю. Кушелев понимает соответствие между третьим нуклеотидом любого кодона и значением композиционного кода (от 1 до 4 по числу вариантов нуклеотидов), которые детерминируют ориентацию присоединяемого аминокислотного остатка относительно растущей полипептидной цепи [10].

Таблица 2

Композиционный генетический код

X	Y	Z	К. код	X	Y	Z	К. код	X	Y	Z	К. код
G	C	C	1	C	C	C	1	G	U	C	1
		A	2			A	2			A	2
		U	3			U	3			U	3
		G	4			G	4			G	4
U	G	C	1	A	A	C	1	A	A	A	2(3)
		U	3			U	3			G	4(1)
G	A	C	1	C	A	A	2(3)	C	A	C	1
		U	3			G	4(1)			U	3
G	A	A	2(3)	A	G	A	2(3)	U	U	C	1
		G	4(1)			G	4(1)			U	3
G	G	C	1	C	G	C	1	A	C	C	1
		A	2			A	2			A	2
		U	3			U	3			U	3
		G	4			G	4			G	4
C	U	C	1	U	C	C	1	A	U	C	1
		A	2			A	2			A	2
		U	3			U	3			U	3
		G	4			G	4			G	4
U	U	A	2	A	G	C	1	U	A	C	1
		G	4(1)			U	3			U	3
U	G	G	4(1)	A	U	G	4(1)	U	A	A	0
										G	0

Примечание: XYZ – триплет нуклеотидов в кодоне; 1, 2, 3 и 4 – значения композиционного кода.

Значения композиционного кода в редакции А.Ю. Кушелева приведены в скобках рядом с уточнениями, принятыми в данной работе. 0 – стоп-кодон.

В отличие от таблицы 2, где кодону аминокислотного остатка соответствует один из четырёх вариантов абстрактных значений композиционного кода (1, 2, 3 или 4), в таблице генетического кода пространственной структуры белка (табл. 3) каждому кодону аминокислотного остатка поставлен в соответствие один из трёх вариантов конфигурации пептидной связи (R, 0 или L). Четвёртый вариант конфигурации пептидной связи в рамках кольцевидной модели белка структурно не предусмотрен.

Для дальнейшей проверки и уточнения таблицы 2, составленной Кушелевым А.Ю. и соавт. на основе корреляционного анализа нуклеотидных последовательностей, детерминирующих белки с надёжно установленными вторичными структурами, была использована схема пространственной структуры триплетного кода [12] (рис. 3). В

схеме 3А каждый кодон заменили соответствующим ему вариантом конфигурации пептидной связи (R, 0, L), по сравнению со значениями композиционного кода (1, 2, 3, 4). Такой приём позволил на основе внутренней симметрии малых кубов в вершинах большого куба уточнить 4 позиции для вариантов конфигурации пептидной связи и 11 позиций для значений композиционного кода (рис. 3С и 3Б).

Таблица 3

Генетический код пространственной структуры белка

Y X	C	A.o.	A	A.o.	U	A.o.	G	A.o.	Z	КПС
C	CCC	Pro	CAC	His	CUC	Leu ^λ	CGC	Arg [†]	C	R
	CCA		CAA	Gln	CUA		CGA		A	0
	CCU		CAU	His	CUU		CGU		U	L
	CCG		CAG	Gln	CUG		CGG		G	R
A	ACC	Thr	AAC	Asn	AUC	Ile	AGC	Ser ^θ	C	R
	ACA		AAA	Lys	AUA		AGA	Arg ^b	A	0
	ACU		AAU	Asn	AUU		AGU	Ser ^θ	U	L
	ACG		AAG	Lys	AUG	Met	AGG	Arg ^b	G	R
U	UCC	Ser ^λ	UAC	Tyr	UUC	Phe	UGC	Cys	C	R
	UCA		UAA	T	UUA	Leu ^θ	UGA	T	A	0
	UCU		UAU	Tyr	UUU	Phe	UGU	Cys	U	L
	UCG		UAG	T	UUG	Leu ^θ	UGG	Trp	G	R
G	GCC	Ala	GAC	Asp	GUC	Val	GGC	Gly	C	R
	GCA		GAA	Glu	GUA		GGA		A	0
	GCU		GAU	Asp	GUU		GGU		U	L
	GCG		GAG	Glu	GUG		GGG		G	R

Примечание: КПС – конфигурация пептидной связи; XYZ – первый, второй и третий нуклеотиды в кодоне; R, 0, L – варианты конфигурации пептидной связи; T – стоп-кодон; Ser^λ и Ser^θ, Leu^λ и Leu^θ, Arg[†] и Arg^b – поворотные изомеры серина, лейцина и аргинина, кодируемые различными дуплетами нуклеотидов кодонов, соответственно. Маркировка конформеров аминокислот соответствует первоисточнику [10].

Экспериментально установлено, что разновидностей среди правых и левых спиралей больше, чем это задаётся структурным шаблоном пептидных связей: 2/7, 3/10, α, π и др. [1]. Поэтому, последовательность конфигураций пептидных связей в спирали определяет только направление закручивания: правая или левая спираль, а разнообразие спиралей определяется свойствами боковых цепей аминокислотных остатков, формирующих виток спирали. Из приведенного положения следует, что на основе структурного шаблона правой спирали, благодаря взаимодействию боковых цепей аминокислотных остатков, может быть сформирована правая спираль 3/10, правая α-спираль, правая π-спираль, правая 2/7 спираль и никогда – β-тяж или любая разновидность левой спирали. Для трансформаций вида: правая спираль ↔ β-тяж ↔

левая спираль ↔ правая спираль необходимо изменение конфигурации пептидных связей, невозможное без их разрыва. В то время как при формировании вариантов спирали с одним направлением закручивания наблюдается лишь некоторая деформация пептидных связей без изменения их конфигурации.

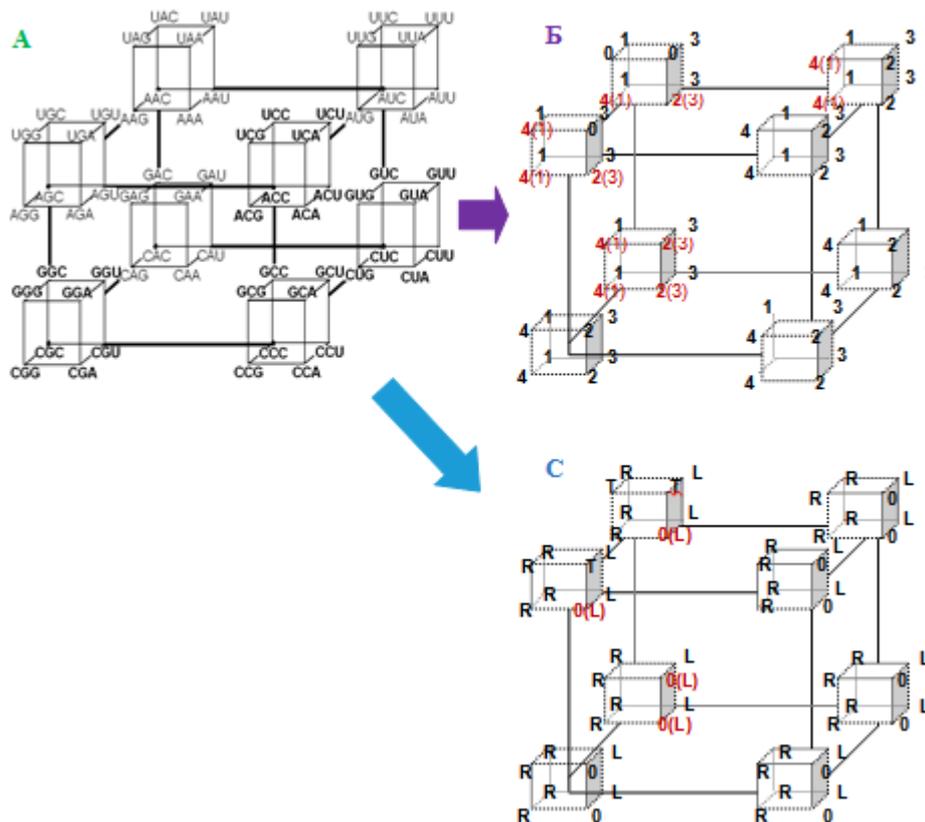


Рис. 3. Пространственная структура триплетного генетического кода для кодонов (А), значений композиционного кода (Б) и вариантов конфигурации пептидной связи (С). Соответствующие значения композиционного кода (1, 2, 3, 4) в редакции А.Ю. Куселева приведены в скобках рядом с уточнениями, принятыми в данной работе. 0 – значение композиционного кода для стоп-кодона (Б). R, 0, L – варианты конфигурации пептидных связей. Т – стоп-кодон (С). Авторские уточнения вариантов КПС выделены цветом.

Таким образом, все рассуждения и экспериментальные свидетельства того, что при денатурации α -спираль может переходить в β -тяж не конструктивны, поскольку изменить конфигурацию пептидной связи не разрушив её физически не возможно, а это уже деградация белковой молекулы. Не исключено, что при очень сильном растяжении спирали положение отдельных атомов в ней перестает соответствовать классическому и может отдалённо напоминать таковое в β -тяже. Однако, это вовсе не означает, что спираль – стабильная вторичная структура с минимумом свободной энергии преобразовалась в стабильную структуру β -тяжа. Мы можем только говорить о предельно напряжённой неустойчивой структуре очень растянутой спирали, которая при первом же удобном случае опять вернётся в своё нативное состояние с характерной пространственной структурой. Точно также никакими воздействиями невозможно фрагмент правой спирали в белке скрутить в левую спираль.

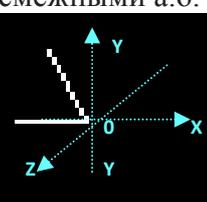
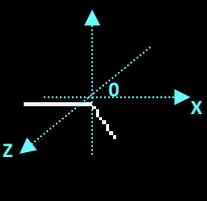
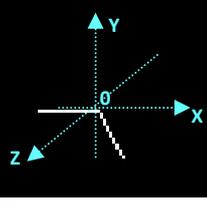
Использование таблицы генетического кода пространственной структуры белка даёт возможность по детерминирующей его нуклеотидной последовательности построить структурный шаблон последовательности конфигураций пептидных связей, а также декодировать положение фрагментов

вторичной структуры. Это значит, что можно построить структурный шаблон индивидуально для любого неизвестного белка лишь “прочитав” детерминирующую его нуклеотидную последовательность.

Визуализировать структурный шаблон, а точнее ход основной полипептидной цепи, можно в графическом редакторе Ggenedit.exe., в основе алгоритма которого лежит соответствие между кодонами и вариантами конфигураций пептидной связи (табл. 3), которые детерминируют углы в композиции смежных аминокислотных остатков (табл. 4). Данные углы являются следствием изгиба полипептидной цепи на границах плоскости пептидных связей, а не отклонения самой планарной пептидной связи от плоскости пептидных групп в результате изгиба. Кроме этого, наблюдается поворот по оси пептидной связи в композиции смежных аминокислотных остатков (табл. 1). Хотя в литературе и обсуждается роль полиморфизма неплоского строения пептидной группы в преимущественном формировании определенных вариантов вторичной структуры белка [13], используемые в нашем исследовании кольцевые модели и графический редактор Ggenedit.exe. опираются на каноническое представление о принадлежности атомов пептидных групп и самой пептидной связи между ними одной плоскости.

Таблица 4

Соответствие между конфигурацией пептидной связи и углами поворота в композиции смежных аминокислотных остатков в системе координат XYZ

КПС	Вариант вторичной структуры	Угол между смежными а.о.	Угол с (OX)	Угол с (OY)	Угол с (OZ)
R	Правая спираль		10	30	97
0	β -тяж		0	120	120
L	Левая спираль		0	30	-60

Примечание: КПС – конфигурация пептидной связи, а.о. – аминокислотный остаток.

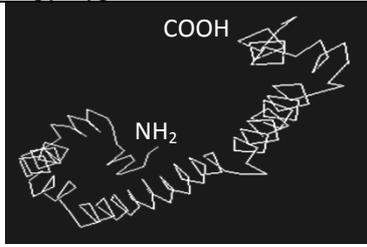
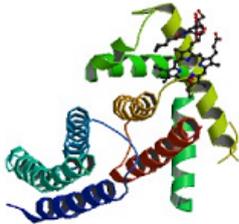
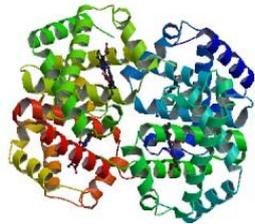
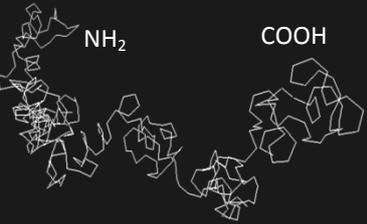
Примеры структурных шаблонов, построенных по нуклеотидным последовательностям, представлены в таблице 5 для α -субъединицы гемоглобина, основного актина и α -актина 1. В таблице также фигурируют основные модели этих белков из депозитария RCSBPDB (Protein Data Bank).

Конфигурация пептидной связи R- или L-вида является только **необходимым условием** формирования витка правой или левой спирали. **Достаточными условиями** существования спиральной структуры есть система водородных связей между атомами пептидных групп соседних витков, стабилизирующих вторичную структуру, наличие 3-

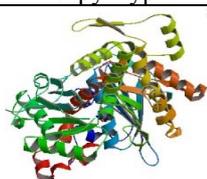
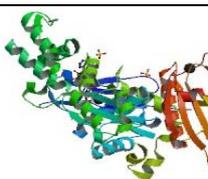
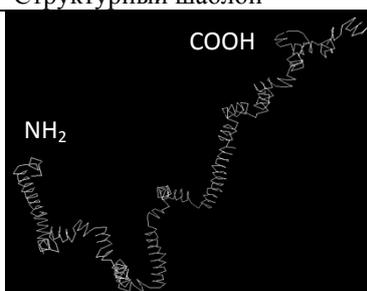
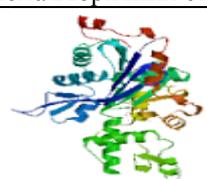
4 и более подряд аминокислотных остатков, кодируемых триплетами с G/C (U) в третьем положении кодона, а также свойства боковых цепей аминокислот. Наиболее полная разработка проблемы достаточного условия для правых α - и 3/10 спиралей представлена в работе Карасёва В.А., где автор привлекает модели графов связности/антисвязности для дифференциации аминокислотных остатков по характеристикам боковых цепей [4]. Там же представлена матрица топологического генетического кода (табл. 6), в которой ведущая роль в кодировании аминокислот связности/антисвязности отводится второму нуклеотиду триплета: для операторов антисвязности это пиримидины (C и U), операторов связности – пурины (A и G). Под топологическим кодом В.А. Карасёв понимает топологическую структуру генетического кода, являющуюся следствием объективной иерархии L-структуры шестимерного булева гиперкуба кодонов [4].

Таблица 5

Структурные шаблоны, построенные по нуклеотидным последовательностям, для α -субъединицы гемоглобина, основного актина и α -актина 1

α-Субъединица гемоглобина: UniProtKB/Swiss-ProtP69905 (HBA_HUMAN)					
Структурный шаблон		Нуклеотидная последовательность (423 нуклеотида)			
		<p>GUGCUGUCUCUCCGCGACAAGACCAACGCUCAAGGCCGCCUGGGUAAGGUCGGCGCGCAC GCUGGCGAGUAUGGU GCGGAGGCCCGGAGAGGAUGUCCUGUCCUCCCCACCACCAAG ACCUACUCCCCGACUUCGACCCUGAGCCACGGCUCUGCCACAGSUUAAGGGCCACGGCAAG AAGGUGGCCGACGCGCUGACCAACGCCGUGGGCGACGUGGACGACAUGCCCAACGCGCUG UCCGCCUGAGCGACCCUGCAGCGGACAAGCUUCGGGUGGACCCGGUCAACUUAAGCUC CUAAGCCACUGCCUGGUGACCCUGGCCGCCACCUCGCCCGAGUUCACCUCUUCGCG GUGCAGCCUCCUGGACAAGUCCUGCUUCUGUGAGCACCCGUGCUGACCUCUAAAUAC CGU</p>			
Код PDB-файла	Схема вторичных структур	Код PDB-файла	Схема вторичных структур		
1Y01 - димер		1A01 - тетрамер			
Основной актин: UniProtKB/Swiss-ProtP07830 (ACTS_HUMAN)					
Структурный шаблон		Нуклеотидная последовательность (1125 нуклеотидов)			
		<p>GACGGU GAAGAUGUCAA GCUUAGUUAUCGAUAACGGUUCUGGUAUGUGUAAA GCCGGU UUUGCUGGUGAUGAU GCC CCA CGUGCUGUUUUU CCAUCAUUGUUGGUCGU CCAAGA CAC ACUGGUGUU AUGGUCGUAUGGGU CAAAAAGACUCAUUGUA GGUGAU GAAGCC CAAUCA AAGAGA GGUUUCUUA ACCCUC AAAUAC CCAAUU GAA CACGGU AUCGUU ACCACUGGGAU GAUAUG GAAAAA AUCUGG CAUCAUACUUCUACA AU GAA CUC CGUGUU GCC CCAGAAGAA CACCCA GUUCUCUUA ACUGAA GCC CCAUUAAA CCA AAGGCCA CAGAGAAAA AUGACC CAAAUU AUGUUU GAA ACCUUAACACG CCA GCAUGUACGUCGCCAUU CAA GCCGUCUUA UCCUUA UAU GCCUCUGGUCGU ACCACC GGUUUGUU AUGGAU UCAGGUGAUGGU GUCUCC CACACU GUACCA AUCUAU GAA GGUUAUGCUU UACCA CAUGCCAUCUUA CGU UUA GAU UUA GCUGGCGGUAUCUUA ACC GAU UACAUGAUG AAA AUU UUAACU GAACGUGGUUACUCAUUC ACCACC ACUGCC GAAAGAGAA AUCGUC AGAGAU AUC AAA GAAAAAUUAGCCUACGUCGCC CUCGACUUU GAA GCU GAA AUG CAAACUGCUGCCUCAUCAUCAGCCUUGAAAAAUCAUAC GAAUU ACCAGACGGU CAA GUCAUCACCAUCGGUA AU GAA CGU UUC CGU UGC CCAGAA GCC CUCUUC CAACCAUCA UUC UUA GGU AUG GAA U CUGCUGGUAUCACC GAA ACCACCUAC AAC UCCAUCAUG AAA UGUGAUGUUGAUUCCGU AAA GAU UUA UAC GGUAACGUCGUCUUA UCC GGUGGU ACAACU AUGUUC CCA GGU AUUGCUGAUCGU AUGAAC AAAGAAUUAACUGCU UUA GCC CAUCAACCAUG AAA AUU AAAAUC AUUGCU CCACCAGAA CGU AAAUACUCUGUCUGG AUUGGU GGAUCU AUC UUA GCU UCACUCUCAACU UUC CAACAA AUGGGAUCUCC AAA GAA GAAUACGAC GAAUCA GCU CCAUCAUUGUUCAC AGAAAA UGUUUC</p>			
Код PDB-файлов	Схема вторичных структур	Фрагмент полипептида	Код PDB-файлов	Схема вторичных структур	Фрагмент полипептида

Способ моделирования пространственной структуры белка по детерминирующей ...

1C0F, 1C0G, 1DEJ		Димер полипептид 127 а.о. + полипептид 375 а.о.	1NLV, 1NM1, 1NMD, 3CI5, 3CIP		Димер полипептид 375 а.о. + полипептид 125 а.о.
α-Актин 1: UniProtKB/Swiss-ProtP68133 (ACTS_HUMAN)					
Структурный шаблон		Нуклеотидная последовательность (1125 нуклеотидов)			
		<p>GACGAAAGACGAGACCACCGCCUCGUGGCGACAAUGGCUCGCGCCUGGUGAAA GCCGGC UUCGCGGGGAUGACGCCCCUAGGGCCGUGUCCCGUCCAUCGUGGGCCGCCCGACAC CAGGGCGUCAUGGUCGUAUGGGUCAGAAAGAUUCCUACGUGGGCGACGAGCCUCAGAGC AAGAGAGGUAUCCUGACCCUGAAGUACCCUAUCGAGCACGGCAUCAACCAACUGGGAU GACAUGGAGAAGAUUGGCACCACACCUUCAACAAGAGCUUCGCGUGGCUCCGAGGAG CACCCACCCUGCUCACCGAGGCCCCCCUCAAUCCCAAGGCCAACCGCGAGAAGAUAGCC CAGAUCAUGUUUGAGACCUUCAACGUGCCCGCCAUAGUACGUGGCCAUCCAGGCCGUGCUG UCCCUCAACGCCUCCGGCAGGACCACCGGCAUCGUGCUGGACUCCGGCGACGGCGUACCC CACACGUGCCCAUUUAUGAGGGCUACGCGUGCCGACGCCAUCAUCGCCUGGACCUG GCGGGCCGCGAUUCACCGACUACCGUAUGAAGAUCUCACUGAGCGUGGCUACUCCUUC GUGACCACAGCUGAGCGCGAGAUCGUGCGCGACAUCAAGGAGAAGCUGGCUACGUGGCC CUGGACUUCGAGAACGAGAUUGGCGACGGCCGCCUCCUCCUCCUCCUGGAAAGAGCUAC GAGCUGCCAGACGGGCAGGUAUCACCAUCGGCAACGAGCGCUCCGCGUCCCGGAGCG CUCUUCAGCCUCCUUAUCGGUAUGGAGUCGCGCGGCAUUCACGAGACCACCUACAAAC AGCAUCAUGAAGUGUACAUCAUCAGGAAGGACCUGUAUGCCCAACAACGUCAUGUCG GGGGCACCCAGAUUACCCUGGGAUCGUGACCGCAUGCAGAAAGAGAUACCCGCGUG GCACCCAGCACAUGAAGAUAAGAUAUCGCCCCGCGGAGCGCAAUAUCGGUGUGG AUCGGCGGCCUCCAUCUGGCCUCGUGUCCACCUCCAGCAGAUUGGAUACCAAGCGC GAGUACGACGAGGCCCGCCUCCAUCGUGCCACCGCAAUUGCUUC</p>			
Код PDB-файлов		Схема вторичных структур		Фрагмент полипептида	
P68133				E6-K373	

Примечание: варианты конфигурации пептидной связи (R, U, L) выделены цветом.

В связи с **ретроградным направлением** боковых цепей аминокислотных остатков в витке **правой спирали** топологический код не распространяется на **первые** 2-3 аминокислотных остатка фрагмента полипептидной цепи. Например, пролин в начале такого фрагмента способствует вхождению полипептидной цепи в правую спираль, тогда как в любой другой позиции вызывает прерывание спирального фрагмента. Аналогичным образом, в связи с **сонаправленной росту спирали** ориентацией боковых цепей аминокислотных остатков в витке **левой спирали**, топологический код не распространяется на **последние** 2-3 аминокислотных остатка фрагмента полипептидной цепи.

Таблица 6

Топологический генетический код

Операторы антисвязности и их кодоны, второй нуклеотид которых -- пиримидин				Операторы связности и их кодоны, второй нуклеотид которых -- пурин			
C	а.о.	U	а.о.	A	а.о.	G	а.о.
CCC(UG)	Pro	UUC(U)	Phe	CAC(U)	His	CGC(UG)	Arg ^t
ACC(UG)	Thr	GUC(UG)	Val	CAG	Gln	GGC(UG)	Gly
UCC(UG)	Ser ^λ	CUC(UG)	Leu ^λ	AAC(U)	Asn	AGC(U)	Ser ^θ
GCC(UG)	Ala	UUG	Leu ^θ	AAG	Lys	AGG	Arg ^b
		AUC(U)	Ile	UAC(U)	Tyr	UGC(U)	Cys

		AUG	Met	GAC(U)	Asp	UGG	Trp
				GAG	Glu		

Примечание: Таблица является частью матрицы топологического кода для 64 кодонов аминокислот [4] и включает только кодоны аминокислотных остатков, определяющие R или L конфигурации пептидной связи в витках правой или левой спирали, соответственно.

В таблице 7 приведены аминокислотные последовательности изучаемых белков с учётом необходимого и достаточного условий для витков спирали, по сравнению со схемами вторичной структуры из базы данных Protein Data Bank. В первой строке таблицы для каждого белка приведены последовательности конфигураций пептидных связей (ПКПС) экстраполированные на соответствующую аминокислотную последовательность, что даёт возможность предсказать наличие и позицию фрагментов вторичной структуры с точностью до 1 аминокислотного остатка, учитывая необходимое условие вторичной структуры (НУВС). Во второй строке представлены схемы вторичной структуры с учётом достаточного условия для спиральных фрагментов (ДУСФ) полипептидной цепи. И, наконец, в третьей строке содержится информация о вторичной структуре (СВС PDB) белка из Protein Data Bank, полученная на основе экспериментальных данных спектроскопии ЯМР и рентгено-структурного анализа. В правом столбце приведены результаты сравнительного анализа схем вторичных структур, составленных нами путём декодирования нуклеотидных последовательностей, детерминирующих данные белки, и построенных по экспериментальным данным (файлы Protein Data Bank).

Таблица 7

Сравнительная характеристика схем вторичной структуры α -субъединицы гемоглобина, основного актина и α -актина 1

Параметры	α-субъединица гемоглобина (141 а.о.)	СИ ⁴ (%)
ПКПС НУ ¹	VLSPADKTNVKAAWGKVGAAHAGEYGAERALERMFLSFPTTKTYFPHFDLSHGSA QVKGHGKQKADALNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCL LVTLAAHLPAEFTPAVHASLKDFLASVSTVLTSKYR	ПКПС ↓ СВС PDB 91%
ДУ ² СФ	VLSPADKTNVKAAWGKVGAAHAGEYGAERALERMFLSFPTTKTYFPHFDLSHGSA QVKGHGKQKADALNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCL LVTLAAHLPAEFTPAVHASLKDFLASVSTVLTSKYR	ПКПС ↓ ДУСФ 100%
СВС PDB ³ P69905	VLSPADKTNVKAAWGKVGAAHAGEYGAERALERMFLSFPTTKTYFPHFDLSHGSA QVKGHGKQKADALNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCL LVTLAAHLPAEFTPAVHASLKDFLASVSTVLTSKYR	ДУСФ ↓ СВС PDB 58%
	основной актин (375 а.о.)	
ПКПС НУ ¹	DGEDVQALVIDNNGSGMCKAGFAGDDAPRAVFPISVGRPRHTGVMVGMGQKDS YVGDEAQSQRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLT EAPLNPKANREKMTQIMFETFNTPAMYVAIQAVLSLYASGRITGIVMDSGDGVS HTVPIYEGYALPHAALRLDLAGRDLTDYMMKILTERGYSFTTTAEREIVRDIKEK LAYVALDFAEMQTAASSALEKSYELPDGQVITIGNERFRCPEALFQPSFLGME SAGIHETTYNSIMKCDVDIRKDLYGNVVLGGTTMFPGIADRMNKELTALAPST MKIKIIPPERKYSVWIGGSILASLSTEQQMWISKEEYDESGPSIVHRKCF	ПКПС ↓ СВС PDB 48%

Способ моделирования пространственной структуры белка по детерминирующей ...

ДУ ² СФ	DGEDVQALVIDN G SGMCKA G FAGDDAPRAV F PS I VGR P RH T GV M VGMG Q KDS YVGDEA Q SKRGILTLKYPIEHGIVTNWDD M EKI W HHTFYNELRV A PEEHPVLLT E A PLNPKAN R EKMTQIMFETFNTPAMYVAIQAVLSLY A SGRT T G I VMD S GD G V S H T VPIY E GY A LPHAILRLDLAGRDLTDYMMKILTERGYSFTTT A E R EIVRD I KE K LAYVALDFAEMQTA A SS S A L E K S Y ELPDGQVITIGNERFR C PEAL F Q P S F L G M E S AGIHETTYNSIM K CDVDIRKDLYGN V LS G GT M FP G IAD R M N KELTAL A PS T MKIK I AP P ERK Y SVWIGGSILASLST F Q Q M W IS K EEY D ES G PS I V H R K CF	ПКПС ↓ ДУСФ 100%
CBC PDB ³ P6813 3	DG E DVQALVIDN G SGMCKA G FAGDDAPRAV F PS I VGR P RH T GV M VGMG Q KDS YVGDEA Q SKRGIL T TLKYPIEHGIVTNWDD M EKI W HHTFYNELRV A PEEHP V LL T E APLNPKAN R EKMTQIMFETFN T PAMYVAIQAVLSLY A SGR T T G I V M D S G D G V S H T V P I Y E GY A LPHAIL R LDLAGRDLTDYMMKILTERGYSFTTT A E R EIVRD I KE K LAYVALDFAEMQTA A SS S A L E K S Y ELPDGQ V ITIGNERFR C PEAL F Q P S F L G M E SAGIHETTYNSIM K CDVDIRKDLYGN V LS G GT M FP G IAD R M N KELTAL A PS T MKIK I AP P ERK Y SVWIGGSILASLST F Q Q M W IS K EEY D ES G PS I V H R K CF	ДУСФ ↓ CBC PDB 9%
α-актин 1 (375 а.о.)		
ПКПС НУ ¹	D E D ETALVCD N SGSL V KAGFAGDDAPRAV F PS I VGR P RH Q GV M VGMG Q KDS YVGDEA Q SK R RGILTLKYPIEHGI T NWDD M EKI W HHTFYNELRV A PEEHP T LL T E A PLNPKAN R EKMTQIMFETFN V PAMYVAIQAVLSLY A SGR T T G I V L D S G D G V T H N V P IY E GY A LPHAIMRLDLAGRDLTDY L M K ILTERGYS F V T A E R E IVRD I KE K L CYVALDFENEMATA A SS S S L E K S Y ELPDGQVITIGNERFR C P E T L F Q P S F I G E S A G I HETTYNSIM K CDIDIRKDL Y ANNVMSGGTT M Y P GIAD R M Q KE I TAL A PS T M K I K I I A PP E R K Y S VWIGGSILASLST F Q Q M W IT K Q E Y D E A G P S I V H R K CF	ПКПС ↓ CBC PDB 37%
ДУ ² СФ	DE D ETALVCD N SGSL V KAGFAGDDAPRAV F PS I VGR P RH Q GV M VGMG Q KDS YVGDEA Q SKRGILTLKYPIEHGI T NWDD M EKI W HHTFYNELRV A PEEHP T LL T E A P L N P K A N R E K M T Q I M F E T F N V P A M Y V A I Q A V L S L Y A S G R T T G I V L D S G D G V T H N V P IY E GY A LPHAIMRLDLAGRDLTDY L M K ILTERGYS F V T A E R E IVRD I KE K L CYVALDFENEMATA A SS S S L E K S Y ELPDGQVITIGNERFR C P E T L F Q P S F I G E S A G I HETTYNSIM K CDIDIRKDL Y ANNVMSGGTT M Y P GIAD R M Q KE I TAL A PS T M K I K I I A PP E RK Y SVWIGGSILASLST F Q Q M W IT K Q E Y D E A G P S I V H R K CF	ПКПС ↓ ДУСФ 100%
CBC PDB ³ P6813 3	DG E DVQALVIDN G SGMCKA G FAGDDAPRAV F PS I VGR P RH T GV M VGMG Q KDS YVGDEA Q SKRGIL T TLKYPIEHGIVTNWDD M EKI W HHTFYNELRV A PEEHP V LL T E APLNPKAN R EKMTQIMFETFN T PAMYVAIQAVLSLY A SGR T T G I V M D S G D G V S H T V P I Y E GY A LPHAIL R LDLAGRDLTDYMMKILTERGYSFTTT A E R EIVRD I KE K LAYVALDFAEMQTA A SS S A L E K S Y ELPDGQ V ITIGNERFR C PEAL F Q P S F L G M E SAGIHETTYNSIM K CDVDIRKDLYGN V LS G GT M FP G IAD R M N KELTAL A PS T MKIK I AP P ERK Y SVWIGGSILASLST F Q Q M W IS K EEY D ES G PS I V H R K CF	ДУСФ ↓ CBC PDB 42%

Примечание: ¹Последовательность конфигураций пептидных связей (ПКПС) -- необходимое условие вторичной структуры (НУ): цветом выделены аминокислотные остатки, присоединённые пептидными связями в **R**, **O** и **L** конфигурации.

²Достаточное условие спирального фрагмента (ДУСФ): цветом выделены аминокислотные остатки, образующие элементы вторичной структуры: правую, левую спираль и β-тяж.

³Схема вторичной структуры из соответствующего файла Protein Data Bank (CBCPDB).

⁴СИ – степень идентичности между (ПКПС и CBC PDB), (ПКПС и ДУСФ) и (ДУСФ и CBC PDB), выраженная в процентах.

Обращает на себя внимание то обстоятельство, что позиционирование спиральных фрагментов и β-тяжей для схем вторичной структуры белков из файлов Protein Data Bank не всегда соответствует варианту конфигурации пептидной связи. Если для глобулярных белков такие ошибки встречаются не часто (9% пула а.о. вторичной

структуры α -субъединицы гемоглобина), то для фибриллярных протеинов погрешность наблюдается в 52% -- основного актина (табл. 7). Сравнительный анализ последовательности конфигурации пептидных связей α -актина 1 со схемой вторичной структуры для гомологичного протеина (основной актин) показал 37% идентичности. Это ещё меньше, чем для основного актина (48% идентичности), что дискредитирует данную схему вторичной структуры из Protein Data Bank для этих гомологичных белков.

В подробном обсуждении нуждается факт установленный в данной работе: два гомологичных белка (основной актин и α -актин 1) со степенью идентичности аминокислотных последовательностей 91% характеризуются только 35% схожести последовательностей конфигураций пептидных связей (табл. 7). Это значит, что такие гомологи обладают различной пространственной структурой и ни о каком общем шаблоне для её моделирования речи идти не может. Структурные шаблоны для основного актина и α -актина 1 должны быть индивидуальными (табл. 5) с учётом специфики последовательности конфигураций пептидных связей (табл. 7).

Схемы вторичной структуры, декодированной по нуклеотидным последовательностям в данном исследовании с учётом достаточного условия спирального фрагмента (ДУСФ), ещё меньше совпадают с аналогичными схемами из Protein Data Bank (табл. 7 правый столбец второй строки для каждого белка): для основного актина – 9%, α -актина 1 – 41% и α -субъединице гемоглобина – 58%, соответственно. Это говорит о том, что местоположение вторичных структур экспериментальными методами позиционируется с ещё меньшей точностью, чем наличие определённой разновидности фрагментов вторичной структуры.

В стремлении исследователей экспериментальными методами с максимальным разрешением (1 – 2.5 Å) «увидеть» каждый атом в белке от их критического рассмотрения ускользнула методология анализа получаемых результатов и моделирования на их основе пространственной структуры белка. Например, известно, что основной актин имеет в своей структуре фрагменты левой спирали преимущественно в N-доме белка [1]. Мы также выявили наличие закодированных витков левой спирали в этом протеине (табл. 7), тогда как на схеме вторичной структуры из Protein Data Bank (файл P07830) показаны β -тяжи. Проблема не в данной конкретной неточности, а в установившемся представлении о том, что, достигнув разрешения в 1 Å при экспериментальном исследовании и построив тысячи схем для большого пула белков, мы получаем модели, адекватные пространственной структуре белка, закодированной в детерминирующей его нуклеотидной последовательности. Это, как иллюстрирует данное исследование, мягко говоря не совсем так: степень идентичности составляет чуть более 50% даже для хорошо изученного гемоглобина с «надёжно» установленной вторичной структурой и сотнями схем её укладки (табл. 5 и 7), что уж говорить о малоизученных белках, для которых известна лишь аминокислотная последовательность, конвертированная по нуклеотидной.

В заключение несколько слов о перспективе развития излагаемого подхода к моделированию пространственной структуры белка. По нуклеотидной последовательности можно декодировать и построить структурный шаблон белка. Но следует учитывать, что мы получаем не абсолютный вариант нативной конформации белка, а тот его прототип, который сходит с рибосомы, поскольку информация о посттрансляционной модификации белка, синтезированного *de novo*, не содержится в детерминирующей его нуклеотидной последовательности. Речь идёт о таких процессах, как гликозилирование, сульфатирование, сиамирование, ограниченный протеолиз, димеризация и др., моделирование которых предполагает другие алгоритмы программ.

Это не умаляет значимости описанного в данной работе подхода, а напротив, свидетельствует о том, что он должен служить фундаментом в моделировании белков по детерминирующей их нуклеотидной последовательности путём построения персонального структурного шаблона. В то время как, хорошо разработанная методология моделирования, опирающаяся на аминокислотные последовательности белков, должна работать надстройкой на этом фундаменте, позволяя оптимизировать пространственную структуру такого шаблона с учётом взаимодействий боковых цепей аминокислотных остатков.

ВЫВОДЫ

1. Поскольку информация о конформации белка в полном объёме содержится в его гене, моделировать пространственную структуру протеинов целесообразно по детерминирующим их нуклеотидным последовательностям.
2. Ротамерные варианты конфигурации пептидной связи (R, 0 и L) детерминированы в гене белка наряду с аминокислотными остатками и определяются третьим нуклеотидом каждого кодона.
3. Используя генетический код пространственной структуры белка можно декодировать индивидуальный структурный шаблон любого белка.
4. Надёжно и однозначно позиционировать фрагменты вторичной структуры белка (спирали и β -тяжи) можно опираясь на закономерности топологического генетического кода.
5. Два гомологичных белка основной актин и α -актин 1 со степенью идентичности аминокислотных последовательностей 91% имеют различные структурные шаблоны и схемы вторичной структуры (только 35% схожести последовательностей конфигураций пептидных связей), что обуславливает индивидуальную пространственную структуру каждого из них.

СПИСОК ЛИТЕРАТУРЫ

1. Финкельштейн А.В., Птицын О.Б. Физика белка. Курс лекций. М.: Книжный Дом Университет. 2002. 376 с.
2. Anfinsen C.V. Structural basis of ribonuclease activity. *FedProc.* 1957. V.16.№3. P.783–791.
3. Карасёв В.А. Генетический код: новые горизонты. СПб.: ТЕССА. 2003. 145 с.
4. Карасёв В.А., Лучинин В.В. Введение в конструирование бионических наносистем. М.: Физматлит. 2009. 463 с.
5. Кондратьев М.С., Кабанов А.В., Комаров В.М. Спиралеобразующие конформеры в структурной организации метиламидов N-ацетил- α -L-аминокислот. *Квантово-химический анализ. Биофизика.* Т.52. №3. С.401—408.
6. Crick F. Codon-anticodon pairing: the wobble hypothesis. *J. Mol. Biol.* 1966. V.19. P.548-555.
7. Соколик В.В. Принципы моделирования 3D-структуры белков-виновников возраст-зависимой конформационной патологии. *Клиническая геронтология.* 2009. Т.15. №8-9. С.119.
8. He Y., Chen Y., Alexander P. et al. NMR structures of two designed proteins with high sequence identity but different fold and function. *Proc. Natl. Acad. Sci. USA.* 2008. №105. P. 14412—14417.
9. Илиел Э. Основы стереохимии. М.: ИЦ «Академия». 2008. 464 с.
10. Кушелев А.Ю., Полищук С.Е., Неделько Е.В. и др. Построение масштабной модели структуры белка. *Актуальные проблемы современной науки.* 2002. №.2(5). С.236—243.
11. Волькенштейн М.В. Конфигурационная статистика полимерных цепей. М.-Л. 1959. 466 с.
12. Jimenez-Montano M.A., Mora-Basanez C.R., Poschel Th. The hypercube structure of the genetic code explains conservative and non-conservative amino acid substitutions in vivo and in vitro. *BioSystems.* 1996. V.39. P.117—125.
13. Самченко А. А., Кабанов А. В., Комаров В. М. Бистабильность неплоской формы пептидной группы в структуре дипептидов L-аминокислот. "Математика. Компьютер. Образование". Сб. трудов XIV международной конференции. Под общей редакцией Г.Ю. Ризниченко Ижевск: Научно-издательский центр "Регулярная и хаотическая динамика". 2007. Т. 2. С. 291—304.