



УДК 577.112.(012.62+0.87)

СТРУКТУРА КОЛЛАГЕНА С НОВОЙ СЕТКОЙ ВОДОРОДНЫХ СВЯЗЕЙ© 1999 г. Ю. В. Мильчевский[#], Б. С. Жоров*, Н. Г. Есипова, В. Г. Туманян*Институт молекулярной биологии им. В.А. Энгельгардта РАН, 117984, Москва, ул. Вавилова, 32;*** Институт эволюционной физиологии и биохимии им. И.М. Сеченова РАН,
194223, Санкт-Петербург, ул. Тореза, 44*

Поступила в редакцию 10.04.98 г. Принята к печати 25.01.99 г.

Методом молекулярной механики исследуется конформационная вариабельность коллагена. Для трипептидов, образующих первичную структуру белка, используется либо модель Рича–Крика с одной водородной связью на трипептид, либо модель с двумя водородными связями. В первом случае во втором положении трипептида находится остаток аминокислоты, а во втором случае – остаток аминокислоты. Исследуются конформации на границах, где за структурой с одной водородной связью следует структура с двумя водородными связями и наоборот. По существу рассмотрены все типы остова коллагена из часто встречающихся в этом белке остатков. Показано, что модель, построенная как сочетание фрагментов модели Рича–Крика и новой модели, стереохимически состоятельна и энергетически более выгодна. Предложенная структура коллагена согласуется с экспериментальными данными. Показано, что в случае коллагена первичная структура фибриллярного белка определяет третичную структуру.

Ключевые слова: белки; коллаген; пространственная структура; водородные связи; молекулярная механика.

ВВЕДЕНИЕ

Вопрос о связи первичной и третичной структуры в случае фибриллярных белков, с определенным типом вторичной структуры, на первый взгляд решается без особых затруднений. Между тем в случае важнейшего фибриллярного белка коллагена, представляющего собой комплекс из трех полипептидных цепей (далее для краткости тройной комплекс) в конформации типа полипролин II (третий фундаментальный тип вторичной структуры [1]), этот вопрос остается достаточно сложным и до сих пор открытым.

Важнейшим моментом в организации структуры тройного комплекса является число и конкретная схема завязывания межцепочечных сеток водородных связей типа остов–остов в тройном комплексе, составляющем макромолекулу коллагена. Завязывание соответствующих водородных связей и плотная упаковка цепей в комплексе достигается за счет искажения регулярных левозакрученных спиралей типа полипролин II. Только после соответствующего анализа – как расчетного, так и экспериментального, можно надеяться определить конкретные конформации, соответствующие данной последовательности аминокислотных остатков и сделать окончательные выводы о структуре этого фибриллярного белка.

В свое время Рич и Крик [2] и Коуан и соавт. [3] независимо предложили для коллагена структуру комплекса из трех полипептидных цепей, стабилизированного сеткой межцепочечных водородных связей с одной водородной связью на трипептид. После проведенных с тех пор рентгенографических исследований фибрилл и кристаллов полипептидов и олигопептидов, моделирующих коллаген [4, 5], можно считать установленным, что эта схема реализуется, когда в каждом из трипептидов, образующих полипептидные цепи коллагена, во втором положении находится аминокислота.

Однако две трети трипептидных фрагментов в цепи коллагена, содержат во втором положении не имино-, а аминокислоту. Казалось бы, для этих участков последовательности должна быть пригодна модель Рамачандрана–Картхи [6], предложенная одновременно с моделями Рича–Крика и Коуан с соавт. Эта модель обеспечивает образование двух водородных связей на трипептид на участках, где во втором положении находится не имино-, а аминокислота. Но для модели Рамачандрана–Картхи невозможно сосуществование с моделью Рича–Крика, поскольку для участков с аминокислотой ею предполагается завязывание иной, чем у Рича–Крика и Коуан с соавт., водородной связи. Кроме того, эта модель имеет стереохимические недостатки и, по-видимому, именно из-за них Рамачандран с сотрудниками отказались от своей старой схемы с двумя пептид-

[#] Автор для переписки (тел.: 137-99-57; e-mail: milch@imb.ac.ru).

пептидными водородными связями на триплет, а в более поздней работе предлагают обеспечение образования второй дополнительной связи с участием двух молекул воды [7].

Как показывают эксперименты по изотопному обмену, в коллагене действительно реализуется более одной (в среднем 1.7) водородной связи на трипептид [8]. Следовательно, необходимо построение модели нового типа, в которой дополнительные к тем, которые следуют из модели Рича–Крика, водородные связи образуются за счет соответствующих участков аминокислотной последовательности. Второй вариант Рамачандрана [7] не может удовлетворительно объяснить экспериментальные данные [8], так как мостики воды, особенно двойные, энтропийно не выгодны и, конечно, подобные схемы не могут объяснить наблюдаемую кинетику изотопного обмена.

Новый тип модели коллагена был построен в работах нашей группы [9–11]. Мы нашли конформации белка, в которых наличие amino-, а не иминокислоты во втором положении трипептида обеспечивает образование второй водородной связи на трипептид. Там, где во втором положении трипептида находится иминокислота, реализуется модель типа Рича–Крика. Далее совокупность водородных связей, соответствующую этой модели, мы будем называть первой сеткой, а дополнительные связи, соответствующие новой модели коллагена, – второй сеткой водородных связей; структуру молекулы коллагена, имеющую одну водородную связь на трипептид, будем называть односвязной, а структуру, имеющую две водородных связи на трипептид, – двусвязной.

Ключевым моментом нашей модели является использование карбонильного кислорода аминокислотного остатка во втором положении трипептида первой цепи в качестве универсального акцептора протонов NH-групп: от аминокислоты во втором положении трипептида второй цепи и от глицина в первом положении трипептида третьей цепи. Образование дополнительной водородной обеспечивается при этом поворотом пептидной группы между глицином и следующей аминокислотой (изменяются двугранные углы ψ глицина и ϕ следующей аминокислоты). Эта схема находится в соответствии с реальной картиной, наблюдаемой в белках. Как стало ясно теперь, идеальные водородные связи, типа один донор–один акцептор, вовсе не универсальны для белков. Часто встречается ситуация, когда на один акцептор приходится два и более доноров [12, 13]; возможны случаи, когда один протон принадлежит двум акцепторам.

Ранее был разработан метод конформационного анализа фибриллярных молекул коллагена [14, 15]. Этот метод, существенной чертой которого является возможность учета гибкости пяти-

членного цикла пролина, в отличие от других попыток расчета [16] позволяет получить структуру политрипептида $(GPP)_n$ с экспериментальными параметрами спирали [5]. С помощью этого метода и была создана новая модель коллагена и продемонстрирована ее стереохимическая состоятельность [9–11].

На гомополитрипептидных моделях было показано, что преимущество структуры с двойной сеткой водородных связей по сравнению со структурой типа Рича–Крика с простой сеткой водородных связей обусловлено не только наличием дополнительной водородной связи, но и более выгодной энергией ван-дер-ваальсовых взаимодействий и более удобным для гидратации расположением полярных групп [9–11].

В настоящей работе мы рассматриваем модели гетерогенных полипептидных последовательностей более близких к природной первичной структуре коллагена, чем гомогенные $[(GPO)_n]_3$ или $[(GAO)_n]_3$ (O – однобуквенный код гидроксипролина). Объектом расчетов являются последовательности, образованные чередованием трипептидов разного типа. Изучаются конформации участков сочленений, где вслед за фрагментом с двумя водородными связями следует участок с одной связью и наоборот. Таким образом, расчетная схема уже не ограничена рассмотрением строго регулярных конформаций и в принципе становится возможным расчет реальной структуры коллагена.

В табл. 1 показаны компоненты энергии для тройных полипептидных комплексов, исследованных в работе. Наименее благоприятная электростатика наблюдается в комплексе $[(GPO)_n]_3$, тогда как для $[(GAO)_n]_3$ она имеет более выгодный характер, и взаимодействие противоположно заряженных боковых групп способствует в большей степени стабилизирующей электростатике. Напротив, энергия ван-дер-ваальсовых взаимодействий в комплексе наиболее благоприятна для полимеров с двумя иминокислотами на триплет, главным образом за счет контакта остатков пролина или гидроксипролина, принадлежащих разным цепям комплекса $[(GPO)_n]_3$. Последние два столбца таблицы дают компоненты энергии для двух вариантов структуры: комплекса $[(GPP)_n]_3$, предложенного Шерагой (код 1BVE в Брукхевенском банке структур), и в работе [15]. Можно видеть, что для последней конформации все компоненты энергии (кроме энергии замыкания цикла, которая в работе Шераги и соотр. отсутствует, так как там цикл фиксирован) более благоприятны; спиральные параметры также ближе экспериментально наблюдаемым. Реализация двойных водородных связей в тройном комплексе $[(GAO)_n]_3$ приводит в этом случае к наиболее выгодной энергии водородных связей. В торсионную энергию почти весь неблаго-

Таблица 1. Сравнение компонент энергии для тройных комплексов в пересчете на один трипептид (ккал/моль)

Компонента энергии	Тройные комплексы						
	[[GAO] ₈] ₃	[[GPO] ₈] ₃	[[GAO] ₄ -(GPO) ₄] ₃	[[GPO] ₄ -(GAO) ₄] ₃	[[GPO] ₂ -(GEK)-(GPO) ₂] ₃	[[GPP] ₈] ₃	
						1BBE*	[15]
Электростат.	0.37	4.65	2.62	2.95	0.238	5.04	3.64
Ван-дер-Ваальсова	-14.23	-18.46	-16.81	-16.79	-17.17	-17.21	-17.32
Н-Связь	-2.12	-0.77	-1.47	-1.43	-0.82	-0.57	-0.61
Торсионная	3.90	7.24	5.39	5.83	6.75	7.33	7.13
Замык. цикл.	0.03	0.06	0.05	0.08	0.12	0.18	0.19
Конф. энерг.	-12.05	-7.27	-10.20	-9.34	-10.87	-5.20	6.96

* Код в банке пространственных структур PDB для структуры, предложенной Шерагой и др. (банк доступен в Интернете по адресу: <http://www.pdb.bnl.gov>).

приятный вклад вносят пирролидиновые кольца. Хотя суммарная конформационная энергия в последней строке табл. 1, естественно, не позволяет сравнивать между собой разные последовательности, она, как свидетельствуют наши предыдущие и настоящие расчеты, демонстрирует преимущество двусвязной структуры по сравнению с односвязной.

Энергия водородных связей в комплексе [[GAO]_n]₃ существенно выгоднее, чем для [[GPO]_n]₃. Это обусловлено образованием между его цепями дополнительной сетки водородных связей. Различия в торсионной энергии объясняются тем, что в последовательности GPO имеется два пирролидиновых кольца на трипептид, а в GAO – одно. Для комбинированных структур после оптимизации все компоненты энергии оказываются весьма близкими, причем эти компоненты принимают промежуточные значения между соответствующими компонентами для комплексов [[GAO]_n]₃ и [[GPO]_n]₃.

В качестве примера в табл. 2 показаны конформационные углы в основных цепях средней части тройного комплекса [[GPO]_n]₃, где происходит переход от односвязной структуры к двусвязной. В соответствии с результатами, полученными на регулярных спиральных, все двугранные углы в обеих структурах практически одинаковы за исключением углов ψ и φ, задающих ориентацию пептидной группы между глицином и следующим остатком. Как видно из табл. 2, эти углы при переходе от одно- к двусвязным структурам меняются приблизительно на 90°, что вызвано поворотом пептидной группы, которая обеспечивает образование дополнительной водородной связи. Как и следовало ожидать, наибольшая нерегулярность возникает на стыке двух типов структур, однако степень этой нерегулярности незначительна.

Спиральные параметры оптимизированных тройных комплексов приведены в табл. 3. Средние

значения для проекции остатка находятся в хорошем согласии с рентгенографическими данными по коллагену и моделирующим его полипептидам [4, 5], хотя наблюдается тенденция к небольшому занижению значений проекции остатка по сравнению с экспериментом. Среднее значение для угла поворота на асимметричную единицу характеризуется некоторым завышением по сравнению с экспериментальным значением (52°), определенным в рентгеноструктурных исследованиях [[GPP]₁₀]₃ [5]. Для коллагена число остатков в обороте спирали, определенное по фибриллярной рентгенограмме, варьировало от 3.33 для спирали 10/3 до 3.5 для спирали 7/2. Для получения более точных данных требуется расчет координат атомов протяженной молекулы в огромной ячейке, параметры которой остаются дискуссионными. Рассмотрение данных табл. 3 приводит к выводу о том, что имеет место небольшое сплющивание и скручивание комплекса. Результаты могут быть обусловлены как некоторой неточностью в использованных энергетических параметрах, так и концевыми эффектами. К вопросу о сплющивании комплекса следует отметить, что в ряде синтетических полимеров, изоморфных коллагену, согласно рентгенографическим данным [4], проекция аминокислотного остатка на ось тройной спирали варьирует, достигая в ряде образцов величины, соответствующей 2.75 Å.

Комплексы [[GAO]_n]₃ и [[GPO]_n]₃ являются регулярными структурами и все три цепи, если пренебречь концевыми эффектами, одинаковы, а их геометрические параметры достаточно точно воспроизводятся при переходе от одного триплета к другому. Для всех составных тройных комплексов таких как [[GPO]₄-(GAO)₄]₃ и [[GAO]₄-(GPO)₄]₃ геометрические параметры принципиально различаются для различных участков молекулы. Для них в табл. 3 показаны спиральные параметры для первой и второй частей молекулы.

Таблица 2. Значения углов ϕ , ψ , ω для четырех центральных трипептидов тройного комплекса $[(GPO)_4-(GAO)_4]_3$

Остаток (номер трипептида)	Угол	1-я цепь	2-я цепь	3-я цепь	Остаток (номер трипептида)	Угол	1-я цепь	2-я цепь	3-я цепь
Gly7 (3)	ϕ	-70.0	-69.8	-70.1	Gly13 (5)	ϕ	-68.6	-66.2	-72.9
	ψ	176.9	177.3	176.3		ψ	-105.8	-87.8	-87.1
	ω	179.5	179.6	179.6		ω	-173.8	-172.3	-166.9
Pro8 (3)	ϕ	-75.0	-75.1	-74.5	Ala14 (5)	ϕ	-148.6	-160.7	-155.4
	ψ	162.7	162.2	161.1		ψ	154.7	158.3	159.0
	ω	-178.4	-178.1	-177.0		ω	161.0	158.2	161.5
Hyp9 (3)	ϕ	-61.1	-60.6	-61.4	Hyp15 (5)	ϕ	-52.9	-52.5	-51.1
	ψ	149.6	149.4	150.2		ψ	131.8	135.5	137.1
	ω	-178.0	-176.0	-179.1		ω	-172.6	-171.3	-173.4
Gly10 (4)	ϕ	-69.1	-70.0	-70.0	Gly16 (6)	ϕ	-71.9	-72.0	-74.1
	ψ	176.2	174.6	-178.3		ψ	-88.8	-85.9	-84.8
	ω	-179.2	-179.5	176.6		ω	-171.0	-171.4	-172.5
Pro11 (4)	ϕ	-75.0	-75.6	-80.1	Ala17 (6)	ϕ	-151.2	-153.9	-155.6
	ψ	161.0	162.6	165.4		ψ	159.0	158.7	157.4
	ω	-177.7	179.1	169.4		ω	162.6	162.2	163.4
Hyp12 (4)	ϕ	-61.1	-62.2	-63.3	Hyp18 (6)	ϕ	-51.7	-51.8	-53.1
	ψ	149.1	145.6	140.7		ψ	137.0	135.9	135.0
	ω	-172.1	-175.8	-171.4		ω	-173.1	-172.0	-175.1

Таблица 3. Сравнение средних спиральных параметров* для тройных комплексов

Тройной комплекс	t , град	h , нм	
$[(GPO)_8]_3$	57.99	0.28	
$[(GAO)_8]_3$	63.05	0.28	
$[(GAO)_4-(GPO)_4]_3^{**}$	Вся молекула	61.17	0.28
	Первая половина молекулы	72.46	0.27
	Вторая половина молекулы	54.38	0.28
$[(GPO)_4-(GAO)_4]_3^{**}$	Вся молекула	63.12	0.28
	Первая половина молекулы	59.56	0.28
	Вторая половина молекулы	66.42	0.28
$[(GPO)_2-(GEK)-(GPO)_2]_3$	57.48	0.24	

* Спиральные параметры: угол поворота на асимметричную единицу (t) и проекция остатка на ось спирали (h). Эти параметры однозначно определяют спиральное преобразование, с помощью которого можно определить переход между асимметричными единицами спирали (здесь это трипептид).

** Первая половина молекулы – это часть тройного комплекса от N -конца до середины каждой из трех цепей, вторая половина – это оставшаяся часть молекулы.

Сетки водородных связей в исследованных структурах подробно показаны в табл. 4–6. Для составного комплекса $[(GAO)_4-(GPO)_4]_3$ (табл. 4) вторая сетка водородных связей реализуется везде, где это позволяет первичная структура. Характерно, что на стыке последовательностей $(GAO)_4$ и $(GPO)_4$ не возникает трудностей с образованием второй сетки. Связи второй сетки заметно длиннее, чем для первой, на участках, где

реализуются обе сетки. Там, где по модели Рича-Крика возможна только одна сетка, водородные связи несколько длиннее, чем для этой же сетки на участках двусвязной структуры. Вполне аналогичная картина наблюдается на стыке инвертированных последовательностей $(GPO)_4$ и $(GAO)_4$ (табл. 5).

Результаты, полученные для тройного комплекса $[(GAO)_8]_3$, хорошо воспроизводят, за ис-

Таблица 4. Атомы кислорода и водорода, участвующие в образовании водородных связей в тройном комплексе $[(\text{GAO})_4-(\text{GPO})_4]_3$

Кислород		Водород		Сетка*	Длина связи, Å
Остаток	Номер цепи	Остаток	Номер цепи		
Ala2	1	Gly1	3	F	1.92
	1	Ala2	2	S	2.01
	2	Gly4	1	F	1.91
	2	Ala2	3	S	2.11
	3	Gly4	2	F	1.96
	3	Ala5	1	S	2.10
Ala5	1	Gly4	3	F	1.95
	1	Ala5	2	S	2.10
	2	Gly7	1	F	1.96
	2	Ala5	3	S	2.08
	3	Gly7	2	F	1.89
	3	Ala8	1	S	2.13
Ala8	1	Gly7	3	F	1.90
	1	Ala8	2	S	2.17
	2	Gly10	1	F	1.92
	2	Ala8	3	S	2.24
	3	Gly10	2	F	1.93
	3	Ala11	1	S	2.15
Ala11	1	Gly10	3	F	1.95
	1	Ala11	2	S	2.15
	2	Gly13	1	F	2.32
	2	Ala11	3	S	2.23
	3	Gly13	2	F	2.14
	3	Ala11	1	S	2.15
Pro14	1	Gly13	3	F	2.31
	2	Gly16	1	F	2.21
	3	Gly16	2	F	2.18
Pro17	1	Gly16	3	F	2.10
	2	Gly19	1	F	2.08
	3	Gly19	2	F	2.11
Pro20	1	Gly19	3	F	2.17
	2	Gly22	1	F	2.01
	3	Gly22	2	F	2.25
Pro23	1	Gly22	3	F	2.37

* F и S – первая и вторая сетка соответственно.

Таблица 5. Атомы кислорода и водорода, участвующие в образовании водородных связей в тройном комплексе $[(\text{GPO})_4-(\text{GAO})_4]_3$

Кислород		Водород		Сетка*	Длина связи, Å
Остаток	Номер цепи	Остаток	Номер цепи		
Pro2	1	Gly1	3	F	1.97
	2	Gly4	1	F	2.17
	3	Gly4	2	F	2.19
Pro5	1	Gly4	3	F	2.09
	2	Gly7	1	F	2.16
	3	Gly7	2	F	2.13
Pro8	1	Gly7	3	F	2.08
	2	Gly10	1	F	2.14
	3	Gly10	2	F	2.05
Pro11	1	Gly10	3	F	2.04
	2	Gly13	1	F	2.12
	3	Gly13	2	F	2.15
Ala14	3	Ala14	1	S	2.56
	1	Gly13	3	F	2.05
	1	Ala14	2	S	2.18
	2	Gly16	1	F	2.00
	2	Ala14	3	S	2.29
	3	Gly16	2	F	1.95
Ala17	3	Ala17	1	S	2.19
	1	Gly16	3	F	1.94
	1	Ala17	2	S	2.09
	2	Gly19	1	F	1.99
	2	Ala17	3	S	2.12
	3	Gly19	2	F	1.99
Ala20	3	Ala20	1	S	2.29
	1	Gly19	3	F	2.06
	1	Ala20	2	S	2.24
	2	Gly22	1	F	2.18
	2	Ala20	3	S	2.06
	3	Gly22	2	F	2.02
Ala23	3	Ala23	1	S	2.37
	1	Gly22	3	F	1.91
	1	Ala23	2	S	2.42
	2	Ala23	3	S	2.12

* См. табл. 4.

ключением концевых эффектов, результаты расчетов в предположении полной регулярности структуры [15]. Для тройного комплекса $[(\text{GPO})_8]_3$ возможна только первая сетка водородных связей, при этом водородные связи имеют несколько большую длину, чем водородные связи первой сетки комплекса $[(\text{GAO})_8]_3$. Длина водородных связей

в $[(\text{GPO})_8]_3$ такая же как на соответствующих участках составных молекул $[(\text{GPO})_4-(\text{GAO})_4]_3$ и $[(\text{GAO})_4-(\text{GPO})_4]_3$.

Образование водородных связей для тройного комплекса $[(\text{GPO})_2-(\text{GPK})-(\text{GPO})_2]_3$, имеющего последовательность, сходную с представленной в работе [17], проиллюстрировано в табл. 6. Сохра-

няются все водородные связи первой сетки и при этом, что весьма важно, образуются все возможные три связи, относящиеся ко второй сетке.

В табл. 7 показаны расстояния между атомами, которые, согласно работе [17], должны образовывать ионные связи в тройном комплексе $[(\text{GPO})_2-(\text{GEK})-(\text{GPO})_2]_3$, причем третья из указанных связей, по мнению авторов, может образоваться только при условии благоприятной локальной деформации. Как показывают наши расчеты, образуются только первые две ионные связи.

Проведенное исследование демонстрирует состоятельность схемы с завязыванием двух сеток водородных связей на соответствующих участках последовательности для ситуаций, максимально приближенных к существующим в природном белке. После оптимизации тройных комплексов $[(\text{GAO})_4-(\text{GPO})_4]_3$ (рис. 1) и $[(\text{GPO})_4-(\text{GAO})_4]_3$ (рис. 2) образуются все водородные связи, которые возможны согласно двусвязной модели. То есть в двух принципиально возможных случаях стыковки на участке $(\text{GAO})_4$ образуются две водородные связи на трипептид, на участке $(\text{GPO})_4$ – одна, а на границе тоже образуется теоретически возможная водородная связь. Конформационная энергия и энергия водородных связей для этих составных комплексов практически одинакова, при этом энергия водородных связей имеет значение, промежуточное в сравнении с энергиями для структур с одной и двумя сетками водородных связей. Длины водородных связей и спиральные параметры для этих составных комплексов различаются для каждой из частей и при этом на каждом участке близки к значениям, полученным для соответствующих гомогенных комплексов. Наличие всех водородных связей, допускаемых первичной структурой, свидетельствует в пользу того, что структуры с одной и двумя сетками водородных связей могут свободно чередоваться без образования существенных локальных деформаций на границе.

Для тройного комплекса $[(\text{GPO})_2-(\text{GEK})-(\text{GPO})_2]_3$, подобному более длинному $[(\text{GPO})_4-(\text{GEK})-(\text{GPO})_5]_3$ [17], получены (рис. 3) ионные пары, совпадающие с парами, предложенными авторами цитируемой работы. Различие в конформационной энергии между тройным комплексом $[(\text{GPO})_2-(\text{GEK})-(\text{GPO})_2]_3$ и тройным комплексом $[(\text{GPO})_5]_3$, по внутренним координатам которого получено начальное приближение для $[(\text{GPO})_2-(\text{GEK})-(\text{GPO})_2]_3$, сводится в основном к различию в электростатической составляющей энергии. Остальные компоненты энергии, так же как и геометрические параметры, близки к тем, что были получены для $[(\text{GPO})_8]_3$. Это представляется естественным ввиду схожести первичной структуры и выбора начального приближения для координат перед оптимизацией энергии комплекса $[(\text{GPO})_2-(\text{GEK})-(\text{GPO})_2]_3$. Примечательно то, что для этого

Таблица 6. Атомы кислорода и водорода, участвующие в образовании водородных связей в тройном комплексе $[(\text{GPO})_2-(\text{GEK})-(\text{GPO})_2]_3$

Кислород		Водород		Сетка*	Длина связи, Å
Остаток	Номер цепи	Остаток	Номер цепи		
Pro2	1	Gly1	3	F	2.02
	2	Gly4	1	F	2.07
	3	Gly4	2	F	2.18
Pro5	1	Gly4	3	F	2.20
	2	Gly7	1	F	2.17
	3	Gly7	2	F	2.22
	3	Glu8	1	S	2.20
Glu8	1	Gly7	3	F	2.13
	1	Glu8	2	S	2.16
	2	Gly10	1	F	2.15
	2	Glu8	3	S	2.18
	3	Gly10	2	F	2.02
Pro11	1	Gly10	3	F	1.98
	2	Gly13	1	F	2.18
	3	Gly13	2	F	2.13
Pro14	1	Gly13	3	F	2.00

* См. табл. 4.

Таблица 7. Расстояния между атомами, способными образовать ионные пары в тройном комплексе $[(\text{GPO})_2-(\text{GEK})-(\text{GPO})_2]_3$

Атом NZ остатка	Атом OE остатка	Расстояние, Å
Lys9(1)	Glu8(2)	5.23
Lys9(2)	Glu8(3)	6.50
Lys9(3)	Glu8(1)	17.53

Для атомов азота и кислорода используются обозначения, принятые в банке пространственных структур белков PDB (банк и соглашения об обозначениях доступны в Интернете по адресу: <http://www.pdb.bnl.gov>).

комплекса все же образовались все три в принципе возможные водородные связи, принадлежащие второй сетке. Весьма показательно, что двусвязная структура для тройного комплекса $[(\text{GPO})_2-(\text{GEK})-(\text{GPO})_2]_3$ образовалась в ходе оптимизации энергии спонтанно. Между тем, до оптимизации конформация этого комплекса соответствовала односвязной модели. Наблюдаемый конформационный переход служит дополнительным доводом в пользу двусвязной модели.

Двусвязная модель дает преимущество в общей суммарной энергии не только за счет выигрыша в энергии водородных связей, но и за счет энергии электростатических взаимодействий.

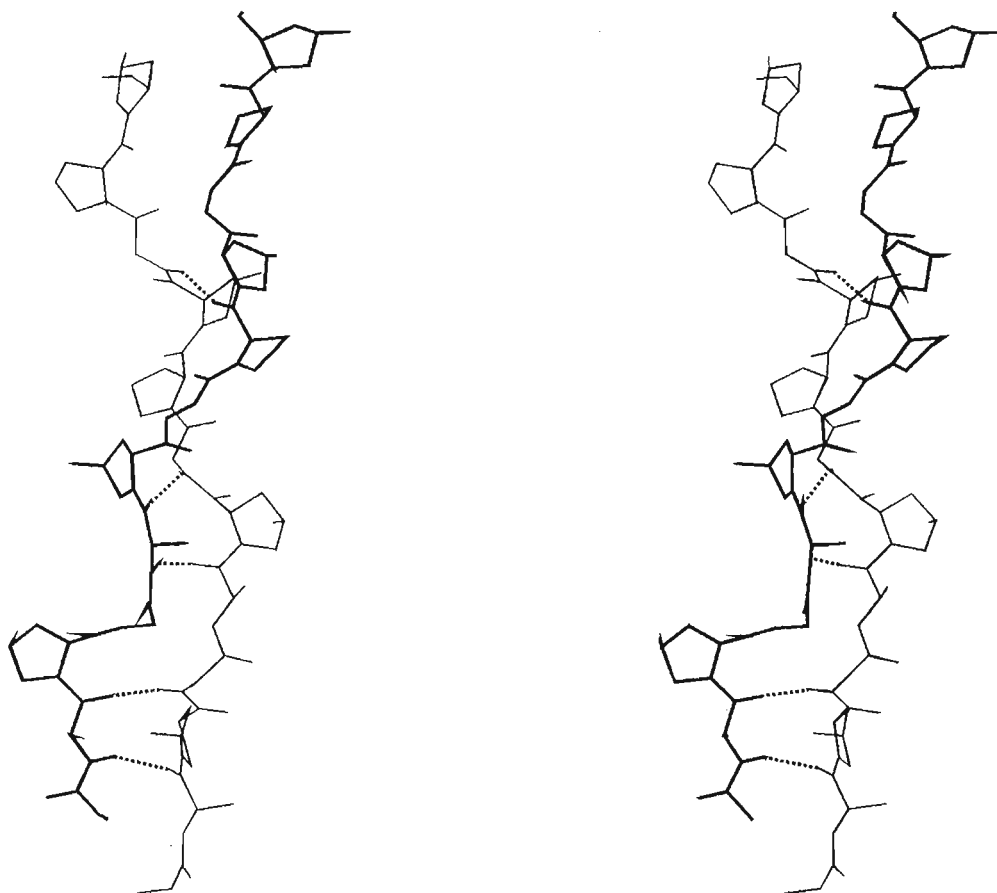


Рис. 1. Переход в центре молекулы от сетки с двумя водородными связями на трипептид к сетке с одной связью для тройного комплекса $[(\text{GAO})_4-(\text{GPO})_4]_3$. Стереорисунки первой и второй (выделена толстой линией) цепи тройного комплекса. Направление цепей снизу вверх. Атомы водорода, кроме участвующих в образовании водородных связей (тонкий пунктир), не показаны.

Двусвязная модель допускает свободное чередование трипептидов, содержащих имино- и аминокислоту во втором положении, без существенного ухудшения энергетических характеристик. Качественная оценка энергии гидратации, в свою очередь, свидетельствует в пользу двусвязной модели, поскольку в этом случае большее число гидрофильных СО-групп имеет наружное расположение и оказывается доступным для гидратации.

Собственно говоря, поворот пептидной группы, необходимый, для формирования второй сетки водородных связей, приводит не только к большей экспонированности в растворитель карбонильного кислорода, но и к меньшей экспонированности NH-группы. Если первое дает выигрыш в свободной энергии гидратации, так как СО-группа весьма гидрофильна, то закрытие для растворителя NH-группы несущественно, поскольку энергия гидратации этой группы [18] близка к нулю. Макромолекула в целом, учитывая аминокислотную последовательность, состоит на 2/3 из двусвязной структуры и на 1/3 из од-

носвязной, типа Рича и Крика. Таким образом, не только формируется конформация, отвечающая минимуму свободной энергии, но и достигается согласие с экспериментальными данными по числу сеток водородных связей в коллагене [8].

Таким образом, проведенное исследование показывает, как конкретно деформируется вторичная структура левой спирали типа полипролин II в коллагене с тем, чтобы произошло насыщение межцепочечных остов-остовных водородных связей. Это продемонстрировано для составных последовательностей, в которых чередуются типы сеток водородных связей, в том числе для минимального трипептидного фрагмента, способного образовать две сетки водородных связей, в окружении участков, способных сформировать только одну сетку. Предложенный подход, основанный на методах молекулярной механики, выявляет для исследуемого белка зависимость структуры от последовательности. Следовательно, в отличие от глобулярных белков для фибриллярного белка коллагена задачу связи первич-

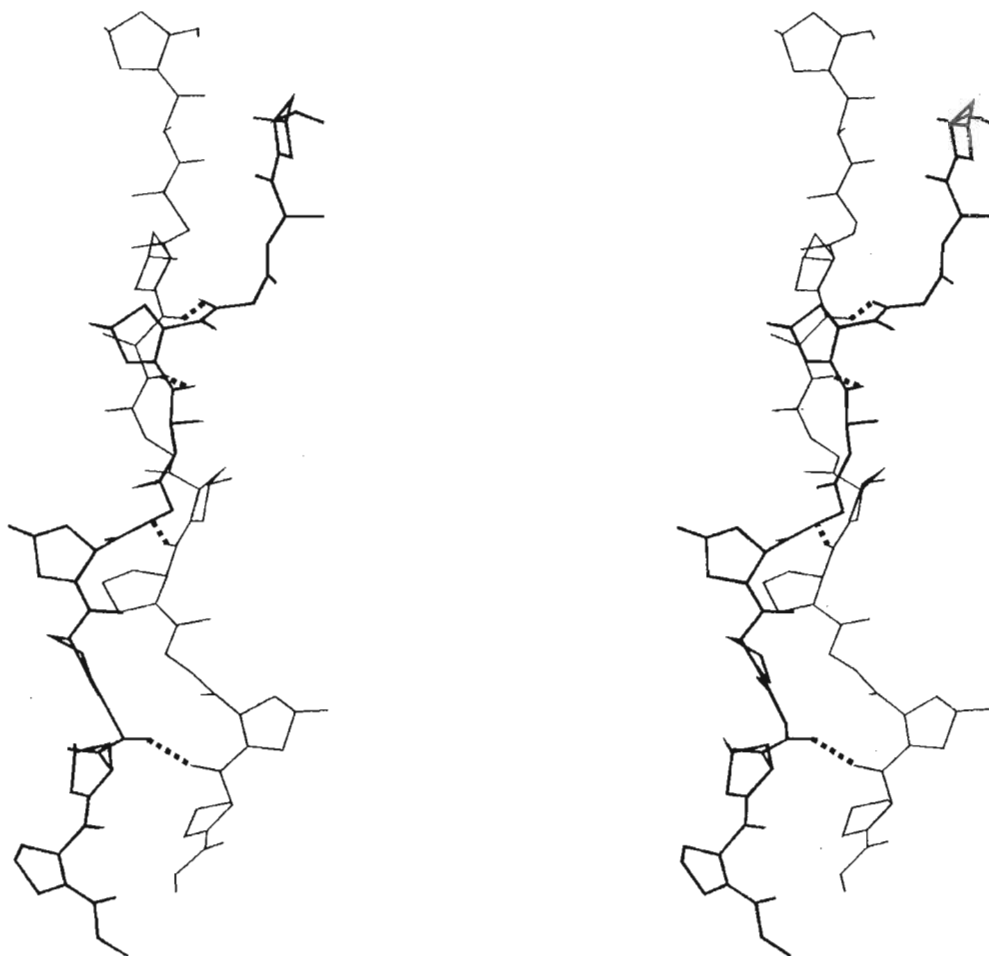


Рис. 2. Переход в центре молекулы от сетки с двумя водородными связями на трипептид к сетке с одной связью для тройного комплекса $[(\text{GPO})_4-(\text{GAO})_4]_3$. Стереорепрезентация первой (выделена толстой линией) и второй цепи тройного комплекса. Направление цепей снизу вверх. Атомы водорода, кроме участвующих в образовании водородных связей (тонкий пунктир), не показаны.

ной и третичной структуры можно считать решенной и нет никаких принципиальных препятствий для получения расчетным путем структуры фибриллярной части макромолекул коллагенов разных типов. Знание точной структуры может быть основой для объяснения многочисленных важнейших биологических функций коллагенов.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Молекулярно-механические расчеты проводили с использованием программы ZMM [19, 20], хорошо зарекомендовавшей себя в конформационных расчетах биологических молекул. Программа оперирует внутренними переменными, включая двугранные и валентные углы. При расчетах использовалась параметризация Шераги [21]. Программа ZMM успешно применялась для различных молекулярно-механических расчетов [22, 23]. Для замыкания кольца пролина или гидроксипролина

использовался стягивающий потенциал, причем валентные углы в пирролидиновом кольце полагались переменными (как описано в работе [20]). В качестве исходного приближения для оптимизации были взяты конформации, соответствующие минимуму потенциальной энергии регулярных политрипептидов $[(\text{GAO})_n]_3$ и $[(\text{GPO})_n]_3$ [9, 11, 15]. Получению этих конформаций предшествовало исследование конформационного пространства исходя из критерия энергии [24] и наличия водородных связей [15]. Эти оптимальные структуры комбинировали друг с другом с тем, чтобы получить стартовые точки для расчета сложных последовательностей. Детали сборки конформаций для начальных приближений описаны ниже. Поскольку структуры после оптимизации энергии становятся в той или иной степени нерегулярными, необходимо вычисление средних параметров спирали.

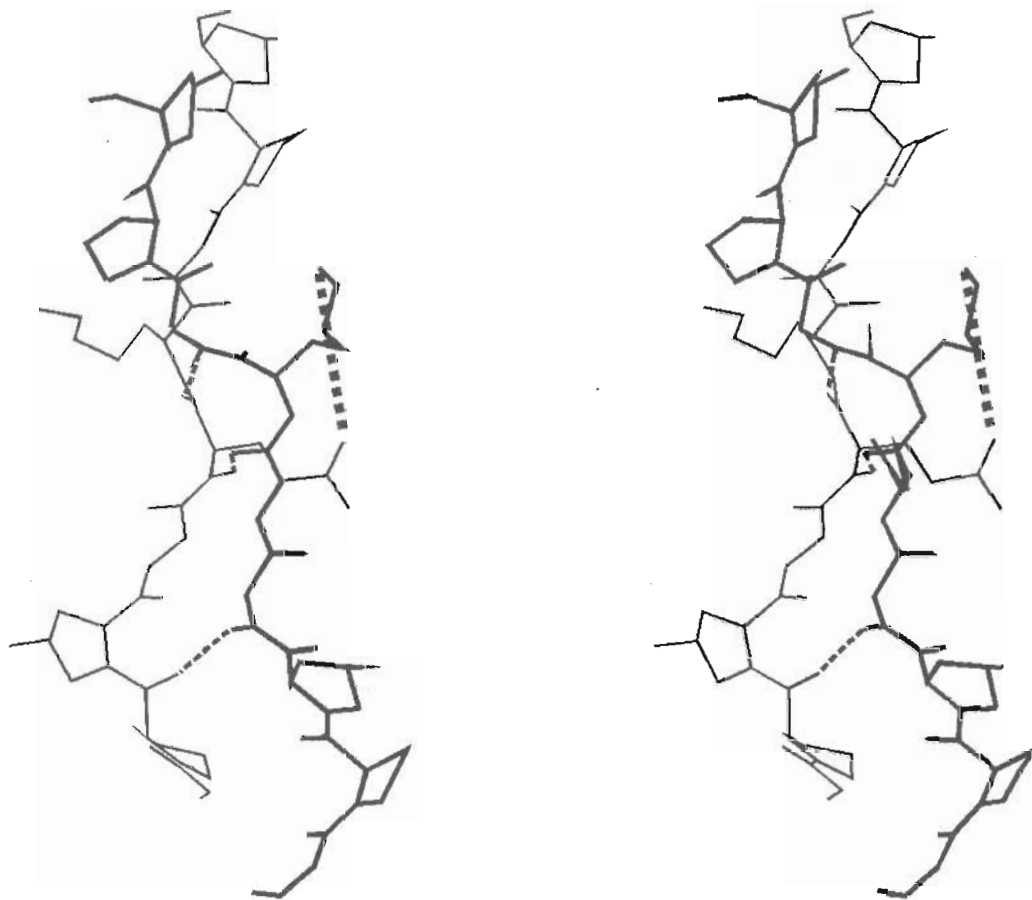


Рис. 3. Центральная часть тройного комплекса $[(\text{GPO})_2-(\text{GEK})-(\text{GPO})_2]_3$. Направление цепей снизу вверх. Показаны первая (выделена толстой линией) и вторая цепи тройного комплекса. Атомы водорода, кроме участвующих в образовании водородных связей (тонкий пунктир), не показаны. Ионная связь обозначена толстым пунктиром.

В работах [9, 11, 15] даны координаты трипептида, представляющего из себя асимметричную единицу одной из цепей тройного комплекса. Для удлинения этой цепи и генерации координат остальных двух цепей применяли спиральные преобразования симметрии к асимметричной единице с известными из [9, 11, 15] параметрами спирали. Так из координат асимметричных единиц $(\text{GAO})_1$ и $(\text{GPO})_1$ были получены исходные координаты тройного комплекса $[(\text{GAO})_n]_3$ и $[(\text{GPO})_n]_3$. Исходные координаты составных тройных комплексов $[(\text{GAO})_4-(\text{GPO})_4]_3$ и $[(\text{GPO})_4-(\text{GAO})_4]_3$ получали путем комбинирования регулярных структур $(\text{GAO})_1$ и $(\text{GPO})_1$ из работы [9]. Минимизацию энергии проводили с помощью программы ZMM.

Тройной комплекс $[(\text{GPO})_2-(\text{GEK})-(\text{GPO})_2]_3$ конструировался по двугранным углам главной цепи $(\text{GPO})_8$. Каждая цепь короче приведенной в статье [17] на три звена GPO на N-конце и на два звена GPO на C-конце.

Использование несколько укороченных последовательностей не должно существенно иска-

жать структуру молекулы, поскольку предварительные расчеты для $[(\text{GAO})_6]_3$, $[(\text{GAO})_8]_3$, $[(\text{GAO})_{15}]_3$ и других комплексов показали, что концевыми эффектами цепей можно пренебречь и структура центральной части тройных комплексов при изменении длины каждой цепи изменится несущественно.

Во всех расчетах на C-конце была карбонильная концевая группа, а на N-конце аминогруппа. Диэлектрическая постоянная во всех расчетах принималась равной 4.0. При минимизации энергии структур на предварительных этапах оптимизации вводились ограничения на максимальную и минимальную длину водородных связей. На конечных этапах оптимизации эти ограничения снимались.

Поскольку комплексы отклоняются от строгой спиральной симметрии, для вычисления спиральных параметров потребовалась специальная процедура. Спиральные параметры вычислялись по координатам атомов главной цепи для каждой из трех цепей отдельно. При вычислении спиральных параметров внутри одной цепи рассмат-

ривалась асимметричная единица N-C^α-C'-N-C^α-C'-N-C^α-C', соответствующая одному трипептиду GXY. При рассмотрении следующей такой асимметричной единицы вычисляли спиральные параметры, соответствующие переходу между ними. Затем брали асимметричную единицу C^α-C'-N-C^α-C'-N-C^α-C'-N и снова вычисляли соответствующие спиральные параметры. Таким образом, продвигались до конца цепи, после чего получали среднее значение спиральных параметров каждой отдельной цепи и, наконец, усредняли их по трем цепям.

Работа частично поддержана грантом РФФИ "Научные школы" № 96-15-98093.

СПИСОК ЛИТЕРАТУРЫ

1. Adzhubei A., Eisenmenger F., Zinke M., Brodzinski S., Tumanyan V., Esipova N. // *Biochem. Biophys. Res. Commun.* 1987. V. 146. P. 934-938.
2. Rich A., Crick F.H.C. // *Nature*. 1955. V. 176. P. 915-916.
3. Cowan P.M., McGavin S., North A.C.T. // *Nature*. 1955. V. 176. P. 1062-1064.
4. Andreeva N.S., Esipova N.G., Millionova M.I., Rogulenkova V.N., Shibnev V.A. // *Conformation in Biopolymers* / Ed. G.N. Ramachandran. New York; London: Acad. Press, 1967. V. 2. P. 469-481.
5. Okuyama K., Okuyama K., Arnott S., Takayanagi M., Kakudo M. // *J. Mol. Biol.* 1981. V. 152. P. 427-443.
6. Ramachandran G.N., Kartha G. // *Nature*. 1955. V. 176. P. 593-595.
7. Ramachandran G.N., Bansal M., Bhatnagar R.S. // *Biochim. Biophys. Acta*. 1973. V. 322. P. 166-171.
8. Yee R.Y., Englander S.W., von Hippel P.H. // *J. Mol. Biol.* 1974. V. 83. P. 1-16.
9. Туманян В.Г., Есипова Н.Г. // *Биофизика*. 1983. Т. 28. С. 962-965.
10. Tumanyan V.G., Rogulenkova V.N., Esipova N.G. // *Modelling and Computer Methods in Mol. Biol. and Genetics* / Eds V.A. Ratner, N.A. Kolchanov. New York: Nova Science Publishers, 1992. P. 197-201.
11. Туманян В.Г., Роголенкова В.Н., Есипова Н.Г., Айзенхабер Ф. // *Биофизика*. 1992. Т. 37. С. 1-4.
12. Blandel T., Barlow D., Borkakoti N. // *Nature*. 1983. V. 306. P. 281-283.
13. James M.N.G., Sielecki A.R. // *J. Mol. Biol.* 1983. V. 163. P. 299-361.
14. Tumanyan V.G. // *Biopolymers*. 1970. V. 9. P. 955-963.
15. Tumanyan V.G., Esipova N.G. // *Biopolymers*. 1982. V. 21. P. 475-497.
16. Miller M.H., Scheraga H.A. // *Polymer Sci.* 1976. Symp. № 54. P. 171-200.
17. Venugopal M.G., Ramshaw J.A.M., Brashwell E., Zhu D., Brodsky B. // *Biochemistry*. 1994. V. 33. P. 7948-7956.
18. Wolfenden R. // *Biochemistry*. 1978. V. 17. P. 201-204.
19. Zhorov B.S. // *J. Struct. Chem.* 1981. V. 22. P. 8-12.
20. Zhorov B.S. // *J. Struct. Chem.* 1982. V. 23. P. 3-9.
21. Momany F.A., McGuire R.F., Burgers A.W., Scheraga H.A. // *J. Phys. Chemistry*. 1975. V. 79. P. 2361-2381.
22. Zhorov B., Ananthanarayanan V.S. // *Biophysical Journal*. 1996. V. 70. P. 22-37.
23. Zhorov B., Ananthanarayanan V.S. // *J. Biomol. Struct. Dyn.* 1993. V. 11. P. 529-540.
24. Miller M.H., Nemethy G., Scheraga H.A. // *Macromolecules*. 1980. V. 13. P. 910-913.

Collagen Structure with a New Net of Hydrogen Bonds

Yu. V. Milchevskii^{##}, B. S. Zhorov^{**}, N. G. Esipova^{*}, and V. G. Tumanyan^{*}

^{*}Engelhardt Institute of Molecular Biology, Russian Academy of Sciences,
ul. Vavilova 32, GSP-1 Moscow, 117984 Russia

^{**}Sechenov Institute of Evolutionary Physiology and Biochemistry, Russian Academy of Sciences,
ul. Toreza 44, St. Petersburg, 194223 Russia

A conformational variability of the collagen triple helix was studied with the methods of molecular mechanics. The Rich-Crick model with one hydrogen bond per tripeptide fragment or the model with two hydrogen bonds per tripeptide fragment were used for tripeptides forming the primary structure of the protein. Imino acid and amino acid residues were located in the second position of the tripeptide fragments in the first and second cases, respectively. Conformations on domain boundaries, which had alternating structures with one and two hydrogen bonds per tripeptide, were particularly studied. Essentially all types of collagen backbone composed of amino acid residues most frequently occurring in this protein were considered. A new model was suggested that combined elements of the Rich-Crick model and our new approach. This was shown to be stereochemically valid, energetically advantageous, and consistent with the experimental data. It was conclusively demonstrated that the primary structure of collagen determines its tertiary structure.

Key words: collagen, hydrogen bonds, molecular mechanics, proteins, spatial structure

[#] To whom correspondence should be addressed; phone: +7 (095) 137-9957; e-mail: milch@imb.imb.ac.ru.