



УДК 577.323.5:543.422.25

РАСШИРЕНИЕ ГЕОМЕТРИЧЕСКОГО АЛГОРИТМА РАСЧЕТА
ПРОСТРАНСТВЕННОЙ СТРУКТУРЫ БЕЛКОВО-ПЕПТИДНЫХ
МОЛЕКУЛ ПО ДАННЫМ ЯМР-СПЕКТРОСКОПИИ

Майоров В. Н., Арсеньев А. С., Быстров В. Ф.

Институт биоорганической химии им. М. М. Шемякина
Академии наук СССР, Москва

Разработано специальное дополнение к программе дистанционного геометрического алгоритма, позволяющее при построении пространственной структуры полипептидов учитывать экспериментальные данные ЯМР, неоднозначно связанные с ограничениями расстояний, — константы спин-спинового взаимодействия вицинальных протонов $\text{H}-\text{NC}^\alpha-\text{H}$ и $\text{H}-\text{C}^\alpha\text{C}^\beta-\text{H}$ и скорости дейтерообмена амидных групп NH основной цепи. Кроме того, результаты анализа констант спин-спинового взаимодействия привлечены для эмпирической калибровки интенсивности кросс-пиков ядерного эффекта Оверхаузера, с тем чтобы повысить их информативность для дистанционного геометрического алгоритма. Расчет, проведенный по расширенной программе при определении спиральной конформации фрагмента $\text{Ala}^{13}-\text{Cys}^{20}$ инсектотоксина I_5A *Buthus eupeus*, показал корректность предложенной методики оценки вводимых ограничений расстояний. Из сравнения конформаций этого участка, полученных исходя из независимых начальных координат, следует, что точность определения пространственной структуры по среднеквадратичному отклонению между структурами (0,3 Å для атомов основной цепи и 1,1 Å для всех атомов) превосходит точность расчета без учета констант спин-спинового взаимодействия (0,5 и 1,3 Å соответственно).

Дистанционный геометрический алгоритм [1] является средством компьютерной обработки данных физико-химических методов с целью построения наиболее вероятной пространственной структуры молекулы. Особенно эффективно геометрический алгоритм применяется в конформационном анализе белково-пептидных веществ в растворе методом ЯМР-спектроскопии [2]. Так, с его помощью определены пространственные структуры блеомицина [3] и инсектотоксина I_5A *Buthus eupeus* [4] в растворе, а также глюкогона, встроенного в мицеллы [2, 5], реконструирована конформация ингибитора трипсина в растворе [6], установленная рентгеноструктурным анализом [7]. При этом в качестве основного источника исходных экспериментальных данных использовали межпротонный ядерный эффект Оверхаузера (ЯЭО). Однако вследствие сложности релаксационных процессов в высокомолекулярных соединениях возникают затруднения при определении расстояний между протонами методом ЯЭО. Поэтому условно принимают [8], что факт экспериментального наблюдения ЯЭО указывает на расстояние между протонами $<5 \text{ \AA}$, а соответственно отсутствие ЯЭО — на расстояние $>5 \text{ \AA}$. Такая приближенность оценки расстояний, во-первых, не позволяет достаточно точно определить локальную конформацию отдельных аминокислотных остатков (двугранные углы ϕ и ψ основной цепи и χ^1 боковой цепи) и, во-вторых, требует для определения пространственной структуры молекулы с достаточной точностью выявления в спектрах ЯМР большого числа межпротонных ЯЭО-контактов, особенно между аминокислотными остатками, принадлежащими различным участкам полипептидной цепи.

В известной степени компенсировать эти недостатки может привлечение констант спин-спинового взаимодействия (КССВ) вицинальных протонов $\text{H}-\text{NC}^\alpha-\text{H}$ и $\text{H}-\text{C}^\alpha\text{C}^\beta-\text{H}$, которые прямо связаны с двугранными углами ϕ и χ^1 [9], а также данных о скорости дейтерообмена амидных протонов основной цепи, характеризующих их участие во внутримолекулярных водородных связях [10] и, следовательно, представляющих до-

полнительные сведения о контактах между аминокислотными остатками. Кроме того, исходя из расстояний, оцененных по КССВ, оказалось возможным эмпирически калибровать интенсивность ЯЭО и тем самым значительно повысить информативность этих экспериментальных данных для дистанционного геометрического алгоритма.

В предлагаемой вниманию работе проведена модификация известной программы дистанционного геометрического алгоритма [2, 3], с тем чтобы она наряду с ЯЭО учитывала перечисленные выше данные. Модифицированная программа апробирована на фрагменте $\text{Ala}^{13}\text{-Cys}^{20}$ молекулы инсектотоксина I_5A *V. eireus*, для которой ранее из спектров ЯМР был получен обширный набор параметров, характеризующих ее пространственное строение [4, 11]. Основанием для выбора этого фрагмента служило то обстоятельство, что это один из наиболее охарактеризованных ЯЭО-контактами фрагмент молекулы, и то, что имеющийся в нашем распоряжении компьютер (HP-3000 Hewlett Packard, объем оперативной памяти 96 кбайт на одного пользователя) в рамках использованного программного обеспечения не позволяет обрабатывать информацию о более чем 92 атомах. В результате проведенных расчетов получен ряд структур для фрагмента $\text{Ala}^{13}\text{-Cys}^{20}$ молекулы инсектотоксина I_5A с координатами атомов, удовлетворяющих данным ЯМР-спектроскопии и стандартным стереохимическим ограничениям. Предложенный расширенный алгоритм повысил точность определения координат атомов белкового фрагмента в растворе до 0,3 Å для атомов основной цепи и 1,1 Å для остальных атомов, в то время как при расчете только на основе ЯЭО-контактов она равна 0,5 и 1,3 Å соответственно.

Модификация дистанционного геометрического алгоритма. Дистанционный геометрический алгоритм преобразует имеющиеся данные об ограничениях, налагаемых на расстояния между атомами молекулы, в декартовы координаты атомов. Ограничения задают нижними (l_{ij}) и верхними (u_{ij}) пределами расстояния d_{ij} между i - и j -атомами. В алгоритм включены два типа ограничений на расстояния. Это, во-первых, априорные ограничения, обусловленные стандартными параметрами ковалентной структуры молекулы. К их числу относятся ограничения расстояний, определяемые длинами валентных связей, в том числе дисульфидных, и значениями валентных углов (для этих расстояний верхние и нижние ограничения совпадают). Для всех расстояний, не фиксированных ковалентной структурой, нижний предел полагается равным сумме ван-дер-ваальсовых радиусов. Ко второму типу ограничений расстояний относятся экспериментальные данные, например полученные по ЯЭО. Эта информация учитывается путем задания нижнего и верхнего пределов расстояний между соответствующими атомами. Для расстояний, о которых не имеется экспериментальных данных, верхний предел полагается равным величине, заведомо превышающей возможные размеры молекулы [2].

Для экономии машинного времени и памяти ЭВМ число центров, рассматриваемых в расчете по дистанционному геометрическому алгоритму, уменьшено за счет представления β -, γ -, δ - и ϵ - CH_2 -групп боковых цепей аминокислотных остатков в виде псевдоатомов с ван-дер-ваальсовыми радиусами 1,6 Å при соответствующей коррекции ограничений на расстояния аналогично работе [5]. В табл. 1 приведены типы представления боковых цепей остатков фрагмента $\text{Ala}^{13}\text{-Cys}^{20}$ инсектотоксина I_5A .

Минимизируемая на стадии уточнения координат квадратичная функция ошибок в модифицированной версии программы отличается от изложенной в работе [2]. Эта функция имеет вид

$$F^{\text{err}}(r_1, \dots, r_N) = D + C + S_I + S_{II}, \quad (1)$$

где

$$D = \sum'_{i < j} (|r_i - r_j|^2 - u_{ij}^2)^2 + \sum'_{k < l} (l_{kl}^2 - |r_k - r_l|^2)^2,$$

N — число атомов, $\{r_i = (x_i, y_i, z_i), i = 1, 2, \dots, N\}$ — координаты атомов; Σ' означает, что суммируются только члены, нарушающие неравенства

Представление боковых цепей аминокислотных остатков,
использованное при расчете фрагмента Ala¹³ - Cys²⁰
инсектотоксина I₅A *

Аминокислотный остаток	Атомы и псевдоатомы
Ala	M ^β
Lys	C ^β , H ^{β1} , H ^{β2} , L ^γ , L ^δ , L ^ε , K ^ε
Cys	C ^β , H ^{β1} , H ^{β2} , S ^γ
Arg	C ^β , H ^{β1} , H ^{β2} , L ^γ , L ^δ , N ^ε , H ^ε , C ^ε , I ^{η1} , I ^{η2}
Asp	C ^β , H ^{β1} , H ^{β2} , C ^γ , O ^{δ1} , O ^{δ2} , H ^{δ2}

* С, Н, О и S — атомы углерода, водорода, кислорода и серы; L, M, I и K — псевдоатомы, аппроксимирующие соответственно метиленовую, метильную, NH₂- и NH₃-группы.

$l_{ij} \leq |r_i - r_j| \leq u_{ij}$; $C(r_1, \dots, r_N)$ — член, учитывающий *L*- или *D*-конформацию аминокислотных остатков [2]. Стартовые координаты атомов для минимизации квадратичной функции ошибок F^{err} находят при помощи формализма метрической *G*-матрицы [2], определяемой через выбранные случайным образом расстояния $|r_i - r_j|$, где $l_{ij} \leq |r_i - r_j| \leq u_{ij}$. В настоящей работе к минимизируемой функции ошибок добавлено два члена — S_I и S_{II} , учитывающие КССВ вицинальных протонов соответственно для фрагментов типа I (с двумя вицинальными протонами) и типа II (с тремя вицинальными протонами), рассмотренные ниже.

Константы спин-спинового взаимодействия вицинальных протонов. В общем виде соотношение между КССВ вицинальных протонов Н-Х-У-Н (3J) и двугранным углом θ ($0^\circ \leq \theta \leq 180^\circ$) между направлениями Н-Х и У-Н описывается уравнением Карплюса [12]:

$${}^3J = A \cos^2 \theta - B \cos \theta + C, \quad (2)$$

где *A*, *B* и *C* — положительные коэффициенты, равные соответственно 9,4; 1,1 и 0,4 Гц для фрагмента Н-NC^α-Н [9] и 10,2; 1,8 и 1,9 Гц для фрагмента Н-C^αC^β-Н [13]. Эффект электроотрицательности заместителей для аминокислотных остатков учитывался введением коэффициента *k*, не зависящего от двугранного угла θ [9]: ${}^3J = k {}^3J^{эксп}$, где ${}^3J^{эксп}$ — экспериментальное измеренное значение КССВ, *k*=1,09 для фрагментов Н-NC^α-Н, кроме глицина (*k*=1,04), *k*=1 для фрагментов Н-C^αC^β-Н, кроме треонина (*k*=1,11), валина и изолейцина (*k*=1,03) и серина (*k*=1,15) [13]. Торсионные углы φ и χ^1 связаны с углом θ в соответствии с номенклатурой IUPAC [14]. Зависимость расстояния между вицинальными протонами от двугранного угла θ описывается уравнением [15]

$$d = (a - b \cos \theta)^{1/2}, \quad (3)$$

где *a* и *b* — коэффициенты, зависящие от параметров ковалентной структуры фрагмента и определяемые по данным библиотеки геометрических параметров ЕСЕРР [16–19].

В общем случае одному значению КССВ вицинальных протонов соответствуют четыре значения торсионного угла φ (или χ^1) и только два значения расстояния *d* (рис. 1). Поэтому учет КССВ вицинальных протонов при расчете координат с помощью дистанционного геометрического алгоритма удобнее выполнить, если аргументами в минимизируемой функции ошибок являются расстояния *d*.

Участки полипептидной цепи можно подразделить на два типа в зависимости от числа вицинальных протонов. Участок типа I содержит два вицинальных протона. К нему относятся фрагменты Н-NC^α-Н всех аминокислотных остатков, за исключением остатков глицина, пролина и N-концевого остатка, а также фрагменты Н-C^αC^β-Н остатков треонина, валина и изолейцина. Для фрагментов этого типа соотношение между КССВ (3J) и расстоянием между вицинальными протонами описывается

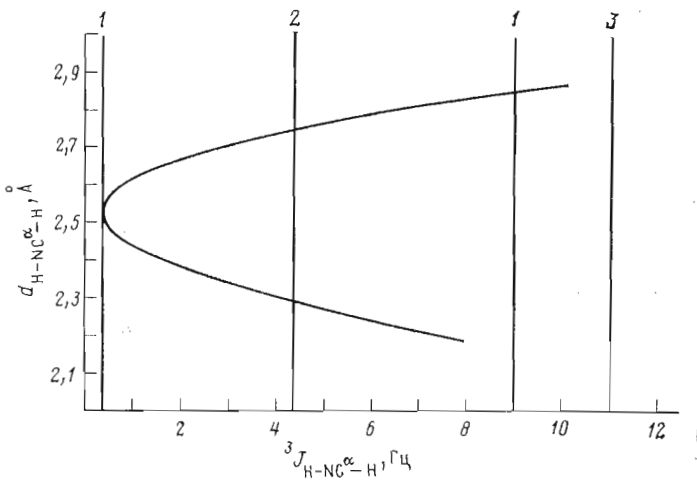


Рис. 1. Зависимость расстояния d между амидным протоном и α -протоном от значения КССВ ${}^3J_{\text{H-NC}^\alpha\text{-H}}$ вицинальных протонов $\text{H-NC}^\alpha\text{-H}$. Одному значению КССВ соответствуют: 1) одно значение d , 2) два значения d , 3) соответствия нет

зависимостью вида $d=f({}^3J)$ (рис. 1), где в качестве скрытого параметра фигурирует двугранный угол θ . Для фрагментов этого типа необходимо различать три случая (рис. 1): 1) имеется однозначное соответствие между 3J и d ; 2) одному значению КССВ соответствуют два значения расстояния d ; 3) соответствия нет. Случай 3 возникает, если наблюдаемое значение 3J оказывается больше максимально возможного значения по уравнению (2), т. е. если коэффициенты A , B и C в этом уравнении неадекватно описывают данную систему взаимодействующих атомов, например из-за специфических стерических напряжений, приводящих к отклонению длин валентных связей и значений валентных углов от стандартных значений. Для инсектотоксина I_{5A} [4] значения всех КССВ вицинальных протонов типа I относятся к случаям 1 и 2.

Фрагмент типа II содержит три вицинальных протона. К нему относится фрагмент $\text{HN-C}^\alpha\text{H}_2$ глицина и фрагменты $\text{HC}^\alpha\text{-C}^\beta\text{H}_2$ боковых цепей всех аминокислотных остатков, кроме остатков треонина, валина, изолейцина и аланина. В этом случае двум значениям вицинальных КССВ ${}^3J_{\text{H}^\alpha\text{-C}^\alpha\text{C}^\beta\text{-H}_A^\beta}$ (${}^3J_{\text{H-NC}^\alpha\text{-H}_A^\alpha}$ для глицина) и ${}^3J_{\text{H}^\alpha\text{-C}^\alpha\text{C}^\beta\text{-H}_B^\beta}$ (${}^3J_{\text{H-NC}^\alpha\text{-H}_B^\alpha}$ для глицина) соответствует одна пара расстояний d_A и d_B между протонами H^α и H_A^β и H^α и H_B^β соответственно.

При введении КССВ в дистанционный геометрический алгоритм необходимо учесть следующие два обстоятельства, которые обуславливают погрешность этого параметра. Во-первых, коэффициенты A , B и C в уравнении Карплюса определяются параметрическим согласованием с экспериментально наблюдаемыми значениями КССВ [9, 13]. Поэтому каждый из них найден с определенной точностью. Кроме того, экспериментальные определяемые из спектров ЯМР значения КССВ измерены с некоторой погрешностью (например, для инсектотоксина I_{5A} обычно $\pm 0,3$ Гц [4]). Это означает, что расстояние d между вицинальными протонами во фрагментах типов I или II имеет допустимые пределы (верхний u и нижний l), определяемые указанными причинами. Во-вторых, молекулы пептидов и белков в растворе представляют собой конформационно подвижные системы, и это отражается на экспериментально измеряемых в спектрах ЯМР параметрах (в том числе КССВ). Поэтому точная интерпретация ЯМР-параметров возможна только для динамической модели молекулы. Для этого необходимо проводить моделирование динамики белка в широком интервале временных параметров и определять относительные населенности ротамеров практически во всем интервале возможных значений углов φ или χ^1 . Такие расчеты в полном объеме пока представляются

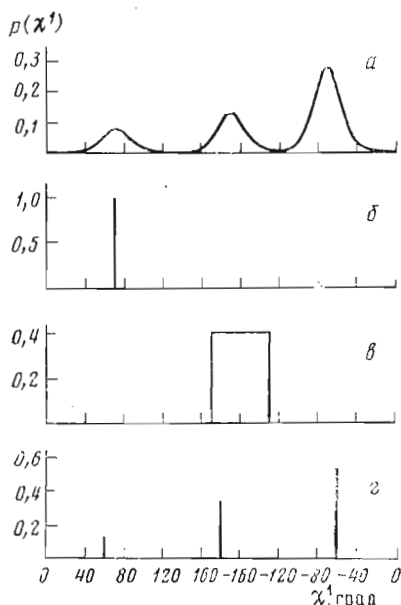


Рис. 2. Модели для интерпретации КССВ $\text{NC}\alpha\text{-C}\beta\text{H}_2$: а — гипотетическое неравновероятное распределение населенностей по состояниям, смещенным по углу χ^1 относительно канонических ротамеров; б — конформационно жесткий фрагмент — одно фиксированное значение угла χ^1 ; в — равновероятное распределение относительно фиксированного значения угла χ^1 в диапазоне $\pm 30^\circ$; г — распределение населенностей по каноническим ротамерам ($\chi^1 = 60, 180, -60^\circ$)

нереальными, так как они требуют большого объема памяти и счетного времени компьютера и, кроме того, в настоящее время еще недостаточно надежно определены эмпирические функции потенциальной энергии меж-атомных взаимодействий.

Приближенный учет конформационной подвижности возможен лишь в отдельных случаях, например при анализе КССВ фрагментов типа $\text{NC}\alpha\text{-C}\beta\text{H}_2$. На рис. 2а показано гипотетическое распределение населенности ротамеров по углу χ^1 . Эта зависимость может быть соотнесена с экспериментально измеренными значениями КССВ следующим образом:

$$^3J^{\text{экс}} = \int_0^{2\pi} {}^3J(\chi^1) p(\chi^1) d\chi^1 / \int_0^{2\pi} p(\chi^1) d\chi^1, \quad (4)$$

где ${}^3J(\chi^1)$ и $p(\chi^1)$ — зависимости величины КССВ (например, по уравнению (2)) и относительной населенности ротамера от угла χ^1 соответственно. Очевидно, что обратную задачу нахождения $p(\chi^1)$ исходя из экспериментальных значений КССВ решить нельзя. Поэтому при структурной интерпретации величин КССВ вицинальных протонов используются упрощенные модели, схематично показанные на рис. 2б—г.

Двум значениям КССВ протонов $\text{H}\alpha\text{-C}\alpha\text{C}\beta\text{-H}_A^\beta$ и $\text{H}\alpha\text{-C}\alpha\text{C}\beta\text{-H}_B^\beta$ ставится в соответствие фиксированное (или доминирующее) значение угла χ^1 (рис. 2б и в) или распределение населенностей ротамеров по стандартным значениям угла χ^1 ($60, 180$ и -60°) (рис. 2г). Фиксированное значение угла χ^1 можно ожидать для аминокислотных остатков, находящихся в стерически затрудненных условиях, например расположенных внутри белковой глобулы, и для остатков цистина, участвующих в образовании дисульфидных связей. Аминокислотные остатки, расположенные на поверхности молекулы, обычно могут менять свою конформацию, переходя из одного вращательного состояния в другое.

В работах [20, 21] для фрагментов $\text{NC}\alpha\text{-C}\beta\text{H}_2$ (тип II) предложен метод распознавания двух динамических ситуаций (конформационно жесткой — рис. 2б и дискретно-вращательной — рис. 2г), основанный на взаимозависимости двух КССВ (${}^3J_{\text{H}\alpha\text{-C}\alpha\text{C}\beta\text{-H}_A^\beta}$ и ${}^3J_{\text{H}\alpha\text{-C}\alpha\text{C}\beta\text{-H}_B^\beta}$), определяемой уравнением (2) и проекционным углом 120° между направлениями $\text{C}\beta\text{-H}_A^\beta$ и $\text{C}\beta\text{-H}_B^\beta$ *. В случае инсектотоксина I₅A проведенный таким

* Соблюдение такого соответствия между измеренными значениями КССВ (с учетом экспериментальной погрешности) означает конформационно жесткую ориентацию боковой цепи с одной парой расстояний между протоном $\text{H}\alpha$ и протонами H_A^β и H_B^β .

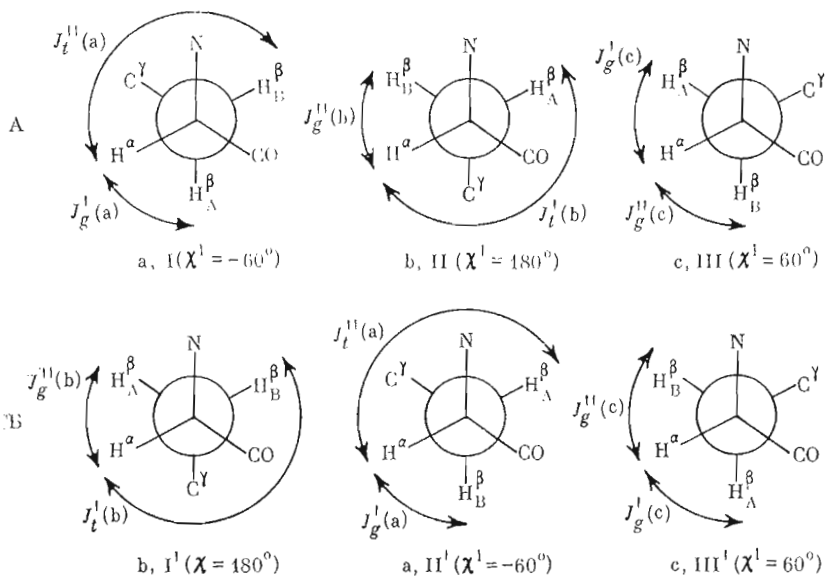


Рис. 3. Канонические ротамеры по связи $C^\alpha-C^\beta$ ($\chi^1=60, 180, -60^\circ$) для двух возможных вариантов *A* и *B* стереоселективного отнесения β -протонов. Ротамеры с индексом «штрих» получают из соответствующих ротамеров без индексов перестановкой атомов, находящихся в *gish*-ориентации по отношению к α -протону

образом анализ КССВ протонов $HC^\alpha-C^\beta H_2$ показал [4], что только три аминокислотных остатка (Cys^2 , Met^3 и Cys^{20}) имеют конформационно жесткие боковые цепи.

Этот подход можно применить для анализа КССВ и при отсутствии строгого соответствия в вышеуказанном смысле, если предположить, что для данного значения угла χ^1 соответствующая пара значений 3J_A и 3J_B получается усреднением при варьировании угла χ^1 в диапазоне $\Delta\chi^1$ (рис. 2). Так, например, с помощью простейшей модели стохастического распределения с $\Delta\chi^1=\pm 30^\circ$ [20] можно согласовать экспериментальные значения КССВ протонов $HC^\alpha-C^\beta H_2$ для аминокислотных остатков $Cys^5, {}^{26}, {}^{33}$. При этом верхние (u_A и u_B) и нижние (l_A и l_B) ограничения на расстояния $H^\alpha-H_A^\beta$ и $H^\alpha-H_B^\beta$ учитывают по уравнению (3), исходя из определяемых так же, как и в предыдущем случае, значений угла χ^1 с учетом его варьирования в диапазоне $\Delta\chi^1$.

Если пару значений КССВ нельзя сопоставить с фиксированным (или доминирующим) значением угла χ^1 , то фрагменты $HC^\alpha-C^\beta H_2$ рассматривают как дискретно-вращательные. Для анализа этих КССВ используют другой подход — рассмотрение относительных населенностей заторможенных ротамеров (рис. 2з) [9, 13]. При таком рассмотрении конформационное состояние относительно связи $C^\alpha-C^\beta$ представляют в виде ротамеров *a, b, c* (рис. 3), соответствующих минимумам потенциальной энергии при вращении вокруг ординарной связи $C^\alpha-C^\beta$. Поскольку скорость вращения относительно связи $C^\alpha-C^\beta$ значительно выше, чем нижний предел шкалы времени для спектров протонного ЯМР, наблюдаемые значения КССВ представляют собой усредненные значения по набору ротамеров с учетом их населенностей. Для фрагментов $HC^\alpha-C^\beta H_2$ с магнитно неэквивалентными протонами H_A^β и H_B^β в спектре наблюдаются два значения для виденных КССВ (3J_A и 3J_B), усредненные по значениям КССВ отдельных ротамеров:

$$\begin{aligned} {}^3J_A &= aJ_g'(a) + bJ_t''(b) + cJ_g'(c) \\ {}^3J_B &= aJ_t''(a) + bJ_g''(b) + cJ_g''(c) \\ a + b + c &= 1 \end{aligned} \quad (5)$$

где *a, b, c* — населенности соответствующих ротамеров (остальные обозначения см. рис. 3).

Измеренные значения КССВ протонов $\text{H}-\text{NC}^\alpha-\text{H}$ (${}^3J_{\text{H}_\text{N}\text{H}^\alpha}$) и $\text{H}-\text{C}^\alpha\text{C}^\beta-\text{H}$ (${}^3J_{\alpha\beta}$) аминокислотных остатков $\text{Ala}^{13}-\text{Cys}^{20}$ инсектотоксина I_5A [4], соответствующие им населенности (по уравнению (7)) заторможенных ротамеров и ограничения расстояний между протонами

Аминокислотный остаток	${}^3J_{\text{H}_\text{N}\text{H}^\alpha}$, Гц	Минимальное и максимальное расстояния $\text{H}-\text{H}^\alpha$, Å*		${}^3J_{\alpha\beta\text{A}}/{}^3J_{\alpha\beta\text{B}}$, Гц [4]**	Населенности заторможенных ротамеров ^{1*}			Минимальное и максимальное расстояния, Å	
		1**	2**		I или I'	II или II'	III или III'	$\text{H}^\alpha-\text{H}_\text{A}^\beta$	$\text{H}^\alpha-\text{H}_\text{B}^\beta$
Ala ¹³	3,7	2,29 2,32	2,72 2,74	5*					
Lys ¹⁴	5,0	2,24 2,27	2,75 2,77	5*					
Lys ¹⁵	4,2	2,27 2,29	2,73 2,75	11,0 3,2	0,72	-0,03	0,31	2,79 2,95	2,19 2,79
Cys ¹⁶	5,2	2,22 2,25	2,75 2,77	5,8 12,3	0,22	0,85	-0,06	2,18 2,78	2,78 2,95
Arg ¹⁷	4,1	2,23 2,26	2,70 2,72	6,3 8,3	0,27	0,46	0,28	2,82 2,98	2,25 2,82
Asp ¹⁸	3,9	2,24 2,28	2,70 2,72	11,6 3,1	0,78	-0,04	0,27	2,81 2,98	2,23 2,81
Cys ¹⁹	3,0	2,30 2,33	2,68 2,71	3,7 11,5	0,01	0,77	0,22	2,18 2,78	2,78 2,95
Cys ²⁰	6,8	2,18 2,21	2,79 2,81	2,0 12,6				2,50** 2,55	2,92** 2,94

* Определение минимального и максимального расстояний произведено согласно уравнениям (2) и (3) с учетом экспериментальной погрешности КССВ.

** См. рис. 1 (случай 2).

** См. табл. 3 в [4].

** Небольшие отрицательные населенности канонических ротамеров отражают неполную адекватность описания КССВ $\text{H}-\text{C}^\alpha\text{C}^\beta-\text{H}$ в рамках рассматриваемой модели, а также погрешность измерения этих КССВ.

** КССВ ${}^3J_{\alpha\beta}$ не измерены из-за совпадения химических сдвигов β -протонов.

** Для Cys²⁰ ограничения расстояний $\text{H}^\alpha-\text{H}_\text{A}^\beta$ и $\text{H}^\alpha-\text{H}_\text{B}^\beta$ получены в предположении о конформационной жесткости (фиксированное значение угла χ') фрагмента $\text{HC}^\alpha-\text{C}^\beta\text{H}_2$ этого аминокислотного остатка.

Обычно принимают (игнорируя ориентационный эффект электроотрицательности заместителя) [9], что все *гаш-* ($\theta=60^\circ$) и *транс-* ($\theta=180^\circ$) КССВ одинаковы:

$$\begin{aligned} J'_g(a) = J''_g(b) = J'_g(c) = J''_g(c) &\equiv J_g, \\ J_t(b) = J''_t(a) &\equiv J_t. \end{aligned}$$

Тогда для населенностей ротамеров получаем

$$\begin{aligned} a &= ({}^3J_\text{B} - J_g) / (J_t - J_g) \\ b &= ({}^3J_\text{A} - J_g) / (J_t - J_g) \\ c &= 1 - a - b \end{aligned} \quad (6)$$

Основное затруднение конформационного анализа боковых цепей аминокислотных остатков состоит в стереохимическом отнесении сигналов протонов H^β и H_B^β *. Без такого отнесения можно оценить только населенность ротамера с (рис. 3), для которого оба β -протона находятся в *гаш-* ориентации по отношению к α -протону. Населенности двух других ротамеров по отдельности оценить нельзя, так как перестановка H_A^β и H_B^β приводит к одному и тому же результату.

* Однозначное отнесение сигналов β -протонов получают с помощью стереоселективного замещения одного из β -протонов на дейтерий [22] или используя вициальные КССВ β -протонов с ядрами ${}^{13}\text{C}$ в цепях $\text{H}-\text{C}^\beta\text{C}^\alpha-{}^{13}\text{C}'(\text{O})$ и $\text{H}-\text{C}^\alpha\text{C}^\beta-{}^{13}\text{C}'$, а также с ${}^{15}\text{N}$ в цепи ${}^{15}\text{N}-\text{C}^\alpha\text{C}^\beta-\text{H}$ [9].

Поскольку в дистанционном геометрическом алгоритме используются ограничения на расстояния, учет КССВ вицинальных протонов $\text{HC}^\alpha\text{—C}^\beta\text{H}_2$ для случая заторможенных ротамеров можно провести без предварительного стереохимического отнесения β -протонов*. Вместо одного набора ротамеров а, б, с введем в рассмотрение два набора: I, II, III и I', II', III' (рис. 3). Ротамеры, обозначенные римскими цифрами с апострофом, получают из ротамеров, обозначенных соответствующими римскими цифрами (отличаясь стереохимическим отнесением β -протонов), путем перестановки атомов, находящихся в *gosh*-ориентации по отношению к H^α -атому. При такой перестановке расстояния $\text{H}^\alpha\text{—H}_\text{A}^\beta$ и $\text{H}^\alpha\text{—H}_\text{B}^\beta$ не меняются (в отличие от значения двугранного угла боковой цепи χ^1) и их можно учитывать в дистанционном геометрическом алгоритме, не проводя стереохимического отнесения β -протонов. В дальнейшем обозначение I будет отвечать двум ротамерам: I и I' и т. п.

Относительные населенности p заторможенных ротамеров I, II и III вычисляются по уравнениям

$$\begin{aligned} p_{\text{I}} &= ({}^3J_{\text{B}} - J_{\text{g}}) / (J_{\text{t}} - J_{\text{g}}) \\ p_{\text{II}} &= ({}^3J_{\text{A}} - J_{\text{g}}) / (J_{\text{t}} - J_{\text{g}}) \\ p_{\text{III}} &= (J_{\text{t}} + J_{\text{g}} - {}^3J_{\text{A}} - {}^3J_{\text{B}}) / (J_{\text{t}} - J_{\text{g}}) \end{aligned} \quad (7)$$

$$(p_{\text{I}} + p_{\text{II}} + p_{\text{III}} = 1)$$

где ${}^3J_{\text{A}}$ и ${}^3J_{\text{B}}$ — экспериментально измеренные значения КССВ. Для всех аминокислотных остатков, имеющих фрагмент $\text{HC}^\alpha\text{—C}^\beta\text{H}_2$, кроме остатка серина, $J_{\text{t}} = 13,90$ Гц и $J_{\text{g}} = 3,55$ Гц [13]. Для остатка серина с учетом электроотрицательности C^β -заместителя $J_{\text{t}} = 12,09$ Гц и $J_{\text{g}} = 3,09$ Гц [13]. Использование уравнений (7) предполагает выполнение очевидного условия

$${}^3J_{\text{A}} + {}^3J_{\text{B}} \leq J_{\text{t}} + J_{\text{g}}$$

В табл. 2 приведены значения относительных населенностей ротамеров I, II и III, рассчитанных исходя из КССВ протонов $\text{HC}^\alpha\text{—C}^\beta\text{H}_2$ аминокислотных остатков участка молекулы инсектотоксина I_5A .

При построении статической модели доминирующей пространственной структуры белка естественно задавать ограничения на расстояния $\text{H}^\alpha\text{—H}_\text{A}^\beta$ и $\text{H}^\alpha\text{—H}_\text{B}^\beta$, соответствующие наиболее населенному ротамеру. В данной работе такие ограничения накладывались, если относительная населенность одного из заторможенных ротамеров превышала 0,65 (обычно 0,80). Чтобы учесть принципиальную возможность отклонения угла χ^1 от канонических значений и флуктуации угла χ^1 в доминирующем ротамере, верхние (*u*) и нижние (*l*) пределы расстояний $\text{H}^\alpha\text{—H}_\text{A}^\beta$ и $\text{H}^\alpha\text{—H}_\text{B}^\beta$ (табл. 2) определяли при варьировании угла χ^1 на $\pm 60^\circ$ от основного значения для соответствующего ротамера. Если населенность ни одного из заторможенных ротамеров не превышает 0,65 (аминокислотные остатки Arg^{17,35} и Asn^{23,34}) или химические сдвиги двух β -протонов совпадают (Lys^{14,25}), то, очевидно, нельзя отдать предпочтение какому-либо одному из ротамеров. В таких случаях КССВ протонов $\text{HC}^\alpha\text{—C}^\beta\text{H}_2$ не учитывались.

Для оценки конформационной подвижности во фрагменте типа I можно воспользоваться сведениями об отсутствии или наличии ЯЭО между амидным протоном $i+1$ -го остатка и соответственно C^αH (d_1 -контакт), NH (d_2 -контакт) и C^βH (d_3 -контакт) протонами i -го остатка. Эти ЯЭО характеризуют конформацию i -го аминокислотного остатка, а именно двугранные углы ψ_i (d_1 -контакт), φ_i и ψ_i (d_2 -контакт) и ψ_i и χ_i^1 (d_3 -контакт) [15]. Принцип рассмотрения ${}^3J_{\text{H—NC}^\alpha\text{—H}}$ следует из сравнения данных рентгеноструктурного анализа и ЯМР-спектроскопии. Во-первых, для конформационно стабильных участков белков наблюдается хорошо

* Стереоселективное отнесение сигналов β -протонов получается в итоге работы дистанционного геометрического алгоритма, если имеются ограничения на расстояния между H^β -атомом, занимающим *gosh*-ориентацию по отношению к H^α -атому, и другими фрагментами молекулы, например амидным протоном данного или последующего в аминокислотной последовательности остатка.

регистрируемый в спектрах ЯМР набор контактов ЯЭО и прямое соответствие между величинами КССВ $N-NC^{\alpha}-H$ в растворе и углами φ основной цепи в кристаллах. Последнее особенно детально обосновано на примерах ингибитора трипсина [23] и лизоцима [24]. Следовательно, для участков полипептидной цепи, где методом ЯМР обнаруживаются систематические наборы d -контактов, можно прямо вычислять углы φ из значений ${}^3J_{N-NC^{\alpha}-H}$ по уравнению (2). Если для конформационно устойчивых фрагментов $N-NC^{\alpha}-H$ значению КССВ соответствуют два расстояния (случай 2 на рис. 1), то их принимают в качестве нижнего и верхнего ограничений на расстояние на стадии получения с помощью формализма G -матрицы [2] координат атомов и псевдоатомов. Эти координаты в дальнейшем уточняют минимизацией квадратичной функции ошибок F^{err} (уравнение (1)) путем оптимизации возможных расстояний, которая на данном этапе минимизации лучше согласуется со всей рассматриваемой структурой молекулы.

Напротив, в спектрах ЯМР не обнаруживают d -контактов методом ЯЭО [25] для участков белковой молекулы (например, N- и C-концевые фрагменты ингибитора трипсина), которые не дают четких рефлексов при дифракции рентгеновских лучей и конформация которых, согласно теоретическому моделированию молекулярной динамики, наиболее лабильна [26–28]. Для таких участков следовало бы, вообще говоря, учитывать усреднение наблюдаемых значений ${}^3J_{N-NC^{\alpha}-H}$ по некоторому набору углов φ (аналогично уравнению (4)). Однако сделать это на данном этапе развития конформационного анализа затруднительно. Если для фрагментов типа I ($N-NC^{\alpha}-H$ всех остатков, кроме остатков глицина, и $N-C^{\alpha}C^{\beta}-H$ остатков Thr, Val и Ile) наблюдаются значения КССВ вицинальных протонов, близкие к максимальному (например, фрагменты $N-NC^{\alpha}-H$ аминокислотных остатков Cys⁵, Phe^{6,27}, Gln³⁰ и Leu³² инсектотоксина I₅A [4]) или минимальному (например, фрагменты $N-NC^{\alpha}-H$ остатков Met¹² и Cys¹⁹ и фрагмент $N-C^{\alpha}C^{\beta}-H$ остатка Thr⁸) теоретическому значению КССВ, определяемому из уравнения (2), то эти фрагменты можно классифицировать как конформационно стабильные.

Математические выражения компонентов функции ошибок (уравнение (1)), учитывающих значения КССВ для фрагментов типов I и II, могут быть представлены в следующем виде:

$$S_I = \sum_{n=1}^{N_1} \min(\Delta_n^{(1)}, \Delta_n^{(2)}), \quad (8)$$

$$S_{II} = \sum_{n=1}^{N_2} (\Delta_n^A + \Delta_n^B), \quad (9)$$

где N_1 и N_2 — число фрагментов типа I и II соответственно, а Δ_n определяется как

$$\Delta_n = \begin{cases} (d_n^2 - u_n^2)^2, & \text{если } d_n > u_n \\ 0, & \text{если } l_n \leq d_n \leq u_n \\ (l_n^2 - d_n^2)^2, & \text{если } d_n < l_n \end{cases} \quad (10)$$

где u_n и l_n — верхний и нижний пределы расстояний d_n между вицинальными протонами n -го фрагмента типа I или II. Выражение $\min(\Delta_n^{(1)}, \Delta_n^{(2)})$ означает, что если для фрагмента типа I значению КССВ соответствуют два расстояния между вицинальными атомами (случай 2 на рис. 1), то оптимизация ведется для того из возможных расстояний, которое на данном этапе минимизации лучше согласуется со структурой всего рассматриваемого участка молекулы.

Внутримолекулярные водородные связи. Спектроскопия ЯМР предоставляет возможность измерения скоростей обмена протонов индивидуальных групп NH белковых молекул с растворителем [10]. Строгая количественная интерпретация этих экспериментальных данных возможна только при учете динамических особенностей пространственной структуры

белка в растворе. Технические и теоретические трудности пока не позволяют это реализовать [29]. Поэтому в настоящее время при интерпретации скоростей дейтериеобмена групп NH с растворителем ограничиваются рассмотрением статической структуры молекулы белка.

В работах [10, 30] показано, что экспериментально наблюдается полуквантитативная корреляция между скоростями дейтериеобмена и доступностью групп NH для молекул растворителя. Группы, участвующие в быстром обмене, как правило, расположены на поверхности белка и хорошо доступны молекулам растворителя. В случае медленного обмена группы имеют низкую доступность и, по данным рентгеноструктурного анализа, участвуют в образовании внутримолекулярных водородных связей [30]. Практически доступность растворителю групп NH может быть определена только после построения полной пространственной структуры всей молекулы. Поэтому в модифицированной программе дистанционного геометрического алгоритма процедура учета экспериментальных данных о скоростях дейтериеобмена реализована путем предоставления возможности группам NH, участвующим в медленном обмене со средой, образовывать внутримолекулярные водородные связи. Скорости дейтериеобмена групп NH могут быть скорректированы на их доступность молекулам среды после построения полной пространственной структуры и полуквантитативно учтены при дальнейшем ее уточнении.

К сожалению, при построении пространственной структуры белково-пептидных молекул учет данных о внутримолекулярных водородных связях осложняется тем, что, если с уверенностью можно указать донор водородной связи (медленно обменивающуюся с растворителем амидную группу), то пока нет надежного способа идентифицировать соответствующую акцепторную группу (например, карбонильные группы аминокислотных остатков)*. Сведения об акцепторных группах могут быть получены уже после построения пространственной структуры молекулы белка.

В используемом в настоящей работе подходе предполагалось, что каждая амидная NH-группа i -го аминокислотного остатка, характеризующаяся большими временами дейтериеобмена (для рассматриваемого участка молекулы инсектотоксина ≥ 60 ч при pH 2,7 и 22° C, см. табл. 1 в работе [4]), образует водородную связь с карбонильной группой j -го аминокислотного остатка ($N_iH \dots O=C_j$) при условии, что $|i-j| \geq 2$. Для этого расстояния максимальной сближенности между атомами амидных и карбонильных групп, способных образовать водородную связь, уменьшены по сравнению с расстояниями, отвечающими ван-дер-ваальсову контакту, согласно критериям образования водородной связи [32]. При расчете пространственной структуры с помощью дистанционного геометрического алгоритма нижние пределы расстояний между атомами N_i и O_j , H_i и O_j и H_j и C_j полагались равными 1,82; 1,45 и 1,96 Å, тогда как соответствующие расстояния, отвечающие ван-дер-ваальсову контакту, равны 2,6; 2,2 и 2,2 Å. Как показали проверочные расчеты, без такого уменьшения ограничений снизу на вышеуказанные расстояния невозможно добиться сходимости на стадии минимизации квадратичной функции ошибок, т. е. получить стереохимически приемлемую структуру, удовлетворяющую данному набору ограничений на расстояния.

Ядерный эффект Оверхаузера (ЯЭО). Экспериментальные данные для введения ограничений на расстояния по ЯЭО взяты из работы [4], где исследованы 2D-NOESY-спектры растворов инсектотоксина I_5A с временами обмена компонент намагниченности (τ_M) 100, 200, 300 и 500 мс. Интенсивность кросс-пиков ЯЭО зависит от многих факторов. При коротких τ_M , исключая спиновую диффузию (в случае инсектотоксина I_5A 100 и 200 мс), интенсивность кросс-пиков ЯЭО в первом приближении обратно пропорциональна шестой степени расстояния между протонами [8]. При увеличении τ_M (300 и 500 мс в случае инсектотоксина I_5A)

* Донор и акцептор водородной связи могут быть идентифицированы с помощью гетероядерного $^{13}C\{^1H\}$ -эффекта Оверхаузера [31]. Однако этот метод пока трудно применить при работе с белками.

ЯЭО-контакты между атомами водорода и псевдоатомами, полученные из анализа NOESY-спектров инсектотоксина I₅A [4], и соответствующие им ограничения расстояний сверху *:

- 2,7 Å – интенсивный кросс-пик при $\tau_M=100$ мс (S),
 3,1 Å – пик средней интенсивности при $\tau_M=100$ мс (M),
 3,8 Å – малоинтенсивный пик при $\tau_M=100$ мс (W1),
 4,2 Å – кросс-пик появляется при $\tau_M=200$ мс (W2),
 5,2 Å – кросс-пик появляется при $\tau_M=500$ мс (W5)

Аминокислотный остаток	Атом, псевдоатом	Аминокислотный остаток	Атом, псевдоатом	Тип контакта	Ограничение расстояния сверху *, Å
Ala ¹³	H α	Cys ¹⁶	H _N	W1	3,8
Ala ¹³	H α	Cys ¹⁶	H β^1	W2	4,2
Ala ¹³	H α	Cys ¹⁶	H β^2	W1	3,8
Ala ¹³	M β	Cys ¹⁶	H β^2	W5	5,9
Ala ¹³	H _N	Lys ¹⁴	H _N	S	2,7
Ala ¹³	M β	Lys ¹⁴	H _N	M	3,9
Lys ¹⁴	H α	Lys ¹⁴	L γ	M	3,7
Lys ¹⁴	H α	Arg ¹⁷	H _N	W1	3,8
Lys ¹⁴	H α	Arg ¹⁷	H β^1	W2	4,2
Lys ¹⁴	H α	Arg ¹⁷	L γ	W2	4,8
Lys ¹⁴	H _N	Lys ¹⁵	H _N	S	2,7
Lys ¹⁴	H β^1	Lys ¹⁵	H _N	M	3,1
Lys ¹⁵	H _N	Lys ¹⁵	L δ	W1	4,4
Lys ¹⁵	H α	Lys ¹⁵	L δ	W1	4,4
Lys ¹⁵	H α	Asp ¹⁸	H _N	W1	3,8
Lys ¹⁵	H α	Asp ¹⁸	H β^1	W1	3,8
Lys ¹⁵	H _N	Cys ¹⁶	H _N	S	2,7
Lys ¹⁵	H β^1	Cys ¹⁶	H _N	W2	4,2
Lys ¹⁵	L γ	Cys ¹⁶	H _N	W1	4,4
Cys ¹⁶	H α	Cys ¹⁹	H β^1	W1	3,8
Cys ¹⁶	H α	Cys ¹⁹	H β^2	W1	3,8
Cys ¹⁶	H β^2	Arg ¹⁷	H _N	M	3,1
Arg ¹⁷	H _N	Arg ¹⁷	L γ	M	3,7
Arg ¹⁷	H α	Arg ¹⁷	L γ	M	3,7
Arg ¹⁷	H α	Cys ²⁰	H β^1	W1	3,8
Arg ¹⁷	H α	Cys ²⁰	H β^2	W1	3,8
Arg ¹⁷	H _N	Asp ¹⁸	H _N	S	2,7
Arg ¹⁷	H β^1	Asp ¹⁸	H _N	M	3,1
Arg ¹⁷	H β^2	Asp ¹⁸	H _N	W1	3,8
Arg ¹⁷	H α	Asp ¹⁸	H _N	W1	3,8
Asp ¹⁸	H _N	Cys ¹⁹	H _N	S	2,7
Asp ¹⁸	H β^1	Cys ¹⁹	H _N	W1	3,8
Asp ¹⁸	H β^2	Cys ¹⁹	H _N	W1	3,8
Asp ¹⁸	H α	Cys ¹⁹	H _N	W1	3,8
Cys ¹⁹	H _N	Cys ²⁰	H _N	S	2,7
Cys ¹⁹	H β^1	Cys ²⁰	H _N	M	3,1
Cys ¹⁹	H α	Cys ²⁰	H _N	W1	3,8

* Ограничения расстояний между атомами и псевдоатомами введены в соответствии с типом контакта и поправкой, учитывающей, что данный протон входит в состав псевдоатома (см. уравнение (11)).

возрастает интенсивность кросс-пиков для сближенных протонов и, следовательно, проявляются ЯЭО между все более удаленными протонами. Однако по мере увеличения τ_M возрастает вклад мешающего эффекта спиновой диффузии – передачи намагниченности между удаленными ядрами по диполь-дипольному механизму через ядра-посредники [8].

Используя в качестве стандартных интенсивности кросс-пиков ЯЭО протонов, для которых расстояние или верхнее ограничение на расстояние известно (например, из анализа КССВ вицинальных протонов или из расчета стереохимически разрешенных расстояний по данным библиотеки ЕСЕРР [16–19]), мы провели полуколичественную классификацию остальных кросс-пиков ЯЭО с учетом времени τ_M , при котором кросс-пик появляется в NOESY-спектрах (см. примечание к табл. 3). При $\tau_M=$

=100 мс интенсивность кросс-пиков, классифицированных как сильные (обозначаются S), превышает характерную интенсивность кросс-пиков между вицинальными протонами $H-C^{\alpha}C^{\beta}-H$, находящимися по отношению друг к другу в *гош*-ориентации (расстояние $\sim 2,7$ Å). Если интенсивность кросс-пика сравнима с интенсивностью кросс-пиков между протонами $H-C^{\alpha}C^{\beta}-H$, находящимися в *транс*-ориентации (расстояние $\sim 2,9$ Å), но меньше, чем интенсивность кросс-пика типа S , такие кросс-пики классифицировались как средние (обозначаются M). Как показали проверочные расчеты, для того чтобы добиться наилучшего соответствия со всей совокупностью ЯЭО-контактов, верхнюю границу для расстояний, отвечающих кросс-пикам этого типа, необходимо увеличить на 0,2 Å. Необходимость такого увеличения обусловлена, по-видимому, динамической природой поведения молекулы белка в растворе и приблизительным характером оценки интенсивностей кросс-пиков. Таким образом, для кросс-пиков M принято $u=3,1$ Å. Классификация остальных кросс-пиков была проведена следующим образом: $W1$ — малointенсивный кросс-пик при $\tau_m=100$ мс, $u=3,8$ Å (интенсивность такого кросс-пика составляет $\leq 10\%$ от S); $W2$ — кросс-пик, появляющийся при $\tau_m=200$ мс ($u=4,2$ Å); $W5$ — кросс-пик, появляющийся при $\tau_m=500$ мс ($u=5,2$ Å). Значения верхних пределов расстояний, отвечающих контактам типа $W1$, $W2$ и $W5$, были подобраны опытным путем в тестовых расчетах таким образом, чтобы получаемые структуры наилучшим образом удовлетворяли всей совокупности стереохимических ограничений и экспериментальных данных. Установление верхнего предела расстояния на основе ЯЭО оправдано тем, что вычисленные по жесткой молекулярной модели значения ЯЭО всегда оказываются больше измеренных величин, по-видимому, вследствие внутренних движений в молекуле [29].

Нижнее ограничение расстояний для контактов всех типов принималось, как обычно, равным сумме ван-дер-ваальсовых радиусов. Интенсивность кросс-пиков ЯЭО между вицинальными протонами для ограничений расстояний не использовали, поскольку в этом случае ограничения более точно определяются из анализа КССВ (см. предыдущий раздел).

Таким образом, при вводе в расчет данных о ЯЭО-контактах верхняя граница u_{ij} расстояния между атомами и псевдоатомами i и j , разделенными более чем одним углом внутреннего вращения, полагалась равной

$$u_{ij}=D_{ij}+\Delta, \quad (11)$$

где D_{ij} — верхняя граница расстояния между протонами, определяемая типом ЯЭО-контакта (табл. 3); Δ — поправка, учитывающая, что данный протон входит в состав псевдоатома (табл. 1). Она равна 1,3 Å для CN_2 -группы и 1,7 Å для CN_3 -группы [5].

Расчет пространственной структуры фрагмента Ala¹³-Cys²⁰ инсекто-токсина I₅A. Проверка разработанной модификации программы дистанционного геометрического алгоритма проведена при полном расчете конформационного строения фрагмента Ala¹³-Cys²⁰ инсектотоксина I₅A. Для построения конформации этого фрагмента использованы упрощенные представления группы атомов в аминокислотных остатках (табл. 1). Такое представление практически не влияет на точность описания конформации, но позволяет уменьшить число рассматриваемых центров. В качестве вводимых в расчет ограничений на расстояния между атомами и псевдоатомами использованы данные, полученные из анализа NOESY-спектров (табл. 3), КССВ протонов $H-NC^{\alpha}-H$ и $H-C^{\alpha}C^{\beta}-H$ (табл. 2) и скорости дейтерообмена амидных протонов [4].

Поскольку данные ЯМР-спектроскопии носят характер сравнительно неточных ограничений на расстояния между атомами, то одному набору экспериментальных данных соответствует не одна, а некоторое множество пространственных структур. Поэтому каждый расчет со случайного независимого начального приближения будет давать новую структуру, более или менее отличающуюся от прежних.

Расчет конформации фрагмента Ala¹³-Cys²⁰ проведен в два этапа. Первоначально было сделано 20 независимых расчетов конформации

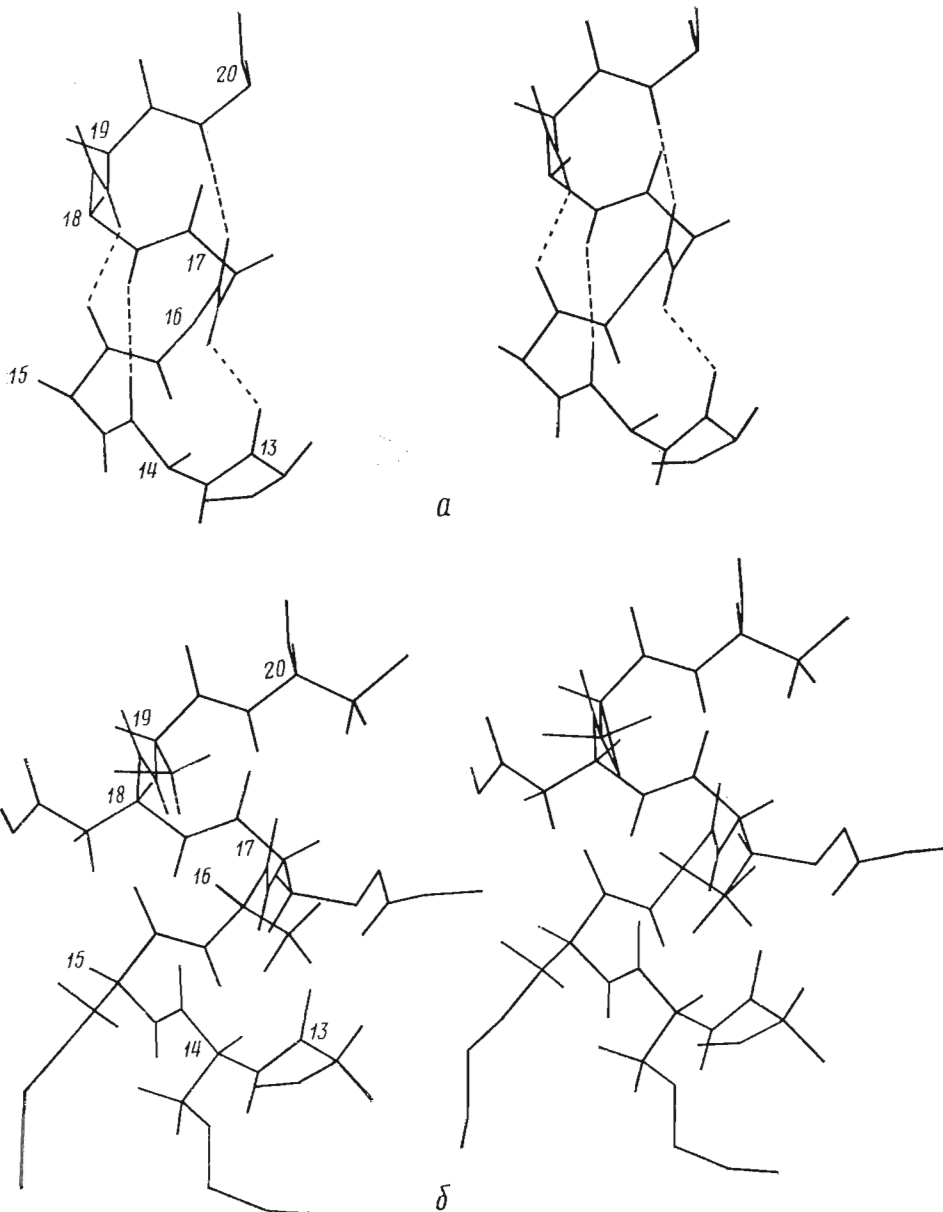


Рис. 4. Стереоскопическое изображение одной из полученных конформаций (№ 3, см. табл. 4) фрагмента Ala¹³-Cys²⁰ инсектотоксина I₅A: а — показаны только атомы основной цепи (водородные связи обозначены пунктиром); б — показаны все атомы (см. табл. 2)

фрагмента. В каждом случае проводили 5 итераций для получения начального приближения с использованием формализма G -матрицы и 500 итераций для его уточнения минимизацией F^{err} (r_1, \dots, r_N) — см. уравнение (1) [2]. Из полученных таким образом конформаций с целью экономии машинного времени отобрано 10, характеризующихся относительно низкими значениями квадратичной функции ошибок ($F^{\text{err}} \leq 0,5 \text{ \AA}^4$). Наш опыт показывает, что отбор, проведенный по такому критерию, в большинстве случаев гарантирует при дальнейшей минимизации квадратичной функции ошибок получение стереохимически приемлемых конформаций (отклонения величин валентных связей и валентных углов от номинальных не превышают соответственно $0,1 \text{ \AA}$ и $\sim 1^\circ$), удовлетворяющих использованному набору экспериментальных данных. Для отобранных конформаций сделано дополнительно 1500 итераций процедуры уточнения координат.

Абсолютные величины отклонений геометрических параметров и значений КССВ в 7 конформациях фрагмента Ala¹³ - Cys²⁰ инсектотоксина I₅A

Конформация	Величина квадратичной функции ошибок ×10 ² , Å ⁴	Максимальное отклонение от номинальной величины **			Отклонение рассчитанных ³ J _{H-NCα-H} от экспериментальных значений, Гц **		S _{II} · 10 ³ , Å ^{4*}
		валентной связи, Å	валентного угла, град	двугранного угла ω, град **	максимальное	суммарное	
1	1,4	0,02	1,07	12,6	0,24	0,40	0,31
2	1,4	0,02	1,23	12,1	0,34	0,84	7,70
3	0,7	0,01	0,53	7,5	0,15	0,48	3,40
4	0,5	0,01	0,60	7,8	0,28	0,34	4,50
5	0,8	0,01	0,84	9,3	0,10	0,28	0,05
6	10,2	0,09	1,10	10,3	0,16	0,33	0,21
7	1,6	0,02	0,97	11,9	0,32	0,63	0,02

* См. уравнение (1).

** Нарушения ограничений расстояний не превышали 0,1 Å.

** Отклонение двугранного угла ω основной цепи от 180° (транспланарная пептидная связь).

** Отклонение рассчитывали следующим образом:

$$^3J - (^3J_{\text{ЭКСП}} + \Delta^3J), \text{ если } ^3J > ^3J_{\text{ЭКСП}} + \Delta^3J;$$

$$0, \text{ если } ^3J_{\text{ЭКСП}} - ^3J \leq ^3J \leq ^3J_{\text{ЭКСП}} + \Delta^3J;$$

$$(^3J_{\text{ЭКСП}} - \Delta^3J) - ^3J, \text{ если } ^3J < ^3J_{\text{ЭКСП}} - \Delta^3J,$$

где ³J, ³J_{ЭКСП} и Δ³J — соответственно рассчитанная согласно уравнению (2) КССВ, экспериментальная КССВ и погрешность в измерении КССВ, равная ±0,3 Гц [4].

Таблица 5

Средние среднеквадратичные отклонения (Å) по характеристическим выборкам атомов для 7 конформаций фрагмента Ala¹³ - Cys²⁰ инсектотоксина I₅A

Вариант расчета	Выборка **			
	1	2	3	4
А *	0,24	0,28	0,46	0,74
Б *	0,34	0,39	0,67	1,09

* Среднеквадратичное отклонение между двумя конформациями рассчитывали по формулам

$$A) \left\{ \sum_{i < j} (d_{ij} - d'_{ij})^2 / (N(N-1)/2) \right\}^{1/2} \quad [34],$$

$$B) \left\{ \sum_k (Rr_k - r'_k)^2 / N \right\}^{1/2} \quad [33],$$

где d_{ij} и d'_{ij} — расстояния между атомами i и j ; r_k и r'_k — радиусы-векторы k -го атома в сравниваемых конформациях; R — матрица преобразования координат, наилучшим образом совмещающая две конформации; N — число центров (атомов и псевдоатомов).

** Среднеквадратичное отклонение рассчитывалось по выборкам: 1) атомы C^α; 2) атомы H_N, N, C^α, H^α, C', O; 3) атомы, перечисленные в п. 2, а также H_A^β, H_B^β и C^β (M^β); 4) все атомы.

В результате отобрано 7 конформаций со значением квадратичной функции ошибок, не превышающим ~0,1 Å⁴, одна из которых показана на рис. 4. Из табл. 4, в которой приведены параметры, характеризующие качество минимизации для этих конформаций, следует, что максимальные отклонения от номинальных значений для валентных связей, валентных углов и двугранных углов основной цепи ω равны соответственно 0,09 Å, 1,23 и 12,6°. Структурные параметры этих конформаций хорошо согласуются с экспериментальными ЯЭО-контактами и КССВ.

Для оценки точности, с которой определена пространственная структура фрагмента, каждую конформацию сравнивали с остальными шестью

Статистические данные о двугранных углах основной цепи ϕ и ψ для 7 конформаций фрагмента $\text{Ala}^{13} - \text{Cys}^{20}$ инсектотоксина I_{5A}

Аминокислотный остаток	Среднее значение ϕ , град	Стандартное отклонение ϕ , град	Среднее значение ψ , град	Стандартное отклонение ψ , град
Ala^{13}	—	—	-41,0	8,9
Lys^{14}	-72,1	2,8	-35,3	9,4
Lys^{15}	-65,3	0,3	-49,0	6,1
Cys^{16}	-70,2	0,3	-24,2	1,1
Arg^{17}	-64,5	1,8	-30,0	4,9
Asp^{18}	-70,6	5,1	-36,0	3,8
Cys^{19}	-62,5	5,8	-45,8	5,1
Cys^{20}	-81,3	0,6	—	—

Таблица 7

Статистические данные о внутримолекулярных водородных связях для 7 конформаций фрагмента $\text{Ala}^{13} - \text{Cys}^{20}$ инсектотоксина I_{5A} *

Амидная группа NH

Карбонильная группа C=O	Ala^{13}	Lys^{14}	Lys^{15}	Cys^{16}	Arg^{17}	Asn^{18}	Cys^{19}	Cys^{20}
	Ala^{13}					1,67 2,17		
Lys^{14}						2,26 3,11		
Lys^{15}						2,09 2,49	2,01	
Cys^{16}							2,65 2,00	1,45
Arg^{17}	7						2,53	2,15
Asn^{18}		1	6					
Cys^{19}			3	5				
Cys^{20}				7				

* В правом верхнем углу таблицы указаны минимальное и максимальное расстояния (Å) между амидным атомом водорода и карбонильным атомом кислорода основной цепи, слева внизу — число конформаций с водородными связями, отвечающими критериям, указанным в тексте [32]. Система внутримолекулярных водородных связей между карбонильной группой i -го остатка и амидной группой $(i+k)$ -го остатка отвечает спирали типа 3_{10} при $k=3$, α -спирали при $k=4$.

двумя способами: по координатам [33] и по расстояниям [34]. Сравнение проводили для четырех характеристических выборок атомов (см. табл. 5). Результаты сравнения показывают, что все конформации по рассмотренным выборкам мало различаются как по расстояниям, так и по взаимной ориентации и расположению атомов. Это, а также малые величины стандартных отклонений значений двугранных углов основной цепи ϕ и ψ (табл. 6) свидетельствуют о сходстве пространственного строения основной цепи и ориентации боковых цепей в сравниваемых конформациях фрагмента $\text{Ala}^{13} - \text{Cys}^{20}$.

Оценку выигрыша в точности определения пространственной структуры фрагмента при расчете с учетом КССВ дают, например, средние величины среднеквадратичных отклонений между конформациями, полученными в расчете без учета КССВ: 0,5 Å для выборки 2 и 1,3 Å для выборки 4 в отличие от 0,3 и 1,1 Å соответственно для конформаций, полученных в расчете с учетом КССВ (см. табл. 5).

Значения двугранных углов ϕ и ψ основной цепи (табл. 6) свидетельствуют о том, что аминокислотные остатки фрагмента $\text{Ala}^{13} - \text{Cys}^{20}$ образуют правую спираль (см. рис. 4) во всех рассматриваемых конформациях.

Идентификация внутримолекулярных водородных связей, образуемых атомами основной цепи, проведена по критериям [32]: $1,45 \text{ \AA} \leq d(\text{O} \dots \text{H}) \leq 2,45 \text{ \AA}$; $105^\circ \leq \text{угол } \text{O} \dots \text{H}-\text{N} \leq 180^\circ$; $85^\circ \leq \text{угол } \text{C}=\text{O} \dots \text{H} \leq 180^\circ$. Обна-

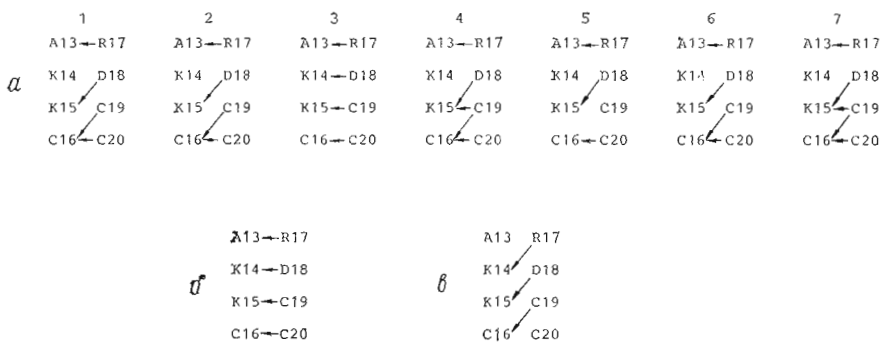


Рис. 5. Схематическое представление водородных связей: *a* — в 7 конформациях фрагмента Ala¹³—Cys²⁰ инсектотоксина I₃A (по критериям [32]); *б* — в идеальной α-спирали; *в* — в Z₁₀-спирали. Стрелками указано направление водородной связи от амидной группы NH к карбонильной группе соответствующих остатков, изображенных в однобуквенном коде

ружено, что в конформациях фрагмента могут образовываться системы водородных связей C_i=O...H—N_{i+k} с *k*, равным 4 (характерно для α-спирали) и 3 (характерно для спирали типа Z₁₀). Конформации отличаются друг от друга системами водородных связей (рис. 5, табл. 7), несмотря на то что среднеквадратичные отклонения по координатам атомов (табл. 5) находятся в пределах неопределенности, ожидаемой для тепловой флуктуации атомов полипептидной цепи белка из расчетов по моделированию молекулярной динамики (см., например, [27, 28]).

Из анализа вторичной структуры ряда кристаллографически исследованных белков следует [35], что реально существующие спирали оказываются промежуточными между α- и Z₁₀-спиралями, причем во многих структурах наиболее искажена в этом смысле конформация аминокислотных остатков, расположенных на концах спиралей. Эти остатки образуют предпочтительно водородные связи, характерные для спирали типа Z₁₀ (*k*=3) (рис. 5). Показательно, что молекулярно-динамические расчеты в пикосекундном диапазоне свидетельствуют о возможности перестройки систем водородных связей в спиральных структурах [27]. Дальнейший анализ и уточнение пространственной структуры может быть проведен, например, расчетами потенциальной энергии и моделированием динамического поведения молекулы.

Полученные в результате расчета конформационные структуры фрагмента Ala¹³—Cys²⁰ инсектотоксина I₃A, удовлетворяющие всей совокупности экспериментальных данных ЯМР, говорят о корректности предложенной в данной работе методики учета в расширенной версии дистанционного геометрического алгоритма значений КССВ и данных о дейтериеобмене амидных протонов, а также обоснованности примененного способа оценки вводимых ограничений расстояний. Значения среднеквадратичных отклонений, характеризующие различия между конформациями, отражают, с одной стороны, погрешности экспериментальных данных ЯМР-спектрологии о пространственной структуре молекулы (в особенности это относится к ЯЭО-контактам) и, возможно, неполную адекватность методики введения их в расчет. С другой стороны, различия между конформациями могут быть обусловлены динамической природой поведения молекулы белка в растворе, вследствие чего измеряемые ЯМР-параметры представляют собой усредненные характеристики отдельных состояний динамической структуры.

Основным недостатком разработанной модификации дистанционного геометрического алгоритма для построения пространственной структуры белков и пептидов по данным спектроскопии ЯМР является относительно небольшое число центров (атомов и псевдоатомов), которые можно учесть в расчете (≤150 [2]). Этот предел обусловлен использованной процедурой получения стартовых приближений рассчитываемой структуры через описанный в [2] формализм метрической *G*-матрицы. Такое ограничение

препятствует детальному исследованию белковых молекул, для которых современная 2D-ЯМР-спектроскопия обеспечивает получение экспериментальных данных об их пространственном строении. Увеличение числа центров, рассматриваемых с помощью дистанционного геометрического алгоритма, и построение конформации белковой молекулы в атомном приближении — цель нашей дальнейшей работы.

Авторы выражают благодарность д-ру В. Брауну (Цюрих, Швейцария) за предоставление программы дистанционного геометрического алгоритма [2].

ЛИТЕРАТУРА

1. *Crippen G. M.* Distance geometry and conformational calculations. Chichester: Research Studies Press, 1981.
2. *Braun W., Bosh C., Brown L. R., Co N., Wuthrich K.* Biochim. et biophys. acta, 1981, v. 667, № 22, p. 377–396.
3. *Crippen G. M., Oppenheimer N. J., Connolly M. L.* Int. J. Peptide and Protein Res., 1981, v. 17, № 2, p. 156–169.
4. *Арсеньев А. С., Кондаков В. Н., Майоров В. Н., Быстров В. Ф., Овчинников Ю. А.* Биоорг. химия, 1983, т. 9, № 12, с. 1667–1689.
5. *Braun W., Wider G., Lee K. H., Wuthrich K. J.* Mol. Biol., 1983, v. 169, № 4, p. 921–948.
6. *Havel T., Wuthrich K.* Bull. Math. Biol., 1984, v. 46, № 4, p. 673–698.
7. *Deisenhofer J., Steigemann W.* Acta cryst., 1975, v. B31, № 1, p. 238–250.
8. *Anil Kumar, Wagner G., Ernst R. R., Wuthrich K. J.* Amer. Chem. Soc., 1981, v. 103, № 13, p. 3654–3658.
9. *Bystrov V. F.* Progress in nuclear magnetic resonance spectroscopy. 1976, v. 10, № 2, p. 41–81.
10. *Wagner G.* Quart. Rev. Biophys., 1983, v. 16, № 1, p. 1–57.
11. *Арсеньев А. С., Кондаков В. Н., Майоров В. Н., Волкова Т. М., Гришин Е. В., Быстров В. Ф., Овчинников Ю. А.* Биоорг. химия, 1983, т. 9, № 6, с. 768–793.
12. *Karplus M. J.* Chem. Phys., 1959, v. 30, № 1, p. 11–15.
13. *Cung M. T., Marraud M.* Biopolymers, 1982, v. 21, № 4, p. 953–967.
14. IUPAC — IUB Commission on biochemical nomenclature.— Biochim. et biophys. acta, 1971, v. 229, № 1, p. 1–17.
15. *Biletter M., Braun W., Wuthrich K. J.* Mol. Biol., 1982, v. 155, № 3, p. 321–346.
16. *Momany F. A., McGuire R. F., Burgess A. W., Scheraga H. A. J.* Phys. Chem., 1975, v. 79, № 22, p. 2361–2381.
17. *Nemethy G., Pottle M. S., Scheraga H. A. J.* Phys. Chem., 1983, v. 87, № 11, p. 1883–1887.
18. *Chuman H., Momany F. A., Schafer L.* Int. J. Peptide and Protein Res., 1984, v. 24, № 3, p. 233–248.
19. *Chuman H., Momany F. A.* Int. J. Peptide and Protein Res., 1984, v. 24, № 3, p. 249–266.
20. *Nagayama K., Wuthrich K.* Eur. J. Biochem., 1981, v. 115, № 3, p. 653–657.
21. *Wyssbrod H. R.* In: Peptides. Synthesis — structure — function (Proc. 7th Amer. Pept. Symp.) / Eds Rich D. H., Gross E. Rockford, Illinois: Pierce Chem. Comp., 1981, v. 261–265.
22. *Bada J. L., Miller S. L. J.* Amer. Chem. Soc., 1970, v. 92, № 9, p. 2774–2782.
23. *Pardi A., Biletter M., Wuthrich K. J.* Mol. Biol., 1984, v. 180, № 3, p. 741–751.
24. *Delepierre M., Dodson C. M., Howarth M. A., Poulsen F. M.* Eur. J. Biochem., 1984, v. 145, № 2, p. 389–395.
25. *Wagner G., Wuthrich K. J.* Mol. Biol., 1982, v. 155, № 3, p. 347–366.
26. *Van Gunsteren W. F., Karplus M.* Nature, 1981, v. 293, № 5834, p. 677–678.
27. *Lewitt M. J.* Mol. Biol., 1983, v. 168, № 3, p. 595–620.
28. *Lewitt M. J.* Mol. Biol., 1983, v. 168, № 3, p. 621–657.
29. *Olejniczak E. T., Dodson C. M., Karplus M., Levy R. M. J.* Amer. Chem. Soc., 1984, v. 106, № 7, p. 1923–1930.
30. *Wagner G., Wuthrich K. J.* Mol. Biol., 1982, v. 160, № 2, p. 343–361.
31. *Khaled M. A., Watkins C. L. J.* Amer. Chem. Soc., 1983, v. 105, № 10, p. 3363–3365.
32. *Hohne E., Kretschmer G.* Stud. Biophys., 1982, v. 87, № 1, p. 23–28.
33. *McLachlan A. D. J.* Mol. Biol., 1979, v. 128, № 1, p. 49–79.
34. *Oobatake M., Crippen G. M. J.* Phys. Chem., 1981, v. 85, № 4, p. 1187–1197.
35. *Бланделл Т., Джонсон Л.* Кристаллография белка. М.: Мир, 1979.

Поступила в редакцию
4.III.1985

MODIFICATION OF DISTANCE GEOMETRY ALGORITHM FOR POLYPEPTIDE SPATIAL STRUCTURE CALCULATION FROM NMR SPECTROSCOPY DATA

MAIOROV V. N., ARSENIYEV A. S., BYSTROV V. F.

M. M. Shemyakin Institute of Bioorganic Chemistry, Academy of Sciences of the USSR, Moscow

A special supplement to the distance geometry algorithm is developed to take into account in polypeptide spatial structure simulation the NMR data indirectly related to the distance constraints — spin-spin coupling constants of the vicinal H—NC^α—H and H—C^αC^β—H protons and deuterium exchange rates for backbone amide NH groups. Moreover, the coupling constants are used to calibrate empirically the nuclear Overhauser effect intensities, thus increasing their usefulness for distance geometry calculation. A test of the modified algorithm in the conformational analysis of the helical Ala¹³—Cys²⁰ fragment of *Buthus eupeus* insectotoxin I₅A proves the adequacy of the proposed method for distance constraints estimation. The comparison of the fragment conformations generated through random starts demonstrates better precision (root — mean — square deviations are equal to 0,3 Å for backbone atoms and 1,1 Å for all atoms) than for conformations calculated without spin-spin coupling constants data (0,5 and 1,3 Å, respectively).