



УДК 547.993.02+541.63

АПРИОРНЫЙ РАСЧЕТ ТРЕХМЕРНОЙ СТРУКТУРЫ  
ИНСЕКТОТОКСИНА И<sub>1</sub> ИЗ ЯДА СКОРПИОНА *BUTHUS EUROPEUS*

Попов Е. М., Швырков В. Н., Спасов В. З.

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

Расчитано пространственное строение инсектотоксина И<sub>1</sub> -- пептидного компонента яда скорпиона *Buthus europaeus*. Молекула состоит из 36 аминокислотных остатков и включает в себя восемь остатков цистеина, образующих четыре дисульфидные связи, локализация которых экспериментально не определена. Конформационный анализ выполнен по известной аминокислотной последовательности инсектотоксина при неполной информации о его ковалентной структуре. Установлены конформационные параметры основных и боковых цепей всех остатков, количественно оценены конформационные возможности и идентифицирована система дисульфидных связей в трехмерной структуре молекулы.

Инсектотоксин И<sub>1</sub> является компонентом яда скорпиона *Buthus europaeus* [1]. Его аминокислотная последовательность состоит из 36 остатков, в том числе восьми остатков Cys [2]. Экспериментальные данные относительно пространственного строения инсектотоксина и его гомологов отсутствуют. Более того, неизвестна даже реализующаяся в молекуле система четырех дисульфидных связей. Поэтому конформационный анализ инсектотоксина выполнен при неполной информации о ковалентной структуре.

Помимо установления структуры молекулы данное исследование имело и более общую задачу, заключающуюся в апробации физической модели и расчетной схемы априорного предсказания пространственной организации природных пептидов и белков [3–5]. Выбор в качестве объекта анализа инсектотоксина продиктован главным образом тем, что расчет его трехмерной структуры при использовании одной только аминокислотной последовательности уже включал в себя надежный внутренний контроль. Согласно конформационной теории белка [3–5] и термодинамической гипотезе Анфинсена [6], анализ линейной последовательности инсектотоксина должен автоматически и в наиболее предпочтительных нециклических конформационных состояниях привести к сближенности соответствующие остатки Cys и тем самым локализовать четыре цистеиновые пары. При многостадийном расчете такого сложного пептида чрезвычайно мала вероятность случайной идентификации единственной системы дисульфидных связей, удовлетворяющей всем положениям теории и принципам образования S–S-мостиков.

Потенциальные функции, параметризация и метод минимизации, используемые для количественной оценки невалентных и электростатических взаимодействий атомов, водородных связей и торсионного вклада, указаны в работах [4, 7]. Там же даны объяснения употребляемым в тексте понятиям и обозначениям.

Полная схема расчета пространственного строения инсектотоксина показана на рис. 1. Первый шаг на пути к установлению трехмерной структуры молекулы заключался в рассмотрении конформационных возможностей относительно коротких фрагментов, содержащих все восемь остатков цистеина, Met<sup>1</sup>–Thr<sup>3</sup>, Cys<sup>16</sup>–Lys<sup>21</sup> и Cys<sup>26</sup>–Cys<sup>33</sup>. На этом этапе конформационный анализ инсектотоксина должен был ответить на вопрос о предрасположенности к образованию дисульфидных связей выделенных фрагментов. Используя полученные здесь результаты, мы исследовали за-

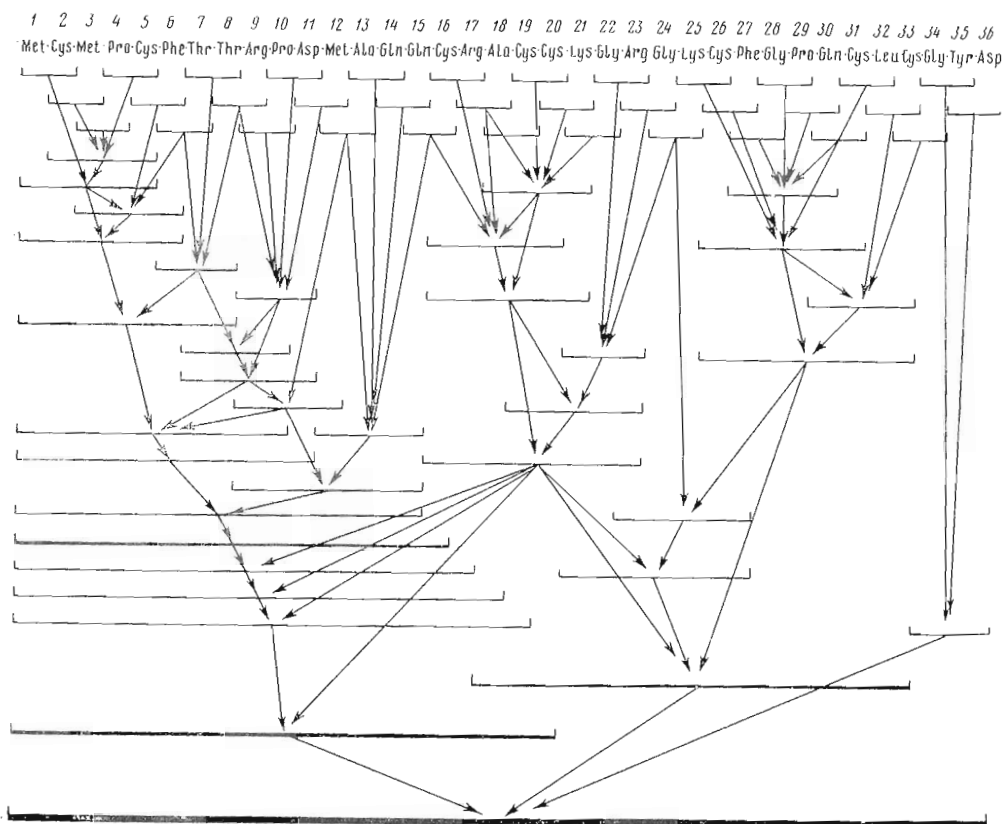


Рис. 1. Схема расчета пространственного строения инсектотоксина II<sub>1</sub>

тем более сложные участки молекулы Met<sup>1</sup>—Cys<sup>20</sup> и Ala<sup>18</sup>—Cys<sup>33</sup> и, наконец, всю последовательность. Исходными в анализе инсектотоксина служили наборы оптимальных конформаций свободных метиламидов N-ацетил- $\alpha$ -аминокислот [3—5]. На их основе были рассчитаны конформационные состояния перекрывающихся дипептидных фрагментов. Низкоэнергетические варианты всех возможных форм основной цепи и шейпов дипептидов использованы в расчете более сложных участков молекулы. Ниже мы остановимся на некоторых узловых моментах конформационного анализа инсектотоксина.

Завершающая стадия расчета октапептида Met<sup>1</sup>—Thr<sup>8</sup> заключалась в рассмотрении 130 исходных приближений 18 различных форм основной цепи. Минимизация выявила резкую дифференциацию конформаций по энергии; в широкий энергетический интервал  $E_{отн} = 0-10$  ккал/моль попали лишь восемь родственных структурных вариантов. Во всех случаях боковые цепи остатков Cys<sup>2</sup> и Cys<sup>5</sup> находятся далеко друг от друга и принудительное сближение атомов S сопровождается значительным повышением энергии.

Исходные 354 структурных варианта гексапептида Cys<sup>16</sup>—Lys<sup>21</sup> сформированы из конформационных состояний перекрывающихся по трем остаткам фрагментов Cys<sup>16</sup>—Cys<sup>20</sup> и Ala<sup>18</sup>—Lys<sup>21</sup> с  $E_{отн} \leq 6,0$  ккал/моль (рис. 1). Они имеют 76 форм основной цепи, распределяющихся по всем в принципе возможным для гексапептида 32 шейпам. Минимизация энергии при варьировании двугранных углов  $\phi$ ,  $\psi$  и  $\chi$  свидетельствует об отсутствии очевидного преимущества какой-либо формы или шейпа пептидного скелета. Относительной энергией от 0 до 8,0 ккал/моль обладают 118 конформаций 36 форм основной цепи 20 различных типов. Сближенность атомов S<sub>(16)</sub> и S<sub>(20)</sub> наблюдается у конформаций следующих трех форм: B<sup>16</sup>-R<sup>17</sup>-R<sup>18</sup>-B<sup>19</sup>-B<sup>20</sup> (effe), B<sup>16</sup>-B<sup>17</sup>-L<sup>18</sup>-R<sup>19</sup>-B<sup>20</sup> (efef) и B<sup>16</sup>-B<sup>17</sup>-B<sup>18</sup>-L<sup>19</sup>-B<sup>20</sup> (eefe) [8]. Однако возможность образования дисульфидной связи S<sub>(16)</sub>—

— $S_{(20)}$  здесь чрезвычайно мала. Все варианты первых двух форм имеют высокую энергию, а у представителей третьей формы создание S—S-мостика неизбежно сталкивается с необходимостью изменения положения боковой цепи Arg<sup>17</sup>, что ведет к резкому повышению энергии. Конформации последнего типа низкоэнергетичны только у линейной последовательности. Циклы с дисульфидной связью между Cys<sup>16</sup> и Cys<sup>19</sup> может быть создан на основе структуры с формами B<sup>16</sup>-B<sup>17</sup>-L<sup>18</sup>-B<sup>19</sup> (*eje*) и B<sup>16</sup>-R<sup>17</sup>-R<sup>18</sup>-R<sup>19</sup> (*eff*) [8]. Состояния шейпа *eje* имеют высокую энергию, а образование мостика в случае *eff* сопровождается нарушением стабилизирующих контактов и даже возникновением неблагоприятных взаимодействий. Поэтому мостик S<sub>(16)</sub>—S<sub>(19)</sub> нереален, как и S<sub>(16)</sub>—S<sub>(20)</sub>. Невозможно образование дисульфидной связи и между смежными остатками Cys<sup>19</sup> и Cys<sup>20</sup> [8, 9]. На малую вероятность S—S-валентного связывания на участке инсектотоксина Cys<sup>16</sup>—Lys<sup>21</sup> указывает также наличие незначительной энергетической дифференциации конформаций.

Иная ситуация имеет место у фрагмента Cys<sup>26</sup>—Cys<sup>33</sup>. В его расчете были использованы конформации гексапептида Cys<sup>26</sup>—Cys<sup>31</sup> всех шейпов, за исключением *eeffe* и *eefff*, имеющих очень высокую энергию, и конформации тетрапептида Gln<sup>30</sup>—Cys<sup>33</sup> с  $E_{отн} \leq 5,0$  ккал/моль. Минимизация 100 исходных структур 64 шейпов пептидного скелета участка Cys<sup>26</sup>—Cys<sup>33</sup> выявила их резкое различие в стабильности. В интервал 0—5,0 ккал/моль попадает всего пять конформаций, причем четыре из них, самые предпочтительные ( $E_{отн} = 0-1,3$  ккал/моль), имеют идентичное конформационное состояние на участке Cys<sup>26</sup>—Cys<sup>31</sup>, отвечающее глобальной структуре свободного гексапептида. В самых выгодных конформациях линейного октапептида Cys<sup>26</sup>—Cys<sup>33</sup> участок Cys<sup>26</sup>—Cys<sup>31</sup> представляет собой жесткую пуклеацию, а Cys<sup>31</sup>—Cys<sup>33</sup> — лабильную часть. В нуклеации расстояние между атомами S боковых цепей Cys<sup>26</sup> и Cys<sup>31</sup> составляет ~3,6 Å. Образование между ними валентной связи S—S, т. е. сближение до 2,04 Å [10] сопровождается незначительным изменением геометрии, которое не приводит к нарушению стабилизирующих невалентных взаимодействий, сложившихся в оптимальной конформации линейной последовательности Cys<sup>26</sup>—Cys<sup>33</sup>. В то же время создание S—S-связи между Cys<sup>26</sup> и Cys<sup>33</sup>, как и между Cys<sup>31</sup> и Cys<sup>33</sup>, стерически невозможно. Таким образом, конформационный анализ фрагментов Met<sup>1</sup>—Thr<sup>8</sup>, Cys<sup>16</sup>—Lys<sup>21</sup> и Cys<sup>26</sup>—Cys<sup>33</sup> сократил количество возможных у инсектотоксина систем дисульфидных связей со 105 до следующих шести (приведены номера остатков Cys; рис. 1):

I	2-16	5-19	20-33	26-31
II	2-16	5-20	19-33	26-31
III	2-19	5-16	20-33	26-31
IV	2-20	5-16	19-33	26-31
V	2-19	5-20	16-33	26-31
VI	2-20	5-19	16-33	26-31

Следующим этапом анализа инсектотоксина явилось рассмотрение конформационных состояний гексадекапептидного фрагмента, Ala<sup>18</sup>—Cys<sup>33</sup> и оценка возможностей образования дисульфидной связи между Cys<sup>33</sup> и Cys<sup>19</sup> или Cys<sup>20</sup>. Ввиду сложности центрального участка Lys<sup>21</sup>—Lys<sup>25</sup> расчет этого фрагмента включал в себя большое число стадий (рис. 1). На последней стадии было рассчитано более 100 конформационных состояний Ala<sup>18</sup>—Cys<sup>33</sup>, представляющих собой комбинации низкоэнергетических вариантов Cys<sup>16</sup>—Lys<sup>21</sup>, Lys<sup>21</sup>—Phe<sup>27</sup> и Cys<sup>26</sup>—Cys<sup>33</sup>. В интервал 0—5,0 ккал/моль попало только три, а в интервал 0—10,0 ккал/моль — 18 конформаций. В четырех из них атом C<sup>β</sup> остатка Cys<sup>33</sup> сближен за счет невалентных взаимодействий менее чем на 10 Å с атомом C<sup>β</sup> остатка Cys<sup>19</sup> или Cys<sup>20</sup> (согласно экспериментальным данным [10], это расстояние при наличии дисульфидной связи составляет 3,8 Å при угле  $\chi^{S-S} \pm 90^\circ$  и 4,8 Å при  $\chi^{S-S} = 180^\circ$ ). Одна из этих конформаций имеет низкую энергию ( $E_{отн} = 2,0$  ккал/моль), а три других — более высокую (6,5—8,0 ккал/моль). Все попытки образовать у них, как и у других конформаций с  $E_{отн} \leq 10$  ккал/моль, дисульфидную связь между остатками Cys<sup>33</sup> и Cys<sup>19</sup> оказались безуспешными. Сближение атомов S<sub>(33)</sub> и S<sub>(19)</sub> на валентное расстояние приводило к

резкому увеличению энергии или вследствие рассогласованности ранее сложившихся стабилизирующих взаимодействий, или из-за возникновения неблагоприятных контактов. Аналогичная ситуация имела место во всех случаях, кроме одного — при образовании дисульфидной связи между остатками Cys<sup>33</sup> и Cys<sup>20</sup>. Единственным исключением явилась конформация гексадекапептида с энергией  $E_{отн}$  2,0 ккал/моль. Ее зацикливание приводит к такой бициклической структуре ( $S_{(20)}-S_{(33)}$ ;  $S_{(26)}-S_{(31)}$ ), абсолютная энергия которой (-103,5 ккал/моль) очень близка энергии исходной конформации (-105,2 ккал/моль). Из сравнения межостаточных взаимодействий и геометрии остатков в би- и моноциклической конформациях Ala<sup>18</sup>-Cys<sup>33</sup> следует, что, во-первых, образование второго S-S-мостика не нарушило стабилизирующих контактов предшествующей конформации и, во-вторых, отклонения двугранных углов основной цепи коснулись лишь C-концевого наиболее лабильного трипептидного участка Cys<sup>31</sup>-Cys<sup>33</sup>. На этом участке локализуются и делаются низкоэнергетическими все изменения пептидной цепи, которые неизбежны при сближении атомов S на валентное расстояние. И наконец, образование связи  $S_{(33)}-S_{(20)}$ , как и  $S_{(26)}-S_{(31)}$ , произошло по строго детерминированному механизму, в основе которого лежат конформационные аспекты. Такой механизм обусловлен предрасположенностью фрагмента к формообразованию, автоматически приводящему к сближенности определенных остатков Cys. Результаты расчета фрагмента Ala<sup>18</sup>-Cys<sup>33</sup> уменьшили число возможных для молекулы инсектотоксина систем дисульфидных связей с шести до двух:

I	2-16	5-19	20-33	26-31
II	2-19	5-16	20-33	26-31

В конформационном анализе фрагмента Met<sup>1</sup>-Cys<sup>20</sup>, однако, нами не принята во внимание возможность образования дисульфидной связи между остатками Cys<sup>20</sup> и Cys<sup>33</sup>. Была рассмотрена только вероятность образования шести S-S-мостиков между остатками Cys<sup>2</sup>, Cys<sup>3</sup> и Cys<sup>16</sup>, Cys<sup>19</sup>, Cys<sup>20</sup>. Остановимся сначала на результатах расчета фрагмента Met<sup>1</sup>-Cys<sup>16</sup>, в пределах которого возможна дисульфидная связь Cys<sup>2</sup>-Cys<sup>16</sup> или Cys<sup>3</sup>-Cys<sup>16</sup>. Расчету этого участка инсектотоксина предшествовало детальное рассмотрение конформационных возможностей целого ряда фрагментов меньшей длины (рис. 1). Кроме отмеченного выше октапептида Met<sup>1</sup>-Thr<sup>8</sup> были рассчитаны Thr<sup>7</sup>-Pro<sup>10</sup> (646), Thr<sup>7</sup>-Asp<sup>11</sup> (495), Arg<sup>9</sup>-Met<sup>11</sup> (1135), Met<sup>12</sup>-Gln<sup>15</sup> (1051), Met<sup>1</sup>-Pro<sup>10</sup> (148) и Met<sup>1</sup>-Asp<sup>11</sup> (48), Arg<sup>9</sup>-Gln<sup>15</sup> (32) и Met<sup>1</sup>-Gln<sup>15</sup> (32) (в скобках указано количество рассмотренных структурных вариантов данного фрагмента).

По мере увеличения длины цепи от окта- до гексадекапептида неизменно возрастала энергетическая дифференциация конформаций и, как следствие, уменьшалось число подлежащих анализу структур. Фактически однозначная детерминация по форме и шейпу пептидного скелета, т. е. образование конформационно жесткой нуклеации, обнаруживается у декапептида Met<sup>1</sup>-Pro<sup>10</sup>. Из 148 рассчитанных для фрагмента конформационных состояний в интервал 0-6,0 ккал/моль попадают только восемь вариантов, принадлежащих шейпу *efefefffe* и имеющих одну и ту же форму основной цепи  $B^1-R^2-B^3-R^4-B^5-R^6-R^7-R^8-B^9-R^{10}$ . К расчету гексадекапептида Met<sup>1</sup>-Cys<sup>16</sup> мы подошли с набором из девяти исходных приближений, которые различались лишь состояниями C-концевого дипептида Gln<sup>15</sup>-Cys<sup>16</sup>. Минимизация энергии выявила всего три предпочтительных линейных конформации шейпа *efefefffefffff*. В самой низкоэнергетической из них актуальные для дисульфидной связи расстояния имели следующие значения (Å):  $C_{(2)}^{\beta} \dots C_{(16)}^{\beta} - 7,8$ ,  $S_{(2)} \dots S_{(16)} - 3,4$  и  $C_{(5)}^{\beta} \dots C_{(16)}^{\beta} - 7,0$ ,  $S_{(5)} \dots S_{(16)} - 10,9$ . Образовать дисульфидную связь удалось лишь между остатками Cys<sup>2</sup> и Cys<sup>16</sup>. Все попытки сблизить остатки Cys<sup>3</sup> и Cys<sup>16</sup> приводили к многочисленным наталкиваниям и рассогласованности взаимодействий задолго до достижения расстояния между атомами S<sub>(5)</sub> и S<sub>(16)</sub>, равного длине валентной связи. В то же время образование связи S<sub>(2)</sub>-S<sub>(16)</sub> не нарушило ни одного стабилизирующего контакта; величины

Энергия (ккал/моль) в внутри- и межостаточных взаимодействиях в глобальной конформации фрагмента Met<sup>1</sup>-Cys<sup>16</sup>

	Met <sup>1</sup>	Cys <sup>2</sup>	Met <sup>3</sup>	Pro <sup>4</sup>	Cys <sup>5</sup>	Phe <sup>6</sup>	Thr <sup>7</sup>	Thr <sup>8</sup>	Arg <sup>9</sup>	Pro <sup>10</sup>	Asp <sup>11</sup>	Met <sup>12</sup>	Ala <sup>13</sup>	Gln <sup>14</sup>	Gln <sup>15</sup>	Cys <sup>16</sup>
Met <sup>1</sup>	-6,2	-2,3	-2,4	-2,6	0,0	0,0	-0,1	0,0	1,3	0,0	-1,4	-0,1	0,0	0,0	-0,1	-0,1
Cys <sup>2</sup>	-1,8	-0,9	-0,8	-0,8	0,0	0,0	0,0	-0,1	-0,3	-0,1	0,2	-0,5	-0,1	-0,1	-0,1	0,0
Met <sup>3</sup>			-3,6	-1,4	-1,9	-0,7	-0,1	-0,4	-2,0	-0,2	-0,1	-3,6	-2,4	-0,1	-0,2	-1,8
Pro <sup>4</sup>			iro	-2,9	-1,4	-0,2	-0,3	-3,5	-1,0	0,1	-0,5	-0,2	-0,1	0,0	0,0	0,0
Cys <sup>5</sup>				Cys <sup>5</sup>	-1,1	-0,9	-3,1	-3,6	-1,1	-0,4	0,2	-0,1	-0,7	-0,1	-0,1	-0,1
Phe <sup>6</sup>					Phe <sup>6</sup>	-3,9	-1,5	-1,0	-1,5	-5,5	0,1	-0,2	-2,6	-0,3	-0,1	-0,1
Thr <sup>7</sup>						Thr <sup>7</sup>	-3,7	-0,7	-0,8	-1,3	0,1	-0,1	-0,1	-0,1	-0,1	-0,1
Thr <sup>8</sup>							Thr <sup>8</sup>	-4,4	-3,8	-1,9	0,2	-0,2	-0,1	-0,1	-0,1	-0,1
Arg <sup>9</sup>								Arg <sup>9</sup>	-5,0	0,1	-13,5	-6,1	-1,0	0,1	0,1	0,0
Pro <sup>10</sup>									Pro <sup>10</sup>	-2,4	-3,2	-1,8	-3,2	-1,8	-0,1	-0,1
Asp <sup>11</sup>										Asp <sup>11</sup>	-2,2	-3,3	-1,2	-1,8	-1,4	-0,1
Met <sup>12</sup>											Met <sup>12</sup>	-4,4	-1,1	1,6	-2,9	-1,2
Ala <sup>13</sup>												Ala <sup>13</sup>	-1,2	0,4	-2,0	-1,8
Gln <sup>14</sup>													Gln <sup>14</sup>	-4,7	-4,0	-0,6
Gln <sup>15</sup>														Gln <sup>15</sup>	-5,5	-1,5
Cys <sup>16</sup>															Cys <sup>16</sup>	-0,9

Примечание. Центральная диагональ отвечает энергии ближних взаимодействий, следующие четыре диагонали – энергии средних и остальная часть матрицы – энергии дальних взаимодействий.

энергии невалентных взаимодействий у глобальной конформации линейной последовательности Met<sup>1</sup>-Cys<sup>16</sup> и соответствующей циклической структуры с дисульфидной связью S<sub>(2)</sub>-S<sub>(16)</sub>, практически совпадают. В табл. 1 приведена матрица энергетических составляющих внутри- и межостаточных взаимодействий, которая практически в равной мере характеризует линейную и циклическую конформации гексапептида.

Найденная конформация фрагмента Met<sup>1</sup>-Cys<sup>16</sup> имеет форму основной цепи B<sup>1</sup>-R<sup>2</sup>-B<sup>3</sup>-R<sup>4</sup>-B<sup>5</sup>-R<sup>6</sup>-R<sup>7</sup>-R<sup>8</sup>-B<sup>9</sup>-R<sup>10</sup>-R<sup>11</sup>-R<sup>12</sup>-R<sup>13</sup>-R<sup>14</sup>-R<sup>15</sup>-B<sup>16</sup> (efefeffeffeffeff), которая обеспечивает сближенность большинства остатков и реализацию между ними эффективных стабилизирующих взаимодействий. В треугольной матрице (табл. 1) выделены значения энергии ближних, средних и дальних взаимодействий; их суммарные вклады в общую конформационную энергию фрагмента (-176,0 ккал/моль) имеют стабилизирующий характер и составляют соответственно -53,9, -102,6 и -19,6 ккал/моль. Конформационные состояния всех остатков отвечают самым предпочтительным оптимальным конформациям свободных монопептидов. Наиболее эффективны в найденной конформации гексадекапептида средние взаимодействия, т. е. взаимодействия каждого остатка с четырьмя предшествующими и четырьмя последующими соседями по цепи. Форма основной цепи обеспечивает сближенность всех остатков и практически полную реализацию средних взаимодействий в пределах каждого перекрывающегося по восьми остаткам нонпептидного участка (Met<sup>1</sup>-Arg<sup>9</sup>, Cys<sup>2</sup>-Pro<sup>10</sup>, Met<sup>3</sup>-Asp<sup>11</sup> и г. д.). Если не считать образования дисульфидной связи S<sub>(2)</sub>-S<sub>(16)</sub>, суммарный эффект дальних взаимодействий сравнительно невелик; наиболее значительные стабилизирующие контакты образуют -Met<sup>3</sup>- с -Arg<sup>9</sup>-, -Met<sup>12</sup>-, -Ala<sup>13</sup>- и -Phe<sup>6</sup>- с -Ala<sup>13</sup>- (табл. 1). В среднем на остаток энергия межостаточных взаимодействий составляет ~-7,6 ккал/моль (~-6,4 и -1,2 ккал/моль соответственно за счет средних и дальних взаимодействий).

Конформационные возможности следующего крупного участка инсектотоксина Met<sup>1</sup>-Cys<sup>20</sup> были исследованы путем поостаточной генерации фрагмента Met<sup>1</sup>-Cys<sup>16</sup> (рис. 1). Присоединяемые остатки в каждом случае рассматривались в поле жесткой циклической структуры гексадекапептида в состояниях, отвечающих низкоэнергетическим структурным вариантам фрагмента Cys<sup>16</sup>-Cys<sup>20</sup>. В свободном состоянии пентапептид обладает большой конформационной свободой, и поэтому были опробованы все возможные для него формы основной цепи. Минимизация энергии показала, что структуре Met<sup>1</sup>-Cys<sup>16</sup> с дисульфидной связью S<sub>(2)</sub>-S<sub>(16)</sub>



комплементарна лишь одна форма фрагмента  $\text{Cys}^{16}-\text{Cys}^{20}$ , а именно  $B^{16}-B^{17}-R^{18}-R^{19}-R^{20}$ . Только в этом случае имеет место полная компенсация отрицательной и положительной энергии электростатических взаимодействий  $-\text{Arg}^{17}$ - соответственно с  $-\text{Asp}^{11}$ - ( $-2,78$  ккал/моль) и с  $-\text{Met}^1$ -,  $-\text{Arg}^9$ - ( $+2,70$  ккал/моль), а также образование целого ряда стабилизирующих контактов между удаленными по цепи остатками. Наиболее существенны из них взаимодействия  $-\text{Met}^3$ - и  $-\text{Ala}^{13}$ - с  $\text{Cys}^{19}-\text{Cys}^{20}$  ( $-9,5$  ккал/моль),  $-\text{Phe}^6$ - с  $\text{Cys}^{19}-\text{Cys}^{20}$  ( $-2,6$  ккал/моль) и  $-\text{Pro}^4$ - с  $-\text{Cys}^{19}$ - ( $-1,0$  ккал/моль). Интересно, что именно в глобальной конформации участка  $\text{Met}^1-\text{Cys}^{20}$  сближенными автоматически оказываются боковые цепи  $-\text{Cys}^5$ - и  $-\text{Cys}^{19}$ -, для которых образование дисульфидной связи не ведет к заметной дестабилизации ранее сложившихся невалентных взаимодействий. В принципе возможен и мостик  $S_{(5)}-S_{(20)}$ , однако его образование сопровождается повышением энергии более чем на  $10$  ккал/моль. Таким образом, независимое рассмотрение конформационных возможностей фрагментов инсектотоксина  $\text{Met}^1-\text{Cys}^{20}$  и  $\text{Ala}^{18}-\text{Cys}^{33}$  приводит к непротиворечивым заключениям о предрасположенности первого к образованию дисульфидной связи  $S_{(5)}-S_{(19)}$ , а второго — к  $S_{(20)}-S_{(33)}$ .

Завершающий этап расчета пространственного строения инсектотоксина свелся к упаковке двух жестких структур (перекрывающихся и связанных между собой мостиками  $S_{(5)}-S_{(19)}$  и  $S_{(20)}-S_{(33)}$  фрагментов  $\text{Met}^1-\text{Cys}^{20}$  и  $\text{Ala}^{18}-\text{Cys}^{33}$ ) и к определению конформационного состояния С-концевого трипептида  $\text{Gly}^{34}-\text{Asp}^{36}$  (рис. 1). При объединении двух половин молекулы в единую структуру возникли некоторые стерические осложнения. Тем не менее рассчитанные конформации свободных фрагментов  $\text{Met}^1-\text{Cys}^{20}$  и  $\text{Ala}^{18}-\text{Cys}^{33}$  следует считать комплементарными друг другу, поскольку: 1) все наталкивания снимаются при незначительной коррекции двугранных углов нескольких боковых цепей, но существу не изменяющей их прежних конформационных состояний; 2) снятие наталкиваний приводит к образованию целого ряда новых стабилизирующих контактов. Оптимальное положение трипептидного участка  $\text{Gly}^{34}-\text{Asp}^{36}$  выбрано путем количественной оценки всех его низкоэнергетических конформаций в потенциальном поле структуры  $\text{Met}^1-\text{Cys}^{33}$ . В табл. 2 и 3 приведены значения двугранных углов основной и боковой цепей и координаты атомов  $\text{C}^\alpha$  всех остатков инсектотоксина в найденной трехмерной структуре молекулы. В табл. 4 даны геометрические параметры дисульфидных связей, которые вполне удовлетворительно соответствуют экспериментальным значениям: длине S—S-связи ( $2,04 \text{ \AA}$ ), расстояниям  $\text{C}^\beta \dots \text{S}$  ( $3,05 \text{ \AA}$ ) и  $\text{C}^\beta \dots \text{C}^\beta$  ( $3,8 \text{ \AA}$  при *гош*- и  $4,8 \text{ \AA}$  при *транс*-ориентациях дисульфидной связи) [10].

Рассмотрим краткую характеристику рассчитанной конформации инсектотоксина и отметим некоторые общие аспекты пространственной организации природных полипептидов. Спонтанная детерминация единственной системы из четырех дисульфидных связей и получение на завершающей стадии расчета стабильной трехмерной структуры молекулы прежде всего свидетельствуют о применимости положенной в основу конформационного анализа инсектотоксина структурной теории белка [3—5], т. е. в конечном счете правильном представлении о принципах его пространственной организации и возможности количественного предсказания нативной конформации исходя только из аминокислотной последовательности. Результаты расчета подтвердили одно из основных положений теории о наличии в плотно упакованной молекулярной структуре белка согласованности между всеми ближними, средними и дальними взаимодействиями. Это свойство присуще только эволюционно отобранной аминокислотной последовательности и ответственно за резкую энергетическую дифференциацию и выделение из множества возможных структурных вариантов стабильной и уникальной в каждом случае конформации.

В найденной структуре инсектотоксина геометрия всех 36 остатков отвечает наиболее предпочтительным конформациям соответствующих свободных монопептидов ( $\text{CH}_3-\text{CONH}-\text{C}^\alpha\text{RH}-\text{CONH}-\text{CH}_3$ ). Ближние взаимодействия определяют конформационные возможности каждого остатка,

Двугранные углы (град) основной и боковых цепей в трехмерной структуре инсектотоксина II<sub>1</sub>

№	Оста-ток	Иден-тифи-катор	Основная цепь		Боковая цепь				
			φ	ψ	χ <sub>1</sub>	χ <sub>2</sub>	χ <sub>3</sub>	χ <sub>4</sub>	χ <sub>5</sub>
1	Met	B <sub>2212</sub>	180,0	146,22	-177,0	175,2	73,7	180,0	
2	Cys	R <sub>2</sub>	-94,0	-61,0	174,0				
3	Met	B <sub>2332</sub>	-129,0	122,1	-176,5	-70,8	-70,4	151,0	
4	Pro	R	-57,6	-69,1	0,0	0,0	0,0	0,0	
5	Cys	B <sub>2</sub>	-152,0	155,3	-146,3				
6	Phe	R <sub>21</sub>	-59,0	-29,7	173,8	97,6			
7	Thr	R <sub>222</sub>	-72,7	-34,5	58,1	163,1	180		
8	Thr	R <sub>322</sub>	-76,4	-64,7	-50,6	-178,7	180		
9	Arg	B <sub>33321</sub>	-122,8	77,1	-57,7	-62,2	-60,8	165,5	0,0
10	Pro	R	-57,6	-31,3	0,0	0,0	0,0	0,0	
11	Asp	R <sub>12</sub>	-69,6	-34,3	59,8	131,7			
12	Met	R <sub>3322</sub>	-82,8	-21,3	-62,2	-64,8	-179,8	180	
13	Ala	R <sub>2</sub>	-74,2	-52,3	-157,5				
14	Gln	R <sub>123</sub>	-49,1	-27,7	56,6	-179,6	-0,5		
15	Gln	R <sub>331</sub>	-91,2	-46,5	-55,7	-57,3	110,5		
16	Cys	B <sub>3</sub>	-100,5	93,5	-67,0				
17	Arg	B <sub>32211</sub>	-66,3	97,3	-127,2	71,3	153,4	155,5	0,0
18	Ala	R <sub>2</sub>	-146,3	-89,7	165,0				
19	Cys	R <sub>2</sub>	-63,7	-35,9	-157,3				
20	Cys	R <sub>2</sub>	-46,0	-39,3	-169,6				
21	Lys	R <sub>32322</sub>	-72,5	-48,8	-59,5	173,4	-64,9	-164,3	180
22	Gly	P	93	-68,0					
23	Arg	B <sub>12222</sub>	-126,0	169,3	59,1	171,8	178,1	177,7	180
24	Gly	R	-61,1	-46,0					
25	Lys	B <sub>21222</sub>	-87,1	54,7	-164,0	67,8	159,1	-179,9	180
26	Cys	B <sub>2</sub>	-145,1	90,0	180,0				
27	Phe	B <sub>11</sub>	-173,0	172,0	61,0	88,9			
28	Gly	P	55,4	-135,4					
29	Pro	R	-57,6	-33,5	0,0	0,0	0,0	0,0	
30	Gln	R <sub>323</sub>	-66,8	-42,9	-57,8	168,8	112,2		
31	Cys	B <sub>3</sub>	-138,3	114,1	-50,9				
32	Leu	B <sub>3322</sub>	-48,7	-42,8	-64,0	-75,9	180	180	
33	Cys	B <sub>2</sub>	-77,0	70,0	177,9				
34	Gly	R	-153,8	-70,1					
35	Tyr	R <sub>312</sub>	-126,2	-66,1	-74,6	68,8	-179,4		
36	Asp	B <sub>12</sub>	-103,7	62,3	54,1	128,5			

Таблица 3

Координаты (Å) атомов C<sup>α</sup> аминокислотных остатков в трехмерной структуре инсектотоксина II<sub>1</sub>

C <sup>α</sup> <sub>i</sub>	x	y	z	C <sup>α</sup> <sub>i</sub>	x	y	z	C <sup>α</sup> <sub>i</sub>	x	y	z
1	0	0	0	13	10,59	3,82	0,36	25	17,64	-9,53	7,96
2	2,31	2,90	0,75	14	12,89	6,00	2,46	26	18,22	-6,30	6,09
3	5,41	2,12	-1,28	15	9,66	7,06	4,17	27	18,75	-3,27	8,30
4	5,67	-0,59	-3,94	16	7,83	3,75	4,12	28	21,22	-1,24	10,32
5	9,43	-1,00	-3,99	17	10,34	1,24	5,44	29	24,84	-2,00	9,42
6	12,49	1,14	-3,38	18	9,25	-1,88	3,61	30	23,55	-4,76	7,16
7	13,26	0,76	-7,07	19	10,73	-2,90	0,27	31	21,31	-2,33	5,31
8	9,84	2,06	-8,00	20	13,60	-0,55	1,02	32	21,79	1,41	5,64
9	9,40	4,86	-5,49	21	14,62	-3,19	3,53	33	18,09	1,72	4,89
10	12,74	5,21	-3,68	22	13,94	-6,09	1,19	34	17,15	0,27	8,26
11	11,43	8,48	-2,30	23	14,25	-8,86	3,74	35	18,34	2,11	11,35
12	8,43	6,65	-0,90	24	13,90	-9,23	7,49	36	21,56	4,02	10,84

т. е. набор дозволённых для него конформационных состояний. Средние взаимодействия только из этих наборов выбирают конформации, сочетание которых оказывается наиболее выгодным. Суммарный эффект средних взаимодействий в полученной конформации инсектотоксина составил -180,1 ккал/моль. На некоторых участках последовательности выбор

Геометрические параметры дисульфидных связей в трехмерной структуре инсектотоксина  $I_1$ 

$\overbrace{\text{Cys}^i - \text{Cys}^j}^{ }$	Расстояние (Å)			
	$s^i - s^j$	$c_i^\beta \dots s^j$	$s^i \dots c_j^\beta$	$c_i^\beta \dots c_j^\beta$
$\overbrace{\text{Cys}^2 - \text{Cys}^{16}}$	2,0	3,2	3,2	3,9
$\overbrace{\text{Cys}^5 - \text{Cys}^{19}}$	2,0	3,3	3,3	4,0
$\overbrace{\text{Cys}^{20} - \text{Cys}^{33}}$	2,0	3,0	3,0	4,4
$\overbrace{\text{Cys}^{26} - \text{Cys}^{31}}$	2,0	3,0	2,9	4,1

конформационного состояния остатков только за счет средних взаимодействий практически однозначен (особенно в отношении формы основной цепи). В этих случаях соответствующие фрагменты образуют конформационно жесткие нуклеации. В аминокислотной последовательности инсектотоксина к таким участкам можно отнести  $\text{Met}^1 - \text{Pro}^{10}$  ( $\text{Met}^1 - \text{Cys}^{16}$ ) и  $\text{Cys}^{26} - \text{Cys}^{31}$ . На других участках последовательности средние взаимодействия в той или иной мере сокращают число возможных конформационных состояний остатков и определяют для них ограниченные наборы приблизительно равновероятных структурных вариантов. Функция дальних взаимодействий по отношению к конформационно жестким по средним взаимодействиям фрагментам сводится к их дополнительной стабилизации, а также к еще большему ограничению свободы боковых цепей и, возможно, к незначительной корректировке формы основной цепи (главным образом на конечных участках). Например, у нуклеации  $\text{Met}^1 - \text{Pro}^{10}$  энергия средних взаимодействий составляет в окончательной конформации молекулы  $-92,1$  ккал/моль, а дальних  $-24,3$  ккал/моль; у нуклеации  $\text{Cys}^{26} - \text{Cys}^{31}$  соответствующие величины равны  $-34,4$  и  $-27,7$  ккал/моль. Таким образом, согласованность имеет место между сопоставимыми по своей эффективности взаимодействиями.

По отношению к конформационно лабильным участкам последовательности функция дальних взаимодействий заключается в выборе из равновероятных по средним взаимодействиям состояний единичных конформаций, обеспечивающих образование между ними и нуклеациями эффективных контактов, а также сближенность комплементарных друг другу нуклеаций. Так, расчет инсектотоксина показал, что значительно большей подвижностью по сравнению с нуклеациями обладают в свободном состоянии фрагменты  $\text{Arg}^{17} - \text{Lys}^{25}$  и  $\text{Gly}^{34} - \text{Asp}^{36}$ . Однако когда они встраиваются в общую цепь, их конформации полностью детерминируются. Энергия средних взаимодействий фрагмента  $\text{Arg}^{17} - \text{Lys}^{25}$  равна  $-31,8$  ккал/моль, а энергия его взаимодействий с предшествующими и последующими остатками составляет соответственно  $-7,5$  и  $-30,1$  ккал/моль. О высокой степени согласованности всех межостаточных взаимодействий свидетельствует тот факт, что у рассмотренных фрагментов инсектотоксина независимо от их длины конформационные состояния, входящие в нативную структуру молекулы, оказывались одними из самых предпочтительных, а во многих случаях глобальными; их энергия была всегда меньше, чем у большинства структурных вариантов, отобранных для последующего анализа.

В расчете инсектотоксина учитывались внутримолекулярные невалентные, электростатические и торсионные взаимодействия и водородные связи применительно к условиям водного окружения [11]. Каковы же вклады этих взаимодействий в энергию найденной конформации молекулы?

Аминокислотная последовательность инсектотоксина включает в себя восемь остатков (около четверти всего состава), несущих шесть положи-



Энергия внутри- и межостаточных взаимодействий  
в трехмерной структуре инсектотоксина

Энергия, ккал/моль	Взаимодействия		Энергия, ккал/моль	Взаимодействия	
	внутриоста- точные	межостаточ- ные		внутриоста- точные	межостаточ- ные
$E_{\text{нес}}$	-93,5	-285,5	$E_{\text{торс}}$	+27,4	-
$E_{\text{в.с.}}$	-1,0	-12,1	$E_{\text{общ}}$	-52,3	-260,7
$E_{\text{эл}}$	+14,8	+36,9			

тельных и три отрицательных целочисленных заряда. Общая энергия межостаточных электростатических взаимодействий равна 36,9 ккал/моль. Дестабилизирующий эффект обусловлен электростатическими взаимодействиями между всеми смежными остатками, вклад которых (65 ккал/моль), однако, можно считать постоянным из-за малой чувствительности этих взаимодействий к конформационным состояниям так близко расположенных остатков. Поэтому суммарный эффект влияющих на конформацию молекулы электростатических взаимодействий между всеми другими остатками имеет стабилизирующий характер (-28,1 ккал/моль). Боковые цепи заряженных остатков обращены в сторону растворителя. В расчете не были сделаны специальные предположения на этот счет, и такая ориентация  $N^+H_3$ - групп  $-Met^1$ - боковых цепей  $Arg^{9,17,23}$ ,  $Lys^{21,25}$ ,  $Asp^{11,36}$ , а также  $Thr^{7,8}$  и  $Gln^{14,15,30}$  получена при фрагментарном анализе и учете всех видов взаимодействий автоматически. Стабилизирующий вклад электростатических взаимодействий заряженных остатков составляет -33,6 ккал/моль, а дестабилизирующий — +27,2 ккал/моль, т. е. имеет место почти полная компенсация противоположных эффектов. При учете электростатических взаимодействий нами были использованы парциальные заряды на атомах, предложенные в работе [12], и величина диэлектрической проницаемости ( $\epsilon_{10}$ ) [11]. Учитывая водное окружение, предполагались также ослабленными водородные связи; их энергию при оптимальной длине принимали равной -1,5 ккал/моль [13, 14]. Суммарный стабилизирующий эффект водородных связей составил -13,1 ккал/моль. Очевидно, что реализуемые в найденной структуре инсектотоксина 20 водородных связей обусловлены сложным комплексом многих межостаточных взаимодействий, среди которых собственно водородные связи играют незначительную роль (имеется в виду их энергетический вклад).

Общая энергия торсионных взаимодействий в пайденной структуре инсектотоксина равна 27,4 ккал/моль, т. е. 0,75 ккал/моль на остаток. Из этого следует, что взаимные расположения практически всех атомных групп основных и боковых цепей аминокислотных остатков отвечают минимумам торсионных потенциалов. В табл. 5 приведены величины общей энергии внутри- и межостаточных взаимодействий и энергетические вклады отдельных видов взаимодействий. Из данных таблицы очевидно, что доминирующую роль в стабилизации структуры играют невалентные, точнее, дисперсионные взаимодействия. На рис. 2 представлено распределение энергии всех межостаточных дисперсионных и электростатических контактов в рассчитанной структуре инсектотоксина.

Опишем процесс самоорганизации пространственного строения молекулы инсектотоксина от полностью развернутой формы до нативной трехмерной структуры, опираясь на полученные в расчете данные о конформационных возможностях конкретной последовательности и ее отдельных фрагментов. Процесс свертывания полипептидной цепи начинается одновременно и независимо на двух участках с образованием за счет средних взаимодействий пуклеаций  $Met^1-Cys^{16}$  и  $Cys^{26}-Cys^{31}$ . В результате оказываются сближенными остатки  $Cys^2$ ,  $Cys^{16}$  и  $Cys^{26}$ ,  $Cys^{31}$ . Образование между ними дисульфидных мостиков не сопровождается заметным изменением

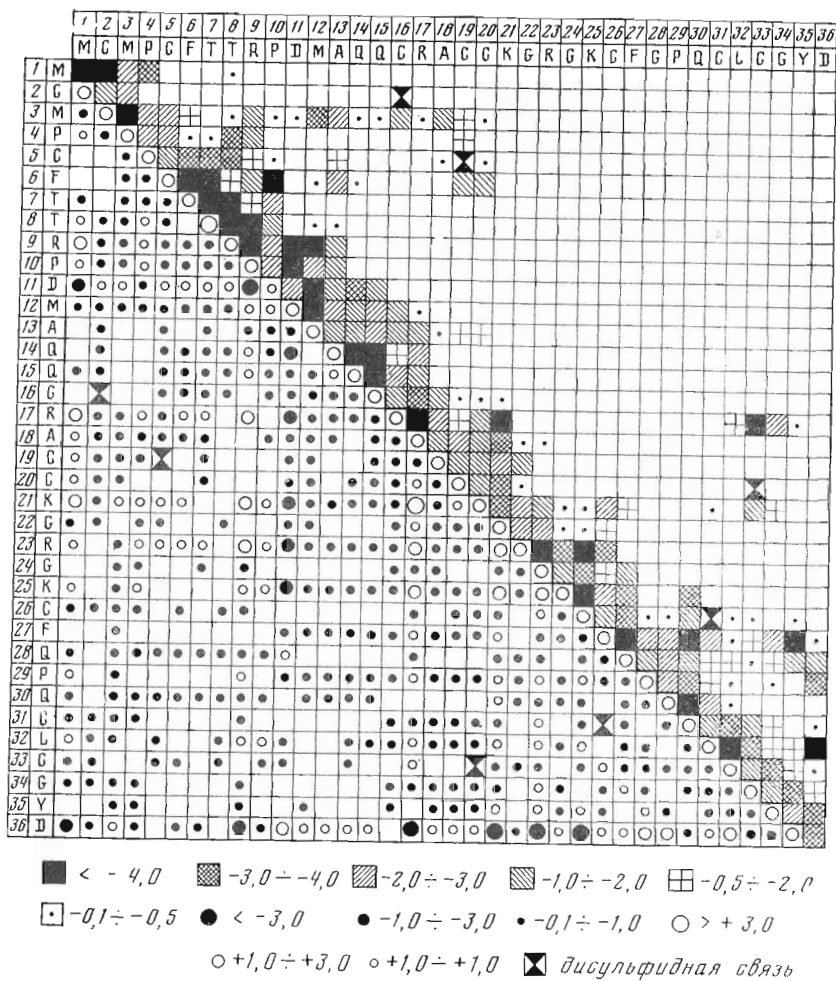
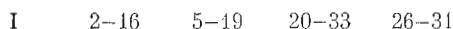


Рис. 2. Распределение энергии невалентных (правая верхняя треугольная матрица) и электростатических (левая нижняя) взаимодействий в трехмерной структуре инсектотоксина  $I_1$

геометрии и, следовательно, не приводит к рассогласованности сложившихся у линейных последовательностей стабилизирующих межостаточных взаимодействий. Обе нуклеации входят в трехмерную структуру инсектотоксина без существенных изменений. Промежуточный Arg<sup>17</sup>-Lys<sup>25</sup> и С-концевой Gly<sup>34</sup>-Asp<sup>36</sup>- участки, более подвижные, но обладающие предпоспособленностью к формам цепи, комплементарным образовавшимся циклическим нуклеациям, детерминируют свои состояния. При этом за счет многочисленных стабилизирующих контактов происходит сближение остатков Cys<sup>5</sup>, Cys<sup>19</sup> и Cys<sup>20</sup>, Cys<sup>33</sup>. Между ними, также не вызывая стерических осложнений, образуются еще две цистиновые пары. Выше отмечалось, что конформационный анализ молекулы инсектотоксина выполнялся для линейной аминокислотной последовательности с опробованием всех возможных способов образования четырех дисульфидных связей. Результаты расчета свидетельствуют о существовании только одной системы цистиновых пар:



Образование каждой пары осуществлялось по одному и тому же механизму [4, 5], в основе которого лежат конформационные свойства аминокислотной последовательности. Во всех случаях сближенность соответствующих остатков Cys происходила в энергетически самых предпочтительных конформационных состояниях линейных участков; создание

дисульфидных связей  $S_{(2)}-S_{(16)}$  и  $S_{(28)}-S_{(31)}$  не только не затруднило, а, напротив, предопределило возможность образования двух других дисульфидных связей.

#### ЛИТЕРАТУРА

1. Гришин Е. В., Солдатов Н. М., Ташмухамедов Б. А., Атакузиев Б. У. Биоорган. химия, 1978, т. 4, № 4, с. 450-460.
2. Жданова Л. П., Адамович Т. Б., Назимов И. В., Гришин Е. В., Овчинников Ю. А. Биоорган. химия, 1977, т. 3, № 4, с. 485-493.
3. Попов Е. М. Молекулярн. биология, 1975, т. 9, № 3, с. 578-593.
4. Попов Е. М. Int. J. Quant. Chem., 1979, v. 16, p. 707-737.
5. Попов Е. М. В сб.: Физика молекул, 1980, вып. 8, с. 67-113.
6. Anfinsen C. B. Science, 1973, v. 181, № 4096, p. 223-230.
7. Попов Е. М., Мельников П. Н. Биоорган. химия, 1979, т. 5, № 10, с. 1471-1493.
8. Спасов В. З., Попов Е. М. Биоорган. химия, 1981, т. 7, № 1, с. 25-44.
9. Chandrasekaran R. Proc. Ind. Acad. Sci., 1968, v. 6, p. 1-10.
10. Jones D. D., Bernal I., Frey M. N., Koetzle T. F. Acta crystallogr., 1974, B30, p. 1220-1229.
11. Липкинд Г. М., Архипова С. Ф., Попов Е. М. Ж. структурн. химии, 1970, т. 11, № 1, с. 121-126.
12. Scheraga H. A. Advances Phys. Org. Chem., 1968, v. 6, p. 103-184.
13. Schellman I. A. Comp. rend. trav. Lab. Carlsberg, ser. chim., 1955, v. 29, p. 223-228.
14. Klotz T. M., Franzen I. S. J. Amer. Chem. Soc., 1962, v. 84, p. 3461-3464.

Поступила в редакцию  
29.VI.1981

#### A PRIORI CALCULATION OF SPATIAL STRUCTURE OF INSECTOTOXIN FROM SCORPION *BUTHUS EUPEUS* VENOM

POPOV E. M., SHVYRKOV V. N., SPASOV V. Z.

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

The three-dimensional structure of insectotoxin I<sub>1</sub>, a peptide component of the scorpion *Buthus eupeus* venom has been calculated. The molecule consists of 36 amino acid residues and includes 8 cysteine residues forming 4 disulfide bridges of unknown localization. Conformational analysis was carried out starting from the amino acid sequence and incomplete information on the insectotoxin covalent structure. The calculations produced all conformational parameters for the backbone and side chains and enabled identification of the disulfide bond in the insectotoxin spatial structure.