

# Компьютерное моделирование электронной и молекулярной структуры фрагментов М-ДНК: стекнинг димеров и тримеров

Ю.В. Рубин, Л.Ф. Белоус, А.А. Якуба, А.Л. Головинский, Ю.В. Савин

С целью изучения молекулярной структуры металлизированной ДНК (М-ДНК) и механизмов проводимости в ней, нами проведен квантово-механический *ab initio* расчет электронной и пространственной структуры, термодинамических характеристик стекнинг димера и тримера пар оснований нуклеиновых кислот с ионами цинка (АТ)(СГ)  $2\text{Zn}^{2+}$  и (СГ)(АТ)(СГ)  $3\text{Zn}^{2+}$ , а также димера и тримера этих пар оснований (АТ)(СГ) и (СГ)(АТ)(СГ) без иона цинка. Такая М-ДНК, с расположенными между основаниями ионами переходных металлов, способна проводить электрический ток и представляет большой интерес для нанотехнологий. Расчет выполнялся с помощью метода MP2 и пакета Gamess на кластере СКИТ-3 Института Кибернетики им. В.М. Глушкова НАН Украины ( $R_{\text{peak}}=7.4$  Tflops).

## Введение

Ввиду отсутствия рентгеноструктурных данных и трудностей экспериментальных исследований с изолированными молекулами металлизированной ДНК (М-ДНК) в последнее время получили развитие теоретические расчеты фрагментов и моделей М-ДНК [1-3].

Для таких расчетов используются DFT-методы с применением кластеров и суперкомпьютеров, занимающих верхние строчки в рейтинге TOP500, которые способны оперировать более чем с 3000 молекулярных орбиталей. К примеру, подобные задачи решаются на кластерах США в DOE/Oak Ridge National Laboratory (5-е место в TOP500, 260 Tflops) и Испании в Barcelona Supercomputing Center (26-е место в TOP500, 94 Tflops).

*Ab initio* расчеты для молекулярных систем, содержащих более чем 30-40 атомов, невозможны без применения параллельных алгоритмов вычислений. В свою очередь, возможность увеличения количества атомов в молекулярных моделях, определяющих достоверность и близость этих моделей к реальным системам, и возможности увеличения размеров используемых в *ab initio* расчетах базисов определяются, прежде всего, современным состоянием суперкомпьютерной техники для высокопроизводительных вычислений (HPC), перешагнувшей петафлопный диапазон производительности.

В 90-е годы канадские исследователи обнаружили, что ионы некоторых двухвалентных металлов ( $\text{Ni}^{2+}$ ,  $\text{Co}^{2+}$ ,  $\text{Zn}^{2+}$ ) способны встраиваться в пары оснований ДНК [4]. Такую структуру М-ДНК экспериментально получают в щелочной среде при  $\text{pH}=8.5$  [4-7]. ЯМР исследования М-ДНК показали отсутствие сигнала имино-протонов, что позволило предположить, что в каждой паре оснований иминный протон замещен ионом переходного металла [4], и таким образом внутри макромолекулы ДНК образуется одномерная цепочка ионов металлов. В прямых исследованиях проводимости М-ДНК, помещенной между золотыми электродами, было показано, что М-ДНК является лучшим проводником электрического тока по сравнению с В-формой ДНК [5]. Аналогичный вывод был сделан на основе флуорометрических исследований электронного транспорта в синтетических олигонуклеотидах с присоединенными по концам донором и акцептором [4,6]. Обсуждается вопрос об использовании М-ДНК в качестве нанопроводника и в качестве биосенсора в нанотехнологиях [4-7].

Недавно опубликованы работы [1,2], в которых с помощью метода DFT проведены расчеты молекулярной и электронной структуры пары оснований нуклеиновых кислот гуанин-цитозин (GC) с ионами  $\text{Zn}^{2+}$ ,  $\text{Co}^{2+}$ ,  $\text{Fe}^{2+}$ . В работе [1] также проведен расчет зонной модели полинуклеотида  $\text{poly}(\text{GC})$  с ионами  $\text{Zn}^{2+}$  и  $\text{Co}^{2+}$ . Ранее мы провели расчет методом MP2 электронной структуры пар аденин-урацил и аденин-тимин с ионами цинка, а также расчет физических свойств пар оснований: аденин-тимин и гуанин-цитозин с расположенными между основаниями ионами цинка, никеля и гидроксидными [3,8,9].

С целью подробного изучения и моделирования электронной и молекулярной структуры М-ДНК и механизма проводимости в ней нами проведен неэмпирический расчет простран-

венной и электронной структуры, термодинамических характеристик стекинг димеров и тримеров пар оснований аденин-тимин и гуанин-цитозин с ионами цинка.

## Методы расчета

Расчет полной энергии и оптимизация пространственной структуры стекинг димера (AT)(CG) с двумя ионами цинка, двумя гидроксидными ионами и двумя молекулами воды, димера (AT)(CG) и тримера (CG)(AT)(CG) проводился методом MP2 с базисом 6-31G\*. Расчет тримера (CG)(AT)(CG) с тремя ионами цинка, тремя гидроксидными ионами и двумя молекулами воды был проведен методом MP2 с базисом 3-21G\*. Расчеты проводились с помощью пакета программ Gamess [10]. Гидроксил OH<sup>-</sup> в рассчитанных системах моделировал щелочную среду. Молекулы воды в изученных стекинг димерах и тримерах, расположенные над и под ионами цинка, моделировали соседние по стопке в макромолекуле ДНК пары оснований нуклеиновых кислот.

Расчеты энергии взаимодействия в стекинг димерах проводились по обычной схеме, в частности, описанной в работе [11], с учетом погрешности BSSE. Визуализация пространственной структуры и молекулярных орбиталей была проведена с помощью программы ChemCraft [12].

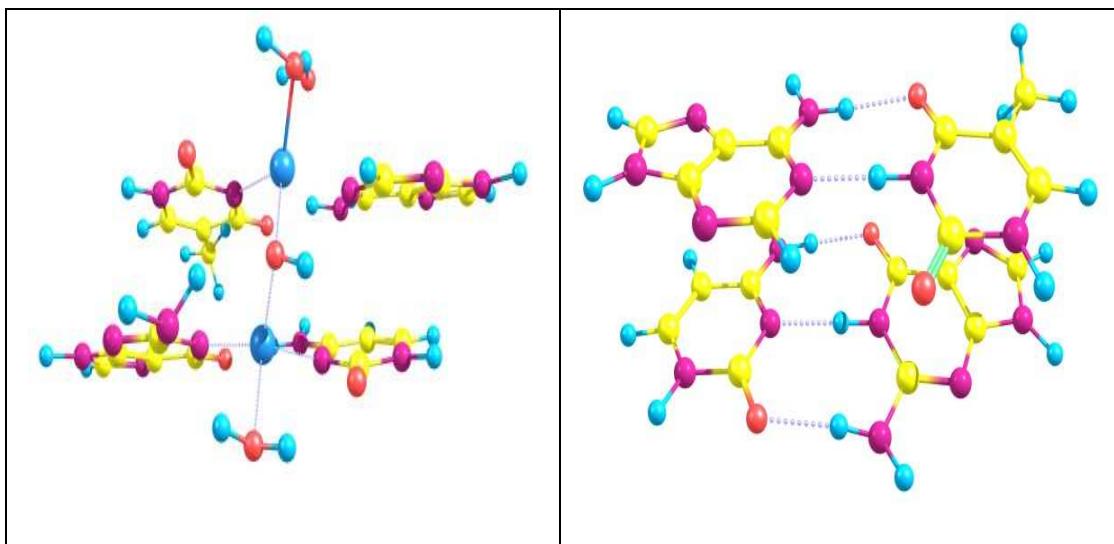
Серьезным недостатком пакета Gamess является повышенные требования к оперативной памяти. Программный пакет Gamess, в частности, его метод MP2, учитывающий корреляцию электронов, ориентирован в идеале на общую оперативную память суперкомпьютера. Для кластерной архитектуры оперативная память реализуется в Gamess как модель распределенной памяти, обслуживаемая имеющимися средствами межузловых соединений, например, Gigabit Ethernet или InfiniBand, что создает «узкое горло» при обмене данными. Эта модель очень требовательна к среде передачи сообщений. Так, при расчете димера и использовании интерфейса Gigabit Ethernet эффективность работы процессоров не превышала 11 процентов, а время вычисления исчислялось месяцами. Поэтому доступная через интернет версия Gamess была модифицирована для использования низколатентного InfiniBand (в СКИТ-3 пропускная способность InfiniBand равна 20 Gbps, а латентность - 2.3 мкс). С этим интерфейсом эффективность работы процессоров достигла 45 процентов. Кроме того, была расширена поддержка Gamess-вычислений на большом количестве процессоров. На сегодняшний день на СКИТ-3 поддерживается до 1000 процессоров. При запуске задачи в максимально возможной конфигурации использовалось 700 процессоров и 1,4 ТБайт оперативной памяти. Для расчета тримера с использованием интерфейса InfiniBand требовалось более 200 Гбайт моделируемой оперативной памяти, что в свою очередь, потребовало использования около 500 процессорных ядер. Последнее свидетельствует об очень неэффективном использовании дорогих процессорных ресурсов кластера.

В дальнейшем планируется существенно улучшить используемые средства вычислений:

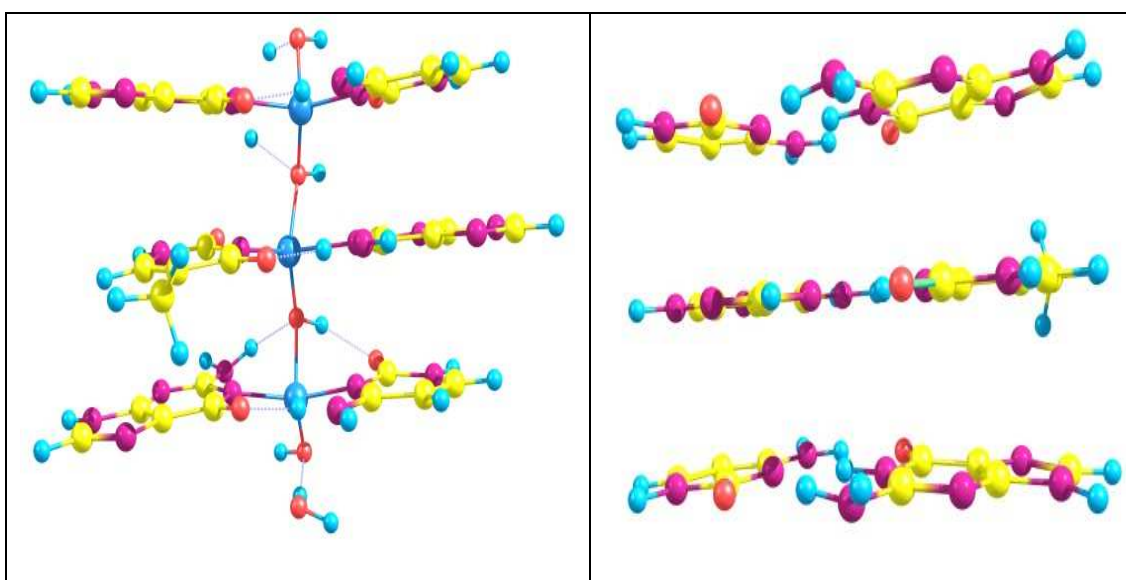
- Увеличить в несколько раз объем установленной в вычислительном узле кластера оперативной памяти (сейчас имеется 2 Гбайта на процессорное ядро, всего до 16 Гбайт на вычислительный узел);
- Проводить расчеты не на программном пакете Gamess, а на пакете NWChem [13]. Для этого пакета характерны лучшая масштабируемость и существенное снижение требований к оперативной памяти по сравнению с пакетом Gamess;
- Через ГРИД перейти на суперкомпьютеры с производительностью в сотни Tflops, где средства межузловой связи имеют лучшие характеристики, чем на используемом кластере СКИТ-3.

## Результаты и обсуждение

На **Рис.1** и **Рис.2** представлены оптимизированные методом MP2 пространственные структуры рассчитанных димеров и тримеров пар оснований нуклеиновых кислот с ионами цинка и без них.



**Рис. 1.** Молекулярные структуры рассчитанных методом MP2/6-31G\* стекинг димера (AT)(CG) с двумя ионами цинка, двумя гидроксилами и двумя молекулами воды (слева) и стекинг димера (AT)(CG) (справа).



**Рис. 2.** Молекулярные структуры рассчитанных стекинг тримера (CG)(AT)(CG) с тремя ионами цинка, тремя гидроксилами и двумя молекулами воды (слева) и стекинг тримера (CG)(AT)(CG) (справа).

Отличительной чертой димеров и тримеров с ионами цинка является наличие мостиков между парами оснований, образованных ионами цинка и гидроксилами. Образование мостиков между ионами обусловлено, в частности, тем, что ионы цинка, как и переходных металлов, склонны образовывать неплоские тетраэдрические комплексы [14]. Это хорошо видно на примере комплексов пар оснований с гидроксидом, ионом цинка и двумя молекулами воды [9].

Во всех изученных димерах и тримерах расстояния между плоскостями пар оснований равны  $3.2\text{\AA}$  (Табл.1). Это расстояние меньше межплоскостного расстояния в В-форме ДНК, но согласуется с данными работы [15], в которой методом MP2 рассчитаны димеры пар оснований. Пары оснований как в димере и тримере с ионом цинка, так димере и тримере без иона являются неплоскими, за исключением средней пары AT в составе тримеров.

Неплоскость для пар оснований в тримере с ионом цинка более выражена, чем в стекинг тримере без иона. Для стекинг димеров имеет место противоположная ситуация - непо-

скость в парах оснований в димерах без иона выражена больше, чем в парах с ионом цинка.

Для всех пар в димерах и тримерах (кроме пары АТ в тримере) характерна пропеллерность (Табл.1). Наличие угла пропеллера и двугранного угла в парах оснований обусловлено  $sp^2$  гибридизацией электронов на атоме азота аминогрупп, участвующих в образовании водородных связей, и в случае димера и тримера с ионами цинка – как было сказано выше, с образованием координационной сферы тетраэдрической формы. Как видно на **Рис.1**, **Рис.2** и **Табл.1** в рассчитанных тримерах средняя пара АТ является более плоской по сравнению с этой парой в димере и парами СГ.

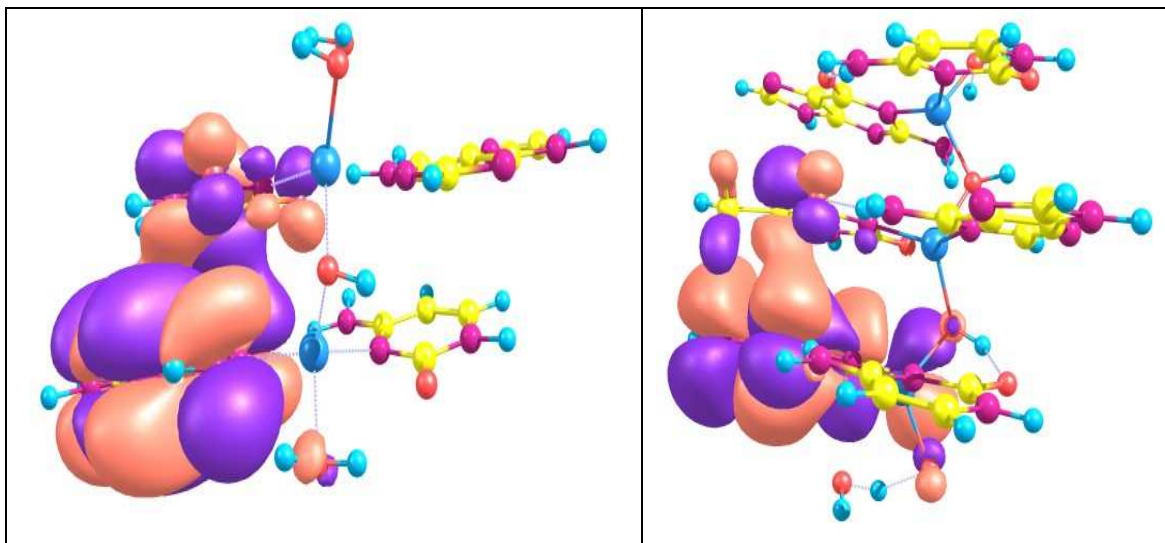
В составе димеров и тримеров ион цинка, входящий в состав пары АТ, образует четыре связи: две связи с атомами N1 аденина и N3 тимина и две связи - с двумя гидроксилами (с гидроксилом своей пары и с гидроксилом соседней пары). В составе димеров ион цинка, входящий в состав пары СГ, образует четыре связи: две связи с атомами N1 гуанина и N3 цитозина, одну связь - с гидроксилом и одну связь с молекулой воды. В составе тримеров ион цинка, входящий в состав второй пары СГ, образует 5 связей: две связи с атомами N1 гуанина и N3 цитозина, две связи - с гидроксилами и одну связь с молекулой воды. В парах оснований, входящих в состав димера и тримера с ионом цинка, имеется по одной Н-связи между кислородами и аминогруппами соседних оснований.

**Табл. 1.** Структурные параметры рассчитанных стекинг димеров и тримеров.

Структурный параметр	(АТ)(СГ)	(АТ)(СГ) 2Zn	(СГ)(АТ)(СГ)	(СГ)(АТ)(СГ)3Zn
Наличие межплоскостных мостиков		+		+
Межплоскостное расст, Å	3.2	3.2	3.2	3.2
Спиральность, град	0	2	0 СГ1 – СГ2 10 СГ1-АТ	0
Пропеллер, град	15-АТ 20-СГ	12 АТ 5 СГ	5 – СГ1 0 - АТ 3 – СГ2	20 –СГ1 0 - АТ 10 – СГ2
Двугр. угол, град	175 -АТ 176 -СГ	178 -АТ 177 -СГ	176 - СГ1 180 - АТ 176 – СГ2	170 - СГ1 178 - АТ 165 - СГ2
Колич-во коорд. связей		4 - 4		4 - 4 - 5
Колич-во Н-св	2 в АТ паре 3 в СГ паре	1 в АТ паре 1 в СГ паре	2 в АТ паре 3 в СГ паре	1 в АТ паре 1 в СГ паре
Энергия взаимодействия, ккал/моль	11.0	37.4		

Уголы спиральности для двух рассматриваемых стекинг димеров, геометрия которых оптимизирована методом МР2, имеют величину 0 и 2°, хотя исходно был задан угол спиральности 40°. Угол спиральности для двух стекинг тримеров имеет величину 0 и 10°.

Энергия взаимодействия между парами оснований в димере с ионом цинка равна 37.4 ккал/моль. Это значительно больше, чем энергия стекинг взаимодействия между парами канонических оснований [15]. Однако надо учесть, что в нашем случае энергия взаимодействия включает также энергию координационной связи между ионом цинка и кислородом гидроксила, входящим в состав соседней пары. Энергия взаимодействия в димере без иона цинка равна 11.0 ккал/моль.



**Рис. 3.** Верхняя заполненная молекулярная орбиталь (НОМО) стекинг димера (АТ)(СG) с двумя ионами цинка, двумя гидроксилами и двумя молекулами воды (слева) и аналогичного стекинг тримера (СG)(АТ)(СG) (справа).

С точки зрения изучения механизмов проводимости в М-ДНК, значительный интерес представляет анализ формы верхней заполненной молекулярной орбитали (НОМО) изученных димеров и тримеров с ионами цинка. Как видно на **Рис. 3**, верхняя заполненная орбиталь димера и тримера включает молекулярные орбитали соседних по стопке оснований, что свидетельствует о возможной делокализации электрона в рассчитанных стекинг структурах

## Выводы

Проведенные методом MP2 неэмпирические расчеты молекулярной и электронной структуры димеров и тримеров пар оснований нуклеиновых кислот с ионами цинка показали следующее:

- Стекинг-димеры и тримеры оснований нуклеиновых кислот с ионами цинка имеют мостики между парами оснований.
- Межплоскостное расстояние в димере и тримере с ионами цинка – 3.2Å.
- Верхняя заполненная молекулярная орбиталь (НОМО) в димере и тримере пар оснований с ионами цинка включает соседние по стопке основания.

## Литература

1. Alexandre S., Soler J., Seijo L., Zamora F., Geometry and electronic structure of *M*-DNA ( $M=\text{Zn}^{2+}$ ,  $\text{Co}^{2+}$ , and  $\text{Fe}^{2+}$ ) // Phys Rev., 2006, B, 73: 205112.
2. Cabrera M., Sumpter B., Sponer Ju., Sponer J., Petit L., Wells J., Theoretical Study on the Structure, Stability, and Electronic Properties of the Guanine-Zn-Cytosine Base Pair in M-DNA//J. Phys. Chem., 2007, B, 111,:p.870-879.
3. Рубин Ю.В., Белоус Л.Ф., Якуба А.А. Молекулярная и электронная структуры фрагмента металлизированной ДНК - пары аденин-тимин с ионом цинка.// 'Доповіди НАН України' 2008, 5, с.186-191

4. Aich P, Labiuk S., Tari L., Delbaere L., Roesler W., Falk K., Steer R., Lee J. M-DNA: A Complex Between Divalent Metal Ions and DNA Which Behaves as a Molecular Wire.// *J. Mol. Biol.* 1999 ,294, p. 477-485.
5. Rakitin A, Aich P, Papadopoulos C, Kobzar Y, Vedenev A., Lee J., Xu J. Metallic Conduction Through Engineered DNA: DNA Nanoelectronic Building Blocks. // *Phys. Rev. Lett.* – 2001, 86, p.3670-3673.
6. Aich P, Skinner R., Wettig S., Steer R., Lee J. Long Range Molecular Wire Behaviour in a Metal Complex of DNA.// *J. Biomol. Struct. Dyn.*- 2002 ,20, p.93-98.
7. Wettig S., Li C., Long Y., Kraatz H., Lee J. M-DNA: a Self-assembling Molecular Wire for Nanoelectronics and Biosensing.// *Anal Sci.* 2003 ,19, p.23-26.
8. Белоус Л.Ф., Гречнев Г.Е., Рубин Ю.В., Степанян С.Г. Нанотехнологии и компьютерное моделирование // Труды Всерос. Научн. Конф. “Научный сервис в сети Интернет: технологии параллельного программирования” – Москва – Изд. МГУ -2006 - с.5-8.
9. Rubin Yu.V., Belous L.F., Yakuba A.A. Molecular and Electronic Structure of M<sub>2</sub>-DNA Fragments. // In book abstracts “NATO Advanced Research Workshop ‘Molecular Self-Organisation...’” Kyiv, June, 2008
10. Schmidt M.W., et al. // *J. Comput. Chem.* 1993- v.14-p.1347-1363; Gamess Version = Jun 2005 (R5) - Iowa State University.
11. Podolyan Ye., Rubin Yu.V., Leszczynski J. An ab initio Post-Hartree-Fock Comparative Study of 5-Azacytosine and Cytosine and their Dimers with Guanine.// *J. Phys. Chem.A.* 2000,104, p.9964-9970.
12. Zhurko G.A., Chem Craft .Tool for treatment of chemical data .Version 5.1.
13. Bylaska E.J., et al "NWChem, A Computational Chemistry Package for Parallel Computers, Version 5.1" (2007), Pacific Northwest National Laboratory, Richland, Washington 99352-0999, USA.
14. Неорганическая биохимия. Ред. Эйхгорн Г. М., Мир 1978, 358с
15. Sponer J., Jurecka P., Marchan I., Leque F J., Orozco M., Hobza P. Nature of Base Stacking: Quantum – Chemical Stacking Energies in Ten Unique B-DNA Base-Pairs Steps // *Chem Eur. Journ.* 2006, 12,p 2854-2865.