

УДК 573.4+001.8

## ОПТИМИЗАЦИЯ МЕТОДА КВАНТОВО-ХИМИЧЕСКИХ РАСЧЕТОВ БИОЛОГИЧЕСКИХ МАКРОМОЛЕКУЛ

*Д.А. Алишева, Д.С. Тарасов, Н.И. Акберова*

### Аннотация

Проведена оптимизация метода теории функционала плотности для биологических макромолекул на примере небольших молекул (глицин, аланин и другие) для того, чтобы в дальнейшем моделировать сложные биологические системы при условии получения достаточно точных результатов за наименьшее время. Молекулы изучали с помощью волновой функции, геометрической оптимизации и молекулярной динамики для нахождения оптимального значения параметра cutoff (отсечения) и наиболее подходящего псевдопотенциала.

### Введение

В последнее время все большую популярность стали приобретать работы по моделированию и прогнозированию механизмов протекания биохимических реакций, а также по предсказанию структуры и свойств белков, лекарственных препаратов [1]. Было проведено большое количество исследований по моделированию процессов синтеза белка, среди которых можно выделить работы по молекулярному моделированию аналогов аргинин-глицин-аспартатной кислоты [2], моделированию трехмерной конформации пептидов, которая влияет на их способность образовывать водородные связи и полярность [3], а также работы по симуляции процесса агрегации пептида с помощью дискретной молекулярной динамики [4].

Ведущие мировые фирмы, такие как IBM, уже на протяжении нескольких лет ведут разработку программного и аппаратного обеспечения для таких исследований. На сегодняшний день суперкомпьютер Blue Gene, созданный фирмой IBM, является лидером в области программного и аппаратного обеспечения. Специальное приложение Blue Matter было разработано для изучения механизмов синтеза белка, начиная от одного маленького пептида, вплоть до белковой мембраны, а также для моделирования лекарственных препаратов [5].

Это направление в науке на сегодняшний день приобретает передовые позиции в политике финансирования ведущих отраслей многих передовых государств. Огромные средства вкладываются в развитие таких исследований правительством США и Японии. Еврокомиссия к 2006 г. выделила из бюджета на развитие и расширение наноисследований и исследований в области компьютерного моделирования биохимических реакции и прогнозирование лекарственных препаратов 1.3 млрд. евро, а к 2015 г. эксперты прогнозируют увеличение объема инвестиций в данной отрасли до 1 трлн. долларов.

Для успешного моделирования и прогнозирования свойств биологических макромолекул необходимо в первую очередь определить наиболее подходящие методы для данных молекул. На сегодняшний день существует три основных типа методов для изучения строения, пространственной структуры, свойств и других параметров биологических молекул: это эмпирические, полуэмпирические и неэмпирические методы. Все расчеты многоатомных молекул основаны на приближенных решениях уравнения Шрёдингера:  $\hat{H}\Psi = E\Psi$ , где  $E$  и  $\Psi$  – полная энергия и волновая функция системы,  $\hat{H}$  – оператор Гамильтона (гамильтониан) системы, представляющий собой сумму операторов кинетической и потенциальной энергии электронов и ядер, входящих в систему [6]. Оператор потенциальной энергии равен сумме операторов кулоновского взаимодействия частиц, которые зависят от расстояний  $r_{ij}$  между парами частиц, а также операторов, взаимодействующих частиц с внешним полем. При учете спина в гамильтониан входят также операторы спин-орбитального взаимодействия и спин-спинового взаимодействия [7]. Волновая функция для молекулярной системы, получаемая в результате решения уравнения Шрёдингера, должна удовлетворять общим требованиям, предъявляемым к волновым функциям в квантовой механике. В частности, для многоэлектронной системы она должна быть антисимметричной относительно перестановки координат (пространственных и спиновых) любой пары электронов [8].

Практика предъявляет два главных требования к уровню приближения и выбору расчетной схемы [9]:

1. достаточное соответствие результатов расчета результатам эксперимента;
2. достаточная экономность расчетов, т. е. разумные затраты времени при выполнении их на быстродействующих ЭВМ.

Полуэмпирические методы квантовой химии – это методы расчета молекулярных характеристик или свойств вещества с привлечением экспериментальных данных. По своей сути полуэмпирические методы аналогичны неэмпирическим методам решения уравнения Шрёдингера для многоатомных молекулярных систем, однако для облегчения расчетов в полуэмпирические методы вводятся дополнительные упрощения. Как правило, эти упрощения связаны с валентным приближением. При изучении макромолекул, например белков, или кристаллических образований нередко пользуются полуэмпирическими методами, в которых электронное строение не анализируется, а определяется непосредственно поверхность потенциальной энергии [10].

Неэмпирические расчеты основаны на точном решении уравнений, которые не включают никаких экспериментальных параметров, кроме фундаментальных физических постоянных. Все основные квантовохимические расчеты используют приближение молекулярных орбиталей Хартри – Фока. При использовании метода Хатри – Фока – Рутаана главной характеристикой расчетной модели является выбор базиса орбиталей [11].

Основной идеей метода теории функционала плотности является описание взаимодействующей системы фермионов посредством их плотности, а не с помощью решения многоэлектронной волновой функции. Для  $N$  электронов в твердом теле, которые подчиняются принципу Паули и взаимодействуют друг с другом согласно закону Кулона, это означает, что базисная переменная сис-

темы зависит только от трех пространственных координат  $x$ ,  $y$ , и  $z$ , а не  $3N$  степеней свободы [12]. В то время как метод теории функционала плотности в общем дает хорошее описание основных структурных свойств, практическое применение этого метода основывается на приближениях для так называемого потенциала обменной корреляции. Потенциал обменной корреляции описывает действие принципа Паули и потенциала Кулона вне чистого электростатического взаимодействия электронов [13].

Целью этой работы явилась оптимизация метода квантово-химических расчетов на примере биологических макромолекул. При выполнении данной работы решались следующие задачи:

1. исследование геометрических параметров биологических и органических макромолекул с помощью этих методов и сравнение полученных результатов с экспериментальными данными;
2. выявление оптимальных параметров метода для биологических макромолекул.

### 1. Методы и объекты

Постановка виртуальных экспериментов проводилась с помощью программы *cpmd* (Car-Parrinello Molecular Dynamics, The CPMD consortium, <http://www.cpmc.org>), которая делала квантовохимические расчеты волновых функций, геометрических оптимизаций и молекулярных динамик на основе теории функционала плотности и заданных псевдопотенциалов. Для обработки и визуализации полученных данных использовалась программа *ChemCraft* и *Excel*. Для обработки и расчета необходимых для исследования параметров была написана программа “*moleculardynamics*” с помощью языка *F# Visual Studio*, которая позволила найти среднее значение углов и длин связей. Для постановки эксперимента использовалась также программа *HyperChem* (<http://www.hyper.com>), которая создавала исследуемую молекулу, задавала координаты и её параметры. Файл с информацией о молекуле преобразовывался в файл, принимаемый на входе программой *cpmd*, для чего была написана программа “*mdf*” на языке *F# Visual Studio*.

Расчеты производились на кластере *MVS* Суперкомпьютерного Центра коллективного пользования КазНЦ РАН (<http://wt.knc.ru>).

Для исследования были выбраны неорганические и биохимические объекты, выбор которых определялся минимальным количеством атомов для создания интересующей связи и большей частотой встречаемости в биологических макромолекулах.

Экспериментально определенные параметры изучаемых молекул взяты из компьютерной базы Национального Института стандартов и технологий США *Computational Chemistry Comparison and Benchmark DataBase* (<http://srdata.nist.gov/cccbdb>).

### 2. Результаты

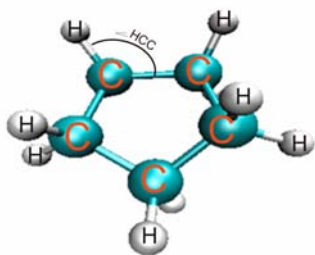
Проведены квантово-химические расчеты методом теории функционала плотности (DFT) при различных значениях *cutoff* для следующих молекул: ди-

метиловый эфир, сероуглерод, циклопентан, метилсилан, глицин и аланин. Молекулы диметилового эфира, сероуглерода, циклопентана и метилсилана были выбраны для исследования и оптимизации метода, так как имеют связи, сходные с таковыми в молекулах аминокислот. Квантовохимический расчет связи C–S в сероуглероде дает возможность прогнозировать результаты для расчетов такой связи в аминокислотах и биологических макромолекулах в целом, содержащих атом S. Квантовохимический расчет длины связи C–C в молекуле циклопентана помогает в выборе псевдопотенциалов для ароматических аминокислот. Квантовохимический расчет величины угла СОН дает ориентировочные величины угла СОН в аминокислотах и других биологических макромолекулах.

Оптимизация метода проводилась в три этапа: 1. расчет волновой функции; 2. расчет геометрической оптимизации; 3. расчет молекулярной динамики.

Ниже приведены рисунки и таблицы, содержащие данные для каждого этапа исследования при двух псевдопотенциалах MT\_PBE (псевдопотенциал Martins-Trouiller с обменно-корреляционным функционалом Perdew-Burke-Ernzerhof) и MT\_GIACanada\_BLYP (псевдопотенциал Martins-Trouiller с гибридным обменно-корреляционным функционалом Беке-Ли-Янга-Парра (Becke-Lee-Yang-Parr)).

**2.1. Циклопентан.** По результатам расчета волновой функции для циклопентана были выбраны значения cutoff от 110 до 190. После расчета геометрической оптимизации погрешность по длине связи C–C и для угла H–C–C



С была наименьшая при псевдопотенциале BLYP. Минимальная погрешность при cutoff = 180 составляет  $-0.00541 \text{ \AA}$  для длины связи C–C и  $2.192^\circ$  для величины угла H–C–C. На основе оптимизированных данных была проведена молекулярная динамика со значением cutoff = 110. Сводные данные по результатам оптимизации представлены в табл. 1.

По результатам расчета молекулярной динамики был сделан вывод, что для получения точных расчетных параметров молекулы необходимо больше времени, чем было задано в начале (рис. 1).

**2.2. Метилсилан.** Результаты расчета волновой функции, геометрической оптимизации и молекулярной динамики методом теории функционала плотности с псевдопотенциалом MT\_PBE и значением cutoff = 110 показали, что геометрические параметры молекулы метилсилана отличаются от экспериментальных в незначительной степени и при увеличении времени расчета можно получить очень точные данные.

**2.3. Диметиловый эфир.** Квантово-химический расчет геометрических параметров молекулы диметиловый эфир так же показал, что при увеличении времени, затрачиваемого на молекулярную динамику возможно получить более точные данные. Величина отклонения от экспериментальных данных приведена в табл. 2.

Табл. 1

Отклонение расчетных параметров от экспериментальных после расчета геометрической оптимизации

cutoff	Изменение длины связи C–C, Å		Изменение величины угла H–C–C, Å		Изменение длины связи C–H, Å	
	PBE	BLYP	PBE	BLYP	PBE	BLYP
110	0.00461		2.259		0.01284	
120	0.00501	–0.00558	2.262	2.199	0.01296	0.01778
130	0.00508	–0.00554	2.261	2.199	0.01301	0.01778
140	0.00506	–0.00565	2.259	2.196	0.01309	0.01778
150	0.00491	–0.00579	2.259	2.198	0.01322	0.01787
160	0.00496	–0.00581	2.26	2.203	0.01334	0.01791
170	0.00498	–0.00581	2.259	2.203	0.01342	0.01793
180	0.00493	–0.00541	2.265	2.192	0.01344	0.01776
190	0.00493	–0.00584	2.263	2.203	0.01347	0.01793

Табл. 2

Величина отклонения расчетных данных от экспериментальных после расчета молекулярной динамики

Δ длины связи C–O	Δ длины связи C–H	Δ величины угла C–O–C
–0.012812455	–0.013594827	–1.145168051

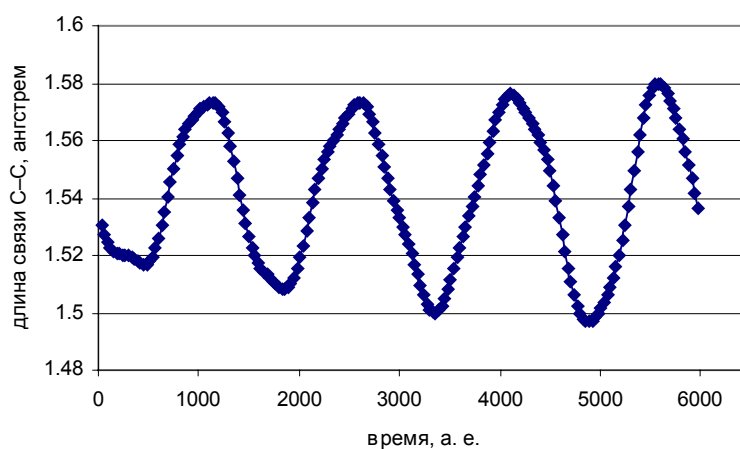


Рис. 1. Изменение длины связи C–C в молекуле циклопентана с течением времени (на основе расчета молекулярной динамики методом DFT с псевдопотенциалом MT\_PBE и значением cutoff = 110), а. е. – атомные единицы

**2.4. Сероуглерод.** На основании результатов, полученных после расчета волновой функции, для молекулы сероуглерода проводился расчет геометрической оптимизации. Данные, полученные геометрической оптимизацией, сравнили с экспериментальными данными. На рис. 2 и 3 представлены результаты сравнения, отражающие величину погрешности для метода теории функционала плотности с двумя различными псевдопотенциалами при различных значениях cutoff.

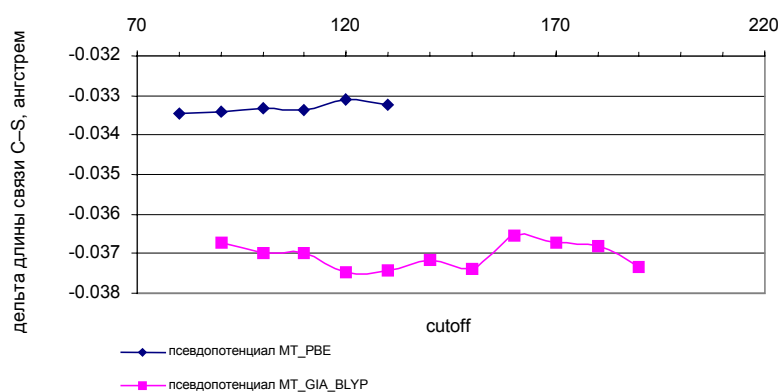


Рис. 2. Изменение величины погрешности длины связи между атомами C–S в молекуле сероуглерода в зависимости от значений cutoff (оптимизация методом DFT при псевдопотенциалах MT\_PBE и MT\_GIA\_BLYP)

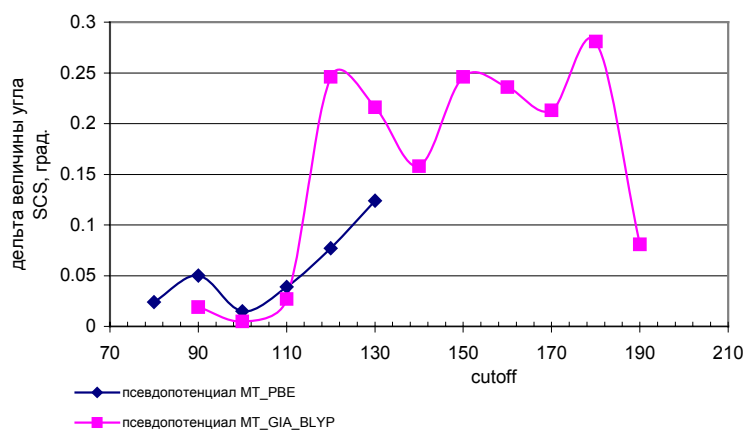


Рис. 3. Изменение величины погрешности для угла между атомами S–C–S в молекуле сероуглерода в зависимости от значений cutoff (оптимизация методом DFT при псевдопотенциалах MT\_PBE и MT\_GIA\_BLYP)

Величина отклонения расчетных данных от экспериментальных после расчета молекулярной динамики составляет  $2.1425319^\circ$  для величины угла C–S–C, а для длины связи C–S равна  $-0.10218329 \text{ \AA}$ .



и молекулярной динамики.

**2.5. Глицин.** Аминокислота глицин является простейшей аминокислотой, состоящей из минимального количества атомов, а также основой для всех остальных аминокислот, поэтому она идеально подходит для исследования и оптимизации метода теории функционала плотности. Для данной молекулы был произведен расчет волновой функции, геометрической оптимизации и молекулярной динамики.

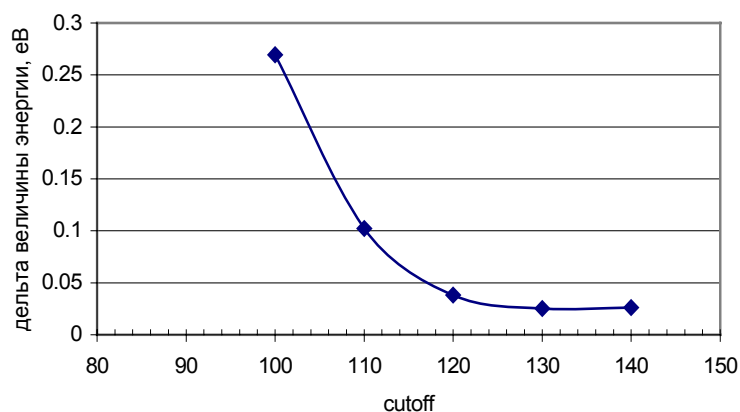


Рис. 4. Зависимость изменения величины энергии от значений cutoff для молекулы глицина (расчет волновой функции методом теории функционала плотности с псевдопотенциалом MT\_PBE)

По результатам расчета волновой функции были выбраны следующие значения cutoff для псевдопотенциала MT\_PBE от 100 до 130 и для псевдопотенциала MT\_GIA\_BLYP от 100 до 140 (рис. 4).

Для молекулы глицина после расчета геометрической оптимизации погрешность для длин связи является весьма незначительной и не превышает  $0.02 \text{ \AA}$ , а для некоторых параметров, таких как длина связи C–O(H) и длина связи C–N, не более  $0.008 \text{ \AA}$  (см. табл. 3).

Расчет молекулярной динамики показал, что данные, полученные методом теории функционала плотности для длин связи незначительно отличаются от экспериментальных. Погрешность в расчетах для длин связи составляет около  $0.01 \text{ \AA}$ , а для величины углов – от  $1$  до  $4^\circ$ .

**2.6. Аланин.** Для молекулы аланина также были проведены расчеты волновой функции, геометрической оптимизации и молекулярной динамики. По результатам волновой функции были выбраны такие же значения cutoff, что и для глицина.

После расчета геометрической оптимизации был выбран cutoff = 110, поскольку наименьшая погрешность наблюдается именно при этом значении. Погрешность для длин связи в молекуле аланина варьирует от  $0.001$  до  $0.02 \text{ \AA}$  в зависимости от связи. Погрешность для величин углов варьирует для молекулы аланина от  $0.6$  до  $3^\circ$ .

После расчета молекулярной динамики методом теории функционала плотности с псевдопотенциалом MT\_PBE и значением cutoff = 110 погрешность для длин связей стала ещё более незначительной и варьировала в пределах от  $0.007$  до  $0.04 \text{ \AA}$  (см. табл. 4). Для величины угла погрешность варьировала в пределах от  $0.47$  до  $3^\circ$ .

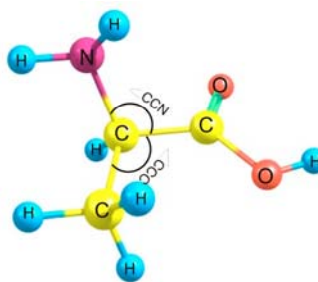


Табл. 3

Отклонение расчетных параметров от экспериментальных после расчета геометрической оптимизации для длин связей

cutoff	Δ длины связи C–N, Å		Δ длины связи C–C, Å		Δ длины связи C–O(H), Å		Δ длины связи C–O, Å	
	PBE	BLYP	PBE	BLYP	PBE	BLYP	PBE	BLYP
100	0.0084	–0.00316	0.01416	0.00633	–0.0060	–0.0169	–0.0088	–0.0101
110	0.00866	–0.00354	0.01414	0.0066	–0.0056	–0.01732	–0.0085	–0.0101
120	0.0086	–0.00355	0.0135	0.00581	–0.0065	–0.01783	–0.0086	–0.0103
130	0.00873	–0.00394	0.01357	0.00581	–0.0064	–0.01669	–0.0088	–0.0103
140	–	–0.00344	–	0.00577	–	–0.01736	–	–0.01
200	0.00875	–	0.01373	–	–0.0063	–	–0.0087	–

cutoff	Δ длины связи O–H, Å		Δ длины связи N–H, Å		Δ длины связи C–H, Å	
	PBE	BLYP	PBE	BLYP	PBE	BLYP
100	–0.00464	–0.00231	–0.00632	–0.00281	–0.00556	0.00018
110	–0.00425	–0.00229	–0.00608	–0.00293	–0.00537	0.00051
120	–0.00406	–0.00196	–0.00594	–0.00302	–0.00497	0.00053
130	–0.00404	–0.00205	–0.00587	–0.00292	–0.00524	0.00067
140	–	–0.00191	–	–0.00297	–	0.00055
200	–0.00385	–	–0.00553	–	–0.00456	–

Табл. 4

Отклонение расчетных данных от экспериментальных (на основе расчета молекулярной динамики)

Δ угла C–C–N, град.	Δ угла C–C–C, град.	Δ угла C–C–O(H), град.	Δ угла C–C–O, град.
1.482775337	–2.997570038	–2.713821348	0.472657067
Δ длины связи C–C, Å	Δ длины связи C–N, Å	Δ длины связи C–O(H), Å	Δ длины связи C–O, Å
–0.007437391	–0.015035288	–0.043945855	–0.023465794

На рис. 5 показано, что величина угла между атомами C–C–N только начинает колебаться, и для достижения точного значения необходимо увеличить время расчетов. Для получения точных значений остальных параметров молекулы аланина также требуется увеличить время расчетов.

### Заключение

Метод теории функционала плотности дает адекватные результаты, сравнимые с экспериментальными данными для исследуемых молекул. Псевдопотенциал MT\_GIA\_BLYP дает более точные результаты, однако требует больше расчетного времени. Для моделирования биохимических систем оптимальным методом является метод теории функционала плотности с псевдопотенциалом MT\_GIA\_BLYP и значением cutoff = 110.



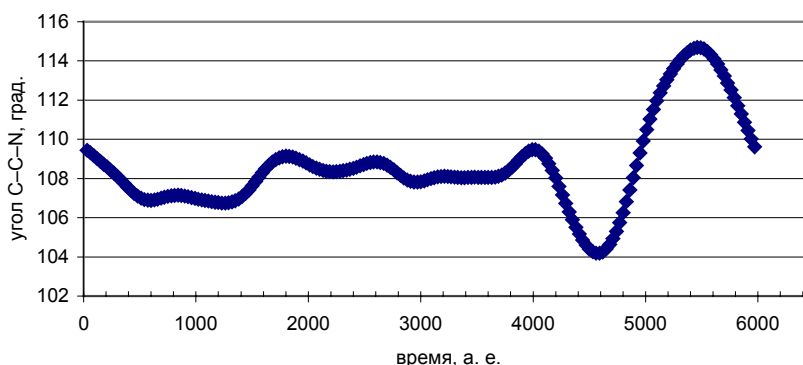


Рис. 5. Изменение величины угла между атомами С–С–N в молекуле аланина с течением времени (расчетом молекулярной динамики методом теории функционала плотности с псевдопотенциалом MT\_PBE и значением cutoff = 110)

Авторы выражают благодарность сотрудникам отдела информационных технологий КазНЦ РАН М.Н. Астафьеву и Д.В. Чачкову за возможность использования ресурсов Суперкомпьютерного Центра коллективного пользования КазНЦ РАН.

### Summary

*D.A. Alisheva, D.S. Tarasov, N.I. Akberova.* Optimization of quantum chemical computations method for biological macromolecules.

This paper presents results of optimization of method DFT (density functional theory) for the biological macromolecules by the example of such molecules as glycine, alanine, dimethyl ether, methylsilane and others.

### Литература

1. *Okamoto Y., Okamoto Y., Sugita Y., Nagahama T.Y., Mitsutake A., Nishikawa T.* Protein folding simulations from the first principles // Annual Report of the Earth Simulator Center. – Tokyo: Institute for Molecular Science, 2005. – P. 169–170.
2. *Sharan R., Rubas W., Kolling W.M., Ghandehari H.* Molecular modeling of arginine-glycine-aspartic acid (RGD) analogs: relevance to transepithelial transport // J. Pharm. Pharmaceut. Sci. – 2001. – V. 4, No 1. – P. 32–41.
3. *Венер М.В., Фомин Д.П., Егорова А.Н., Цирельсон В.Г.* От молекулярных кристаллов с Н-связями к самоорганизующимся ансамблям макромолекул. Квантово-топологический анализ // Тез. докл. IV Нац. кристаллохим. конф., Черногловка, 26–30 июня 2006 г. – М.: Рос. хим.-технол. ун-т им. Д.И. Менделеева, 2006. – С. 87 (URL: <http://www.icp.ac.ru/conferences/new/NCCC2006/thfiles/5/060317-0554376-Vener.doc>).
4. *Peng S., Ding F., Urbanc B., Buldyrev S.V., Cruz L., Stanley H.E., Dokholyan N.V.* Discrete molecular dynamics simulations of peptide aggregation // Phys. Rev. E. – 2004. – V. 69. – P. 041908-1–041908-7.
5. *Fitch B.G., Germain R.S., Mendell M., Pitera J., Pitman M., Rayshubskiy A., Sham Y., Suits F., Swope W., Ward T.J.C., Zhestkov Y., Zhoua R.* Blue Matter, an application framework for molecular simulation on Blue Gene // J. of parallel and distributed computing. – 2003. – V. 63, No 7–8. – P. 759–773.

6. *Фларри Р., Германа Э.Д., Розенберга Е.Л.* Квантовая химия / Пер. с англ. – М.: Мир, 1985. – 256 с.
7. *Foresman J.B.* Exploring chemistry with electronic structure methods. – Pittsburgh, PA: USA Gaussian Inc., 1996.
8. *Минкин В.И., Симкин Б.Я., Миняев Р.М.* Теория строения молекул. – Ростов н/Д: Феникс, 1997. – 560 с.
9. *Rogers D.W.* Computational Chemistry using the PC. – New Jersey, Canada: Wiley-Interscience, 2003. – 363 p.
10. *Yong Duan, Lu Wang, Kollman P.* The early stage of folding of villin headpiece subdomain observed in a 200-nanosecond fully solvated molecular dynamics simulation // PNAS USA. – 1998. – V. 95, No 17. – P. 9897–9902.
11. *Sherrill C.D.* An Introduction to Hartree-Fock Molecular Orbital Theory. – School of Chemistry and Biochemistry Georgia Institute of Technology, 2000  
(URL: <http://vergil.chemistry.gatech.edu/notes/hf-intro/hf-intro.pdf>).
12. *Язев О.В., Устынюк Ю.А.* Теоретическое исследование строения и свойств тетрааминоэтилена неэмпирическими методами и методом функционала плотности // Вест. Моск. ун-та. Сер. Химия. – 2000. – Т. 41, № 5. – С. 315–318.
13. *Burke K.* The ABC of DFT. – Rutgers University, 2003  
(URL: <http://dft.rutgers.edu/kieron/beta>).

Поступила в редакцию  
15.03.07

---

**Алишева Диана Айратовна** – студент кафедры биохимии Казанского государственного университета.

**Тарасов Денис Станиславович** – аспирант кафедры генетики Казанского государственного университета.

E-mail: [dtarasov@compnera.com](mailto:dtarasov@compnera.com)

**Акберова Наталья Ивановна** – кандидат биологических наук, старший научный сотрудник кафедры биохимии Казанского государственного университета.

E-mail: [nakberov@ksu.ru](mailto:nakberov@ksu.ru)