

Научно-исследовательский журнал «**Chemical Bulletin**»
<https://cb-journal.ru>
2024, Том 7, № 3 / 2024, Vol. 7, Iss. 3 <https://cb-journal.ru/archives/category/publications>
Научная статья / Original article
УДК 544.18, 004.42
DOI: 10.58224/2619-0575-2024-7-3-49-63

Комплекс возможных аппроксимативных методов эффективного учета вклада кулоновских интегралов для кардинального ускорения расчетов DFT гигантских биомолекул: сведение к быстровычисляемым короткодействующим двухцентровым сплайнам плюс FMM дальнего кулона

Аникин Н.А.*,

Институт органической химии им. Н.Д. Зелинского Российской академии наук,

* Ответственный автор E-mail: nikan53@ioc.ac.ru

Аннотация: предложен комплекс аппроксимативных методов для кардинального ускорения расчета вклада кулоновских интегралов при расчетах DFT гигантских биомолекул - лимитирующей стадии таких актуальных, но крайне ресурсоемких расчетов, включая расчеты тысяч докинг-комплексов из тысяч атомов. Предложенный комплекс включает через быструю и аккуратную аппроксимацию вклада громадного числа 4-центровых кулоновских интегралов через линейную комбинацию 3-центровых интегралов, и далее через комбинацию 2-центровых интегралов. Немультиполные короткодействующие компоненты этих 2-центровых интегралов очень быстро считаются заранее заготовленными сплайнами от межцентровых расстояний. Остающиеся дальнодействующие мультипольные вклады быстро считаются для гигантских молекул в стиле FMM (разбиение громадного пространства на области и подобласти, исходно было разработано для динамики галактик). Везде максимально экономятся расчеты за счет предварительно отобранных комбинаций интегралов. Все двухцентровые компоненты (в т.ч. аппроксимация двухцентровых перекрываний базисных функций через линейные комбинации одноцентровых вспомогательных функций плотности) быстро считаются за счет сплайнов от межядерных расстояний из специально приготовленной БД. Для новых базисов БД легко и быстро пополняется за счет разложения нового базиса по набору универсальных экспонент и БД с ними.

Ключевые слова и фразы: быстродействующие квантовохимические методы, большие молекулы, DFT, кулоновские интегралы, докинг-комплексы

Для цитирования: Аникин Н.А. Комплекс возможных аппроксимативных методов эффективного учета вклада кулоновских интегралов для кардинального ускорения расчетов DFT гигантских биомолекул: сведение к быстровычисляемым короткоткодействующим двухцентровым сплайнам плюс FMM дальнего кулона // Chemical Bulletin. 2024. Том 7. № 3. С. 49 – 63. DOI: 10.58224/2619-0575-2024-7-3-49-63

Поступила в редакцию: 24 мая 2024 г.; Одобрена после рецензирования: 25 июля 2024 г.; Принята к публикации: 28 августа 2024 г.

**A set of possible approximative methods for effectively accounting for the contribution
of Coulomb integrals to dramatically accelerate the calculations of DFT giant biomolecules:
reduction to fast-computable short-range two-center splines plus FMM long-range Coulomb**

Anikin N.A.,*

Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences,

** Corresponding author E-mail: nikan53@ioc.ac.ru*

Abstract: a set of approximative methods is proposed to radically accelerate the calculation of the contribution of Coulomb integrals in the calculations of DFT giant biomolecules - the limiting stage of such relevant but extremely resource-intensive calculations, including calculations of thousands of docking complexes of thousands of atoms. The proposed complex includes, through a quick and accurate approximation of the contribution of a huge number of 4-center Coulomb integrals through a linear combination of 3-center integrals, and then through a combination of 2-center integrals. The non-multi-complete short-range components of these 2-center integrals are very quickly considered pre-prepared splines from the center-to-center distances. The remaining long-range multipole contributions are quickly calculated for giant molecules in the FMM style (splitting a huge space into regions and subdomains, was originally developed for the dynamics of galaxies). Calculations are saved as much as possible everywhere due to pre-selected combinations of integrals. All two-center components (including the approximation of two-center overlaps of basic functions through linear combinations of single-center auxiliary density functions) are quickly calculated due to splines from internuclear distances from a specially prepared database. For new bases, the database is easily and quickly replenished by decomposing the new basis into a set of universal exponents and a database with them.

Keywords: high-speed quantum chemical methods, large molecules, DFT, Coulomb integrals, docking complexes

For citation: Anikin N.A. A set of possible approximative methods for effectively accounting for the contribution of Coulomb integrals to dramatically accelerate the calculations of DFT giant biomolecules: reduction to fast-computable short-range two-center splines plus FMM long-range Coulomb. Chemical Bulletin. 2024. 7 (3). P. 49 – 63. DOI: 10.58224/2619-0575-2024-7-3-49-63

The article was submitted: May 24, 2024; Approved after reviewing: July 25, 2024; Accepted for publication: August 28, 2024.

Введение

1.1. Квантовохимические методы позволяют эффективно расчитывать разнообразные свойства молекул, включая актуальные гигантские биомолекулы, требующие, однако, громадных компьютерных ресурсов, тем более при массовых расчетах. Поэтому для них наряду с использование суперкомпьютеров актуальна (см., например, [1]) разработка специальных методов (в т.ч. со специфическими аппроксимациями) для кардинального ускорения расчета ресурсоемких стадий, включая вклад кулоновских интегралов.

Для эффективного расчета двухцентровых кулоновских интегралов, возникающих при решении квантовохимических уравнений для многоатомных молекул используются наборы базисных функций, экспоненциально убывающие с расстоянием от ядер атомов – атомные орбитали (АО) или базисные орбитали – базисные функции (БФ), в виде произведения экспоненциально убывающего радиального множителя и множителя, зависящего от углов в сферических координатах:

$$\phi(r, \Theta, \varphi) = Y_{lm}(\Theta, \varphi) r^l R(r), \quad (1)$$

Радиальный множитель при неэмпирических (включая DFT) расчетах обычно задается в виде линейная комбинация (ЛК) гауссовых функций для эффективного расчета большого числа двухэлектронных 4-центровых интегралов:

$$R(r) = \sum_k \beta_k e^{-\alpha_k r^2} \quad (2)$$

Материалы и методы исследований

1.2. Ресурсоемкость квантовохимических расчетов (особенно время расчета) сильно растет с

ростом числа N используемых АО или БФ при росте размера молекулы. При отсутствии дополнительных аппроксимаций число двухэлектронных 4-центровых кулоновских интегралов в расчетах DFT растет очень сильно, вплоть до N^4 , требуя чрезмерного времени расчета очень больших и гигантских молекул из сотен и тысяч атомов (N типа 10 000 и более).

После расчета молекулярных интегралов от АО/БФ и их вклада в одноэлектронный гамильтониан (фокиан) время расчета последующих стадий растет медленнее, как N^3 и менее.

Существуют гораздо более быстрые полуэмпирические квантовохимические методы расчета гигантских молекул (ориентировано примерно в 1000 раз быстрее, чем широко используемый DFT [2]), у них время расчета расчет только как $O(N^3)$, и даже менее, см. например разработанный нами новый метод LocalSCF [3] для быстрых полуэмпирических расчетов гигантских биомолекул, и его применение для сверхбыстрых массовых расчетов тысяч докинг-комплексов из тысяч атомов [4]. Поэтому полуэмпирические методы уже ранее использовались для крайне ресурсоемких квантовохимических расчетов гигантских биомолекул (например расчет биомолекулы из 100 тысяч атомов на 1024 процессорах [5], и расчеты полных докинг-комплексов на суперЭВМ Ломоносов [6]). Однако полуэмпирические расчеты уступают по точности широко распространенным расчетам DFT. Нами предложен [7] метод быстрого полуэмпирического моделирования гамильтониана DFT.

Результаты и обсуждения

Оптимизация геометрии гигантских молекул типа белков требуются еще значительно больше времени, например оптимизация геометрии всего лишь фрагмента белка из 600 атомов на уровне DFT на суперЭВМ Stampede потребовал более ча-

са времени на 256 ядрах (ускорение в 114 раз.)[8], на рис. 1. представлена зависимость времени расчета от числа используемых процессорных ядер, со снижение распараллеливания при большем их числе.

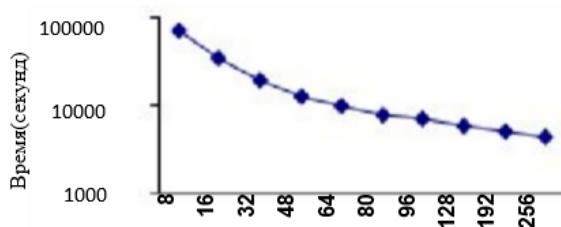


Рис. 1. Зависимость времени расчета DFT белка от числа использованных процессорных ядер.

Fig. 1. Dependence of protein DFT calculation time on the number of processor cores used.

Иногда для суперускорения расчетов гигантских бимолекул квантовохимически (QM) рассчитывают лишь небольшой наиболее активный фрагмент громадного белка, а большую часть рассчитывают радикально более быстрыми классическими методами молекулярной механики (MM) – метод QM/MM [9, 10, 11], однако при этом возникают проблемы на границе квантовой и классической части молекулы, иногда решение – с применением фрагментных молекулярных орбиталей FMO [11].

Один из классов актуальных расчетов тысяч молекул (молекулярных комплексов) каждый из которых из тысяч атомов – это массовые расчеты докинг-комплексов из малых органических молекул (лигандов) с огромными молекулами белков (например при разработке будущих лекарств – пока лигандов). Пока их чаще проводят более простыми и быстрыми (но менее точными, без явного учета квантовых эффектов) методами молекулярной механики (MM), так как некоторые авторы считают неприемлемо ресурсоемкими (и потому массово

практическим нереальными) расчеты таких систем (тем более их тысяч) на полностью квантовом уровне за разумное время [11].

Поэтому сейчас наша цель перейти от ускорения полуэмпирических расчетов к кардинальному ускорению более точных расчетов DFT гигантских (био)молекул. Но при этом необходимо кардинально ускорить расчеты гигантского числа более сложных (чем в полуэмпирике) интегралов и их вклад в одноэлектронный гамильтониан (фокиан). Для этого необходим оптимальный комплекс эффективных аппроксимаций, сочетающих высокое быстродействие (и оптимальную масштабируемость с ростом N числа БФ) с хорошей точностью аппроксимации, что и предлагается в данной статье для лимитирующей стадии расчета DFT гигантских молекул – расчет кулоновского вклада в одноэлектронный гамильтониан (фокиан), в том числе при массовых расчетах тысяч докинг-комплексов из тысяч атомов. Наряду с другими методами для радикального ускорения расчета

кулоновских взаимодействий в гигантских молекулах из тысяч атомов и достижения линейного масштабирования актуально использование в расчетах метода быстрых мультиполей FMM [12].

2. Кардинальное ускорение расчета кулоновского вклада больших молекул путем сведения вклада четырехцентровых интегралов к вкладу трехцентровых и далее двухцентровых

2.1. При крайне ресурсоемких неэмпирических/DFT расчетах гигантских молекул лимитирует расчет вклада двухэлектронных четырехцен-

тровых кулоновских интегралов от БФ ϕ_j (число таких интегралов расчет с ростом размера молекулы как $O(N^4)$). Для начала ускорение этой стадии может быть достигнуто аппроксимацией двухцентровых перекрываний БФ $\phi_i(\mathbf{r}_1)\phi_j(\mathbf{r}_1)$ через ЛК так называемых вспомогательных функций плотности (ВФП) (на английском Auxiliary basis, см. например [13]) – специально подобранных однокентровые гауссовые функции $\chi_p(\mathbf{r}_1)$, центрированные (как и БФ) на ядрах атомов:

$$\phi_i(\vec{r}_1)\phi_j(\vec{r}_1) = f_n(\vec{r}_1) \approx g_n(\vec{r}_1) = \sum_p B_{ijp} \chi_p(\vec{r}_1) \quad (3)$$

При этом четырехцентровые кулоновские интегралы (их число растет как $O(N^4)$) сводятся к комбинации трехцентровым, число которых растет уже только как $O(N^3)$, что существенно ускоряет

расчет DFT (особенно его кулоновского вклада – лимитирующей части) достаточно больших молекул – см. табл. 1, и это ускорение еще больше для гигантских молекул.

$$\iint \frac{\phi_i(\vec{r}_1)\phi_j(\vec{r}_1)\phi_k(\vec{r}_2)\phi_l(\vec{r}_2)}{r_{12}} dV_1 dV_2 = \iint \frac{f_n(\vec{r}_1)f_m(\vec{r}_2)}{r_{12}} dV_1 dV_2 = \langle f_n | f_m \rangle, \quad (4)$$

укороченное обозначение двухэлектронных интегралов .

Таблица 1
Table 1

Молекула/Симметрия	$N_{БФ}^a$	$N_{ВФП}^b$	Память ^c	t_{prep}^d	$t_{RL-j}^e / type^f$	t_{quad}^g	t_j^h
Fe(Cp) ₂ /D _{5d}	184	492	9	3.91	0.13/i	8.53	28.98
	184	492	9	3.93	0.13/i	15.08	28.90
Mn ₂ (CO) ₁₀ /D _{4d}	328	874	53	17.17	0.86/i	37.29	138.18
P(Ph) ₃ /C ₃	300	781	93	33.80	1.50/i	59.36	466.39
	300	781	93	33.75	1.49/i	106.51	463.43
Mn ₂ Cu ₂ (CO) ₁₀ (NH ₃) ₂ /C _{2h}	416	1124	107	49.53	31.87/sd	137.88	894.76

Продолжение таблицы 1
Continuation of Table 1

	672	1290	240	91.0	11.53/i	151.40	1732.95
	672	1290	97	56.1	73.41/sd	151.40	1732.95
	672	1290	11	3.51	178.95/d	151.40	1732.95
CH ₃ CO(gly)4N HCH ₃ /C ₁	389	929	56	60.43	179.93/sd	228.01/2	1630.23
	389	746 ⁱ	146	91.08	81.08/sd	228.77/2	1630.23

^a Число базисных функций БФ.

^b Число ВФП.

^c Оперативная Память , МВ.

^d Время расчета до итераций ССП 3-центровых интегралов от БФ с ВФП и 2-центровых от ВФП, и обработка последних.

^e Время на одну итерацию ССП с ВФП.

^f Типы алгоритмов работы с оперативной памятью In-core (i), semi-direct (sd) или direct (d) при работе с ВФП.

^g Время расчета обменно-корреляционной энергии DFT E_{xc}.

^h Время обычного расчета без ВФП (на 1ой итерации ССП).

ⁱ Расчеты с альтернативным базисом ВФП aux-c.

a Number of basis functions of the BF.

b Number of VFP.

c RAM, MB.

d Calculation time before the iterations of the BCF of 3-center integrals of the BF with VFP and 2-center integrals of the VFP, and processing of the latter.

e Time for one iteration of the BCF with VFP.

f Types of algorithms for working with RAM In-core (i), semi-direct (sd) or direct (d) when working with VFP.

g Calculation time of the exchange-correlation energy DFT Exc.

h Time of a regular calculation without VFP (at the 1st iteration of the BCF).

i Calculations with an alternative basis of the VFP aux-c.

Произведение двух БФ (и гауссовых функций в их составе) на двух соседних атомах приближенно аппроксимируется ЛК ВФП. Для повышения точности расчета кулоновских интегралов при аппроксимации ЛК ВФП (3) можно использовать

сведение всех четырехцентровых интегралов от БФ ($f_n|f_m$) к трехцентровым интегралам ($f_n|g_m$) (их число растет медленнее, как $O(N^3)$) с точностью до малых второго порядка по отклонению f от g :

$$\begin{aligned}
 (f_n|f_m) &= (g_n + (f_n - g_n)|g_m + (f_m - g_m)) = \\
 &= (g_n|g_m) + (f_n - g_n|g_m) + (g_n|f_m - g_m) + (f_n - g_n|f_m - g_m) \approx \\
 &\approx (f_n|g_m) + (g_n|f_m) - (g_n|g_m)
 \end{aligned} \tag{5}$$

Здесь $(f_n|g_m)$ включает трехцентровые интегралы, так как g_m – линейная комбинация одноцентровых ВФП χ_p , а f_n – двухцентровые функции (3).

При этом даже специально подобранных ВФП χ_p для хорошей точности аппроксимации требуется примерно в два-три раза больше, чем БФ, но это много меньше $N^2/2$ числа произведений пар БФ $\phi_i(r)\phi_j(r)$, а для повышения точности используется приближение (5).

Для ориентировочной иллюстрации размеров массивов кулоновских интегралов – см. табл. 2 – приведем типичный относительно небольшой белок из ~ 1700 атомов (его изображение – см. рис. 2) – комплекса человеческого иммуноглобулина с иммунным супрессором IMMUNOPHILIN FKBP-12 из Protein Data Bank [14].

Ориентировочное число кулоновских интегралов для типичного относительно небольшого белка из 1700 атомов.

Таблица 2

Table 2

Approximate number of Coulomb integrals for a typical relatively small protein of 1700 atoms.

Число БФ при использовании часто применяемого в расчетах DFT базиса 6-31G*,	$\sim 14\ 000$
Число ВФП $\chi_p(r)$	$\sim 30\ 000$
Число непренебрежимых 4-центровых кулоновских интегралов $(f_n f_m)$ только от f_n и f_m	$\sim 2 \times 10^{14}$
Число непренебрежимых 3-центровых кулоновских интегралов $(f_n \chi_p)$ в составе $(f_n g_m)$.	$\sim 2 \times 10^{11}$

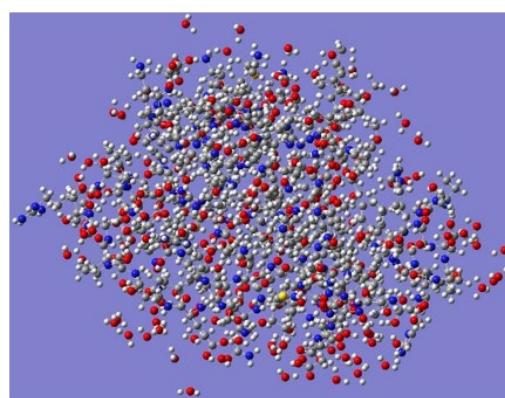


Рис. 2. Типичный относительно небольшой белок IMMUNOPHILIN FKBP-12 из примерно 1700 атомов.

Fig. 2. A typical relatively small protein IMMUNOPHILIN FKBP-12 of approximately 1700 atoms.

Разумеется, вычисление каждого 3-4-центрового кулоновского интеграла требует много арифметических операций, тем более – с БФ из нескольких гауссовых функций (от которых, собственно, исходно и считаются кулоновские интегралы), тогда число операций растет уже по порядку величины (по сравнению с очень быстрыми сплайнами – см. ниже). Наглядно видна необходимость максимального комплекса возможных методов-аппроксимаций для кардинального ускорения расчета кулоновского вклада DFT гигантских биомолекул.

$$f_n(\vec{r}_1) \approx g'_n(\vec{r}_1) = \sum_p B'_{ijp} \chi'_p(\vec{r}_1) \quad (6)$$

Поскольку здесь уже точность аппроксимации получается в первом порядке малости по разности $(f_n - g'_n)$, а не во втором порядке малости, как было в ф-ле (5), то поэтому здесь нужна более точная аппроксимация с расширенным набором ДВФП, большим, чем ВФП, за счет дополнительных функций МЦВФП (МежЦентровые ВФП).

Произведение $f_n(r_1) = \phi_i(r_1)\phi_j(r_1)$ двух БФ (и составляющих их гауссовых функций) на двух разных центрах в пространстве относительно велико (хотя по величине относительно мало, но не пренебрежимо) как раз *Между* этими центрами, а хороший оптимизированный набор ВФП (их в 2-3 раза больше, чем БФ) хорошо аппроксимирует *вблизи* этих двух центров. При этом амплитуда этих меж-центровых произведений БФ, мала, и быстро убывает с увеличением расстояний между этими двумя центрами БФ (далее они становятся пренебрежимыми), поэтому достаточно повышения разумно-приближенной-экономичной аппрок-

2.2. Дальнейшее ускорение расчетов кулоновского вклада мы предлагаем сведением указанных трехцентровых интегралов ($f_n|\chi_p$) в составе ($f_n|g_m$) (их число растет как $O(N^3)$, в отличие от двухцентровых интегралов ($\chi_p|\chi_q$), число которых растет медленнее, как $O(N^2)$) к линейной комбинации двухцентровых интегралов путем более точной аппроксимации двухцентровых функций $f_n(r_1) = \phi_i(r_1)\phi_j(r_1)$ в виде линейной комбинации Дополненными межцентровыми ВФП (ДВФП) $\chi'_p(r_1)$:

симации приблизительно посередине (или около того) между данными двумя центрами БФ.

Поэтому здесь нужна более точная аппроксимация с расширением набора ВФП к набору ДВФП $\chi'_p(r_1)$, дополненному за счет дополнительных ограниченного (небольшого) числа МежЦентровые ВФП (МЦВФП, тоже с угловыми частями типа s, px, py, pz, dyz и т.д.), *между* центрами БФ ϕ_i и ϕ_j , где наиболее велика погрешность аппроксимации (6) через ВФП (центрированных только на центрах ϕ_i и ϕ_j).

Предварительные оценки показывают достаточность добавления небольшого числа МЦВФП между центрами БФ ϕ_i и ϕ_j . Для БФ с очень малыми гауссовыми экспонентами (относительно расстояния между центрами БФ) меж-центровое произведение БФ уже неплохо аппроксимируется ЛК ВФП на этих центрах (а с нашими добавками – тем более точно). При больших гауссовых экспонентах или больших расстояниях между центрами

БФ, это меж-центровое произведение БФ становится пренебрежимым. При большой экспоненте на одном центре БФ и малой на другом – их произведение относительно велико вблизи центра БФ с большой экспонентой, где оно неплохо аппроксимируется хорошим набором ВФП этого центра. Остается хорошо аппроксимировать двухцентровое произведение БФ вдали от этих центров (например посередине между ними или около того) и при экспонентах в БФ средней (для данного расстояния между центрами БФ) величины. Этому и отвечает ограниченный (небольшой) набор гауссовых функций МЦВФП на трех центрах между центрами БФ ϕ_i и ϕ_j (на равных расстояниях между ними) на каждом центре с набором немногих отдельных гауссовых функций (с показателями экспонент, образующими геометрическую прогрессию), зависящими от расстояния между центрами БФ ϕ_i и ϕ_j (больше расстояния – меньше показатели гауссовых экспонент).

2.3. Дополнительные меры для уменьшения числа вычисляемых компонент интегралов в больших молекулах.

При использовании МЦВФП все кулоновские интегралы от трехцентровых в составе $(f_n|g_m)$ сводятся к ЛК более простых двухцентровых кулоновских интегралов $(\chi'_p|\chi_q)$ в составе $(g'_n|g_m)$. При этом хотя число межцентровых МЦВФП (их не надо только для сильно удаленных центров, где все двухцентровые произведения БФ пренебрежимы) довольно велико (хотя и невелико для каждой пары центров), но при вычислении кулоновских интегралов от них имеются дополнительные существенные возможности ускорения.

2.3а. В двухцентровых кулоновских интегралах $(\chi'_p|\chi_q)$ для каждого χ'_p/χ_q в зависимости от его угловой части выделяем немногие мультиполи (мо-

нополь-заряд для s , диполь для $(px\,py\,pz)$, квадруполь для d , и т.д.). После этого дальнодействующая часть таких интегралов сводится к их простым взаимодействиям их немногих мультиполей (функционально не зависящий от тонкостей радиальных частей всех функций на данном центре) для медленно-неэкспоненциально (как $1/R$ для монополь-монополь, $1/R^2$ для монополь-диполь, $1/R^3$ для монополь-квадруполь/диполь-диполь и т.д.) затухающих взаимодействий мультиполей, и быстро экспоненциально затухающую (уже на относительно небольших расстояниях) остальную не-мультипольную часть кулоновских интегралов (уже явно зависящую от деталей МЦВФП функций χ'_p и χ_q).

2.3б. Суммируем вклады от всех МЦВФП $\chi'_p(r_1)$ на данном центре, что кардинально ускоряет расчет кулоновского вклада для больших молекул. Далее для больших молекул Мультипольные взаимодействия (с добавлением вклада монополей-зарядов ядер) группируем в стиле метода быстрых мультиполей FMM [12] путем сложения одинакового типа мультиполей в одной области пространства большой молекулы, и объединения (в т.ч. многократное) этих областей и их мультиполей в громадной молекуле по типу FMM. Такое суммирование мультиполей особенно актуально в нашем случае из-за увеличения числа мультиполей от МЦВФП.

2.3в. Не-мультипольная часть ограниченного числа быстро (экспоненциально) затухающих короткодействующих двухцентровых кулоновских интегралов задаем в виде заранее подобранных (кусочно-полиномиальных от межцентровых расстояний) Сплайнов (см. например [15]) их заранее заготовленной Базы Данных (БД). Более того, здесь богатые возможности использования ЛК

сплайнов = новому одному сплайну. В частности, можно задавать ВФП (и МЦВФП) в виде ЛК любого числа гауссовых функций – время расчета останется как с одной гауссовой функцией (что дает возможность существенно повысить их гибкость и точность без замедления расчета), т.к. для двухцентровых интегралов надо будет просто заранее сделать сплайн из ЛК сплайнов с отдельными гауссовыми функциями, а для мультипольной компоненты число гауссовых функций неважно. Учет угловой направленности отдельных ВФП/МЦВФП (и БФ) (типа $r_x r_y dxz$ и т.д.) в двухцентровых интегралах сводится к ЛК таких компонент (типа $r_x' r_y' dxz'$ и т.д.) вдоль и поперек связи между этими двумя центрами (интегралы от последних зависят только от расстояний между центрами).

2.3г. Для облегчения и ускорения конструирования Сплайнов от большого разнообразия базисов с разнообразными экспонентами мы можем аппроксимировать с высокой точностью каждую гауссову функцию на каждом центре с произвольной экспонентой как ЛК нескольких гауссовых функций с близкими к ней фиксированными экспонентами, образующими универсальную тесную геометрическую прогрессию. Сплайн, необходимый при нахождения двухцентровых интегралов от произвольных экспонент будет сведен к не-большой комбинации сплайнов от фиксированных экспонент из универсального набора. Это позволит для всех новых базисов быстро генерировать сплайны от расстояний как короткие комбинации сплайнов из универсальной БД. Таким образом можно быстро дополнить БД для новых базисов на основе универсальной БД от универсальных фиксированных экспонент.

3. Переход от отдельных интегралов к их вкладу в одноэлектронный гамильтониан (фокиан) и полную энергию.

Поскольку в реальности не обязательно предварительно проводить громоздкие (и ресурсоемкие) вычисления отдельных 4 и 3-центровых кулоновских интегралов, то можно перейти к нахождению вкладов 2-центровых кулоновских интегралов ($\chi_p'|\chi_q$) в составе ($g_n'|g_m$) и отдельно их мультипольной и не-мультипольной компонент в одноэлектронный гамильтониан (фокиан) и полную энергию.

Но вначале надо найти разложения двухцентровых функций $f_n(r)$ по ДВФП χ_p' в составе $g_n'(r)$, и с опорой на уже ранее сгенерированный массив не-мультипольных компонент двухцентровых кулоновских интегралов ($\chi_p'|\chi_q$) в составе ($g_n'|g_m$) – см. п. (2.3в).

3.1. Вначале аппроксимируем все непренебрежимые двухцентровые функции $f_n(r_1) = \phi_i(r_1)\phi_j(r_1)$ в виде ЛК ДВФП (ВФП на этих двух центрах, дополненными МЦВФП между этим центрами) $\chi_p'(r_1)$.

Поскольку здесь все зависит только от данных двух центров и их базисных наборов, то наиболее эффективный вариант – заранее заготовить сплайны (от межцентрового расстояния) для коэффициентов разложения каждой $f_n(r_1)$ по всем подходящим по симметрии (сигма по сигма, пи по пи, и т.д.) $\chi_p'(r_1)$ сначала ориентируясь относительно направления данного межцентрового расстояния, а потом вращая их переходом к общемолекулярной системе координат, аналогично п. (2.3в), см. выше.

Эти сплайны (совсем иные, чем для собственно кулоновских интегралов) заранее запасаем в подходящей БД, и опять можно быстро пополнить эту

БД для новых базисов, используя разложение нового базиса по фиксированным (универсальным, припасенных заранее) экспонентам, аналогично п. (2.3г) выше.

3.2. При оптимизации ЛК разложения всех функций $f_n(r_1) = \phi_i(r_1)\phi_j(r_1)$ в виде линейных комбинаций всех ВФП данной молекулы (а не только ВФП данных двух центров $f_n(r)$) мы получаем ВФП-аппроксимацию всей полной одноэлектронной матрицы плотности (МП) $\rho'(r)$:

$$\rho'(r) = \sum_p B_p \chi_p(r) \quad (7)$$

и далее используем относительно компактный массив B_p разложения полной МП $\rho'(r)$ по ВФП χ_p . Важно отметить, что на данной стадии (получения $\rho'(r)$) достаточно использования относительно компактного массива ВФП без большого массива МЦВФП ввиду погрешности второго порядка малости по точности аппроксимации, см. п. (2.1) выше [13] или [16].

3.3. При такой оптимизации (3.2) и использовании (см. п. (2.2)) разложения двухцентровых функций $f_n(r_1) = \phi_i(r_1)\phi_j(r_1)$ в виде линейной комбинации Дополненными межцентровыми ВФП (ДВФП) $\chi'_p(r_1)$, вклады 2-центровых кулоновских интегралов в одноэлектронный гамильтониан (фокиан) сводится к ЛК быстро-легко получаемым (через сплайны) двухцентровых интегралов ($\chi'_p | \chi_q$). Для экономии при расчетах с очень большим числом ДВФП (ввиду изобилия МЦВФП) мы разбиваем эти интегралы и их вклады на мультипольные (см. выше п. (2.3) и (2.3б)) и Не-мультипольные (см. выше п. (2.3в)) составляющие, как и указано выше:

3.3а. Мультипольные вклады (см. п. (2.3б)) хотя и дальнодействующие, но группируются в стиле метода быстрых мультиполей FMM [12] с фак-

тической группировкой мультиполей в каждой области пространства большой молекулы, и объединения (в т.ч. многократное) этих областей и их мультиполей в громадной молекуле. В первую очередь это относится общему мультипольному вкладу от всей МП $\square' +$ соотв. мультипольный вклад зарядов ядер. Это дает дальнодействующий электростатический потенциал всей громадной молекулы, в т.ч. за ее пределами (например, для потенциала для лигандов в полости громадного протеина (белка) или при взаимодействии нескольких громадных белков). Разумеется, также заранее находим мультипольной вклад для каждой ДВФП $\chi'_p(r_1)$. После этого все мультипольные взаимодействия находятся относительно быстро для громадных молекул.

3.3б. Не-мультипольные вклады (см. выше п.(2.3в)) ввиду их короткодействия тоже можем находить относительно быстро для громадных молекул (можно заранее выделить области пространства гигантской молекулы, где могут быть непрерывные Не-мультипольные вклады).

3.4. После этого, используя (см. п. (2.2)) уточненную (с добавкой МЦВФП) аппроксимацию двухцентровых функций $f_n(r_1) = \phi_i(r_1)\phi_j(r_1)$ в виде ЛК $\chi'_p(r_1)$, мы относительно быстро получаем кулоновский вклад громадной молекулы в одноэлектронный гамильтониан (фокиан).

3.5. Полная энергия считается либо стандартно, суммируя орбитальные энергии занятых МО плюс вклад оставного гамильтониана, либо кулоновская энергия (включая потенциал ядер) тем же способом стилем FMM плюс не-мультипольные короткодействующие вклады.

4. Предложенный комплекс мер (методов) радикального ускорения расчета вклада кулоновских интегралов – одной из лимитирующих стадий

крайне ресурсоемких расчетов (в т.ч. массовых) гигантских (био)молекул из сотен и тысяч атомов широко распространенным методом DFT, включает через быструю и аккуратную аппроксимацию:

4.1. вклада громадного (см. выше п. (2.1)) числа 4-центровых кулоновских интегралов через ЛК 3-центровых интегралов;

4.2. и далее через ЛК 2-центровых интегралов;

4.3. немультиполные короткодействующие компоненты этих 2-центровых интегралов очень быстро считаются заранее заготовленными сплайнами от межцентровых расстояний;

4.4. плюс дальнодействующие мультипольные вклады ,

4.5. быстро считаемые для гигантских молекул в стиле FMM (разбиение громадного пространства на области и подобласти, исходно было разработано для динамики галактик).

4.6. Везде максимально экономятся расчеты за счет предварительно отобранных комбинаций интегралов.

4.7. Все двухцентровые компоненты (в т.ч. аппроксимация двухцентровых перекрываний БФ через ЛК одноцентровых ВФП и ДВФП) быстро считаются за счет сплайнов от межядерных расстояний из специально подготовленной БД.

4.8. Для новых базисов БД легко и быстро пополняется за счет разложения нового базиса по набору универсальных экспонент и БД с ними.

Выводы

Предложенный комплекс мер (методов) позволяет кардинально ускорить расчеты гигантских (био)молекул методом DFT, в т.ч. массовые расчеты тысяч комплексов докинг-комплексов протеин-лиганд из тысяч атомов. Дополнительные возможности ускорения – за счет уменьшения числа непренебрежимых кулоновских интегралов даст использование ортогонализованных локализованных базисных функций [17].

Список источников

1. Jørgensen P., Kjaergaard T., Kristensen K., Baudin P., Ettenhuber P., Eriksen J.J., Wang Y.M., Bykov D. Quantum chemistry on the supercomputers of tomorrow / Smoky Mountains Computational Sciences and Engineering Conference (August 31 – September 2, 2015, Gatlinburg, Tennessee, USA).
2. Li A., Muddana H.S., Gilson M.K. Quantum mechanical calculation of noncovalent interactions: A large-scale evaluation of PMx, DFT, and SAPT approaches // J. Chem. Theory and Computation. 2014. № 10:4. P. 1563 – 1575. <https://doi.org/10.1021/ct401111c>
3. Anikin N.A., Anisimov V.M., Bugaenko V.L., Bobrikov V.V., Andreyev A.M. LocalSCF method for semiempirical quantum-chemical calculation of ultralarge biomolecules // J. Chem. Phys. 2004. Vol. 121. № 3. P. 1266 – 1270. <https://doi.org/10.1063/1.1764496>
4. Аникин Н.А., Андреев А.М., Кузьминский М.Б., Мендкович А.С. "Быстродействующий метод для массовых полуэмпирических расчетов докинг-комплексов // Изв. АН. Сер.: Хим. 2008. № 9. С. 1759 – 1764.
5. Hennemann M., Clark T. EMPIRE: a highly parallel semiempirical molecular orbital program: 1: self-consistent field calculations // Journal of Molecular Modeling. 2014. № 20:7. P. 2331. <https://doi.org/10.1007/s00894-014-2331-4>

6. Oferkin I.V., Katkova E.V., Sulimov A.V., Kutow D.C., Sobolev S.I., Voevodin V.V., Sulimov V.B. Evaluation of docking target functions by the comprehensive investigation of protein-ligand energy minima // Advances in Bioinformatics. 2015. № 2015. P. 126858. <https://doi.org/10.1155/2015/126858>
7. Аникин Н.А., Бугаенко В.Л., Кузьминский М.Б., Мендкович А.С. Быстродействующий метод для квантово-химических расчетов больших молекул с аппроксимацией гамильтониана DFT // Изв. АН. Сер.; Хим. 2014. № 2. С. 346 – 349.
8. Zhang J., Weisman A.L., Saitta P., Friesner R.A. Efficient simulation of large materials clusters using the jaguar quantum chemistry program: Parallelization and wavefunction initialization // Int. J. Quant. Chem. 2016. № 116:5. P. 357 – 368. <https://doi.org/10.1002/qua.25043>
9. Horvath I., Jeszenoi N., Balint M., Paragi G., Hetenyi C. A fragmenting protocol with explicit hydration for calculation of binding enthalpies of target-ligand complexes at a quantum mechanical level // Int. J. Molec. Sci. 2019. № 20:18. P. 4384. <https://doi.org/10.3390/ijms20184384>
10. Brunk E., Rothlisberger U. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States // Chemical Reviews. 2015. № 115 (12). P. 6217 – 6263. <https://pubmed.ncbi.nlm.nih.gov/25880693/> doi:10.1021/cr500628b. PMID 2588069
11. Nemukhin A.V., Polyakov I.V., Moskovsky A.I. Multi-scale supercomputing of large molecular aggregates: A case study of the light-harvesting photosynthetic center // Supercomputing Frontiers and Innovations. 2016. № 2:4. P. 48 – 54. <https://doi.org/10.14529/jsfi150403>
12. Le H.A., Shiozaki T. Occupied-orbital fast multipole method for efficient exact exchange evaluation // Journal of Chemical Theory and Computation. 2018. № 14:3. P. 1228 – 1234. <https://doi.org/10.1021/acs.jctc.7b00880>
13. Eichkorn K., Treutler O., Ohm H., Haser M., Ahlrichs R. Auxiliary basis sets to approximate Coulomb potentials // Chem. Phys. Letters. 1995. № 240. P. 283 – 290.
14. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. The Protein Data Bank // Nucleic Acids Research. 2000. № 28:1. P. 235 – 242.
15. Вержбицкий В.М. Численные Методы. Математический Анализ и Обыкновенные Дифференциальные Уравнения. Москва. ОНИКС 21 век. 2005, гл. 4.
16. Лайков Д.Н. Развитие экономного подхода к расчету молекул методом функционала плотности и его применение к решению сложных химических задач: дис. ... канд. физ.-мат. наук по спец. 02.00.17 – квантовая химия. Москва, 2000.
17. Anikin N.A., Bugaenko V.L., Frash M.V., Gorb L., Leszczynski J. Localized Basis Orbitals: Minimization of 2-Electron Integrals Array and Orthonormality of Basis Set // J. Comput. Chem. Vol. 24. 2003. № 9. P. 1132 – 1141.

References

1. Jørgensen P., Kjaergaard T., Kristensen K., Baudin P., Ettenhuber P., Eriksen J.J., Wang Y.M., Bykov D. Quantum chemistry on the supercomputers of tomorrow. Smoky Mountains Computational Sciences and Engineering Conference (August 31 – September 2, 2015, Gatlinburg, Tennessee, USA).
2. Li A., Muddana H.S., Gilson M.K. Quantum mechanical calculation of noncovalent interactions: A large-scale evaluation of PMx, DFT, and SAPT approaches. *J. Chem. Theory and Computation*. 2014. No. 10:4. P. 1563 – 1575. <https://doi.org/10.1021/ct401111c>
3. Anikin N.A., Anisimov V.M., Bugaenko V.L., Bobrikov V.V., Andreyev A.M. LocalSCF method for semiempirical quantum-chemical calculation of ultralarge biomolecules. *J. Chem. Phys.* 2004. Vol. 121. No. 3. P. 1266 – 1270. <https://doi.org/10.1063/1.1764496>
4. Anikin N.A., Andreev A.M., Kuzminsky M.B., Mendkovich A.S. "A high-speed method for mass semiempirical calculations of docking complexes. *Izv. AN. Ser.: Khim.* 2008. No. 9. P. 1759 – 1764.
5. Hennemann M., Clark T. EMPIRE: a highly parallel semiempirical molecular orbital program: 1: self-consistent field calculations. *Journal of Molecular Modeling*. 2014. No. 20:7. P. 2331. <https://doi.org/10.1007/s00894-014-2331-4>
6. Oferkin I.V., Katkova E.V., Sulimov A.V., Kutov D.C., Sobolev S.I., Voevodin V.V., Sulimov V.B. Evaluation of docking target functions by the comprehensive investigation of protein-ligand energy minima. *Advances in Bioinformatics*. 2015. No. 2015. P. 126858. <https://doi.org/10.1155/2015/126858>
7. Anikin N.A., Bugaenko V.L., Kuzminsky M. B., Mendkovich A.S. High-speed method for quantum-chemical calculations of large molecules with DFT Hamiltonian approximation. *Izv. AN. Ser.; Khim.* 2014. No. 2. P. 346 – 349.
8. Zhang J., Weisman A.L., Saitta P., Friesner R.A. Efficient simulation of large materials clusters using the jaguar quantum chemistry program: Parallelization and wavefunction initialization. *Int. J. Quant. Chem.* 2016. No. 116:5. P. 357 – 368. <https://doi.org/10.1002/qua.25043>
9. Horvath I., Jeszenoi N., Balint M., Paragi G., Hetenyi C. A fragmenting protocol with explicit hydration for calculation of binding enthalpies of target-ligand complexes at a quantum mechanical level. *Int. J. Molec. Sci.* 2019. No. 20:18. P. 4384. <https://doi.org/10.3390/ijms20184384>
10. Brunk E., Rothlisberger U. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chemical Reviews*. 2015. No. 115 (12). P. 6217 – 6263. <https://pubmed.ncbi.nlm.nih.gov/25880693/> doi:10.1021/cr500628b. PMID 2588069
11. Nemukhin A.V., Polyakov I.V., Moskovsky A.I. Multi-scale supercomputing of large molecular aggregates: A case study of the light-harvesting photosynthetic center. *Supercomputing Frontiers and Innovations*. 2016. No. 2:4. P. 48 – 54. <https://doi.org/10.14529/jsfi150403>
12. Le H.A., Shiozaki T. Occupied-orbital fast multipole method for efficient exact exchange evaluation. *Journal of Chemical Theory and Computation*. 2018. No. 14:3. P. 1228 – 1234. <https://doi.org/10.1021/acs.jctc.7b00880>

13. Eichkorn K., Treutler O., Ohm H., Haser M., Ahlrichs R. Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Letters.* 1995. No. 240. P. 283 – 290.
14. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. The Protein Data Bank. *Nucleic Acids Research.* 2000. No. 28:1. P. 235 – 242.
15. Verzhbitsky V.M. Numerical Methods. Mathematical Analysis and Ordinary Differential Equations. Moscow. ONIX 21st Century. 2005, Ch. 4.
16. Laikov D.N. Development an economical approach to calculating molecules using the density functional method and its application to solving complex chemical problems: dis. ... candidate of physical and mathematical sciences in the specialty 02.00.17 – quantum chemistry. Moscow, 2000.
17. Anikin N.A., Bugaenko V.L., Frash M.V., Gorb L., Leszczynski J. Localized Basis Orbitals: Minimization of 2-Electron Integrals Array and Orthonormality of Basis Set. *J. Comput. Chem.* Vol. 24. 2003. No. 9. P. 1132 – 1141.

Информация об авторе

Аникин Н.А., кандидат химических наук, научный сотрудник, Институт органической химии им. Н.Д. Зелинского Российской академии наук, email: nikan53@ioc.ac.ru

Information about the author

Anikin N.A., PhD in Chemistry, Scientific Researcher, N.D. Zelinsky Institute of Organic Chemistry of the Russian Academy of Sciences, email: nikan53@ioc.ac.ru

© Аникин Н.А., 2024