

Научно-исследовательский журнал «**Chemical Bulletin**»

<https://cb-journal.ru>

2025, Том 8, № 3 / 2025, Vol. 8, Iss. 3 <https://cb-journal.ru/archives/category/publications>

Научная статья / Original article

УДК 544.18, 004.42

DOI: 10.58224/2619-0575-2025-8-3-3

Комплекс возможных аппроксимативных методов эффективного многократного пересчета вклада кулоновских интегралов в элементы одноэлектронного гамильтониана на итерациях ССП для кардинального ускорения крайне ресурсоемких расчетов DFT гигантских биомолекул

¹ Аникин Н.А. *,

¹ Институт органической химии им. Н.Д. Зелинского Российской академии наук,

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

Аннотация: целью проведенного исследования является определение комплекса возможных аппроксимативных методов эффективного многократного пересчета вклада кулоновских интегралов в элементы одноэлектронного гамильтониана на итерациях ССП для кардинального ускорения крайне ресурсоемких расчетов DFT гигантских биомолекул. В процесс проведенного исследования решались следующие задачи: а) оценить возможность применения полуэмпирических методологий громадных молекул для квантовохимического анализа; б) обосновать перспективность многочисленных удаленных фрагментов гигантских молекул для ускорения вычисления вклада кулоновских взаимодействий; в) проанализировать имеющиеся подходы к одноточечным расчетам молекул с фиксированной геометрией; г) предложить комплекс методов ускоренного вычисления вклада кулоновских интегралов для DFT-расчетов актуальных гигантских биомолекул.

Для существенного ускорения вычисления вклада кулоновских интегралов в элементы одноэлектронного гамильтониана в рамках теории функционала плотности (DFT) применяется комплексный подход к аппроксимативным вычислениям. Данная методология направлена на преодоление ключевого узкого места – лимитирующей стадии итерационного процесса самосогласованного поля (ССП) при анализе крупномасштабных биологических структур, включая тысячи докинг-комплексов, состоящих из тысяч атомов.

Инновационный подход включает эффективную и точную аппроксимацию изменений вклада колоссального количества 4-центровых кулоновских интегралов при переходе между итерациями ССП. Это достигается путем трансформации задачи в линейную комбинацию 3-центровых интегралов с использованием вспомогательных функций плотности, а затем – в комбинацию 2-центровых интегралов. Изменения вклада немultipольных короткодействующих составляющих этих 2-центровых интегралов эффективно вычисляются через модификацию предварительно сформированных сплайнов от межцентровых расстояний. Изменения остаточных далекодействующих multipольных вкладов рассчитываются для гигантских молекулярных структур по методологии быстрого multipольного метода (ФММ), предусматривающего сегментацию обширного пространства на области и подобласти – подход, изначально разработанный для моделирования галактической динамики.

На каждой итерации ССП происходит максимальная оптимизация вычислений благодаря предварительному отбору значимых (существенно ненулевых) комбинаций интегралов, особенно с учетом прогрессивного уменьшения приращения матрицы плотности при приближении к каждой последующей итерации сходящегося процесса ССП. Учитываются специфические особенности конкретной макромолекулы или их обширного множества, например, тысяч докинг-комплексов одного крупного белка с тысячами различных малых органических молекул-лигандов.

Расчёт всех бицентрических элементов (в том числе аппроксимация перекрытий двухцентровых произведений базисных функций через совокупность одноцентровых вспомогательных функций плотности) осуществляется динамически с применением специфической базы данных, содержащей сплайновые интерполяции межатомных дистанций. При возникновении потребности во внедрении новых базисных наборов происходит мгновенное расширение информационного хранилища путём декомпозиции вводимого базиса на универсальные экспоненциальные составляющие и соответствующую референсную структуру.

Ключевые слова: быстросействующие квантовохимические методы, очень большие молекулы, DFT, итерации ССП, приращение кулоновского вклада в одноэлектронный гамильтониан, тысячи докинг-комплексов

Для цитирования: Аникин Н.А. Комплекс возможных аппроксимативных методов эффективного многократного пересчета вклада кулоновских интегралов в элементы одноэлектронного гамильтониана на итерациях ССП для кардинального ускорения крайне ресурсоемких расчетов DFT гигантских биомолекул // Chemical Bulletin. 2025. Том 8. № 3. 3. DOI: 10.58224/2619-0575-2025-8-3-3

Поступила в редакцию: 7 июня 2025 г.; Одобрена после рецензирования: 13 августа 2025 г.; Принята к публикации: 16 сентября 2025 г.

A set of possible approximative methods for efficiently recalculating the contribution of coulomb integrals to the elements of the single-electron hamiltonian at SCF iterations to dramatically speed up extremely resource-intensive DFT calculations of giant biomolecules

¹ Anikin N.A. *,
¹ Zelinskii Institute of Organic Chemistry, Russian Academy of Sciences,
* Corresponding author E-mail: nikan53@ioc.ac.ru

Abstract: the investigation aims to identify potential approximative methodologies for expediting repeated calculations of Coulomb integral contributions to single-electron Hamiltonian elements during self-consistent field (SCF) iterations, thereby dramatically accelerating computationally intensive density functional theory (DFT) analyses of massive biomolecular structures. The research addressed several challenges: a) evaluating semi-empirical approaches for quantum chemical examination of enormous molecular systems; b) exploring how numerous distant molecular fragments could facilitate faster computation of Coulomb interaction contributions; c) examining contemporary approaches to fixed-geometry single-point molecular calculations; d) developing innovative methodologies for accelerated Coulomb integral contribution computation in DFT analyses of substantial biomolecular entities.

We present a novel suite of approximation techniques designed to substantially expedite calculations of Coulomb integral contributions to one-electron Hamiltonian elements in conventional DFT methodologies during SCF iterations-typically the rate-limiting phase of these essential yet computationally demanding calculations for extensive biomolecular systems, including thousands of docking complexes comprising thousands of atoms.

Our integrated approach features rapid and precise approximation of contribution modifications across innumerable 4-center Coulomb integrals between successive SCF iterations through auxiliary density function-mediated transformation into linear combinations of 3-center integrals, subsequently converted to combinations of 2-center integrals. Contribution variations from non-multipole short-range components of these 2-center integrals are swiftly determined by modifying pre-computed spline contributions based on inter-atomic separations. The remaining multipole-based long-range contributions undergo rapid computation for expansive molecular systems using a fast multipole method (FMM) framework, which strategically partitions extensive spatial domains into hierarchical regions (a technique originally pioneered for galactic dynamics simulations).

Each SCF iteration employs sophisticated screening to identify exclusively non-negligible integral combinations, particularly accounting for the progressively diminishing density matrix increments characteristic of converging SCF processes. The framework accommodates the unique characteristics of specific massive molecular systems or extensive collections thereof, such as thousands of docking arrangements between substantial protein structures and diverse small organic ligand molecules.

All bimolecular components-including approximations of two-center basis function overlaps via linear combinations of single-center auxiliary density functions-undergo efficient computation utilizing specialized database-stored inter-nuclear distance splines. For novel basis sets, the reference database can be promptly augmented through decomposition into universal exponential components with corresponding database enrichment.

Keywords: fast quantum chemical methods, very large molecules, DFT, SCF iterations, Coulomb contribution to the one-electron Hamiltonian, thousands of docking complexes

For citation: Anikin N.A. A set of possible approximative methods for efficiently recalculating the contribution of coulomb integrals to the elements of the single-electron hamiltonian at SCF iterations to dramatically speed up extremely resource-intensive DFT calculations of giant biomolecules. Chemical Bulletin. 2025. 8 (3). 3. DOI: 10.58224/2619-0575-2025-8-3-3

The article was submitted: June 7, 2025; Approved after reviewing: August 13, 2025; Accepted for publication: September 16, 2025.

Введение

Квантовохимические уравнения и их решения открывают широкие возможности для детального анализа разнообразных характеристик органических и биоорганических молекулярных структур. Однако реализация соответствующих вычислительных алгоритмов требует слишком значительных вычислительных мощностей при работе с крупными молекулярными системами, особенно с макромолекулярными биологическими комплексами. Ситуация существенно усложняется при необходимости проведения масштабных серий расчетов тысяч докинг-комплексов белок-лиганд, каждый из которых включает тысячи атомов – задача, имеющая фундаментальное значение для разработки фармацевтических препаратов нового поколения. В данном контексте критическую роль играет применение суперкомпьютерных систем (согласно источнику [1]). Параллельно с этим, исключительную актуальность приобретает разработка специализированных методологических подходов, направленных на ускорение квантовохимических вычислений для молекулярных гигантов и оптимизацию ресурсозатратных этапов, в частности, расчета энергоемкого кулоновского вклада в методе теории функционала плотности (DFT) [2-9].

В актуальной области компьютерных расчётов повсеместно применяются специфические функциональные выражения, демонстрирующие экспоненциальное убывание при возрастании дистанции от нуклеарных центров – именуемые атомно-орбитальными конструкциями (АО) или элементами базисного набора (БН), которые в математическом формализме выражаются произведением радиального и углового компонентов внутри сферической координатной системы:

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

Радиальный множитель элементов базисного пространства, характеризующийся экспоненциальным затуханием при увеличении дистанции от нуклеарных центров, в контексте неэмпирических вычислительных методологий (включая повсеместно применяемую теорию функционала плотности) преимущественно формулируется посред-

ством суперпозиции гауссианов различной протяжённости с соответствующими весовыми коэффициентами:

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

Использование гауссовых функций сильно упрощает и ускоряет ресурсоемкий расчет очень большого числа 4-центровых кулоновских интегралов.

Материалы и методы исследований

Вычислительная трудоемкость расчетов крупных и сверхкрупных молекулярных систем (включая гигантские структуры, насчитывающие сотни и тысячи атомов с числом базисных функций N , превышающим 10 тысяч) стремительно возрастает с увеличением их размерности. Особенно это проявляется при неэмпирических вычислениях, таких как широко применяемый метод теории функционала плотности (DFT). Критическим фактором, лимитирующим эффективность вычислений, становится необходимость многократного (на каждой итерации самосогласованного поля – ССП) расчета вклада кулоновских интегралов (в особенности многочисленных 4-центровых) в элементы одноэлектронного гамильтониана (фокиана), что обусловлено их дальнедействующей природой. Последующие этапы вычислений на каждой итерации характеризуются менее резким ростом времени с увеличением молекулярного размера и не являются критически лимитирующими для чрезвычайно крупных и гигантских молекулярных структур.

Для квантовохимического анализа громадных молекул существует возможность применения полуэмпирических методологий, обеспечивающих значительно более высокое быстродействие, однако уступающих в точности распространенному неэмпирическому методу DFT.

В связи с этим, исключительную актуальность приобретает задача принципиального ускорения именно для громадных молекулярных структур более точных, чем полуэмпирические, распространенных методов класса DFT, и в особенности их лимитирующей для гигантских молекул стадии – расчета на каждой итерации ССП кулоновского

вклада в элементы одноэлектронного гамильтониана (фокиана). Для оптимизации расчёта вклада электростатических взаимодействий между многочисленными дистанционно разнесёнными структурными компонентами макромолекулярных систем исключительно перспективным подходом (наряду с альтернативными методологиями) представляется имплементация алгоритма быстрых мультиполей FMM [10].

Указанные полуэмпирические методологии (требующие вычисления исключительно бицентрических интегралов в противоположность тетрацентрическим, необходимым для теории функционала плотности) при применении к крупномасштабным молекулярным системам демонстрируют приблизительно тысячекратное превосходство в вычислительной эффективности (при определённых компромиссах в точности и надёжности результатов) относительно широко распространённой методологии DFT [11].

Как следствие, полуэмпирические вычислительные схемы ранее успешно применялись для экстремально ресурсоёмких квантово-химических исследований макромолекулярных биологических систем, в частности, для параллельных вычислений структуры, включающей 10^5 атомов, с использованием 1024 процессорных ядер [12], а также для квантово-химического моделирования интегральных докинг-комплексов на высокопроизводительной вычислительной системе «Ломоносов» [13].

Нашим исследовательским коллективом разработан инновационный алгоритм LocalSCF [14], характеризующийся рациональным применением специфически сконструированных и оптимизируемых локализованных молекулярных орбиталей для сверхбыстрых полуэмпирических вычислений макромолекулярных биосистем. Примечательно, что реальная трёхмерная белковая структура, насчитывающая 120 тысяч атомов, или линейная неупорядоченная протеиновая молекула, содержащая свыше миллиона (!) атомов, у нас поддаётся полному квантово-химическому анализу в течение нескольких суток на стандартной персональной вычислительной системе (!). На основе разработанного метода LocalSCF сформирован комплексный подход для высокоскоростных массовых полуэмпирических расчётов тысяч многоатомных докинг-комплексов [15]. В парадигме интеграции вычислительной эффективности полуэмпирических методов с прецизионностью теории функционала плотности нами первоначально предложен [16] алгоритм оперативного моделирования одноэлектронного гамильтониана DFT в полуэмпирическом формате.

Вытянутые фибриллярные белковые структуры обрабатываются специализированными программными комплексами для крупных молекул значительно эффективнее по сравнению с обычными глобулярными белками благодаря возможности более агрессивного отсечения взаимодействий между удалёнными фрагментами молекулы. Так, расчет методом DFT для белка, насчитывающего приблизительно 14000 атомов, с использованием специализированного для крупных молекулярных систем программного комплекса ONETEP потребовал 7 часов [17] на суперкомпьютере Iridis 4, в то время как в исследовании [18] DFT-вычисления белковых структур на суперкомпьютере Cray XC30 проводились с применением альтернативных базисных функций, представленных в виде пространственно развернутых рядов Фурье.

Проанализированные до этого момента подходы относились к одноточечным расчетам молекул с фиксированной геометрией. Практически значимые вычисления с оптимизацией молекулярной геометрии требуют существенно больших вычислительных ресурсов, что особенно критично для гигантских молекулярных структур (например, белков) с чрезвычайно большим числом степеней свободы геометрических вариаций. Показательно, что оптимизация геометрии всего лишь фрагмента белка, включающего 600 атомов ($N = 10\,000$ базисных функций), методом DFT на суперкомпьютере Stampede с эффективным распараллеливанием на 256 ядер посредством специализированной для крупных молекул программы Jaguar потребовала более 70 минут [19].

Для минимизации вычислительных затрат часто имплементируется стратегия, при которой высокоточные (но ресурсоёмкие) квантово-химические (QM) вычисления реализуются исключительно для локализованного фрагмента крупномасштабной молекулярной системы, в то время как доминирующая часть структуры анализируется значительно более производительным (хотя и менее прецизионным) методом молекулярной механики (ММ), что в совокупности конституирует гибридную методологию QM/ММ [20–22]. На интерфейсе между квантово-химической и молекулярно-механической доменами целесообразна интеграция фрагментированных молекулярных орбиталей FMO [22].

Для акселерации экстремально ресурсоёмких вычислений для макромолекулярных ансамблей применяются программные комплексы со специализированными алгоритмическими решениями, в том числе оптимизированные для высокопроизводительных вычислительных систем, в частности, DFT-вычисления посредством программных

платформ NTChem [23] или QChem [24, 25], а также других специализированных вычислительных инструментов [26, 27].

Исключительную актуальность также представляют массовые расчеты докинг-комплексов белок-лиганд, включающие тысячи разнообразных небольших органических молекул-лигандов (в том числе потенциальных прототипов лекарственных препаратов) в активном центре одного крупного белка. В данном контексте для квантовохимических вычислений особенно необходимы специализированные методы и алгоритмы, хотя некоторые исследователи рассматривали подобные квантовохимические расчеты как чрезмерно ресурсоемкие [20].

Ранее нами был предложен комплекс методов ускоренного вычисления вклада кулоновских интегралов для DFT-расчетов актуальных гигантских биомолекул [28], однако вопрос экономичного пересчета кулоновского вклада при переходе к последующим итерациям многочисленного цикла самосогласованного поля не рассматривался и требует отдельного детального анализа.

Результаты и обсуждения

На многочисленных итерациях самосогласованного поля значительно экономичнее пересчитывать все уменьшающуюся поправку (при возрастающей сходимости итераций) к предыдущей итерации, нежели производить полный перерасчет. Это приобретает особую значимость для гигантских молекул, где неоднородность различных участков молекулярной структуры на каждой итерации позволяет максимально эффективно отбрасывать значительное число практически нулевых поправок-приращений для многих областей молекулы, избирательно концентрируясь на участках с существенными изменениями относительно предыдущей итерации.

Дополнительный фактор, способствующий рационализации критически ресурсоемких вычислений – архитектурная организация макромолекулярных биосистем из множественных гомотипических структурных элементов (аминокислотных остатков в протеиновых структурах и дискретных низкомолекулярных лигандов в интермолекулярных докинг-комплексах белок-лиганд). Особую практическую значимость данный аспект приобретает при квантово-химическом моделировании тысяч докинг-комплексов с гетерогенными низкомолекулярными органическими лигандами в активном центре единичной крупномасштабной (включающей тысячи атомов) белковой структуры: электростатическое поле для масштабных, максимально дистанцированных от активного центра и лигандов доменов белковой молекулы

подвергается минимальным модификациям в процессе итераций, а инкремент вклада кулоновских взаимодействий практически нивелируется для лиганда и его непосредственного координационного окружения. Для умеренно удаленных участков может наблюдаться быстрая сходимость на итерациях самосогласованного поля, и соответствующие приращения также стремительно приближаются к нулю. Более того, для удаленных фрагментов белка влияние различных лигандов, вероятно, проявляется однотипно (например, пропорционально дипольному моменту лиганда, в то время как остальные кулоновские вклады быстрее затухают по мере удаления от белковых областей). Все эти аспекты будут детально рассмотрены в последующих разделах, также см. использования метода быстрых мультиполей FMM [10].

Колоссальное повышение производительности при вычислении модификаций кулоновской составляющей в компонентах одноэлектронного гамильтониана в процессе последовательных итераций самосогласованного поля для масштабных молекулярных структур достигается посредством последовательной трансформации комплексных четырёхцентровых интегралов через каскадную редукционную схему – первоначально к трёхцентровым, а впоследствии к двухцентровым интегральным выражениям, что кардинально оптимизирует вычислительную процедуру.

При чрезвычайно ресурсоемких неэмпирических/DFT вычислениях макромолекулярных систем лимитирующим фактором становится процедура определения изменений вклада в матричные элементы одноэлектронного гамильтониана на итерациях самосогласованного поля от двухэлектронных четырёхцентровых кулоновских интегралов базисных функций ϕ_i (количество которых возрастает пропорционально $O(N^4)$). Одна из эффективных стратегий ускорения в данном контексте – использование аппроксимации двухцентровых перекрытий базисных функций $f_n(r) = \phi_i(r)\phi_j(r)$ через линейную суперпозицию вспомогательных функций плотности (ВФП) χ_p (также именуемых Auxiliary basis [29]) - специализированных одноцентровых гауссовых функций $\chi_p(r)$, позиционированных (аналогично базисным функциям) на атомных ядрах:

$$\phi_i(\vec{r}_1)\phi_j(\vec{r}_2) = f_{ij}(\vec{r}_{12}) \approx g_n(\vec{r}_{12}) = \sum_p B_{np} \chi_p(\vec{r}_{12}) \quad (3)$$

При этом 4-центровые кулоновские интегралы (см. ф-лу (4), их число растёт быстро, как $O(N^4)$) сводятся к комбинации 3-центровым, число, которых растёт уже медленнее, как $O(N^3)$, это существенно ускоряет расчет кулоновского вклада K лимитирующей части DFT [29] больших молекул.

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

Здесь в конце приведено укороченное обозначение двухэлектронных интегралов (таблица, численно иллюстрирующая ускорение расчета DFT больших молекул за счет ВФП - см. [28]).

Для повышения точности расчета 4-центровых кулоновских интегралов при аппроксимации ЛК

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

В составе $(f_n | g_m)$ 3-центровые интегралы, а g_m – ЛК 1-центровых ВФП χ_p , а f_n – 2-центровые (3). Хотя для хорошей точности аппроксимации ВФП χ_p требуется в 2-3 раза больше, чем БФ ϕ_i , но для очень больших молекул число ВФП χ_p много меньше $N^2/2$ числа произведений пар БФ $\phi_i(r)\phi_j$.

Все наши аппроксимации направлены на уменьшение скорости роста времени счета и инте-

ВФП (3) часто используют комбинацию 3-центровых интегралов $(f_n | g_m)$ с 2-центровыми $(g_n | g_m)$, для сведения погрешности аппроксимации к малым второго порядка по отклонению f от g :

гралов в гигантских молекулах (даже при введении умеренного объема дополнительных расчетов ради эффективных аппроксимаций, сочетающих быстроедействие с хорошей точностью аппроксимаций). Здесь важна скорость (порядок) роста числа интегралов разного вида с ростом размера молекулы и числа БФ N – см. табл. 1.

Таблица 1

Масштабируемость интегралов с ростом размера молекулы.

Table 1

Scalability of integrals with increasing molecular size.

4-центровые интегралы $(f_n f_m)$	N^4
3-центровые интегралы $(f_n g_m)$	N^3
2-центровые интегралы $(g_n g_m)$ (вместо $(f_n g_m)$)	$\sim N^2$
при введении дополнительных аппроксимаций и FMM	$\sim N^1$

Для численной иллюстрации размеров различных массивов по для генерации кулоновских интегралов – см. табл. 2 – представлена примерная оценка для типичного небольшого протеина из ~ 1700 атомов (см. рис. 1; он меньше многих других

протеинов) – комплекса человеческого иммуноглобулина с иммунным супрессором IMMUNOPHILIN FKBP-12 из Protein Data Bank [30].

Таблица 2

Примерная оценка размеров различных массивов для относительно небольшого протеина (примерно 1700 атомов).

Table 2

Approximate estimates of the sizes of various arrays for a relatively small protein (approximately 1700 atoms).

БФ для распространенного базиса 6-31G*	$\sim 14\,500$
ВФП $\chi_p(r)$	$\sim 31\,000$
Существенно ненулевых 4-центровых кулоновских интегралов $(f_n f_m)$	$\sim 3.5 \times 10^{14}$
Существенно ненулевых 3-центровых интегралов $(f_n \chi_p)$	$\sim 4 \times 10^{11}$

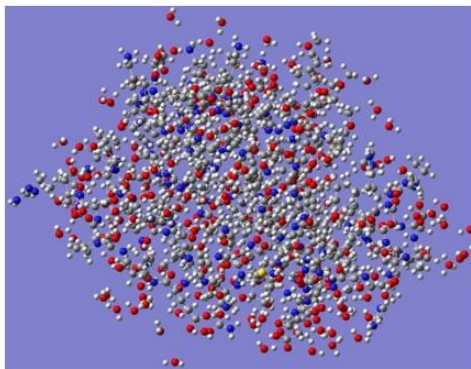


Рис. 1. Типичный небольшой протеин IMMUNOPHILIN FKBP-12, 1700 атомов.
Fig. 1. Typical small protein IMMUNOPHILIN FKBP-12, 1700 atoms.

Кулоновский вклад в элементы одноэлектронного гамильтониана равен интегралу с кулоновским потенциалом $U(r)$:

$$\int \phi_i(\vec{r}) \phi_j(\vec{r}) U(\vec{r}) dV = \langle f_{ij} | U(\vec{r}) \rangle, \quad (6)$$

а изменение кулоновского вклада в данный элемент одноэлектронного гамильтониана равен интегралу с изменением кулоновского потенциала $dU(r)$:

$$\int \phi_i(\vec{r}) \phi_j(\vec{r}) dU(\vec{r}) dV = \langle f_{ij} | dU(\vec{r}) \rangle \quad (7)$$

где: $dU(r)$ выражается через ЛК изменения элементов Матрицы Плотности (МП) dp_{kl} :

$$dU(\vec{r}) = \sum_{kl} dp_{kl} U_{kl}(\vec{r}) \quad (8)$$

где: $U_{kl}(r)$ – кулоновский потенциал, создаваемый перекрыванием БФ $f_n(r) = \phi_k(r)\phi_l(r)$.

В данном контексте особую результативность демонстрирует прирост производительности при работе с масштабными молекулярными структурами наподобие докинг-комплексов протеин-лиганд, включающих тысячи атомов. Ключевой фактор ускорения заключается в том, что после первых итераций в значительных сегментах макромолекулы формируются зоны со стабилизирующим кулоновским потенциалом $U(r)$ и прогрессивно уменьшающимся дифференциалом dU относительно предшествующей итерации. Это существенно расширяет массив пренебрежимо малых интегралов с данным потенциалом – феномен, недоступный для стандартных молекул с десятками атомов из-за отсутствия достаточного пространства для формирования разнообразных областей с минимальным $dU(r)$.

Для реализации максимальной вычислительной эффективности требуется разработка дополнительных специализированных алгоритмических решений, оптимально использующих уникальные характеристики макромолекулярных систем. Это представляет собой значительно более комплексную задачу по сравнению с алгоритмической оп-

тимизацией для компактных молекулярных структур.

Дальнейшее будет в зависимости от способа генерации ЛК разложения $f_n(r_1) \approx g_n(r_1)$ по ВФП $\chi_p(r)$:

а) Если эти ЛК разложения $g_n(r_1)$ по ВФП $\chi_p(r_1)$ получаем минимизацией функционала $(\rho' - \rho | \rho' - \rho)$ – кулоновского отличия нашей аппроксимации МП ρ' (разложения МП по $g_n(r_1)$ и $\chi_p(r_1)$) от исходной МП ρ (разложения МП по $f_n(r_1) = \phi_i(r_1)\phi_j(r_1)$), то $\langle f_n - g_n | U \rangle = 0$, и тогда

$$\langle \phi_i(r)\phi_j(r) | dU(r) \rangle = \langle f_n | dU(r) \rangle \approx \langle \phi_i(r)\phi_j(r) | dU''(r) \rangle = \langle f_n | dU''(r) \rangle,$$

где: $dU''(r)$ получаем на основе ЛК экономичных ВФП (а не намного более громоздкого массива $f_n = \phi_i(r)\phi_j(r)$).

Данный подход упрощает текущий этап вычислений, однако требует значительных ресурсов на предшествующей стадии формирования разложения Локализованных Компонент $dU(r)$ через потенциалы вспомогательных функций подгонки $\chi_p(r)$. Для масштабных молекулярных систем решение соответствующих систем линейных уравнений с разреженными матрицами оптимизируется применением специальных математических методов, в частности метода сопряженных градиентов [31, 32], что позволяет существенно экономить вычислительные ресурсы.

б) В альтернативном сценарии, когда Локализованные Компоненты генерируются мгновенно (и одновременно перед запуском самосогласованного поля) с использованием сплайн-технологий [33], оптимизированных через минимизацию локального двухцентрового функционала (в противоположность учету всей совокупности центров молекулы, как предлагалось в пункте а), необходимо интегрировать дополнительные корректирующие члены для повышения точности расчета матричных элементов $\langle f_n | dU(r) \rangle$ согласно математическому выражению, представленному в

формуле (5):

$$\begin{aligned} \langle f_n | dU(r) \rangle &\approx \langle f_n | dU''(r) \rangle \\ + \langle g_n | dU(r) \rangle &- \langle g_n | dU''(r) \rangle \end{aligned}$$

Для дальнейшего ускорения расчетов мы предлагаем более точную аппроксимацию двухцентровых функций $f_n(r_1) = \phi_i(r_1)\phi_j(r_1)$ в виде линейной комбинации Дополненными межцентровыми ВФП (ДВФП) $\chi'_p(r_1)$:

$$f_n(\vec{r}_1) \approx \sum \mathbf{B}'_{in} \chi'_p(\vec{r}_1) \quad (9)$$

Поскольку в данном случае точность аппроксимации достигается в первом порядке малости относительно разностного выражения $(f_n - g'_n)$, а не во втором порядке малости, как представлено в формуле (5), возникает необходимость в более прецизионной аппроксимационной схеме с расширенным комплексом ДВФП, превосходящим по объёму стандартный набор ВФП за счёт интеграции дополнительных функциональных компонентов МЦВФП (межцентровых вспомогательных функций плотности).

Предлагаемая методология полностью элиминирует необходимость вычисления трёхцентровых интегралов $(f_n|\chi_p)$, редуцируя всю вычислительную процедуру к оперативно рассчитываемым двухцентровым интегралам с привлечением предварительно сгенерированных сплайновых аппроксимаций, параметризованных по межъядерным расстояниям.

Произведение базисных функций $f_n(r) = \phi_i(r)\phi_j(r)$, локализованных на различных центрах, характеризуется значительной амплитудой погрешности при переходе к g'_n именно в промежуточной пространственной области между этими центрами (хотя количественно эта величина относительно невелика, но достаточна для существенного вклада в общую точность). Тщательно оптимизированный ансамбль вспомогательных аппроксимирующих функций (превышающий по кардинальности набор базисных функций в 2-3 раза) обеспечивает высокоточную аппроксимацию непосредственно в околоядерных областях этих двух центров. Принципиально важно учитывать, что амплитуда межцентровых произведений базисных функций экспоненциально затухает с увеличением межъядерной сепарации (достигая пренебрежимо малых значений на значительных дистанциях), следовательно, достаточно реализовать рационально-приближённую экономичную аппроксимацию в пространственной области, приблизительно эквидистантной относительно рассматриваемых центров.

Для повышения аппроксимационной точности рекомендуется экспансия комплекса вспомога-

тельных аппроксимирующих функций до расширенного набора дополненных вспомогательных функций подгонки $\chi'_p(r)$, интегрирующего дополнительный компактный ансамбль межцентровых вспомогательных функций плотности (МЦВФП, функционально эквивалентных стандартным ВФП), дислоцированных в интермедиальной области между центрами базисных функций ϕ_i и ϕ_j , где манифестируется максимальная погрешность аппроксимации (9) при использовании исключительно конвенциональных ВФП, локализованных на центрах ϕ_i и ϕ_j .

Предварительная аналитическая оценка демонстрирует, что достаточно инкорпорировать ограниченное количество МЦВФП между центрами базисных функций ϕ_i и ϕ_j , располагая их приблизительно в медианной позиции и на дистанциях, соответствующих 1/4 межцентрового расстояния от каждого ядра, с экспоненциальными параметрами, демонстрирующими закономерное уменьшение по мере возрастания межъядерной сепарации между центрами базисных функций ϕ_i и ϕ_j [28].

Имплементация межцентровых вспомогательных функций плотности (МЦВФП) обеспечивает трансформацию трёхцентровых кулоновских интегралов $(f_n|g_m)$ в линейные суперпозиции упрощённых двухцентровых кулоновских интегралов $(\chi'_p|\chi_q)$, интегрированных в структуру $(g'_n|g_m)$. Несмотря на значительную многочисленность МЦВФП (которые, однако, нерелевантны для существенно дистанцированных центров с пренебрежимо малыми двухцентровыми произведениями базисных функций), существуют дополнительные стратегии кардинального ускорения вычислений кулоновских интегралов.

В двухцентровых кулоновских интегралах $(\chi'_p|\chi_q)$ для каждой функциональной единицы ВФП χ'_p/χ_q в зависимости от её угловой компоненты идентифицируются немногочисленные мультипольные составляющие (монополь-заряд для s-функций, диполь для (px, py, pz)-функций, квадруполь для d-функций и т.д.). Это позволяет дифференцировать дальнедействующую часть подобных интегралов на элементарные взаимодействия ограниченного числа мультиполей (инвариантные относительно специфики радиальных компонентов функций на данном центре), характеризующиеся неэкспоненциальным затуханием ВФП χ_q с различной угловой частью (s, p, d, f, g) – см. табл. 3, и быстро экспоненциально затухающую немultipольную составляющую, явно детерминированную спецификой функций χ'_p и χ_q .

Таблица 3

Скорость убывания мультипольных взаимодействий в зависимости от типа ВФП χ_q (s, p, d, f, g) на двух центрах с ростом межцентрового расстояния R.

Table 3

The rate of decrease of multipole interactions depending on the type of the VFP χ_q (s, p, d, f, g) at two centers with increasing intercenter distance R.

s - s	$\sim 1/R^1$
s - p	$\sim 1/R^2$
s - d	$\sim 1/R^3$
s - f	$\sim 1/R^4$
s - g	$\sim 1/R^5$
p - p	$\sim 1/R^3$
p - d	$\sim 1/R^4$
p - f	$\sim 1/R^5$
p - g	$\sim 1/R^6$
d - d	$\sim 1/R^5$
d - f	$\sim 1/R^6$
d - g	$\sim 1/R^7$
f - f	$\sim 1/R^7$
f - g	$\sim 1/R^8$
g - g	$\sim 1/R^9$

Ключевой аспект оптимизационной стратегии – агрегирование приращений вкладов от всех МЦВФП $\chi'_p(r)$ на конкретном центре в процессе каждой итерации самосогласованного поля (демонстрирующих затухание с возрастанием итерационного индекса), что фундаментально интенсифицирует расчёт кулоновского вклада для масштабных молекулярных ансамблей. В контексте макромолекулярных систем мультипольные взаимодействия структурируются согласно методологии быстрых мультиполей (FMM) [10] посредством аккумуляции мультиполей идентичного типа в конкретной пространственной локации крупной молекулы с последующей консолидацией (включая многоуровневую) этих локаций и ассоциированных мультиполей. Такая агрегация инкрементов мультиполей демонстрирует исключительную эффективность вследствие увеличенной численности мультиполей от расширенного ансамбля МЦВФП, причём учитываются исключительно значимые, нетривиальные модификации мультиполей при транзите к последующей итерации ССП в специфической области макромолекулярной структуры.

Инкременты вкладов немultipольных компонентов ограниченного числа экспоненциально затухающих короткодействующих двухцентровых кулоновских интегралов репрезентируются посредством предварительно сгенерированных сплайновых аппроксимаций (сегментированных полиномиальных функций межъядерных дистанций) [33] из заблаговременно скомпилированной информационной базы. Существенное преимуще-

ство данного подхода заключается в том, что любая линейная комбинация нескольких сплайновых функций трансформируется в единый консолидированный сплайн. Это обеспечивает возможность формулировки ВФП и МЦВФП через линейные суперпозиции произвольного количества гауссианов без инкрементации вычислительных затрат, поскольку для двухцентровых интегралов достаточно предварительно сконструировать композитный сплайн из линейных комбинаций элементарных сплайнов с дискретными гауссовыми функциями, а для мультипольной составляющей cardinalность гауссовых компонентов вообще не влияет на производительность вычислительного процесса.

Учёт угловой анизотропии индивидуальных ВФП/МЦВФП и базисных функций (категорий r_x , r_y , dx_z и т.д.) в двухцентровых интегралах редуцируется к линейным суперпозициям таких компонентов (r_x' , r_y' , dx_z' и т.д.) параллельно и перпендикулярно интерцентровому вектору, причём интегралы от последних детерминируются исключительно межцентровым расстоянием. По мере конвергенции самосогласованной процедуры достигается значительная и прогрессивно возрастающая экономия вычислительных ресурсов благодаря редукции инкрементов при переходе к последующим итерациям.

Дополнительная стратегия акселерации вычислений кулоновского вклада в одноэлектронный гамильтониан (фокиан) для макромолекулярных ансамблей базируется на представлении аппроксимации инкремента потенциала $dU(r)$ на текущей

итерации самосогласованного поля через дихотомическую декомпозицию: медленно затухающую, но аналитически редуцированную и легко агрегируемую мультипольную составляющую, и более комплексную, но экспоненциально затухающую немультимольную компоненту приращения потенциала. Немультимольная составляющая формализуется через сплайновые функции как функциональная зависимость от дистанции до центра генерации потенциала.

Как для мультипольной, так и для немультимольной компоненты имплементируется консолидация вкладов от единичного центра (с дифференцированной обработкой различных мультиполей и угловых сферических немультимольных составляющих), что фундаментально оптимизирует аппроксимацию инкремента потенциала. Это особенно эффективно для атомов в расширенных базисных наборах, поскольку реализуется принцип, согласно которому линейная суперпозиция произвольного количества сплайнов эквивалентна единичному новому сплайну (зависящему от итерационного индекса, поскольку он интегрирует приращение матрицы плотности на текущей итерации).

Для дальнейшей интенсификации вычислительной эффективности предварительно элиминируются те мультипольные и немультимольные вклады, которые достигли пренебрежимо малых величин для инкрементов на текущей итерации самосогласованного поля в анализируемой области макромолекулярной структуры.

Все значимые линейные комбинации аппроксимации $f_n(r) \approx g'_n(r)$ детерминируются однократно перед инициализацией самосогласованного процесса с применением сплайновых функций от дистанций (или их квадратичных выражений) между центрами базисных функций ϕ_i и ϕ_j .

Для оптимизации и акселерации процедуры конструирования сплайнов для широкого спектра базисных наборов с гетерогенными экспоненциальными параметрами применяется высокоточная аппроксимация через линейные комбинации сплайнов для универсальных фиксированных экспонент. Этот методологический подход обеспечивает универсальность алгоритма и его аппликабельность к разнообразным молекулярным системам без существенной деградации вычислительной эффективности [28].

Поскольку в практической реализации нет необходимости в предварительном выполнении громоздких (и ресурсоёмких) вычислений дискретных 4- и 3-центровых кулоновских интегралов, возможен непосредственный переход к определению вкладов 2-центровых кулоновских инте-

гралов ($\chi'_p|\chi_q$) в структуре ($g'_n|g_m$) и отдельно их мультипольной и немультимольной компонент в инкремент элементов одноэлектронного гамильтониана (фокиана) на очередной итерации ССП.

Изначально необходимо детерминировать разложения двухцентровых функций $f_n(r)$ по ДВФП χ'_p в составе $g'_n(r)$, опираясь на предварительно сгенерированный массив немультимольных компонент 2-центровых кулоновских интегралов ($\chi'_p|\chi_q$) в структуре ($g'_n|g_m$) - см. п. (2.3).

На первом этапе аппроксимируем все нетривиальные 2-центровые функции $f_n(r) = \phi_i(r)\phi_j(r)$ в форме линейной комбинации ДВФП (ВФП на этих двух центрах, дополненными МЦВФП в интермедиальной области между этими центрами) $\chi'_p(r)$.

Поскольку в данном контексте все детерминировано исключительно спецификой этих двух центров и их базисных наборов, наиболее эффективная стратегия - предварительно скомпилировать сплайны (от межцентрового расстояния) для коэффициентов разложения каждой $f_n(r)$ по всем симметрично-совместимым (сигма по сигма, пи по пи, и т.д.) $\chi'_p(r)$, первоначально ориентируясь относительно направления данного межцентрового вектора, а затем трансформируя их в общемолекулярную систему координат посредством ротационных преобразований, аналогично п. (2.3), рассмотренному выше.

Эти сплайновые функции (принципиально отличные от сплайнов для собственно кулоновских интегралов) предварительно интегрируются в соответствующую базу данных, причём возможно оперативное дополнение этой базы для новых базисных наборов через разложение инновационного базиса по стандартизированным (универсальным, предварительно подготовленным) экспонентам, аналогично методологии п. (2.3), описанной выше.

При оптимизации линейно-комбинационного разложения всех функций $f_n(r) = \phi_i(r)\phi_j(r)$ в форме линейных суперпозиций всех ВФП данной молекулярной системы (а не только ВФП данных двух центров $f_n(r)$) мы получаем ВФП-аппроксимацию инкремента на текущей итерации ССП интегральной одноэлектронной матрицы плотности (МП) $d\rho'(r)$:

$$d\rho'(\vec{r}) = \sum_p B_p \chi_p(\vec{r}). \quad (10)$$

На последующих этапах используется относительно компактный массив B_p разложения приращения полной матрицы плотности $d\rho'(r)$ по вспомогательным функциям подгонки χ_p на текущей итерации самосогласованного поля. Важно подчеркнуть, что на данной стадии (получения $d\rho'(r)$) достаточно применения сравнительно компактного набора ВФП без обширного массива МЦВФП,

поскольку возникающая погрешность имеет второй порядок малости по точности аппроксимации, как указано в работах [29, 34].

С каждой последующей итерацией сходящегося процесса ССП прогрессивно увеличивается количество пренебрежимо малых элементов приращения V_r (при одновременном сокращении числа значимых ненулевых элементов), что приводит к повышению разреженности матрицы приращения V_r . Этот процесс адаптивно учитывает структурные особенности конкретной макромолекулы и специфику данного процесса ССП.

При указанной оптимизации и использовании разложения двухцентровых функций $f_n(r) = \phi_i(r)\phi_j(r)$ в виде линейной комбинации дополненных межцентровыми вспомогательными функциями плотности (ДВФП) $\chi'_p(r1)$, вклады двухцентровых кулоновских интегралов в приращение одноэлектронного гамильтониана (фокиана) на очередной итерации ССП сводятся к линейным комбинациям быстро вычисляемых (через сплайны) двухцентровых интегралов ($\chi'_p|\chi_q$). Для оптимизации вычислений с обширным набором ДВФП (вследствие многочисленных МЦВФП) интегралы и их вклады разделяются на мультипольные и немультимпольные составляющие, как указано выше в п. (2.3). На каждой последующей итерации ССП количество существенно ненулевых элементов уменьшается с учетом специфики макромолекулы, состоящей из многочисленных крупных фрагментов, и характеристик текущего процесса ССП.

Приращения мультипольных вкладов на очередной итерации ССП (п. (2.3)), несмотря на их далекодействующий характер, эффективно группируются согласно методологии быстрых мультиполей (FMM) [10] с консолидацией приращений мультиполей в каждой пространственной области макромолекулы и последующим объединением (включая многоуровневое) этих областей и их мультиполей. Прежде всего это относится к общему мультипольному вкладу от всей матрицы плотности и соответствующему мультипольному вкладу зарядов ядер, что формирует далекодействующий электростатический потенциал всей макромолекулы, включая области за её пределами (например, для расчета потенциала для лигандов в полости крупного протеина или при взаимодействии нескольких макромолекулярных белков). Предварительно определяется мультипольный вклад для каждой ДВФП $\chi'_p(r1)$, после чего все мультипольные взаимодействия вычисляются с высокой эффективностью для масштабных молекулярных систем.

По мере сходимости итераций ССП прогрессивно сокращается количество непренебрежимых

элементов приращений мультиполей (особенно актуально высококачественное начальное приближение матрицы плотности для минимизации приращений уже на начальных итерациях).

Приращения немультимпольных вкладов при переходе к следующей итерации ССП, благодаря их короткодействующему характеру, также вычисляются эффективно для макромолекул (с предварительным определением пространственных областей с потенциально значимыми немультимпольными вкладами).

Использование уточненной (с включением МЦВФП) аппроксимации двухцентровых функций $f_n(r) = \phi_i(r)\phi_j(r)$ в виде линейных комбинаций $\chi'_p(r)$ позволяет оперативно вычислять приращение (затухающее по мере сходимости итераций ССП) кулоновского вклада макромолекулы в приращение одноэлектронного гамильтониана.

Высококачественное начальное приближение матрицы плотности $\rho(r)$ не только сокращает количество итераций ССП до достижения сходимости, но и минимизирует необходимое приращение этой матрицы, что ведет к уменьшению числа значимых элементов в аппроксимации приращения кулоновского потенциала и повышению производительности начальных итераций в предлагаемом методе.

Для дальнейшего повышения эффективности возможно использование структурных особенностей конкретной макромолекулы (например, различных протеинов из одготипных аминокислот) или множества сходных молекулярных систем - например, тысяч докинг-комплексов одного крупного протеина с различными относительно небольшими молекулами лигандов.

Учет специфики, включающий наличие качественного начального приближения в виде комбинации крупного протеина и отдельного лиганда (малой биоорганической молекулы), дополнительно повышает эффективность использования только приращений элементов одноэлектронного гамильтониана, поскольку при таком начальном приближении эти приращения минимальны с самого начала процесса ССП. Дополнительная возможность оптимизации заключается в учете поляризации протеина под влиянием дипольного момента лиганда. Также возможно улучшение начального приближения (здесь в рамках неэмпирического метода DFT) на основе формирования специальных групп атомов компактно расположенных лигандов, расположенных в разных частях полости протеина совместно с атомами протеина прилегающими (выделены черным цветом) к атомам данной группы лигандов (рис. 2).

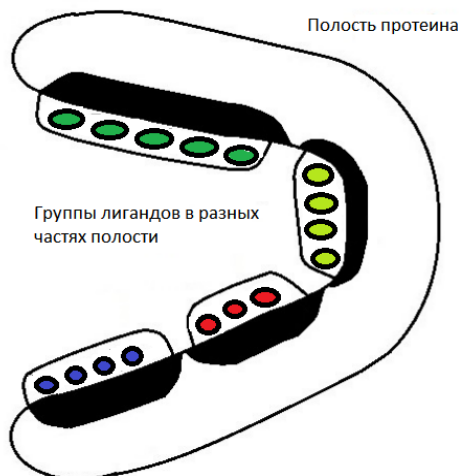


Рис. 2. Использование специальной группировки лигандов докинг-комплексов для дополнительного ускорения крупномасштабных массовых расчетов докинг-комплексов при использовании дополнительных возможностей предлагаемого нами комплекса мер-методов.

Fig. 2. Using a special grouping of docking complex ligands to further accelerate large-scale mass calculations of docking complexes using the additional capabilities of the proposed set of methods.

Ранее мы частично уже использовали такую специфику для кардинального ускорения более простых (чем DFT) полуэмпирических расчетов тысяч докинг-комплексов [15, 35].

Выводы

Использованный комплекс методов обеспечивает кардинальное ускорение расчета приращения вклада кулоновских интегралов в элементы одноэлектронного гамильтониана (фокиана) на очередной итерации самосогласованного поля (затухающего по мере сходимости итераций) - одной из ресурсолимитирующих стадий вычислительно-интенсивных расчетов макромолекулярных систем из сотен и тысяч атомов широко применяемым методом теории функционала плотности (DFT), включая массовые расчеты тысяч докинг-комплексов протеин-лиганд.

Данный методологический комплекс включает следующие компоненты для быстрой и точной аппроксимации приращения кулоновского вклада на каждой итерации ССП:

1. На начальном этапе осуществляется переход от вклада колоссального числа четырехцентровых кулоновских интегралов (см. п. 2.1) к аппроксимации вклада линейной комбинации трехцентровых интегралов.

2. Далее производится аппроксимация трехцентровых кулоновских интегралов через линейные комбинации двухцентровых интегралов с уточненной аппроксимацией двухцентровых перекрытий базисных функций с привлечением минимального числа межцентровых вспомогательных

функций подгонки между рассматриваемыми центрами.

3. Двухцентровые кулоновские интегралы разделяются на малочисленные универсальные легко группируемые дальнедействующие мультипольные компоненты и короткодействующие немультимпульные компоненты.

4. Компоненты приращений всех мультиполей на данном центре суммируются, что обеспечивает существенную экономию при использовании стандартных расширенных базисных наборов БФ и ВФП.

5. Дальнедействующие мультипольные компоненты приращений на текущей итерации ССП группируются в каждой пространственной области макромолекулы с последующим многоуровневым объединением этих областей и их мультиполей по методологии быстрых мультиполей (FMM) [10], с затуханием этих приращений по мере сходимости итераций, что позволяет радикально ускорить расчет кулоновского вклада с учетом структурной специфики каждой молекулярной системы.

6. Немультимпульные короткодействующие компоненты двухцентровых интегралов (эффективно отсеиваемые с увеличением расстояния в макромолекулах) вычисляются с высокой скоростью через предварительно сформированные сплайны от межцентровых расстояний, с затуханием их вклада по мере сходимости итераций.

7. Все двухцентровые компоненты (включая аппроксимацию двухцентровых перекрытий БФ через линейные комбинации одноцентровых ВФП

и ДВФП) быстро вычисляются с использованием сплайнов от межъядерных расстояний из специализированной базы данных.

8. На всех этапах максимально оптимизируются вычисления за счет предварительно отобранных комбинаций интегралов для экономичного расчета приращения элементов одноэлектронного гамильтониана на текущей итерации ССП.

9. Активно используется сходимость процесса ССП с уменьшением приращения элементов одноэлектронного гамильтониана и сокращением используемых элементов промежуточных вспомогательных массивов с прогрессивным отсеиванием пренебрежимо малых элементов при переходе к каждой последующей итерации.

10. Высококачественное начальное приближение матрицы плотности $\rho(r)$ не только сокращает количество итераций до достижения сходимости, но и уменьшает число значимых элементов промежуточных вспомогательных массивов.

11. Метод предусматривает активное использование структурных особенностей конкретной макромолекулы (например, протеинов, состоящих из аминокислот) или массива сходных молекулярных систем - в частности, тысяч докинг-комплексов одного протеина.

12. Для новых базисных наборов БФ и ВФП база данных легко и оперативно дополняется за счет

разложения нового базиса по набору универсальных экспонент.

13. Дополнительные возможности ускорения для макромолекулярных систем обеспечиваются за счет сокращения числа значимых кулоновских интегралов при использовании ортогонализированных локализованных базисных функций [36] - дальнедействующие монополи двухцентровых перекрытий БФ обнуляются, а сами двухцентровые перекрытия БФ быстрее затухают с увеличением межцентровых расстояний, что сокращает количество значимых компонент.

В результате проведенного исследования предложен большой комплекс методов по кардинальному ускорению ССП (особенно формирования элементов одноэлектронного Гамильтониана (фокиана) - лимитирующей стадии крайне ресурсоемких расчетов сверхбольших-гигантских молекул из тысяч атомов распространенным квантовохимическим методом DFT, в т.ч. актуальных биомолекул (в т.ч. массового расчета тысяч докинг-комплексов из тысяч атомов каждый ради лекарств будущего). Более подробное описание отдельных компонент этого большого комплекса методов предполагается в дальнейших публикациях.

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. Zaleśny R., Papadopoulos M.G., Mezey P.G., Leszczynski J. (Eds.). (2011). Linear-Scaling Techniques in Computational Chemistry and Physics: Methods and Applications (Vol. 13). Springer Science & Business Media.
3. Ochsenfeld C., Kussmann J., Lambrecht D.S. Linear-scaling methods in quantum chemistry // *Reviews in computational chemistry*. 2007. № 23. P. 1.
4. Kussmann J., Beer M., Ochsenfeld C. Linear-scaling self-consistent field methods for large molecules // *Wiley Interdisciplinary Reviews: Computational Molecular Science*. 2013. № 3 (6). P. 614 – 636.
5. Vitale Valerio Computational methods for first-principles molecular dynamics with linear-scaling density functional theory. Diss. University of Southampton, 2017.
6. Niklasson, Anders MN. «Density matrix methods in linear scaling electronic structure theory». *Linear-Scaling Techniques in Computational Chemistry and Physics: Methods and Applications*. Dordrecht: Springer Netherlands, 2011. P. 439 – 473.
7. Hu, Wei, Mohan Chen *Advances in density functional theory and beyond for computational chemistry* // *Frontiers in Chemistry*. 2021. № 9. P. 705762.
8. Nakai H., Kobayashi M., Yoshikawa T., Seino J., Ikabata Y., Nishimura Y. Divide-and-conquer linear-scaling quantum chemical computations // *The Journal of Physical Chemistry A*. 2023. № 127 (3). P. 589 – 618.
9. Nakata A., Baker J.S., Mujahed S.Y., Poulton J.T., Arapan S., Lin J., Bowler D.R. Large scale and linear scaling DFT with the CONQUEST code // *The Journal of chemical physics*. 2020. № 152 (16).
10. H. A. Le T. Shiozaki 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>.
11. Li A., Muddana H.S., Gilson M.K. Quantum mechanical calculation of noncovalent interactions: A large-scale evaluation of PMx, DFT, and SAPT approaches // *Journal of Chemical Theory and Computation*. 2014. № 10:4. P. 1563 – 1575. <https://doi.org/10.1021/ct401111c>

12. 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>
13. Oferkin I.V., Katkova E.V., Sulimov A.V. 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>
14. 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 // *Journal of Chemical Physics*. 2004. Vol. 121. № 3. P. 1266 – 1270. <https://doi.org/10.1063/1.1764496>
15. Аникин Н.А., Андреев А.М., Кузьминский М.Б., Мендкович А.С. Быстродействующий метод для массовых полуэмпирических расчетов докинг-комплексов // *Известия Академии наук. Серия химическая*, 2008. № 9. С. 1759 – 1764.
16. Аникин Н.А., Бугаенко В.Л., Кузьминский М.Б., Мендкович А.С. Быстродействующий метод для квантово-химических расчетов больших молекул с аппроксимацией гамильтониана DFT // *Известия Академии наук. Серия химическая*. 2014. № 2. С. 346 – 349.
17. Womack J.C., Mardirossian N., Head-Gordon M., Skylaris Ch.-K. Selfconsistent implementation of meta-GGA functionals for the ONETEP linear-scaling electronic structure package // *Journal of Chemical Physics*. 2016. № 145:20. P. 204114. <https://doi.org/10.1063/1.4967960>
18. Higgins J.E., Probert M.I.J., Hasnip P.J., Refson K., Bush I.J. Hybrid OpenMP and MPI within the CASTEP code, 2015. <http://www.archer.ac.uk/community/eCSE/eCSE01-017/eCSE01-017.php>
19. 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 // *International Journal of Quantum Chemistry*. 2016. № 116:5. P. 357 – 368. <https://doi.org/10.1002/qua.25043>
20. 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 // *International Journal of Molecular Sciences*. 2019. № 20:18. P. 4384. <https://doi.org/10.3390/ijms20184384>
21. 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 – 6163. <https://pubmed.ncbi.nlm.nih.gov/25880693/> doi:10.1021/cr500628b. PMID 2588069,
22. 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>
23. Nakajima T., Katouda M., Kamiya M., Nakatsuka Yu. NTChem: A highperformance software package for quantum molecular simulation // *I. J. Quant. Chem.* 2015. № 115:5. P. 349 – 359. https://www.researchgate.net/publication/270223372_NTChem_A_High-Performance_Software_Package_for_Quantum_Molecular_Simulation
24. Shao Y., Gan Z., Epifanovsky E., Gilbert A.T., Wormit M., Kussmann J., Rassolov V.A. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package // *Molecular Physics*. 2015. № 113 (2). P. 184 – 215.
25. Van Voorhis Troy Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. 2021.
26. de Freitas Rodrigues S.B., de Araújo, R.S.A., de Mendonça T.R.D., Mendonça-Júnior F.J.B., Zhan P., da Silva-Júnior E.F. Quantum chemistry in drug design: density function theory (DFT) and other quantum mechanics (QM)-related approaches. *Applied Computer-Aided Drug Design: Models and Methods*. 2023. № 258.
27. Sarker N. Evaluation of computational chemistry software and density functional theory methods for electronic structure computation of perovskites (Master's thesis, Itä-Suomen yliopisto). 2025.
28. Аникин Н.А. Комплекс возможных аппроксимативных методов эффективного учета вклада кулоновских интегралов для кардинального ускорения расчетов DFT гигантских биомолекул: сведение к быстро-вычисляемым короткодействующим двухцентровым сплайнам плюс FMM дальнего кулона // *Chemical Bulletin*. 2024. Т. 7. № 3. С. 49 – 63. DOI: 10.58224/2619-0575-2024-7-3-49-63
29. Eichkorn K., Treutler O., Ohm H., Haser M., Ahlrichs R. Auxiliary basis sets to approximate Coulomb potentials // *Chemical Physics Letters*. 1995. № 240. P. 283 – 290.
30. 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.

31. Баландин М.Ю., Шурина Э.П. Методы решения СЛАУ большой размерности. Новосибирск: Изд-во НГТУ, 2000. 70 с.
32. Кротова Е.Л., Цылова Е.Г., Осипов, Н.Р., Филиппов М.А. Обзор и исследование области применимости численных методов для решения СЛАУ // Наука и образование: современные тренды. 2015. (X). С. 39 – 56.
33. Вержбицкий В.М. «Численные Методы. Математический Анализ и Обыкновенные Дифференциальные Уравнения». Москва. ОНИКС 21 век. 2005, гл. 4.
34. 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. М., 2000.
35. Аникин Н.А., Андреев А.М., Кузьминский М.Б., Мендкович А.С. Новый подход к ускорению массовых квантово-химических расчетов докинг-комплексов // Известия Академии наук. Серия химическая. 2018. № 6. С. 1100.
36. 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 // Journal of Computational Chemistry. 2003. Vol. 24. № 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. Zaleśny R., Papadopoulos M.G., Mezey P.G., Leszczynski J. (Eds.). (2011). Linear-Scaling Techniques in Computational Chemistry and Physics: Methods and Applications (Vol. 13). Springer Science & Business Media.
3. Ochsenfeld C., Kussmann J., Lambrecht D.S. Linear-scaling methods in quantum chemistry. Reviews in computational chemistry. 2007. No. 23. P. 1.
4. Kussmann J., Beer M., Ochsenfeld C. Linear-scaling self-consistent field methods for large molecules. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2013. No. 3 (6). P. 614 – 636.
5. Vitale Valerio Computational methods for first-principles molecular dynamics with linear-scaling density functional theory. Diss. University of Southampton, 2017.
6. Niklasson, Anders MN. “Density matrix methods in linear scaling electronic structure theory.” Linear-Scaling Techniques in Computational Chemistry and Physics: Methods and Applications. Dordrecht: Springer Netherlands, 2011. P. 439 – 473.
7. Hu, Wei, Mohan Chen Advances in density functional theory and beyond for computational chemistry. Frontiers in Chemistry. 2021. No. 9. P. 705762.
8. Nakai H., Kobayashi M., Yoshikawa T., Seino J., Ikabata Y., Nishimura Y. Divide-and-conquer linear-scaling quantum chemical computations. The Journal of Physical Chemistry A. 2023. No. 127 (3). P. 589 – 618.
9. Nakata A., Baker J.S., Mujahed S.Y., Poulton J.T., Arapan S., Lin J., Bowler D.R. Large scale and linear scaling DFT with the CONQUEST code. The Journal of chemical physics. 2020. No. 152 (16).
10. H. A. Le T. Shiozaki 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>.
11. Li A., Muddana H.S., Gilson M.K. Quantum mechanical calculation of noncovalent interactions: A large-scale evaluation of PMx, DFT, and SAPT approaches. Journal of Chemical Theory and Computation. 2014. No. 10:4. P. 1563 – 1575. <https://doi.org/10.1021/ct401111c>
12. 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>
13. Oferkin I.V., Katkova E.V., Sulimov A.V. 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>
14. 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. Journal of Chemical Physics. 2004. Vol. 121. No. 3. P. 1266 – 1270. <https://doi.org/10.1063/1.1764496>

15. Anikin N.A., Andreev A.M., Kuzminsky M.B., Mendkovich A.S. High-speed method for mass semiempirical calculations of docking complexes. *Bulletin of the Academy of Sciences. Chemical Series*. 2008. No. 9. P. 1759 – 1764.
16. Anikin N.A., Bugaenko V.L., Kuzminskii M.B., Mendkovich A.S. A Fast Method for Quantum Chemical Calculations of Large Molecules with DFT Hamiltonian Approximation. *Bulletin of the Academy of Sciences. Chemical Series*. 2014. No. 2. P. 346 – 349.
17. Womack, J. C., Mardirossian, N., Head-Gordon, M., and Skylaris, Ch.-K. “Self-consistent Implementation of Meta-GGA Functionals for the ONETEP Linear-Scaling Electronic Structure Package”. *Journal of Chemical Physics*. 2016. No. 145:20. P. 204114. <https://doi.org/10.1063/1.4967960>
18. Higgins J.E., Probert M.I.J., Hasnip P.J., Refson K., Bush I.J. Hybrid OpenMP and MPI within the CASTEP code, 2015. <http://www.archer.ac.uk/community/eCSE/eCSE01-017/eCSE01-017.php>
19. 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. *International Journal of Quantum Chemistry*. 2016. No. 116:5. P. 357 – 368. <https://doi.org/10.1002/qua.25043>
20. 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. *International Journal of Molecular Sciences*. 2019. No. 20:18. P. 4384. <https://doi.org/10.3390/ijms20184384>
21. 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 – 6163. <https://pubmed.ncbi.nlm.nih.gov/25880693/> doi:10.1021/cr500628b. PMID 2588069,
22. 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>
23. Nakajima T., Katouda M., Kamiya M., Nakatsuka Yu. NTChem: A highperformance software package-age for quantum molecular simulation. *I.J. Quant. Chem.* 2015. No. 115:5. P. 349 – 359. https://www.researchgate.net/publication/270223372_NTChem_A_High-Performance_Software_Package_for_Quantum_Molecular_Simulation
24. Shao Y., Gan Z., Epifanovsky E., Gilbert A.T., Wormit M., Kussmann J., Rassolov V.A. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Molecular Physics*. 2015. No. 113 (2). P. 184 – 215.
25. Van Voorhis Troy Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. 2021.
26. de Freitas Rodrigues S.B., de Araújo, R.S.A., de Mendonça T.R.D., Mendonça-Júnior F.J.B., Zhan P., da Silva-Júnior E.F. Quantum chemistry in drug design: density function theory (DFT) and other quantum mechanics (QM)-related approaches. *Applied Computer-Aided Drug Design: Models and Methods*. 2023. No. 258.
27. Sarker N. Evaluation of computational chemistry software and density functional theory methods for electronic structure computation of perovskites (Master's thesis, Itä-Suomen yliopisto). 2025.
28. Anikin N.A. A set of possible approximate methods for efficiently taking into account the contribution of Coulomb integrals for dramatic acceleration of DFT calculations of giant biomolecules: reduction to fast-computable short-range two-center splines plus long-range Coulomb FMM. *Chemical Bulletin*. 2024. T. 7. No. 3. P. 49 – 63. DOI: 10.58224/2619-0575-2024-7-3-49-63
29. Eichkorn K., Treutler O., Ohm H., Haser M., Ahlrichs R. Auxiliary basis sets to approximate Coulomb potentials // *Chemical Physics Letters*. 1995. No. 240. P. 283 – 290.
30. 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.
31. Balandin M.Yu., Shurina E.P. Methods for solving large-scale linear equations. Novosibirsk: NSTU Publishing House, 2000. 70 p.
32. Krotova E. L., Tsylova E. G., Osipov N. R., Filippov M. A. Review and study of the applicability area of numerical methods for solving linear equations. *Science and education: modern trends*. 2015. (X). P. 39 – 56.
33. Verzhbitsky V.M. Numerical Methods. Mathematical Analysis and Ordinary Differential Equations. Moscow: ONIX 21st Century. 2005, Ch. 4.
34. Laikov D. N. Development of an economical approach to calculating molecules by the density functional method and its application to solving complex chemical problems: Cand. Phys.-Math. Sciences in Specialty 02.00.17 – Quantum Chemistry. Moscow, 2000.

35. Anikin N.A., Andreev A.M., Kuzminsky M.B., Mendkovich A.S. A New Approach to Accelerating Mass Quantum Chemical Calculations of Docking Complexes. Bulletin of the Academy of Sciences. Chemical Series. 2018. No. 6. P. 1100.

36. 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. Journal of Computational Chemistry. 2003. Vol. 24. 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, nikan53@ioc.ac.ru

© Аникин Н.А., 2025

Information the author

Anikin N.A., PhD (Chemistry), Researcher, N.D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, nikan53@ioc.ac.ru

© Anikin N.A., 2025