

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
им. М.В. ЛОМОНОСОВА

ХИМИЧЕСКИЙ ФАКУЛЬТЕТ

На правах рукописи



ШУЛЬГА ДМИТРИЙ АЛЕКСАНДРОВИЧ

**РАЗРАБОТКА И ПАРАМЕТРИЗАЦИЯ МЕТОДОВ РАСЧЕТА
ТОПОЛОГИЧЕСКИ СИММЕТРИЧНЫХ АТОМНЫХ ЗАРЯДОВ ДЛЯ ЦЕЛЕЙ
МОЛЕКУЛЯРНОГО МОДЕЛИРОВАНИЯ**

02.00.03 – органическая химия

Автореферат диссертации на соискание ученой степени
кандидата химических наук

Москва – 2009

Диссертационная работа выполнена на кафедре органической химии Химического факультета Московского государственного университета им. М.В. Ломоносова

Научный руководитель:

кандидат химических наук,
ведущий научный сотрудник

Владимир Александрович Палюлин

Официальные оппоненты:

доктор химических наук

Александр Владимирович Немухин

доктор биологических наук,
кандидат физико-математических наук

Владимир Васильевич Поройков

Ведущая организация:

Институт органической химии имени Н.Д. Зелинского РАН

Защита состоится « 29 » апреля 2009 года в 11:00 час на заседании диссертационного совета Д 501.001.69 по химическим наукам при Московском государственном университете имени М.В. Ломоносова по адресу: 119991, Москва, Ленинские горы, 1/3, МГУ, Химический факультет, кафедра органической химии, аудитория 446.

С диссертацией можно ознакомиться в библиотеке Химического факультета МГУ им. М.В. Ломоносова

Автореферат разослан « 27 » марта 2009 года.

Ученый секретарь диссертационного совета,
доктор химических наук, профессор

Магдесиева Т.В.

Актуальность темы. Межмолекулярные взаимодействия являются ключевыми как при исследовании живых систем на молекулярном уровне, так и при разработке новых перспективных материалов с заданными свойствами. Такие исследования – через направленный органический синтез – приводят к соединениям, имеющим широкое практическое применение и большое значение в фармацевтической химии, медицине, промышленности и сельском хозяйстве. Атомные заряды оценивают наиболее важный – электростатический – вклад в межмолекулярные взаимодействия, поэтому они нашли широкое применение в различных областях молекулярного моделирования. Опыт применения атомных зарядов в различных областях молекулярного моделирования (молекулярная динамика, докинг, конформационный поиск, корреляции «структура-свойство/структура-активность») сформировал целый ряд в чем-то противоречивых требований к зарядам и схемам их расчета: высокая скорость расчета, хорошее воспроизведение электростатического потенциала вокруг молекул (молекулярного электростатического потенциала – МЭП), соблюдение топологической симметрии зарядов для интенсивного исследования конформационного пространства. Предложенные к настоящему моменту схемы не удовлетворяют части этих требований. Кроме того, во многих случаях теория, положенная в их основу не вполне прозрачна физически, что препятствует их последовательной модификации для достижения оптимального соотношения точности описания и скорости вычисления зарядов – важного дополнительного требования при разработке перспективных методов многоуровневого моделирования, действующих на масштабах от небольших органических структур до наноразмерных ассоциатов макромолекул.

Таким образом, существующая потребность в эмпирических схемах расчёта атомных зарядов, удовлетворяющих всем вышеописанным требованиям, делает несомненной актуальность задач, связанных с разработкой подобных методов расчёта.

Цель и задачи исследования. Целью данной работы является построение эмпирических зарядовых схем, которые позволяли бы получить заряды с корректной топологической симметрией, а практическая реализация которых воспроизводила бы молекулярный электростатический потенциал с точностью, удовлетворяющей требованиям молекулярного моделирования.

Для достижения поставленной цели необходимо было решить следующие задачи:

- Предложить рациональные способы последовательного построения эмпирических зарядовых схем, гарантирующих соблюдение топологической симметрии зарядов;
- Оптимизировать параметры предложенных эмпирических схем для воспроизведения МЭП, рассчитанного квантово-химически;
- Продемонстрировать применимость полученных схем в молекулярном моделировании путем воспроизведения физико-химических свойств молекулярных систем, зависящих от зарядов. Для сравнения провести аналогичное исследование с известными зарядовыми схемами;

- На основе предложенных теоретических подходов к построению зарядовых схем продемонстрировать общность некоторых зарядовых схем, известных из литературы и считающихся основанными на различных принципах.

Научная новизна. Предложены два новых общих подхода для получения эмпирических схем расчета топологически симметричных атомных зарядов: а) полное выравнивание ЭО с топологически симметричной функцией энергии системы (ТСФЭ) и б) динамическая релаксация ЭО (ДРЭО) по связям. Показана возможность адекватного воспроизведения МЭП и дипольных моментов молекул различных структурных классов посредством топологически симметричных атомных зарядов, рассчитанных при помощи предложенных эмпирических зарядовых схем. Показана математическая эквивалентность некоторых известных зарядовых схем уравнениям метода ДРЭО и, таким образом, предложена удобная классификация зарядовых схем.

Практическая значимость работы. Предложенная классификация эмпирических подходов к построению схем расчета атомных зарядов, в том числе топологически симметричных включает в себя известные схемы, а также открывает возможности систематической разработки новых. Показаны пути дальнейшего развития и модификации описанных базовых подходов к построению эмпирических схем путем постепенного увеличения уровня детализации. Создан программный комплекс для оптимизации параметров эмпирических зарядовых схем и анализа результатов. Предложенные зарядовые схемы с параметрами, оптимизированными для воспроизведения квантово-химически рассчитанного [ХФ/6-31G(d)] МЭП, могут быть непосредственно применены в прикладном молекулярном моделировании с использованием классических неполяризуемых силовых полей, а также в виртуальном скрининге, молекулярном докинге и количественных исследованиях связи «структура-свойство» и «структура-активность» (QSPR/QSAR).

Апробация работы. Результаты работы были представлены на следующих конференциях: XII Международной конференции «Математика. Компьютер. Образование» (г. Пущино, 2005), 4-й Всероссийской конференции «Молекулярное моделирование» (Москва, 2005), 9-й Всероссийской школе-конференции по квантовой и вычислительной химии им. В.А.Фока (Великий Новгород, 2005), XIII Международной конференции «Математика. Компьютер. Образование» (г. Дубна, 2006), 10-й Международной конференции им. В.А. Фока (Казань, 2006), XIV международной конференции «Математика. Компьютер. Образование» (г. Пущино, 2007), 5-й Всероссийской конференции "Молекулярное моделирование" (Москва, 2007), 4-м международном симпозиуме «Вычислительные методы в токсикологии и фармакологии, включающие Интернет-ресурсы» (СМТПИ-2007, Москва, 2007), 3-м международном семинаре «Молекулярные вычислительные исследования в материаловедении и биологических науках» (MSSMBS-2008, г. Дубна, 2008).

Публикации. По материалам диссертации опубликовано 14 печатных работ, в том числе 3 статьи в отечественных и зарубежных журналах.

Структура и объем работы. Диссертация состоит из введения, четырех глав, выводов, приложения и списка цитируемой литературы. Работа изложена на 197 страницах, включает 59 рисунков и 40 таблиц. Список литературы включает 163 наименования.

Содержание работы

Во введении кратко изложена актуальность темы диссертации, сформулированы цели работы, отражены научная новизна и практическая значимость.

В главе 2 «Литературный обзор» показана важность требования хорошего воспроизведения молекулярного электростатического потенциала (МЭП) для того, чтобы заряды имели как весомое теоретическое обоснование, так и широкое практическое применение в молекулярном моделировании (рис. 1). Обоснован выбор метода расчета МЭП [ОХФ/6-31G(d)] для автоматического согласования зарядов с существующими методами моделирования процессов в конденсированной фазе посредством классических силовых полей молекулярной механики и аналогичных методов. Кроме того, рассмотрено требование топологической симметрии (химической связности) атомов в значениях зарядов для получения качественно верных результатов моделирования. Проанализированы существующие подходы к расчету атомных зарядов с выявлением их сильных и слабых сторон. На основании литературного обзора выдвинуты требования к эмпирическим методам расчета атомных зарядов, необходимых в молекулярном моделировании, которые бы обеспечивали одновременно качественно верные результаты и допускали оптимизацию параметров для воспроизведения МЭП.

Основные результаты работы

В главе 3 «Теоретическая часть» изложены два предлагаемых теоретических подхода для построения топологически симметричных зарядовых схем, генерирующих топологически симметричные заряды, и их реализации. Приведен сравнительный анализ предлагаемых подходов с существующими в литературе методами.

Первый подход основан на полном выравнивании электроотрицательностей (ЭО) атомов с использованием топологически симметричной функции энергии системы (ТСФЭ), второй – на принципе динамической релаксации ЭО (ДРЭО).

Оба метода основаны на положениях концептуальной теории функционала электронной плотности (DFT), согласно которой полная энергия молекулярной системы является функционалом электронной плотности с условием нормировки количества электронов. В случае эмпирических схем энергия системы представляется функцией от величин зарядов на атомах с условием сохранения полного заряда системы:

$$E_{mol} \cong E_{mol}(\mathbf{q}) \quad (1)$$

$$\sum_i^{Nat} q_i = Q_{tot} \quad (2),$$

где q_i – заряд на атоме i ; \mathbf{q} – вектор атомных зарядов q_i для всей структуры; Q_{tot} – полный заряд системы, Nat – число атомов в молекуле.

Несмотря на то, что эмпирическая функция энергии системы может иметь сколь угодно сложный вид, для приближенных вычислений эту функцию обычно представляют в виде разложения в ряд по степеням атомных зарядов, причем первые несколько коэффициентов этого разложения имеют строгое определение в рамках

концептуальной теории функционала плотности. Как правило, ограничиваются несколькими первыми слагаемыми разложения, например

$$\chi_i = \left(\frac{\partial E}{\partial q_i} \right)_{q_i=0} \quad (3)$$

$$\eta_i = \left(\frac{\partial^2 E}{\partial q_i^2} \right)_{q_i=0} \quad (4),$$

где χ_i – электроотрицательность (ЭО) и η_i – жесткость атома i .

Численные результаты по реализации схем, оптимизации их параметров для воспроизведения квантово-химически рассчитанного МЭП, а также валидация полученных параметров в прикладном молекулярном моделировании приведены в последующих главах.

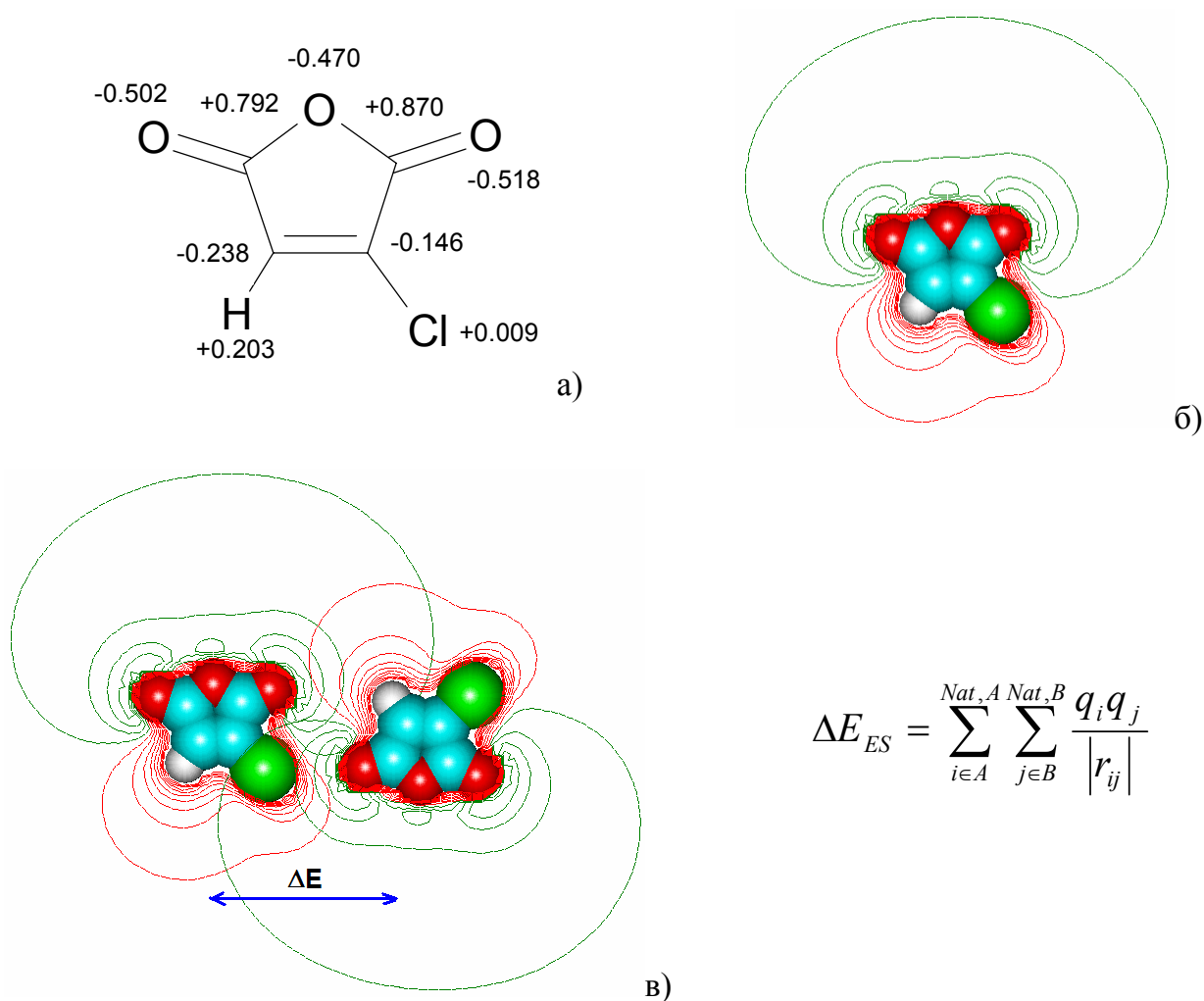


Рисунок 1. Иллюстрация использования атомных зарядов в моделировании межмолекулярных взаимодействий: а) органическая структура с рассчитанными зарядами; б) рассчитанный электростатический потенциал; в) использование зарядов для оценки электростатических взаимодействий.

Полное выравнивание ЭО атомов с использованием топологически-симметричной функции энергии системы (ТСФЭ). Основная идея метода – построение функции энергии системы, которая бы изначально отражала топологию (связность) органической молекулы. Одним из наиболее распространенных эмпирических способов получения атомных зарядов является минимизация функции энергии системы, зависящей от значений атомных зарядов, с условием сохранения полного заряда системы, что эквивалентно полному выравниванию эффективных ЭО атомов. В большинстве известных подходов используется выражение (5) для функции энергии системы, состоящей из заряженных атомов

$$E_{mol}(\mathbf{q}) = \sum_i^{Nat} \left(\chi_i^0 q_i + \frac{1}{2} \eta_i^0 q_i^2 + \frac{1}{2} \sum_{j \neq i}^{Nat} \varphi_{ij} q_i q_j \right) \quad (5),$$

где φ_{ij} , χ_i^0 и η_i^0 – параметры взаимодействия между зарядами q_i и q_j , ЭО и жесткости атома i .

Первые два слагаемых выражения (5) отвечают за притяжение заряда к данному атому, третье слагаемое – за взаимодействие между зарядами. По сути φ_{ij} – параметр электростатического взаимодействия, обычно аппроксимирующийся выражением $1/R_{ij}$ при достаточно больших расстояниях (R_{ij}) между атомными центрами i и j . При использовании только первых двух слагаемых (5) в процедуре полного выравнивания ЭО получаемые значения зарядов не зависят от химического окружения: все атомы одного типа имеют одно значение заряда. Непосредственное использование $1/R_{ij}$ проблематично: а) необходима геометрия структуры (в некоторых приложениях может потребоваться ее генерация), б) величины зарядов плохо определены при наличии близких межатомных контактов – как правило, имеют большие по абсолютной величине значения, в) получаемые заряды не обладают топологической симметрией, так как отражают (через набор конкретных R_{ij}) геометрию лишь одного конформера.

В рамках данной работы предложен подход к оценке φ_{ij} , лишенный приведенных выше недостатков: оценить параметр взаимодействия исходя из свойств молекулярного графа (топологии химических связей), а точнее, путей минимальной длины (ПМД), определяемых как минимальное количество ребер, разделяющее две заданные вершины графа. Для этого была подобрана простая зависимость между средним межатомным расстоянием в органических молекулах и соответствующей ему величиной ПМД. Для выборки оптимизированных [ОХФ/6-31G(d)] органических структур, описанных в главе «Детали вычислительного эксперимента», были вычислены расстояния в пространстве между парами атомов, разделенных 1, 2, 3, 4, 5 и 6 связями. Распределения расстояний для всех шести групп были аппроксимированы гауссовыми кривыми со значениями среднего и дисперсии μ и σ^2 .

Средние величины аппроксимированных кривых хорошо коррелируют со значениями ПМД. Для выражения $\mu(\text{ПМД}) = 0.84 \cdot \text{ПМД} + 0.46$ получены следующие результаты: $R^2 = 0.9993$, $s = 0.04$.

Полученная аппроксимация межатомного расстояния использована для получения оценки параметра межзарядового взаимодействия (6)

$$\varphi_{ij} = \alpha \cdot K(\text{ПМД}_{ij}) \frac{1}{0.84 \cdot \text{ПМД}_{ij} + 0.46} \quad (6),$$

где ПМД_{ij} – путь минимальной длины (в связях) между атомами i и j ; $K(\text{ПМД}_{ij})$ – набор коэффициентов (для $\text{ПМД}_{ij} = 1, 2, 3, 4, 5$ и 6) для сглаживания кулоновского потенциала в области малых расстояний; α – коэффициент преобразования единиц размерности для согласования с величинами χ_i^0 и η_i^0 .

При использовании оценки (6) функция энергии системы (5) становится топологически симметричной, и, следовательно, заряды, получаемые минимизацией этой функции, также отражают топологическую симметрию и, кроме того, зависимость от ближайшего химического окружения. Такой подход также позволяет в некоторой степени учесть взаимную поляризацию зарядов в молекуле.

Для практического использования схемы необходимо численно решить систему линейных уравнений (7) с квадратной матрицей размера $(N_{at} + 1)$.

$$\begin{pmatrix} \eta_1^0 & \cdots & \frac{1}{2} \varphi_{1,Nat} & 1 \\ \vdots & \eta_l^0 & \vdots & \vdots \\ \frac{1}{2} \varphi_{Nat,1} & \cdots & \eta_{Nat}^0 & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} q_1 \\ \vdots \\ q_{Nat} \\ \lambda \end{pmatrix} = \begin{pmatrix} -\chi_1^0 \\ \vdots \\ -\chi_{Nat}^0 \\ Q_{tot} \end{pmatrix} \quad (7),$$

где λ – множитель Лагранжа, имеющий здесь смысл эффективной ЭО молекулы; Q_{tot} – полный заряд молекулы.

В качестве начальных параметров были взяты значения ЭО и жесткости атомов по Годдарду, выраженные в эВ. В этом случае коэффициент согласования размерности α имеет численное значение $14.4 [\text{эВ} \cdot \text{\AA}]$. Безразмерные коэффициенты K могут принимать значение меньше 1 для коррекции кулоновского потенциала на малых расстояниях. В качестве одного из шагов в повышении уровня детализации можно использовать более сложные правила оценки φ_{ij} , с учетом типов взаимодействующих атомов.

Результаты оптимизации параметров метода представлены в главе «Обсуждение результатов». Значения коэффициентов для итоговой схемы: $K(1..6)=[0.652; 0.903; 0.967; 0.990; 1.000; 1.000]$.

Динамическая релаксация ЭО (ДРЭО) атомов. Второй подход к получению атомных зарядов основан на принципе динамической релаксации эффективных значений атомных или орбитальных ЭО в процессе гипотетического образования молекулы из изолированных атомов. Этот процесс не приводит к полному выравниванию ЭО атомов или орбиталей в молекуле. Наиболее известным примером этой группы зарядовых методов является метод частичного выравнивания орбитальных электроотрицательностей (РЕОЕ) Гастайгера-Марсили. Однако некоторые детали этого метода не вполне прозрачны с физической точки зрения, что препятствует его последовательной модификации.

В данной работе предлагается последовательно рассмотреть процесс получения атомных зарядов, исходя из «динамических принципов». Подход основан на использовании уравнений локального сохранения заряда в системе (8), сохраняющие ее полный заряд, заданный начальным распределением атомных зарядов, в процессе эволюции:

$$\frac{dq_i}{dt} = \sum_j^{Nat} \Delta q_{ij} \cong \sum_{j \in \{i, j\}}^{Nbi} \Delta q_{ij} \quad (8),$$

где $\frac{dq_i}{dt}$ – изменения заряда q_i на атоме i в процессе эволюции в эффективном времени t ; Δq_{ij} – величина элементарного переноса заряда от атома i к атому j ; правая часть – предположение, что заряд может переноситься только между непосредственно связанными атомами; Nbi – число связей, исходящих из атома i ; потребуем также детального баланса $\Delta q_{ij} = -\Delta q_{ji}$.

После принятия допущения о пропорциональности (с положительной константой c) величины элементарного переноса заряда понижению энергии системы в момент начала переноса, получается система обыкновенных дифференциальных уравнений (9), которая может быть записана в матричном виде (10):

$$\frac{dq_i}{dt} = -c \cdot \sum_{j \in \{i, j\}}^{связи} \left(\frac{\partial E_{mol}(\mathbf{q})}{\partial q_i} - \frac{\partial E_{mol}(\mathbf{q})}{\partial q_j} \right) = -c \cdot \sum_{j \in \{i, j\}}^{связи} (\chi_i^{eff}(\mathbf{q}) - \chi_j^{eff}(\mathbf{q})) \quad (9)$$

$$\frac{d\mathbf{q}}{dt} = -c \cdot L\chi^{eff}(\mathbf{q}) \quad (10),$$

где $L = D - A$ – матрица Лапласа молекулярного графа; D – матрица порядков вершин графа; A – матрица связности графа.

Хорошо изученные математические свойства спектра матрицы Лапласа применены в работе в приложении к динамическому выравниванию ЭО для получения атомных зарядов и привели к ряду интересных наблюдений. Так, например, стационарное решение системы («конечное» состояние эволюции) отвечает *принципу полного выравнивания ЭО*, поскольку эффективные ЭО всех атомов системы равны.

При анализе особенностей эволюции систем согласно рассматриваемому подходу был сформулирован *принцип динамической релаксации ЭО*: для получения топологически симметричных значений заряда с использованием приближенной функции энергии системы целесообразно провести некоторую релаксацию системы по уравнениям (10), не доводя до стационарного решения, т.е. остановить эволюцию на конечных временах. При этом система эволюционирует в сторону стационарного решения, а наиболее энергетически невыгодные составляющие, отвечающие существенным разностям эффективных ЭО на связанных атомах, редуцируются. Предположительно, эволюцию системы можно остановить в тот момент, когда величина падения энергии системы будет сопоставима с грубостью приближения для описания энергии системы.

Для практических вычислений зарядов целесообразно рассмотреть конечно-разностное интегрирование уравнений (10). Заменяя производные конечными

приращениями в левой части уравнений (10), можно получить две простейшие схемы конечно-разностного интегрирования обыкновенных дифференциальных уравнений: явную (12) и неявную (13) схемы (известные как схемы Эйлера). Разница состоит в том, с какого шага k используются значения зарядов $\mathbf{q}^{(k)}$ в правой части (11):

$$\mathbf{q}^{(n+1)} - \mathbf{q}^{(n)} = -c \cdot (L\chi_0 + L\eta_0 \mathbf{q}^{(k)}) \cdot \Delta t \quad (11),$$

где Δt – шаг дискретизации по времени; $\eta^0 = \text{diag}(\eta_1^0, \dots, \eta_{\text{Nat}}^0)$ – диагональная матрица жесткости; $\mathbf{q}^{(i)}$ – значения зарядов на итерации i , причем $\mathbf{q}^{(0)} = \mathbf{q}_0$.

$$\mathbf{q}^{(n+1)} = (I - c\Delta t \cdot B_0) \cdot \mathbf{q}^{(n)} - c\Delta t \cdot \mathbf{a}_0 \quad (12)$$

$$\mathbf{q}^{(n+1)} = (I + c\Delta t \cdot B_0)^{-1} \cdot (\mathbf{q}^{(n)} - c\Delta t \cdot \mathbf{a}_0) \quad (13),$$

где $B_0 = L\eta_0$, а $\mathbf{a}_0 = L\chi_0$.

В работе проанализированы преимущества и недостатки каждого из методов интегрирования. Явный метода Эйлера вычислительно более прост. Однако в случае реальных химических структур получаются так называемые «жесткие системы» уравнений, требующие для сохранения точности и устойчивости применения неявных методов. Более сложные методы интегрирования нецелесообразны из-за приближенной природы самого вычисления, некоторого произвола в точке остановки релаксации, а также с целью сохранения разумного соотношения вычислительных затрат и качества получаемых результатов. Реализованная нами версия метода ДРЭО, параметры которой были впоследствии оптимизированы, основана на использовании неявной схемы интегрирования.

В реализованной схеме в качестве стартовых значений ЭО и жесткости были использованы значения Годдарда, эффективный шаг интегрирования, $c\Delta t$, был принят равным 0.05 [(зар. эл.)²/эВ], и использовалось две итерации ($n = 2$).

Сравнение предложенного динамического подхода с существующими методами. В результате проведенных исследований была предложена новая методика классификации зарядовых схем на основании выражений метода ДРЭО в виде конечно-разностных схем. Кроме того, удалось обнаружить общность в некоторых методах оценки зарядов, основанных, как правило, на неполном выравнивании ЭО, считавшихся до сих пор основанными на совершенно различных принципах.

Так, например, показана общность с методами связевых инкрементов, таких как метод расчета зарядов силового поля MMFF94 и части составного метода AM1-BCC. При принятии достаточно разумных допущений показана общность решений уравнений ДРЭО при помощи явной схемы интегрирования с методом РЕОЕ Гастайгера. Кроме того, показана математическая общность метода ДРЭО с неявным интегрированием со стадией выравнивания предложенных ранее в нашей лаборатории методов, основанных на представлении структуры в виде молекулярного (МГ) и орбитального (ОГ) графов. При разработке этих методов использовалась аналогия графов органических структур с замкнутой электрической цепью, позволившей воспользоваться законами Кирхгофа. Впоследствии, этим методам была дана более строгая интерпретация в терминах дискретных решений уравнения Лапласа с краевыми условиями Дирихле. В данной работе получена новая интерпретация шага

выравнивания, а также приведены дополнительные проясняющие соображения относительно природы неполного выравнивания ЭО.

Еще одной важной отличительной чертой как стадии выравнивания методов МГ и ОГ, так и предлагаемого в работе динамического подхода, является одновременный учет влияния всех атомов системы на итоговое значение ЭО при использовании неявной схемы. Для иллюстрации этого рассмотрим разложение обратной матрицы неявной схемы в ряд (14), существующее при условии $c\Delta t \cdot \|\eta^0 L\|_2 < 1$:

$$(I + c\Delta t \eta^0 L)^{-1} \cong I - c\Delta t \cdot \eta^0 L + (c\Delta t)^2 (\eta^0 L)^2 - (c\Delta t)^3 (\eta^0 L)^3 + (c\Delta t)^4 (\eta^0 L)^4 \dots \quad (14)$$

В данном разложении первые два слагаемых в точности дают матрицу явного метода. При этом учитывается только влияние непосредственно связанных соседей, поскольку первая степень матрицы Лапласа содержит нулевые элементы для несвязанных атомов. Третье и последующие слагаемые представляют собой осциллирующие поправки, быстро сходящиеся, если выполнено условие существования разложения. Степени матрицы Лапласа отражают влияние ЭО атомов, разделенных двумя, тремя и т.д. связями. Таким образом, при использовании даже одной итерации неявной схемы происходит учет влияния удаленных по связям атомов молекулы, тогда как для явной схемы приходится производить несколько итераций для этих целей. Так, например, в схеме Гастайгера для учета индуктивного эффекта используется шесть итераций явной схемы.

Поэтому для расчета зарядов целесообразно использовать методы, основанные на неявной схеме интегрирования. Из-за существенной разреженности матрицы Лапласа численное обращение для матриц больших молекулярных систем можно оптимизировать путем использования итеративных методов решения систем линейных уравнений и хранения ненулевых элементов в специализированных разреженных массивах (матрицах).

В главе 4 «Детали вычислительного эксперимента» описана методика исследования воспроизведения МЭП различными атомными зарядами при помощи специально созданного программного комплекса, а также методика оптимизации параметров предложенных зарядовых схем для повышения качества надежного воспроизведения электростатического потенциала на широком спектре органических соединений. Описанные в работе результаты получены на выборке из 226 разнообразных органических структур, включающей спирты, амины, тиолы, карбоновые кислоты, галогениды, амиды, нитро- и нитрозосоединения, а также структуры, сочетающие различные группы – аминспирты, тиоамиды, оксикарбоновые кислоты, енолы, диазины, соединения фосфора и серы в различных степенях окисления и другие. Геометрии всех структур были оптимизированы в приближении ХФ/6-31G(d). Для всех структур был рассчитан МЭП (в том же приближении) на решетках вокруг них, а также заряды схем сравнения: Мерца-Коллмана (МК-ESP), RESP-заряды, заряды методов MMFF94, AM1-BCC и заселенностей Малликена. В последующих исследованиях использованы различные подвыборки из описанного полного набора структур. Для целей оптимизации параметров использовался алгоритм симплексного отжига, сочетающий возможности глобальной и локальной оптимизации.

Глава 5 «Обсуждение результатов» состоит из трех разделов: «Воспроизведение RESP-зарядов схемами МГ и ОГ», «Воспроизведение МЭП» и «Результаты оптимизации параметров для воспроизведения МЭП».

Воспроизведение RESP-зарядов схемами МГ и ОГ. Основными задачами этого промежуточного этапа являлись: а) отработка методики оптимизации параметров для математически схожих зарядовых схем МГ и ОГ; б) получение предварительных результатов для анализа. В рамках данного исследования была проведена оптимизация параметров этих двух схем с использованием RESP-зарядов в качестве эталонных, поскольку они наилучшим образом согласно методу наименьших квадратов (МНК) описывают МЭП вокруг каждой структуры. В исследовании использованы первые 174 структуры полной выборки. Для оптимизации использовалась целевая функция F , определяемая формулой (15), для оценки достигнутого качества – стандартное отклонение $RMSD$ (16):

$$F_M(\chi_M, \eta_M, \dots) = \sum_i^{N_{mol}} \sum_j^{N_{am,i}} (q_{M,ij}^{расч} - q_{ij}^{RESP})^2 \quad (15),$$

$$q_{M,ij}^{расч} = q_{M,ij}^{расч}(\chi_M, \eta_M, \dots) \quad M = \{MG, OG\}$$

где N_{mol} – количество молекул в выборке; $N_{am,i}$ – количество атомов в молекуле i ; q_{ij} – частичный заряд на атоме j в молекуле i , а варьируемыми параметрами являются: вектор значений атомных или орбитальных электроотрицательностей χ и вектор значений орбитальных жесткостей для модели ОГ η ;

$$RMSD(\chi, \eta, \dots) = \left(\frac{1}{N_{Mol}} \sum_i D_i^2 \right)^{1/2}$$

$$D_i = \left(\frac{1}{N_{am,i}} \sum_j (q_{ij}^{расч} - q_{ij}^{RESP})^2 \right)^{1/2} \quad (16),$$

где $RMSD(\chi, \eta, \dots)$ – стандартное отклонение рассчитываемых зарядов от эталонных RESP зарядов для всей выборки структур; D_i – стандартное отклонение заряда в рамках одной структуры i .

В работе описаны две стадии данного этапа: 1) оптимизация параметров с использованием всех структур; 2) оптимизация параметров с использованием только тех структур, МЭП которых достаточно адекватно воспроизводится рассчитанными RESP-зарядами.

На первой стадии выборка структур была разделена на обучающую и контрольную. Оптимизация параметров происходила только на структурах обучающей выборки. Значения отклонений, достигнутых при оптимизации параметров, представлены в табл. 1, где приведены также среднеквадратические отклонения зарядов, полученные при использовании оригинальных (неоптимизированных) параметров. Оптимизированные параметры продемонстрировали схожие результаты на контрольной выборке структур, что подтверждает их устойчивость и переносимость.

Таблица 1. Остаточные расхождения [$RMSD(\chi, \eta, \dots)$, в а.е. заряда] для полной выборки. В скобках указано количество структур.

Параметры	Обучающая выборка (166)		Контрольная выборка (8)	
	МГ	ОГ	МГ	ОГ
Оригинальные	0.314	0.244	0.320	0.232
Оптимизированные	0.224	0.127	0.207	0.148

Значения RESP-зарядов для обучающей выборки лежат в диапазоне $[-1.2 \dots +1.4]$.

На второй стадии были использованы два критерия, позволивших выделить структуры с надежными RESP-зарядами: 1) величина относительной остаточной невязки в воспроизведении расчетного МЭП RESP-зарядами $RRMS$ (17) и 2) статистическая обусловленность всех зарядовых центров структуры (18)

$$RRMS = \left[\chi^2 / \sum_{\text{решетка}, i}^{N_T} (V_i)^2 \right]^{1/2} \quad (17),$$

где V_i – значения квантово-химического электростатического потенциала, оцененного в узлах решетки из N_T точек, χ^2 – остаточная дисперсия метода RESP.

$$\sigma(q_k) = \left[\chi^2 / (N_T - N_A - N_{Ozp}) \right]^{1/2} (A^{-1})_{kk} \quad (18),$$

где q_k – заряд на k -ом атоме, N_T – число точек в решетке, N_A – число атомов, N_{Ozp} – число ограничений (в данном случае используется единственное ограничение: сумма атомных зарядов равна нулю), $(A^{-1})_{kk}$ – диагональный элемент обратной матрицы нормальных уравнений метода наименьших квадратов (МНК) для нахождения значений зарядов. Величина σ имеет смысл среднеквадратического отклонения и является мерой выпуклости функции χ^2 в окрестности значения атомного заряда q_k , определяя его значимость для воспроизведения МЭП.

В плохо обусловленных центрах величины зарядов могут значительно варьироваться, оказывая небольшое влияние на общее качество воспроизведения потенциала. Мы отказались от попытки воспроизводить RESP-заряды для структур, в которых содержатся центры с величиной $\sigma \geq 0.1$ (в долях заряда электрона). Также из рассмотрения были исключены структуры с $RRMS > 0.2$.

После исключения структур по представленным критериям была проведена повторная оптимизация и оценка предсказательной способности, результаты которой представлены в табл. 2 и на рис. 2.

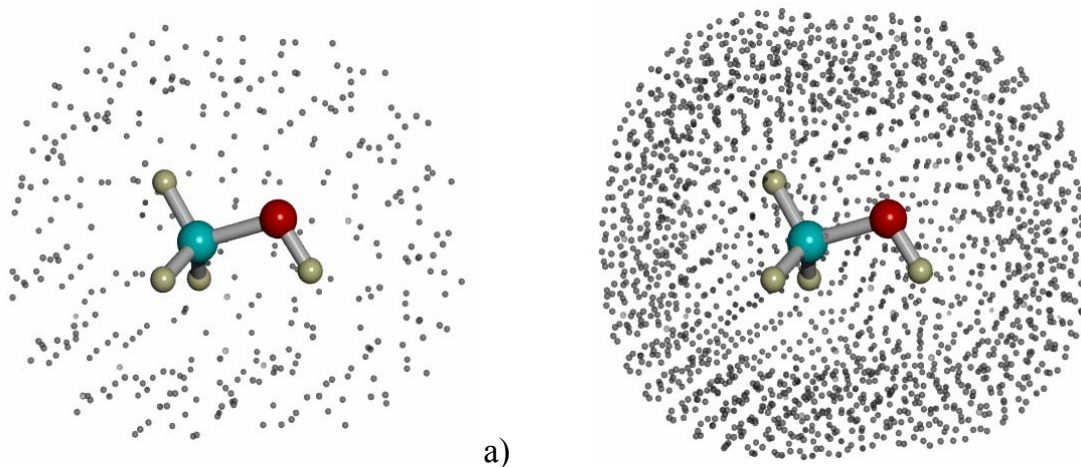


Рисунок 3. Структура метилового спирта с решеткой точек: а) метод Мерца-Колмана, б) предложенная в данной работе модификация. Для наглядности часть точек, расположенных ближе к наблюдателю, не отображена.

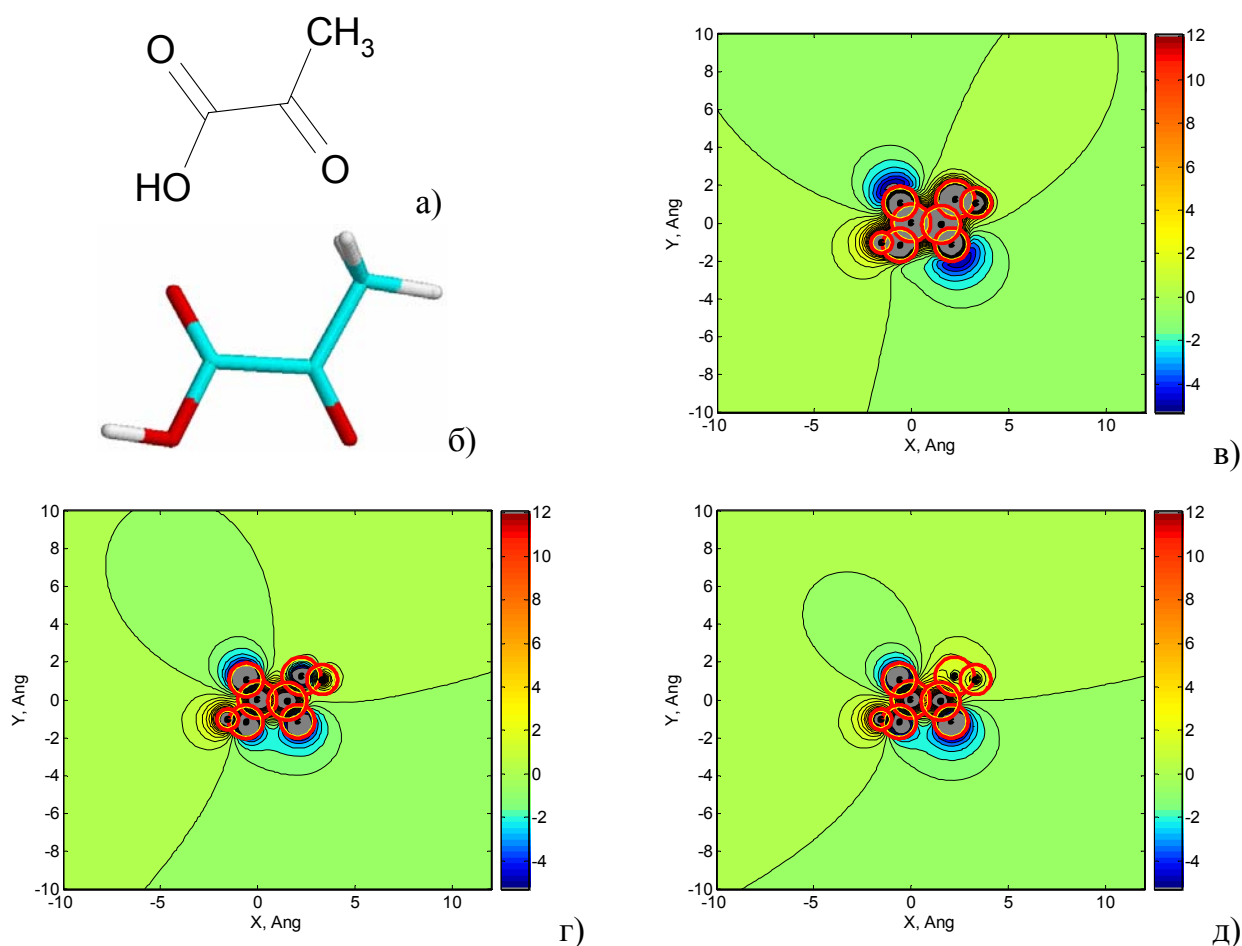


Рисунок 6. Пример сопоставления МЭП [В] в плоскости структуры пировиноградной кислоты контрольной выборки (а, б): в) расчетный квантово-химический МЭП; г) МЭП с использованием RESP-зарядов; д) МЭП с использованием ДРЭО зарядов с оптимизированными на обучающей выборке параметрами.

Исключенные структуры можно разбить на три класса: малополярные молекулы (I), структуры с выраженной анизотропией электронной плотности (II), а также структуры с тяжелыми элементами (сера, бром, фосфор) и с полигалогенметильными группами (III). Объяснены причины отличия в описании МЭП для этих классов, а также приведены возможные пути улучшения описания за счет учета анизотропии распределения зарядовой плотности.

Таблица 2. Остаточные расхождения (в а.е.) для уменьшенной выборки.

Параметры	Обучающая выборка (84)		Контрольная выборка (8)	
	МГ	ОГ	МГ	ОГ
Оригинальные	0.366	0.251	0.320	0.232
Оптимизированные	0.176	0.108	0.151	0.136

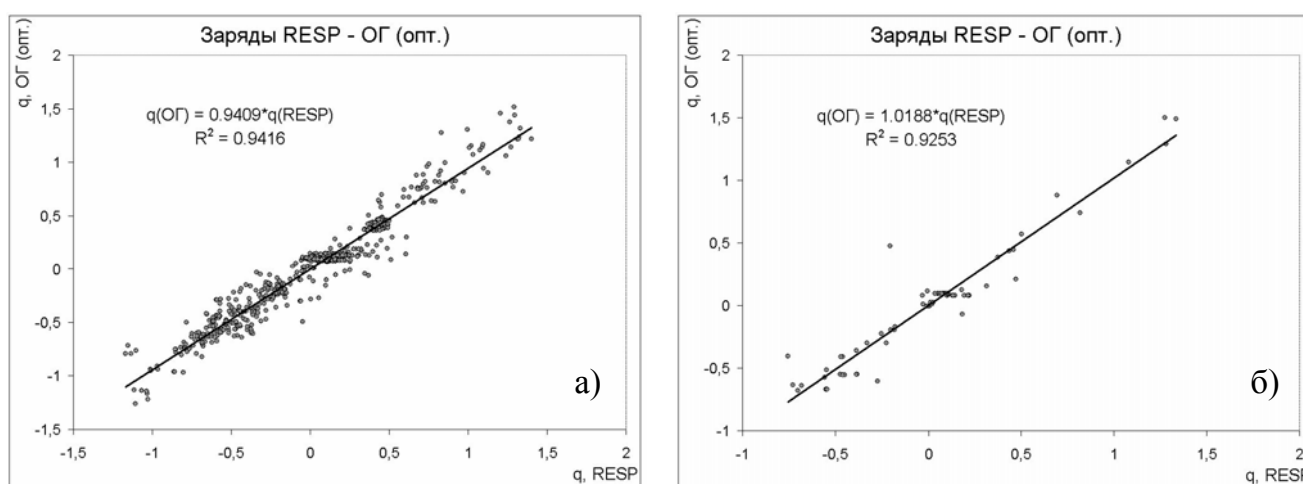


Рисунок 2. Воспроизведение RESP-зарядов эмпирической схемой ОГ после оптимизации параметров на: а) обучающей и б) контрольной выборках.

Выявлен недостаток подхода оптимизации, основанного на воспроизведении RESP-зарядов, состоящий в том, что многие атомные центры, за исключением атомов на поверхности, недостаточно хорошо обусловлены в процедуре МНК подгона зарядов к расчетному МЭП. Значения зарядов значительно варьируются даже в схожих химических окружениях, а их величины, как правило, достаточно высоки. К этой группе соединений относятся структуры, содержащие sp^3 атомы углерода. В таких, а также в структурах со значительной анизотропией электронной плотности (кратные связи, неподеленные электронные пары) заряды, получаемые прямым подгоном к МЭП, принимают некоторые компромиссные значения, призванные уменьшить среднеквадратическую ошибку описания МЭП на выбранной решетке точек вокруг структуры. Такие значения нередко не поддаются разумной интерпретации в терминах химического влияния окружения и не могут быть воспроизведены при помощи эмпирических зарядовых схем. Попытка принять их в расчет приводит к ухудшению параметров для описания других структур. Естественным выходом из данной ситуации видится оптимизация значений параметров для непосредственного воспроизведения МЭП, рассчитанного неэмпирически. В этом случае можно получить параметры, отражающие лишь

существенные вариации в качестве воспроизведения МЭП (при этом величины зарядов могут численно отличаться от RESP-зарядов), поскольку именно хорошее воспроизведение МЭП и является критерием успешной применимости получаемых атомных зарядов.

В разделе «Воспроизведение МЭП» описывается как качество воспроизведения квантово-химически рассчитанного потенциала известными зарядовыми схемами, так и процесс оптимизации параметров эмпирических зарядовых схем для его лучшего воспроизведения при помощи созданного программного комплекса.

Две эмпирические зарядовые схемы, предложенные в данной работе – ТСФЭ и ДРЭО, – а также зарядовые схемы МГ и ОГ были исследованы на предмет их параметризации для непосредственного одновременного воспроизведения расчетного [ХФ/6-31G (d)] электростатического потенциала вокруг молекул выборки структур.

Описанная выше полная выборка структур была разбита на: а) обучающую (196 структур) и б) контрольную (30 структур). На основании выводов предыдущего этапа и дополнительных исследований предложена модификация стандартного алгоритма размещения точек в пространстве вблизи структуры: а) количество слоев увеличено с 4 до 5 и б) увеличена плотность точек на каждом слое в 4 раза по сравнению с процедурой Мерца-Коллмана. В дальнейшем исследовании использовалась именно эта модификация решетки, в которой точки расположены на пяти сферах с радиусами 1.4, 1.6, 1.8, 2.0, 2.2 значений ван-дер-ваальсовых радиусов атомов при плотности в 4 точки на квадратный Å поверхности) (рис. 3).

Для оптимизации параметров эмпирических зарядовых схем была использована целевая функция вида суммы квадратов разностей (19) и средней абсолютной ошибки (21), тогда как для оценки достигнутого качества воспроизведения МЭП использовалась функция (20) в виде среднеквадратического отклонения

$$F(\chi^0, \eta^0, \dots) = \sum_{i=1}^{N_{cmp.}} \sum_{j=1}^{N_{y3l.,i}} \left(V_{ij}^{кв.} - \sum_{l=1}^{N_{am.,i}} q_{il}^{расч.} / r_{jl} \right)^2 \quad (19)$$

$$q_{ij}^{расч.} = q_{ij}^{расч.}(\chi^0, \eta^0, \dots)$$

$$D_V = \left[\frac{1}{N_{cmp.}} \sum_{i=1}^{N_{cmp.}} D_i^2 \right]^{1/2} \quad (20)$$

$$D_i = \left[\frac{1}{N_{y3l.,i}} \sum_j \left(V_{ij}^{кв.} - \sum_{l=1}^{N_{am.,i}} q_{il} / r_{jl} \right)^2 \right]^{1/2}$$

$$AD_V = \frac{1}{N_{cmp.}} \sum_{i=1}^{N_{cmp.}} \left(\frac{1}{N_{y3l.,i}} \sum_j \left| V_{ij}^{кв.} - \sum_{l=1}^{N_{am.,i}} q_{il} / r_{jl} \right| \right) \quad (21),$$

где $N_{cmp.}$ – число структур в выборке; $N_{y3l.,i}$ – число точек решётки, в которых МЭП вычислен квантово-химически и сравнивается с классическим потенциалом, для структуры i ; $V_{ij}^{кв.}$ – квантово-химический электростатический потенциал в точке j

решётки для структуры i ; $N_{am,i}$ – число атомов в молекуле i , r_{jl} – расстояние между атомом l и точкой решётки j ; $q_{il}^{расч.}$ – заряд на атоме l в молекуле i , который в эмпирических схемах (ТСФЭ, ДРЭО, МГ и ОГ) зависит от параметров: χ^0 – вектора ЭО и η^0 – вектора жесткостей для используемых атомных типов; D_i – среднеквадратическая ошибка описания квантово-химического МЭП на решетке вокруг структуры i ; D_V , AD_V – среднеквадратическая и средняя абсолютная ошибки описания МЭП для всей выборки структур одновременно.

Оптимизация параметров осуществлялась поэтапно: а) оптимизация для воспроизведения эталонных RESP-зарядов (15), вычисленных для тех же решеток и потенциала, б) оптимизация среднеквадратичной функции (19) с использованием быстрого суммирования, в) оптимизация среднего абсолютного отклонения (21) для получения робастных оценок параметров.

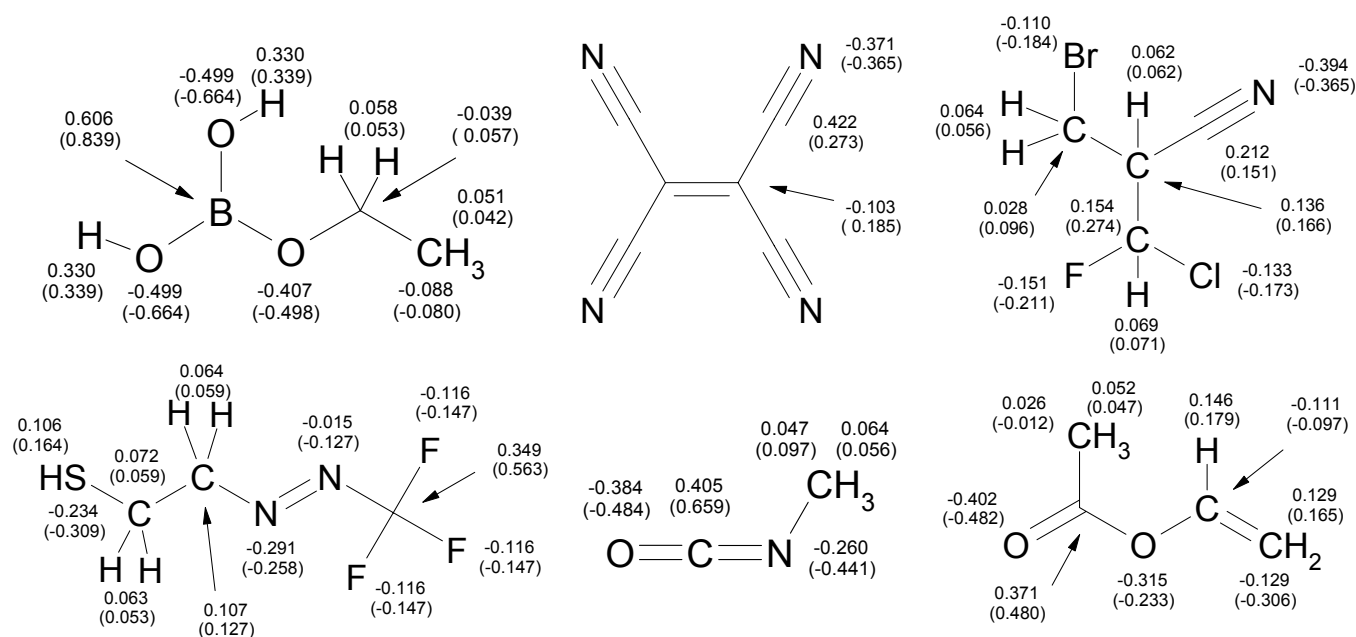


Рисунок 4. Примеры органических структур с рассчитанными значениями зарядов ТСФЭ (без скобок) и ДРЭО (в скобках). Параметры оптимизированы для воспроизведения МЭП.

В работе проанализированы причины значительной корреляции между параметрами, а также существенно различного влияния параметров различных типов атомов, приводящих в итоге к неуникальным параметрам и осложняющих ход самой оптимизации. Так, например, для повышения надежности результатов в работе использована концепция экономного введения атомных типов, в которой новые – более детальные – атомные типы вводятся только, если это статистически оправданно и приводит к качественно лучшему описанию МЭП.

Примеры расчета зарядов методами ТСФЭ и ДРЭО с оптимизированными значениями параметров для некоторых органических структур приведены на рис. 4 и 5. Паттерн распределения величин зарядов, как правило, весьма схож для двух методов. Наибольшие отличия в величинах зарядов наблюдаются на фрагментах,

МЭП вокруг которых воспроизводится не вполне однозначно при помощи зарядовых схем.

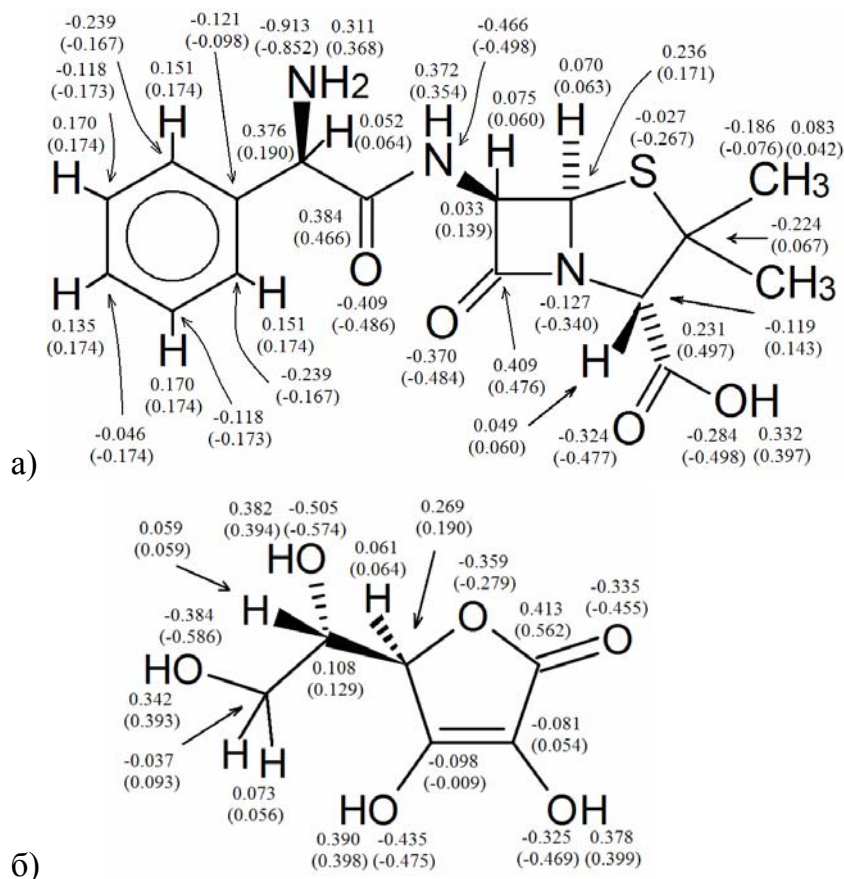


Рисунок 5. Примеры органических структур с зарядами, рассчитанными по методам ТСФЭ и ДРЭО (в скобках) с оптимизированными значениями параметров: а) ампициллин, б) аскорбиновая кислота (витамин С).

В этой части также отмечены выбросы и случаи сложного воспроизведения МЭП исследуемыми зарядовыми схемами в процессе оптимизации их параметров. Их анализ приведен в следующем разделе.

В результате проделанной работы удалось достичь параметров эмпирических зарядовых схем, дающих заряды, адекватно воспроизводящие МЭП.

В разделе «**Общий анализ результатов**» приведено сравнение достигнутого качества воспроизведения МЭП различными методами. Характеристики зарядов, полученных при помощи эмпирических зарядовых схем с найденными в результате оптимизации значениями их параметров, были разносторонне исследованы на предмет их применимости в прикладном молекулярном моделировании. Для этого был проведен ряд расчетов физико-химических свойств структур, входящих как в обучающую, так и в контрольную выборки, непосредственно зависящих от зарядового распределения, а результаты были сопоставлены с рассчитываемыми значениями на предмет воспроизведения: а) МЭП в точках решетки вокруг молекул и б) компонентов вектора дипольного момента. Полученные результаты сопоставлены с аналогичными вычислениями, проведенными с использованием зарядов известных

схем, что позволяет полнее охарактеризовать эмпирические зарядовые схемы, исследованные в рамках данной работы.

Таблица 3. Результаты воспроизведения МЭП [В] для обучающей выборки из 196 структур, рассчитанные при помощи зарядов различных методов.

Зарядовая схема	AD_V	D_V	$D_{i, \min}$	$D_{i, \max}$	μ	σ
МК-ESP	0.038	0.061	0.013	0.130	0.055	0.026
RESP	0.043	0.066	0.014	0.134	0.061	0.027
AM1-BCC	0.079	0.113	0.035	0.383	0.104	0.045
MMFF94 ¹	0.098	0.135	0.037	0.375	0.126	0.049
Анализ по Малликену	0.145	0.202	0.019	0.436	0.184	0.084
МГ (опт.)	0.089	0.124	0.029	0.378	0.116	0.046
ОГ (опт.)	0.070	0.101	0.016	0.291	0.094	0.036
ТСФЭ (опт.)	0.097	0.136	0.026	0.348	0.126	0.053
ДРЭО (опт.)	0.081	0.113	0.016	0.321	0.106	0.041
МГ (ориг.)	0.206	0.281	0.035	0.584	0.259	0.110
ОГ (ориг.)	0.220	0.301	0.020	0.624	0.271	0.131
ТСФЭ (ориг.)	0.268	0.367	0.036	0.939	0.328	0.166
ДРЭО (ориг.)	0.225	0.309	0.039	0.608	0.284	0.122

¹Из выборки исключены 7 соединений бора из-за отсутствия параметров; расчетные значения МЭП лежат в диапазоне $[-2.752; 2.393]$, В; $D_{i, \min}$, $D_{i, \max}$ – минимальное и максимальное значение D_i (20) для структур выборки; μ , σ – параметры гауссиан, аппроксимирующих реальное распределение ошибок D_i по выборке структур.

Таблица 4. Результаты воспроизведения МЭП [В] для контрольной выборки из 30 структур, рассчитанные при помощи зарядов различных методов.

Зарядовая схема	AD_V	D_V	$D_{i, \min}$	$D_{i, \max}$	μ	σ
МК-ESP	0.044	0.066	0.024	0.100	0.064	0.018
RESP	0.047	0.070	0.024	0.105	0.067	0.018
AM1-BCC	0.089	0.119	0.078	0.191	0.115	0.032
MMFF94	0.109	0.146	0.069	0.248	0.140	0.042
Анализ по Малликену	0.155	0.203	0.087	0.291	0.192	0.066
МГ (опт.)	0.092	0.124	0.075	0.239	0.119	0.035
ОГ (опт.)	0.080	0.112	0.043	0.247	0.105	0.037
ТСФЭ (опт.)	0.099	0.131	0.084	0.213	0.128	0.029
ДРЭО (опт.)	0.085	0.113	0.064	0.171	0.110	0.026
МГ (ориг.)	0.219	0.303	0.103	0.718	0.279	0.122
ОГ (ориг.)	0.241	0.317	0.170	0.672	0.297	0.111
ТСФЭ (ориг.)	0.268	0.346	0.134	0.542	0.332	0.099
ДРЭО (ориг.)	0.233	0.323	0.087	0.666	0.298	0.129

Результаты по воспроизведению квантово-химически рассчитанного МЭП вокруг структур обучающей выборки приведены в табл. 3. В верхней части таблицы представлены результаты, полученные с использованием различных зарядовых схем, известных из литературы: МК-ESP – схема подгона к МЭП Мерца-Коллмана; RESP – схема Бейли с наложением топологических ограничений; AM1-BCC – схема с начальным расчетом зарядов полуэмпирическим методом AM1 с последующим применением табулированных значений связевых инкрементов; MMFF94 – метод связевых зарядовых инкрементов, используемый в одноименном силовом поле; и анализ заселенностей по Малликену. В средней части приведены результаты, полученные с оптимизированными, а в нижней – с оригинальными (стартовыми) значениями параметров ЭО и жесткости атомов (орбиталей для метода ОГ).

Единицы измерения потенциала – В (Вольт), соответствует работе в 1 эВ по приведению единичного положительного заряда из бесконечности в точку с данным потенциалом. Множитель для преобразования к а) атомной системе единиц потенциала $3.674932 \cdot 10^{-2}$ [а.е.п./В], б) системе СИ – 96.4845044 [(кДж/моль·зар.эл.)/В].

Для проверки переносимости полученных параметров и непереученности исследуемых эмпирических моделей аналогичные исследования были проведены также на контрольной выборке из 30 структур (табл. 4), не используемых при оптимизации параметров.

На рис. 6 приведен пример контурной карты МЭП для структуры пировиноградной кислоты (структура контрольной выборки). Видно, что распределение потенциала, рассчитанное при помощи зарядов предлагаемой схемы ДРЭО с оптимизированными значениями параметров (рис. 4д) аналогично результатам с RESP-зарядами. Первые быстро вычисляются эмпирической схемой, тогда как вторые требуют трудоемких квантово-химических расчетов. Успешное использование параметров, подобранных на обучающей выборке, для расчета зарядов на структурах контрольной выборки подтверждает переносимость самих параметров, а также непереученность модели. На рисунке приведены окружности половины ван-дер-ваальсовых радиусов атомов, показывающие, что в моделировании две структуры не могут приблизиться на расстояния, меньшие, чем касание внешней огибающей всех окружностей (из-за резко возрастающего обменно-корреляционного отталкивания). Воспроизведение потенциала имеет значение только с внешней стороны этих окружностей.

Сопоставление табл. 3 и 4, а также рис. 6 и 7 позволяет сделать следующие заключения. Первое, даже для методов МК-ESP, RESP, использующих прямой подгон зарядов для воспроизведения МЭП, остаточная среднеквадратическая ошибка достаточно велика и может быть существенной для структур, электростатический потенциал которых неадекватно воспроизводится в рамках приближения точечных атомно-центрированных зарядов. Второе, наложение топологической симметрии на заряды в процедуре МНК метода RESP приводит к незначительному увеличению ошибки МЭП по сравнению со схемой МК-ESP, где эти ограничения отсутствуют. Третье, заряды по Малликену дают среднюю ошибку в 3 раза большую, чем с использованием методов прямого подгона, несмотря на использование той же волновой функции и геометрии молекул. Четвертое, эмпирические схемы на основе зарядовых инкрементов AM1-BCC, MMFF94 занимают промежуточное положение со средней ошибкой в 2 раза большей, чем для методов прямого подгона. Пятое,

предлагаемые эмпирические подходы расчета зарядов ТСФЭ, ДРЭО после оптимизации параметров демонстрируют результаты, аналогичные с методами на основе зарядовых инкрементов AM1-BCC, MMFF94, причем схема ДРЭО по остаточным ошибкам МЭП сопоставима с гораздо более трудоемкой схемой AM1-BCC, требующей задания начальной геометрии, ее оптимизации полуэмпирическим методом AM1 для получения начальных значений зарядов с последующим применением зарядовых инкрементов. Шестое, характеристики ошибки не изменились при переходе к контрольной выборке структур, что доказывает хорошую переносимость полученных параметров.

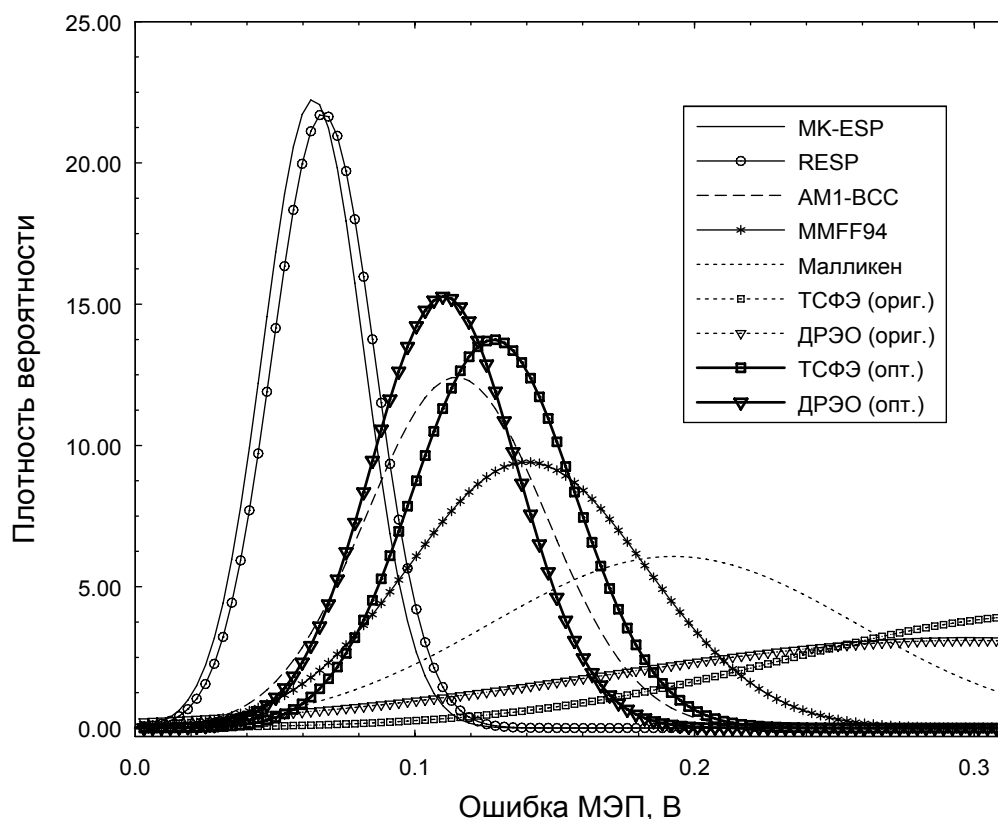


Рисунок 7. Аппроксимации распределения ошибок воспроизведения МЭП (D_i) по структурам контрольной выборки при помощи известных зарядовых схем, а также схем ТСФЭ и ДРЭО с оригинальными (пунктирные линии) и оптимизированными (жирные линии) значениями параметров.

На рис. 7 приведены гауссианы, аппроксимирующие реальное распределение среднеквадратических ошибок D_i для структур контрольной выборки, параметры которых представлены также в табл. 3 и 4 для обучающей и контрольной выборок структур соответственно.

В качестве теста на применимость получаемых зарядов было исследовано воспроизведение квантово-химически рассчитанных дипольных моментов. Компоненты дипольных моментов были рассчитаны для всех структур обучающей и контрольной выборок из зарядов, полученных эмпирическими методами, и сравнены с результатами вышеуказанных известных методов. Этот тест можно считать достаточно независимым, поскольку величины компонентов дипольных моментов не учитывались при оптимизации параметров для всей выборки. Результаты представлены в

табл. 5 и 6, где приведены величины среднеквадратической и средней абсолютной ошибок в воспроизведении квантово-химически рассчитанных компонентов дипольных моментов.

Таблица 5. Характеристики ошибки воспроизведения проекции дипольного момента на ось x , рассчитанной квантово-химически [6-31G(d)] для обучающей выборки структур, Дебай*.

Зарядовая схема	Стандартное отклонение	Беззнаковое среднее	Минимальное значение	Максимальное значение
МК-ESP	0.05	0.03	-0.21	0.15
RESP	0.09	0.06	-0.60	0.30
AM1-BCC	0.34	0.24	-0.92	1.24
MMFF94	0.46	0.33	-1.56	1.57
Малликен	0.80	0.59	-2.45	2.82
МГ (опт.)	0.40	0.25	-2.15	1.61
ОГ (опт.)	0.31	0.20	-1.56	1.21
ТСФЭ (опт.)	0.49	0.30	-2.77	1.93
ДРЭО (опт.)	0.39	0.26	-1.79	1.47
МГ (ориг.)	1.26	0.87	-4.32	4.65
ОГ (ориг.)	1.46	0.97	-4.88	6.02
ТСФЭ (ориг.)	2.02	1.31	-5.54	8.80
ДРЭО (ориг.)	1.38	0.96	-3.84	5.51

* Расчетные значения проекции дипольного момента на ось x для всей выборки лежат в диапазоне $[-7.20; 5.28]$, стандартное отклонение (от нуля) составляет 2.00, а беззнаковое среднее 1.48.

Таблица 6. Характеристики ошибки воспроизведения проекции дипольного момента на ось x , рассчитанной квантово-химически [6-31G(d)] для контрольной выборки структур, Дебай*.

Зарядовая схема	Стандартное отклонение	Беззнаковое среднее	Минимальное значение	Максимальное значение
МК-ESP	0.07	0.05	-0.16	0.15
RESP	0.08	0.05	-0.15	0.22
AM1-BCC	0.50	0.36	-1.04	1.31
MMFF94	0.60	0.48	-1.65	1.04
Малликен	1.10	0.88	-2.10	2.08
МГ (опт.)	0.35	0.23	-0.65	1.21
ОГ (опт.)	0.46	0.30	-1.67	0.78
ТСФЭ (опт.)	0.49	0.39	-1.15	0.97
ДРЭО (опт.)	0.35	0.27	-0.76	0.89
МГ (ориг.)	1.52	0.98	-5.26	3.28
ОГ (ориг.)	1.84	1.40	-4.90	3.51
ТСФЭ (ориг.)	1.66	1.28	-3.46	3.70
ДРЭО (ориг.)	1.53	1.04	-4.81	3.51

* Расчетные значения проекции дипольного момента на ось x для всей выборки лежат в диапазоне $[-4.69; 5.54]$, стандартное отклонение (от нуля) составляет 2.21, а беззнаковое среднее 1.73.

По качеству воспроизведения дипольных моментов схемы расчета заряда выстроились в ряд, совпадающий с рядом воспроизведения МЭП. Аналогично,

ошибка воспроизведения дипольного момента исследуемыми зарядовыми схемами значительно уменьшилась после оптимизации параметров.

В этом же разделе проанализирована относительная важность влияния различных эффектов на качество воспроизведения МЭП на решетках вокруг молекул, представлен анализ структур, для которых воспроизведение МЭП сталкивается с различными трудностями; проанализирована вычислительная сложность алгоритмов и предложены пути дальнейшего увеличения точности описания МЭП, а также развития самого подхода.

Таблица 7. Итоговое сравнение методов расчета атомных зарядов.

Метод	ТСФЭ	ДРЭО	МГ	ОГ	MMFF94	AM1-BCC	RESP	MK-ESP	Малликен
Воспроизведение МЭП	+	++	+	++	+	++	+++	+++	±
Простота, ясность, скорость расчета	++	++	++	+	+++	±	--	-	-
Нетребовательность к наличию 3D геометрии	+	+	+	+	+	-	--	--	--
Топологическая симметрия	+	+	+	+	+	-	+	-	-
Количество параметров. N – число атомных типов	~N	~N	~N	~N	~N ²	~N ²	нет	нет	нет
Заряженные: структуры / фрагменты	+/-	+/+	-/-	+/-	+/+	+/+	+/+	+/+	+/+
Масштабируемость	+	+	+	±	±	-	нет	нет	нет
Возможность дальнейшего улучшения	+	++	±	±	±	±	нет	нет	нет

По совокупности (табл. 7) качественных характеристик и возможности параметризации для адекватного воспроизведения МЭП наиболее перспективной для прикладного использования в молекулярном моделировании и дальнейшего развития является предложенная в данной работе эмпирическая зарядовая схема, основанная на динамической релаксации электроотрицательности (ДРЭО) с использованием неявной схемы интегрирования, позволяющая быстро рассчитывать топологически симметричные значения зарядов для разнообразных органических соединений. В практическом плане предложенные схемы расчета зарядов могут быть непосредственно использованы в молекулярном моделировании для описания электростатических взаимодействий широкого спектра органических соединений.

Выводы

1. Предложено два подхода к построению эмпирических схем расчета топологически-симметричных атомных зарядов в органических структурах, основанные а) на минимизации топологически-симметричной функции энергии системы (ТСФЭ) и б) на динамической релаксации электроотрицательности (ДРЭО) вдоль связей. Предложенные подходы масштабируемы: возможно последовательное изменение детализации описания – от небольших органических структур до наноразмерных ассоциатов макромолекул – с требуемым отношением точности и скорости описания электростатических взаимодействий.
2. В рамках предлагаемого принципа динамической релаксации электроотрицательности вдоль связей (ДРЭО) предложена новая интерпретация ряда известных зарядовых методов, обнаруживающая их теоретическую общность, ранее неочевидную. Таким образом, предложена новая возможность как для классификации уже существующих, так и для разработки новых эмпирических зарядовых схем.
3. Разработана процедура оптимизации параметров исследуемых эмпирических методов для наилучшего воспроизведения молекулярного электростатического потенциала (МЭП), рассчитанного квантово-химически, вокруг совокупности разнообразных органических структур выборки.
4. Применение предложенной процедуры подбора параметров предлагаемых эмпирических зарядовых схем позволило приблизиться к качеству устойчивого воспроизведения МЭП, характерного для существенно более трудоемких и детализированных методов, что подтверждает теоретические предпосылки, заложенные в основы подходов.
5. Заряды реализованных методов ТСФЭ и ДРЭО с оптимизированными значениями параметров для воспроизведения МЭП, рассчитанного в приближении ХФ/6-31G(d), представляют практическую ценность и могут быть непосредственно использованы в молекулярном моделировании.
6. Создан программный комплекс, позволяющий: а) быстро реализовать новые зарядовые схемы (или модифицировать старые); б) подбирать параметры методов, исходя из условия оптимального воспроизведения свойств, зависящих от атомных зарядов, на различных выборках структур; в) анализировать результаты воспроизведения интересующих свойств при помощи внешних атомных зарядов.

Список публикаций

1. Шульга Д.А., Олиференко А.А., Писарев С.А., Палюлин В.А., Зефирова Н.С. Моделирование атомных RESP-зарядов с помощью топологических схем расчета. Доклады Академии Наук, **2006**, 408(3), 340-343.
2. Shulga D.A., Oliferenko A.A., Pisarev S.A., Palyulin V.A., Zefirov N.S. Fast tools for calculation of atomic charges well suited for drug design. SAR and QSAR in Environmental Research, **2008**, 19(1-2), p. 153-165.
3. Шульга Д.А., Олиференко А.А., Писарев С.А., Палюлин В.А., Зефирова Н.С. Параметризация эмпирических схем расчета частичных атомных зарядов для воспроизведения молекулярного электростатического потенциала. Доклады Академии Наук, **2008**, 419(2), 214-218.
4. Шульга Д.А., Олиференко А.А., Писарев С.А., Палюлин В.А., Зефирова Н.С. Исследование параметризации топологических методов расчета частичных атомных зарядов для воспроизведения *ab initio* молекулярного электростатического потенциала. Сборник научных трудов конференции «Математика. Компьютер. Образование» Ижевск: Научно-издательский центр "Регулярная и хаотическая динамика", **2005**, выпуск 12, Том 3., 1101-1112.
5. Шульга Д.А., Олиференко А.А., Писарев С.А., Палюлин В.А., Зефирова Н.С. Исследование параметризации топологических методов расчета частичных атомных зарядов для воспроизведения *ab initio* молекулярного электростатического потенциала. Тезисы XII Международной конференции «Математика. Компьютер. Образование», г. Пущино. Москва, Ижевск: Научно-издательский центр "Регулярная и хаотическая динамика", **2005**, выпуск 12, с. 229.
6. Шульга Д.А., Олиференко А.А., Писарев С.А., Палюлин В.А., Зефирова Н.С. Моделирование атомных RESP-зарядов с помощью топологических схем расчета. Сборник тезисов 4-й Всероссийской конференции «Молекулярное моделирование», **2005**, с. 118.
7. Shulga D.A., Oliferenko A.A., Pisarev S.A., Palyulin V.A., Zefirov N.S. Reproduction of the MEP by means of the topological atomic charge models. Сборник тезисов докладов 9-й Всероссийской школы-конференции по квантовой и вычислительной химии им. В.А.Фока, **2005**, № 1248.
8. Шульга Д.А., Олиференко А.А., Писарев С.А., Палюлин В.А., Зефирова Н.С. Топологические схемы расчета частичных атомных зарядов. Описание квантово-химического молекулярного электростатического потенциала. Тезисы XIII Международной конференции «Математика. Компьютер. Образование», г. Дубна. Москва, Ижевск: Научно-издательский центр "Регулярная и хаотическая динамика", **2006**, выпуск 13, с. 237.
9. Shulga D.A., Oliferenko A.A., Pisarev S.A., Palyulin V.A., Zefirov N.S. Empirical atomic charge models trained to reproduce molecular electrostatic potential (MEP): quality of dipole moment. Book of abstracts of the 10-th V.A. Fock Meeting on Quantum and Computational Chemistry, Kazan, **2006**, p. 61.
10. Шульга Д.А., Олиференко А.А., Палюлин В.А., Зефирова Н.С. Динамическое выравнивание электроотрицательности как способ получения топологических

- атомных зарядов. Тезисы XIV международной конференции «Математика. Компьютер. Образование», г. Пущино. Москва, Ижевск: Научно-издательский центр "Регулярная и хаотическая динамика", **2007**, выпуск 14, с. 194.
11. Шульга Д. А., Олиференко А. А., Палюлин В. А., Зефирова Н. С. Топологические атомные заряды: подходы к построению схем расчета. Тезисы 5-й Всероссийской конференции "Молекулярное моделирование", **2007**, доклад № 20, с. 34.
 12. Shulga D.A., Oliferenko A.A., Pisarev S.A., Palyulin V.A., Zefirov N.S. Fast tools for calculation of atomic charges well suited for drug design. Fourth International Symposium in Computational Methods in Toxicology and Pharmacology Integrating Internet Resources (СМТПИ-2007), Moscow, **2007**, abstract book, PO-79, p. 149.
 13. Олиференко А.А., Яковенко А.Я., Шульга Д.А., Писарев С.А., Палюлин В.А., Зефирова Н.С. Масштабируемое силовое поле для моделирования наноразмерных биологических структур. I Всероссийская конференция «Многомасштабное моделирование процессов и структур в нанотехнологиях» (ММПЧН–2008), Москва, МИФИ, **2008**, стенд № 130.
 14. Shulga D.A., Oliferenko A.A., Pisarev S.A., Palyulin V.A., Zefirov N.S. A route to molecular electrostatic through atomic charges generated by means of fast and robust empirical schemes. 3rd International workshop “Molecular Simulation Studies in Material and Biological Sciences” (MSSMBS-2008), Dubna, **2008**.