

ИССЛЕДОВАНИЕ ПАРАМЕТРИЗАЦИИ ТОПОЛОГИЧЕСКИХ МЕТОДОВ РАСЧЕТА ЧАСТИЧНЫХ АТОМНЫХ ЗАРЯДОВ ДЛЯ ВОСПРОИЗВЕДЕНИЯ *ab initio* МОЛЕКУЛЯРНОГО ЭЛЕКТРОСТАТИЧЕСКОГО ПОТЕНЦИАЛА

Шульга Д. А., Олиференко А. А., Писарев С. А.,
Палюлин В. А., Зефирова Н. С.

(Россия, Москва)

Исследована возможность описания молекулярного электростатического потенциала (МЭП) при помощи кулоновского потенциала с атомными зарядами, рассчитанными топологическими методами, разработанными ранее [1,2]. Для этого проведена оптимизация параметров для минимизации невязки рассчитываемых зарядов и RESP-зарядов [3,4], полученных из МЭП в приближении Хартри-Фока (6-31G), которые хорошо воспроизводят МЭП на выборке из 46 разнообразных органических структур. Полученная параметризация подтверждается на контрольной выборке из 9 структур, не вошедших в обучающую выборку. Полученные заряды могут использоваться для расчета электростатического компонента для многих силовых полей, применяемых в биомолекулярном моделировании.*

Введение. Силовые поля, представляющие энергию системы как функцию координат [5,6], широко используются в молекулярном моделировании. Чаще всего в силовых полях взаимодействия разделяются на связевые и несвязевые. Связевые взаимодействия ответственны за локальную геометрию, тогда как несвязевые отвечают за взаимодействие через пространство и представляют собой сумму вандерваальсова и кулоновского электростатического потенциалов. Электростатический вклад нередко доминирует в межмолекулярном взаимодействии, кроме того, он ответственен за «молекулярное распознавание», так как кулоновский потенциал относительно медленно убывает как

функция расстояния. Поэтому как для расчета энергий связывания методами молекулярной механики, Монте-Карло и молекулярной динамики, так и для быстрого скрининга больших баз данных необходимы заряды, верно воспроизводящие электростатическое взаимодействие.

Для расчета атомных зарядов существует ряд методов [7–10], значительно отличающихся по трудоемкости и качеству. Наиболее физически обоснованными и потому часто используемыми являются заряды, получаемые подгоном под молекулярный электростатический потенциал (МЭП), рассчитанный из *ab initio* волновых функций, — ESP-заряды [11,12]. Имея преимущество перед другими методами в том, что получаемые заряды лучше воспроизводят молекулярные диполи и высшие мультиполи, ESP-схемы обладают и недостатками: а) необходимостью трудоемкого неэмпирического расчета для каждой новой структуры, б) зависимостью получаемых зарядов от конформации молекулы, вследствие чего химически эквивалентные атомы могут стать неэквивалентными, в) плохой переносимостью для схожего химического окружения и г) некоторой произвольностью значений заряда для экранированных атомных центров. Неэквивалентные заряды на эквивалентных атомах могут привести к артефактам в конформационном анализе и молекулярно-динамическом моделировании. Для исправления нежелательных свойств ESP-зарядов был разработан метод RESP [3,4], в котором ограничиваются статистически плохо определенные заряды, а для эквивалентных атомов допускается лишь одинаковый заряд при подгоне. Несмотря на то, что сама процедура ресурсоемка и нелегко автоматизируема, RESP-заряды нашли широкое применение в моделировании (стандартные заряды при моделировании в пакете AMBER [13], моделирование конденсированной фазы [14]).

Весьма популярными являются также методы, основанные на принципе *выравнивания* электроотрицательностей (ЭО) атомов при образовании молекулы [15,16]. При этом атомные заряды выражаются как количества заряда, переместившиеся по системе связей от менее электроотрицательных атомов к более электроотрицательным. Таким образом, из информации о струк-

туре необходима только связность или молекулярная топология, поэтому методы называются *топологическими*.

Некоторые топологические методы исходят не из принципа полного уравнивания ЭО при образовании молекулы, а из предположения, что, хотя ЭО несколько изменяются при смещении электронной плотности, их значения не выравняются полностью. В таком случае уместно говорить об *уравновешенных* значениях ЭО. Наиболее известной является схема Гастайгера-Марсили [17].

Ранее нами был разработан ряд методов [1,2], основанных на уравнивании ЭО при образовании молекулы по законам, аналогичным уравниванию электрических потенциалов в электрических цепях по законам Кирхгоффа (кирхгоффовская зарядовая модель — КЗМ [18]). Эти методы достаточно быстры для интерактивного применения, а получаемые заряды могут быть использованы как для качественных объяснений, так и для количественного моделирования и QSAR [1,19].

В данной работе мы исследуем, насколько хорошо схемы, основанные на КЗМ, описывают электростатический потенциал, а также возможности оптимизации параметров метода для лучшего воспроизведения МЭП. Для простоты в качестве начальной меры воспроизведения электростатического потенциала мы используем близость рассчитываемых зарядов к «эталонным» RESP-зарядам для небольшой, но представительной выборки из 46 органических соединений, содержащих гетероатомы (O, N, S, Cl) и кратные связи. В результате этого и последующих исследований мы надеемся получить модель, позволяющую быстро рассчитывать заряды, приближающиеся по качеству к RESP-зарядам.

Экспериментальная часть

Для оценки способности методов описывать эталонные заряды было создано две выборки нейтральных структур: обучающая — 46 структур и контрольная — 9 структур. В обучающей выборке представлены наиболее типичные органические функциональные группы и их сочетания, ароматические и заряженные структуры не представлены. Оптимизация и последующий

расчет значений МЭП (на вандерваальсовых поверхностях Коннолли [11]) проводились в приближении RHF/6–31G* при помощи программы GAMESS [20]. RESP-заряды были получены программой resp 2.3 из пакета AMBER 6.0.

При помощи созданного программного комплекса были рассчитаны заряды каждым топологическим методом, исходя из значений соответствующих параметров метода (ЭО, жесткости, атомные радиусы) для всей выборки. Минимизация целевой функции приводит к оптимальным значениям параметров для данной выборки.

Описание топологических (КЗМ) моделей. В рамках данного исследования проводилась оптимизация параметров следующих методов: 1) *молекулярного графа* (МГ) [1,2] с атомными ЭО по Полингу [21], 2) *орбитального графа* (ОГ) [2] с орбитальными ЭО и жесткостями по Бергману и Хинце [22], 3) *декопированного молекулярного графа* (ДМГ) [18] с орбитальными ЭО и жесткостями [22].

Исходной предпосылкой методов является предположение, что при образовании молекулы из атомов электроотрицательности (ЭО) атомов «уравновешиваются» таким образом, что к атомам с большим значением ЭО перетекает часть электронной плотности от атомов с меньшими значениями ЭО. При этом по мере перетекания электронной плотности к атому, его эффективная ЭО уменьшается, и другие атомы начинают конкурировать за электронную плотность. Таким образом, «уравновешенные» значения ЭО отражают некоторый электронный баланс молекулярной системы.

Математически процедура получения «уравновешенных» значений ЭО для всех методов сводится к решению системы линейных уравнений с квадратной матрицей. Далее заряды извлекаются дополнительной обработкой. Для повышения эффективности в условиях многократного расчета ЭО и зарядов (итерации) мы использовали библиотеку ATLAS [23], основанную на широко известных кодах LAPACK, но позволяющую провести подстройку во время компиляции для получения максимальной производительности для данной платформы.

Расчет базовых RESP-зарядов. Вместо исходной двухэтапной подгонки зарядов к МЭП [3,4] мы используем только одну подгонку со «слабым ограничением» (weak restraint, qwt = 0.0005). В оригинале, на первом этапе подгоняются заряды на всех атомах, на втором — только на плохо статистически определенных $-\text{CH}_2$ - и $-\text{CH}_3$ - группах с одновременным требованием эквивалентности зарядов на топологически эквивалентных атомах.

Получаемые нашими методами заряды автоматически одинаковы на эквивалентных атомах, поэтому мы не проводили вторую стадию подгона с дополнительными ограничениями, ухудшающими качество воспроизведения МЭП в силу уменьшения числа степеней свободы. Кроме того, это избавило от необходимости создания нетривиальных входных файлов для программы resp в пакетном режиме расчета.

Однако при этом следует иметь в виду, что некоторая часть невязки при минимизации целевой функции будет вызвана попыткой подгона эквивалентных значений зарядов в КЗМ методах к пространственно неэквивалентным RESP-значениям.

Целевая функция. Целевая функция для выборки структур отражает среднеквадратичное отклонение заряда на каждом атоме выборки:

$$F(\chi, \eta, \dots) = \left(\frac{1}{N_{mol}} \sum_i^{N_{mol}} D_i^2 \right)^{1/2}$$
$$D_i = \left(\frac{1}{n_{at}} \sum_j^{n_{at}} (q_{ij}^{calc} - q_{ij}^{RESP})^2 \right)^{1/2}$$

Здесь $\bar{q}^{calc} = \bar{q}^{calc}(\chi, \eta, \dots)$, χ — параметры электроотрицательности и η — жесткости, а D_i — среднеквадратичное отклонение для данной молекулярной структуры, величину которого можно использовать для анализа качества описания МЭП этой структуры системой атомных зарядов.

Методы оптимизации. Мы решили исследовать задачу глобальной оптимизации целевой функции в пространстве параметров метода, поскольку аналитическая зависимость заряда на атоме от параметров метода для КЗМ-методов не всегда извест-

на, поэтому нет уверенности в существовании единственного минимума.

Для оптимизации параметров мы использовали метод «моделируемого отжига» (simulated annealing) [24,25]. Этот метод (в использованной форме [26]) при нулевой температуре отжига превращается в симплекс-метод локальной оптимизации. При глобальной оптимизации использовался метод гиперболического снижения температуры за указанное количество итераций, после чего начиналась фаза локальной оптимизации.

Для каждой модели (МГ, ОГ, ДМГ) значения целевой функции были рассчитаны в точке с оригинальными параметрами, которые использовались ранее с этими методами (см. выше описание КЗМ-методов). После этого производилась локальная оптимизация, начиная с точек оригинальных параметров. В результате численных экспериментов по глобальной оптимизации и наших ранних исследований выяснилось, что значение целевой функции не очень чувствительно к довольно существенным изменениям параметров ЭО и жесткости. По-видимому, поверхность целевой функции в окрестности минимума не сильно выпукла. В таком случае, найденные в результате глобальной оптимизации параметры будут существенно зависеть от пути, по которому они пришли в долину. Поэтому мы использовали средние значения параметров для нескольких десятков запусков глобальной оптимизации.

Обсуждение результатов

В таблице 1 представлены результаты оптимизации целевой функции для обучающей и контрольной выборок. Поскольку в нашей работе использовалась небольшая обучающая выборка структур, то мы использовали минимально возможную типизацию атомов — атомный тип отвечает химическому элементу. Это означает, что для каждого типа атомов используется один набор параметров. Более детальная типизация, учитывающая гибридизацию атома и его ближнее, а иногда и дальнее окружение, должна повысить качество предсказания благодаря увеличению числа степеней свободы при оптимизации. Метод ОГ изначально использует различные значения орбитальных ЭО

для различно гибридизованных атомов, однако, в данном случае сужение типизации необходимо для возможности сравнить потенциалы методов в одинаковых условиях.

Сводные данные по всем результатам регрессии приведены в таблице 2. В результате оптимизации не только улучшается значение R^2 , но и коэффициент наклона приближается к единице.

Модель МГ до оптимизации недооценивает в среднем значения зарядов на атомах в 4 раза (коэффициент наклона равен 0.24). После оптимизации параметров значения зарядов гораздо ближе к эталонным. Результат глобальной оптимизации (таблица 1) говорит о том, что в этом случае минимум, получаемый в результате локальной оптимизации оригинальных полинговских ЭО, является глобальным.

Таблица 1. Остаточные невязки (в долях заряда электрона на атом)

Параметры	Обучающая выборка			Контрольная выборка		
	МГ	ОГ	ДМГ	МГ	ОГ	ДМГ
Оригинальные	0.2476	0.1851	0.1929	0.2293	0.2062	0.2280
Лок. опт.	0.1588	0.1013	0.0939	0.1519	0.1429	0.1254
Глоб. опт. с усреднением	0.1588	0.1028	0.0892	0.1519	0.1323	0.1267

Модель ОГ дает приемлемые заряды и до оптимизации, несмотря на то, что она используется с обобщенным набором орбитальных параметров для каждого типа атомов, а после оптимизации демонстрирует довольно хорошие результаты. Как было отмечено, потенциал этого метода не может быть раскрыт на столь небольшой обучающей выборке структур из-за статистической недоопределенности параметров. Поэтому в дальнейшем планируется исследование параметризации модели на больших выборках. При 32 запусках глобальной оптимизации итоговое значение невязки колебалось в пределах 2% от среднего значения 0.100. При этом оптимальные значения параметров довольно сильно отличались от запуска к запуску, в пределах от 20 до 87% от среднего. Множественные запуски показали разную степень определенности различных параметров. В некоторых случаях параметры принимали отрицательные значения. Такие результаты могут говорить о множественности минимумов и/или недостаточном размере обучающей выборки. С другой стороны,

поскольку минимальные значения невязки были всегда близки, можно считать, что минимум один, но пологий.

Таблица 2. Данные по корреляции RESP-зарядов с рассчитанными зарядами

Параметры		Обучающая			Контрольная		
		МГ	ОГ	ДМГ	МГ	ОГ	ДМГ
Оригинальные	R ²	0.6451	0.6835	0.8109	0.6505	0.6216	0.7591
	k	0.2442	0.8772	1.2171	0.2182	0.9189	1.3072
Лок. опт.	R ²	0.7359	0.8854	0.9084	0.7271	0.7629	0.8318
	k	0.7447	0.9097	0.9075	0.7549	0.8173	0.9479
Глоб. опт.	R ²	0.7345	0.8846	0.9161	0.7272	0.7946	0.8247
	k	0.7450	0.9149	0.9161	0.7547	0.8496	0.9235

Недавно разработанная модель ДМГ продемонстрировала наилучшие показатели с точки зрения воспроизведения RESP-зарядов, однако при глобальной оптимизации в точках минимума значения параметров зачастую принимали неоправданно высокие, а часто и отрицательные значения. Причина этого явления заключается в том, что в методе используются значения разностей ЭО, но не абсолютные значения, таким образом, параметры уже не являются независимыми друг от друга при оптимизации и могут свободно дрейфовать при сохранении нужных разностей. Сочетание наилучшей предсказательной способности с неординарным поведением при оптимизации заставляет обратить большее внимание на эту перспективную модель с целью ее лучшего исследования и возможной доработки.

Для оценки способности полученных моделей воспроизводить первые члены ряда мультипольного разложения электростатического потенциала нами были рассчитаны значения дипольных моментов для соединений контрольной выборки с параметрами, полученными в результате глобальной оптимизации (таблица 3).

Несмотря на то, что в контрольной выборке присутствуют «сложные» для топологических методов сопряженные структуры, такие как имидазол и акролеин, наблюдается неплохое вос-

произведение значений дипольного момента, лучшие результаты демонстрирует модель ДМГ.

Таблица 3. Дипольные моменты для контрольной выборки, Дебан

Структура	RHF/6-31G*	RESP ^a	МГ	ОГ	ДМГ	Гаст. ^b
Бутан	0.000015	0.00411	0.000012	0.000022	0.000022	0.000005
транс-N-метил ацетамид	4.008295	3.97269	4.807083	3.949022	4.117649	1.499849
цис-N-метил ацетамид	4.175271	4.21993	7.044352	5.137945	3.650508	1.923796
бропиламин	1.543467	1.52114	2.693934	1.194799	1.763703	0.766422
имидазол	3.863127	3.80153	2.811437	2.222903	2.138707	0.906733
тиаацетальдегид	2.860094	2.91659	3.758937	1.990974	2.051407	0.959982
2-метил-акрилонитрил	4.296046	4.29645	6.306297	0.398779	2.922275	1.098641
акролеин	3.507325	3.51106	6.157641	2.053376	3.872002	1.743358
2-хлор-ацетальдимин	1.320545	1.29939	1.977104	1.565858	1.804588	0.728694

^a вес = 0.0005, одна стадия; ^b метод Гастайгера-Марсили [17] без модификаций для малых циклов и сопряженных систем

Возможно, при получении RESP-зарядов по двухстадийной схеме улучшилось бы качество моделей, поскольку вторая стадия метода RESP призвана разумно ограничить разброс на экранированных атомах. Другой альтернативой является непосредственная минимизация отклонений электростатического потенциала, рассчитанного в классическом кулоновском приближении, от потенциала, полученного из квантово-химического расчета.

Выводы

Задача относительно простого получения зарядов, хорошо воспроизводящих особенности электростатического потенциала молекул, до сих пор является одним из краеугольных камней молекулярного моделирования, особенно когда речь идет о расчете зарядов для большой базы данных структур, где квантово-

химические расчеты в адекватном приближении оказываются слишком ресурсоемкими. Нами показано, что топологические методы расчета зарядов, основанные на принципе «уравновешивания» ЭО, могут служить приемлемым компромиссом между качеством воспроизведения МЭП и скоростью расчета зарядов. При этом очевидны некоторые трудности, которые предстоит преодолеть на пути повышения качества зарядов, такие как обработка сопряженных и ароматических соединений, учет поляризации и работа с заряженными формами. Проведенные исследования демонстрируют перспективность дальнейшей работы в выбранном направлении.

Список литературы:

1. Oliferenko A.A., Palyulin V.A., Pisarev S.A., Neiman A.V., Zefirov N.S. Novel point charge models: reliable instruments for molecular electrostatics // *J. Phys. Org. Chem.* 2001. V.14. P.355–369.
2. Олиференко А.А., Палюлин В.А., Зефирова Н.С. Метод расчета частичных атомных зарядов на атомах в молекуле, основанный на новом принципе выравнивания электроотрицательностей // *ДАН.* 1999. Т.368. № 1. С.63–67.
3. Bayly C.I., Cieplak P., Cornell W.D., and Kollman P.A. Electrostatic Potential Based Method Using Charge Restraints For Determining Atom-Centered Charges: The RESP Model // *J. Phys. Chem.* 1993. V.97. P.10269.
4. Cornell W.D., Cieplak P., Bayly C.I., and Kollman P.A. Application of RESP Charges to Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation // *J. Am. Chem. Soc.* 1993. V.115. P.9620.
5. Cornell W.D., Cieplak P., Bayly C.I., Gould I.R., Merz K.M.Jr., Ferguson D.M., Spellmeyer D.C., Fox T., Caldwell J.W., and Kollman P.A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules // *J. Am. Chem. Soc.* 1995. V.117. P.5179–5197.
6. MacKerell A.D., Bashford D., Bellott M., Dunbrack R.L. et al All-atom empirical potential for molecular modeling and dynamics studies of proteins // *J. Phys. Chem. B* 1998. V.102. P.3586–3617.

7. Mulliken R.S. Electronic population analysis on LCAO-MO molecular wave functions // *J. Chem. Phys.* 1955. V.23. P.1833–1840.
8. Löwdin P.O. On the nonorthogonality problem // *Adv. Quantum Chem.* 1970. V.5. P.185–199.
9. Reed A., Weinstock R., Weinhold F. Natural Population Analysis // *J. Chem. Phys.* 1985. V.83. 735–746,
10. Bader R.F.W. *Atoms in Molecules: A Quantum Theory.* — Clarendon Press: Oxford, 1990. 458p.
11. Singh U.C., Kollman P.A. An Approach to Computing Electrostatic Charges for Molecules // *J. Comp. Chem.* 1984. V.5. P.129–145.
12. Chirlian L.E., Francl M.M. Atomic Charges Derived from Electrostatic Potentials: A Detailed Study // *J. Comp. Chem.* 1987. V.8. P.894–903.
13. AMBER home page: <http://amber.scripps.edu/>
14. Woods R.J., Chappelle R. RESP-partial charges for condensed phase MD simulations // *Theochem.* 2000. V.527. P.149–156,
15. Mortier W.J., Genechten K.V., Gasteiger J. Electronegativity Equalization: Application and Parametrization // *J. Am. Chem. Soc.* 1985. V.107. P.829–835.
16. Mullay J. A simple method for calculating atomic charge in molecules // *J. Am. Chem. Soc.* 1986. V.108. P.1770.
17. Gasteiger J., Marsili M. Iterative Partial Equalization of Orbital Electronegativity — A Rapid Access to Atomic Charges // *Tetrahedron.* 1980. V.36. P.3219–3288.
18. Oliferenko A.A., Pisarev S. A., Palyulin V.A., Zefirov N.S. Graph Theoretical Techniques in the Calculation of Atomic Charges // *Adv. Quantum Chem.* 2005 (in press). V.43.
19. Oliferenko A.A., Krylenko P.V., Palyulin V.A., Zefirov N.S. A new scheme for electronegativity equalization as a source of electronic descriptors: application to chemical reactivity // *SAR QSAR Environ. Res.* 2002. V.13. P.297–305.
20. Schmidt M.V., Baldridge K.K., Boatz J.A., Elbert S.T., Gordon M.S., Jensen J.H., Koseki S., Matsunaga N., Nguyen K.A., Su S., Windus T.L., Dupuis M., Montgomery J.A. General atomic and molecular electronic structure system // *J. Comp. Chem.* 1993. V.14. P.1347–1363.

21. Паулинг Л. Природа химической связи. — М., Л.: Госхимиздат, 1947. 440с.
22. Bergman D., Hinze J. Electronegativity and Charge Distribution // Structure and Bonding. 1987. V.66. 146–189.
23. ATLAS home page: <http://math-atlas.sourceforge.net/>
24. Kirkpatrick S. Gelatt C.D. and Vecchi M.P. Optimization by Simulated Annealing // Science. 1983. V.220. P.671–680.
25. Metropolis N., Rosenbluth A., Rosenbluth M., Teller A. and Teller E. Simulated annealing // J. Chem. Phys. 1953. V.21. P.1087–1092.
26. Press W.H., Flannery B.P., Teukolsky S.A., Vetterling W.T. Numerical Recipes in C: The Art of Scientific Computing. — Cambridge University Press; 2 edition, 1992.

**RESEARCH ON PARAMETRIZATION OF THE
TOPOLOGICAL SCHEMES FOR PARTIAL ATOMIC
CHARGES DERIVATION TO REPRODUCE *ab initio*
MOLECULAR ELECTROSTATIC POTENTIAL**

**Shulga D. A., Oliferenko A. A., Pisarev S. A.,
Palyulin V. A., Zefirov N. S.**

(Russia, Moscow)

The possibility of the molecular electrostatic potential (MEP) description by means of the Coulombic potential (inherent in many force fields), with the partial charges derived from previously developed topological methods [1,2], has been studied. To that end, the parameters of the methods have been optimized to minimize the residual function between the derived charges and the RESP-charges [3,4] calculated from the MEP at RHF/6-31G level, which well reproduce the MEP for a set of 46 diverse organic structures. The resulting parameterization has been validated using a test set of 9 structures, which were not included into the training set. The atomic charges thus obtained can be used as an electrostatic part of many force fields being used in biomolecular modeling.*