

ОПРЕДЕЛЕНИЕ ИЗМЕНЕНИЯ ДИПОЛЬНОГО МОМЕНТА ПРИ ВОЗБУЖДЕНИИ В ХРОМОФОРЕ ЗЕЛЕННОГО ФЛЮОРЕСЦЕНТНОГО БЕЛКА ИЗ ТРАЕКТОРИЙ МОЛЕКУЛЯРНОЙ ДИНАМИКИ С ПОТЕНЦИАЛАМИ КВАНТОВОЙ МЕХАНИКИ/МОЛЕКУЛЯРНОЙ МЕХАНИКИ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

© 2024 г. Т. М. Захарова^a, А. М. Кулакова^a, М. А. Криницкий^{b, c, d}, М. И. Варенцов^b, М. Г. Хренова^{a, e, *}

^aХимический факультет МГУ имени М. В. Ломоносова, Москва, Россия

^bНаучно-исследовательский вычислительный центр МГУ имени М. В. Ломоносова, Москва, Россия

^cМосковский физико-технический институт, Долгопрудный, Россия

^dИнститут океанологии им. П. П. Ширшова РАН, Москва, Россия

^eФИЦ Биотехнологии РАН, Москва, Россия

*e-mail khrenovamg@my.msu.ru

Поступила в редакцию 17.11.2023 г.

После доработки 17.11.2023 г.

Принята к публикации 15.01.2024 г.

Проведены расчеты молекулярно-динамических траекторий с потенциалами квантовой механики / молекулярной механики (КМ/ММ) для белка EYFP семейства зеленого флуоресцентного белка с последующим построением моделей машинного обучения для установления взаимосвязи между геометрическими параметрами хромофора в кадрах траектории и свойствами его электронного возбуждения. Показано, что недостаточно использовать в качестве геометрического параметра только мостиковые связи между фенильным и имидазolidоновым фрагментами хромофора, а необходимо добавлять в модель еще, по крайней мере, две соседние связи. Предложенные модели позволяют определять величину изменения дипольного момента при возбуждении со средней ошибкой 0.11 а.е.

Ключевые слова: флуоресцентные белки, машинное обучение, молекулярное моделирование, КМ/ММ

DOI: 10.31857/S0044453724110152, **EDN:** EYLEJK

Зеленый флуоресцентный белок и его аналоги являются наиболее популярным средством визуализации в живых системах [1]. Они широко применяются для селективного окрашивания клеток и тканей, для мониторинга клеточных процессов [2, 3], спектроскопии сверхвысокого разрешения [4, 5] и в качестве сенсоров [6–8]. Эти белки представляют собой бочонки, состоящие из β -листов (рис. 1); внутри каждого такого бочонка расположен хромофор, который формируется автокаталитически из трех аминокислотных остатков, входящих в полипептидную цепь белка [9]. Изменение спектральных свойств флуоресцентных белков осуществляется как за счет изменения размера сопряженной π -системы хромофора, так и за счет его окружения [10]. Белки с одинаковым хромофором и разным аминокислотным окружением могут иметь спектры поглощения, различающиеся положениями максимумов на 30–40 нм.

В литературе на протяжении многих лет обсуждается взаимосвязь геометрических параметров хромофора, а также свойств, зависящих от электронной плотности, и спектральных свойств [11–17]. В частности, наиболее важными геометрическими характеристиками являются длины связей на мостике между фенильным и имидазolidоновым фрагментами хромофора (рис. 1) [11]. Во флуоресцентном состоянии, которое наиболее интересно для прикладных исследований, хромофор находится в анионной форме, и существует равновесие между двумя таутомерными формами: Р-формой с фенолят-анионом и I-формой с отрицательным зарядом, локализованным на кислороде пятичленного цикла (рис. 1). В литературе для описания таутомерного равновесия в хромофоре зеленого флуоресцентного белка используется понятие чередования длин связей BLA (от англ. bond length alternation), которое определяется как

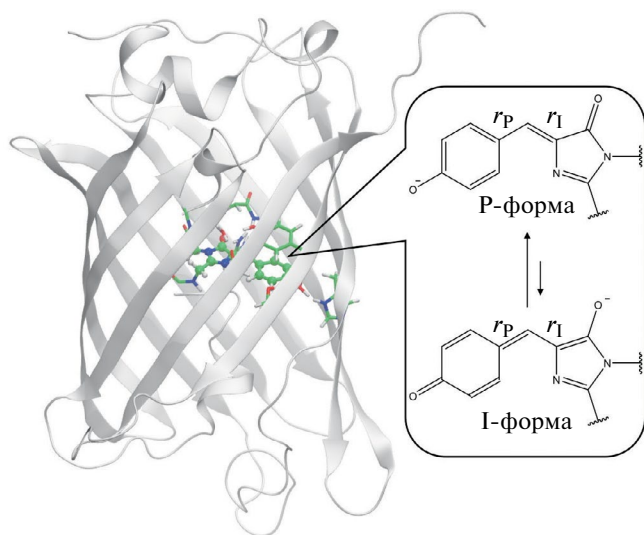


Рис. 1. Структура β -бочонка белка EYFP показана лентами. Хромофор и хромофорсодержащая область показаны шаростержневым и стержневым представлениями соответственно. Во врезке показано равновесие между двумя таутомерными формами анионного состояния хромофора.

разность длин связей на мостике $r_P - r_I$. Показано, что BVA во многом определяет положение максимума спектральной полосы поглощения для белков с хромофором зеленого флуоресцентного белка [11], однако для белков, содержащих более протяженные π -системы, эта закономерность выполняется хуже [12].

Другой дескриптор, характеризующий электронный переход в хромофоре, связан с перераспределением электронной плотности при возбуждении и определяется как изменение дипольного момента при возбуждении ($\Delta\mu$ или DMV, от англ. dipole moment variation upon excitation) и был предложен Дробыжевым и соавт. [13]. Изменение дипольного момента при возбуждении определяется произведением переносимого заряда между основным и возбужденным электронными состояниями, q_{CT} , и расстоянием между центрами масс положительной и отрицательной разностной электронной плотности ($R_+ - R_-$):

$$\Delta\mu = (R_+ - R_-) \times q_{CT}$$

В соответствии с эффектом Штарка $\Delta\mu$ линейно или квадратично связано с изменением разности энергий между основным и возбужденным электронными состояниями. В случае слабых электромагнитных полей эффект линейный, в сильных полях — квадратичный. Первоначально это было показано в экспериментальных работах [11, 13–17]. В теоретических работах было показано, что в случае изолированных комплексов хромофора

зеленого флуоресцентного белка с ароматическими молекулами наблюдается линейная зависимость [18, 19]. Для хромофоров, находящихся в β -бочонках, зависимость становится квадратичной [12, 20].

До сих пор все исследования взаимосвязи электронных и геометрических характеристик для хромофоров флуоресцентных белков проводились для отдельных состояний — кристаллических структур или рассчитанных минимумов на поверхности потенциальной энергии. При проведении расчетов методом молекулярной динамики в молекулах происходят колебания связей, что может влиять на взаимосвязь геометрических параметров и свойств электронной плотности. Однако определение энергий вертикальных переходов для каждой структуры молекулярно-динамической траектории позволяет восстанавливать форму спектральной полосы за счет построения распределения энергий переходов для всех рассчитанных структур. Прецизионный расчет энергии вертикального электронного перехода является сложной задачей и может не дать требуемой точности. Возможной альтернативой является расчет изменения дипольного момента при возбуждении для набора структур, поскольку известна его взаимосвязь со спектральными свойствами хромофора согласно закону Штарка, однако расчет $\Delta\mu$ для большого набора кадров также является вычислительно затратной задачей. Альтернативой является разработка моделей, позволяющих определять величину $\Delta\mu$ из структуры хромофора в каждом кадре траектории. Поэтому в данной работе проведено исследование взаимосвязи геометрических параметров хромофора и рассчитанных значений $\Delta\mu$ с применением машинного обучения на примере флуоресцентного белка EYFP.

МОДЕЛИ И МЕТОДЫ

Полноатомная модельная система флуоресцентного белка EYFP построена на основании кристаллической структуры белка SHardonnay [21] с заменой аминокислотного остатка фенилаланина на тирозин в 203 положении. [22] Модельная система сольватировалась в прямоугольном параллелепипеде из молекул воды и доводилась до нейтрального заряда путем добавления противоионов. Для описания хромофора и белковой макромолекулы использовалось силовое поле CHARMM36 [23, 24], а для молекул воды — TIP3P [25]. Для подготовки структуры проводился расчет методом классической молекулярной динамики (МД) в программе NAMD [26]. Все расчеты методами МД с классическими и комбинированными потенциалами проводились в каноническом ансамбле NPT при давлении 1 атм и температуре 300 К с шагом интегрирования 1 фс. Длина классической траектории составила 5 нс. Далее проводился выбор

репрезентативного кадра, который использовался в качестве стартовой структуры для проведения молекулярно-динамических расчетов с комбинарованными потенциалами квантовой механики / молекулярной механики (КМ/ММ). Длина КМ/ММ МД траектории составила 11 пс. КМ подсистема состояла из хромофора, аминокислотных остатков и молекул воды хромофорсодержащей области, всего 117 атомов вместе со связующими атомами водорода на границе КМ и ММ подсистем. КМ подсистема описывалась методом Кона–Шэма с гибридным функционалом PBE0 [27] с поправками на дисперсионные взаимодействия D3 и корреляционно-согласованным двухэкспонентным базисом cc-pvdz. Расчеты энергий и сил в КМ подсистеме проводились в программе TeraChem [28] и подавались в программу NAMD с помощью специального интерфейса [29].

Расчеты изменения дипольного момента при возбуждении $\Delta\mu$ проводились нестационарным вариантом метода функционала электронной плотности (TDDFT) с гибридным функционалом, рекомендованным для расчетов электронных переходов ω B97X-D3 [30], успешно примененным для расчета изменения дипольного момента при возбуждении для хромофоров типа хромофора зеленого флуоресцентного белка [12,19]. Для расчета энергий возбужденных состояний также использовался базисный набор cc-pvdz. Расчеты возбужденных состояний проводились в программе ORCA [31]. Величина $\Delta\mu$ рассчитывалась для перехода из основного синглетного состояния в низшее возбужденное состояние с большим значением силы осциллятора. Для расчета изменения дипольного момента при возбуждении выбирались 400 кадров КМ/ММ МД траектории из последних 10 пс равноотстоящие друг от друга по времени.

Расчеты КМ/ММ МД в основном электронном состоянии, а также расчеты вертикальных электронных переходов проводились в варианте электронного внедрения, т.е. заряды ММ окружения вносили вклад в одноэлектронную часть гамильтониана.

Для анализа взаимосвязи между рассчитанными значениями $\Delta\mu$ и соответствующими им геометрическими параметрами хромофора (15 длин связей и два двугранных угла, рис. 2) использовался регрессионный анализ, включая метод случайных лесов (Random Forests), относящийся к группе методов машинного обучения. Метод случайного леса в форме регрессии (Random Forests Regression) использовался для построения статистической модели, аппроксимирующей значения $\Delta\mu$ на основе геометрических параметров хромофора, число деревьев в ансамбле составляло 1000. При этом набор данных случайным образом делился на обучающую (320 точек) и тестовую (80 точек) выборки. При обучении модели в качестве функции

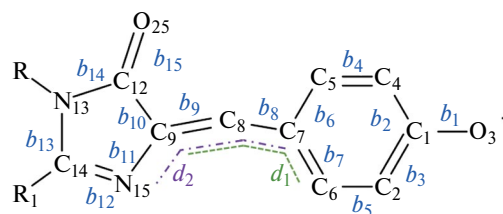


Рис. 2. Параметры моделей: связи (обозначены b с соответствующим индексом) и двугранные углы (d_1 и d_2 показаны штриховой и штрих-пунктирной линиями соответственно).

потерь использовалась средняя абсолютная ошибка (MAE).

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

На рис. 3 представлены значения дескриптора BVA и изменения дипольного момента при возбуждении $\Delta\mu$, рассчитанные для 400 кадров из КМ/ММ МД траектории длиной 10 пс. В отличие от литературных данных, полученных для минимумов на поверхности потенциальной энергии, взаимосвязи BVA и $\Delta\mu$ в молекулярно-динамической траектории не наблюдается ($R^2 = 0.28$). По всей видимости, это связано с колебаниями длин связей, которое происходит в динамике. Это предположение дополнительно подкрепляется распределениями значений мостиковых длин связей в 400 кадрах траектории (рис. 3б, в). Для двух мостиковых связей (C_7-C_8 и C_8-C_9), а также двух связей C-O ($C_{12}-O_{25}$ и C_1-O_3), удлинение одной из связей не всегда приводит к укорачиванию другой. Таким образом, можно заключить, что использование простой регрессионной модели с использованием одного параметра BVA хотя и подходит для описания набора систем с одинаковыми хромофорами в минимумах, однако не может быть использовано для описаний МД-траекторий.

Для определения важности геометрических параметров и их вклада в значение изменения дипольного момента при возбуждении $\Delta\mu$ был использован метод случайных лесов. Отличительной особенностью данного метода является возможность получения оценки важности используемых предикторов (feature importance). На первом этапе модель случайных лесов построена с использованием всех геометрических параметров — 15 длин связей сопряженной системы хромофора и двух двугранных углов, определяющих скрученность структуры по мостиковым связям. Средняя абсолютная ошибка определения $\Delta\mu$ для тестовой выборки составила 0.11 а.е. и $R^2 = 0.97$ (рис. 4). По результатам анализа наиболее важными параметрами являются длины мостиковых связей, однако их общий вклад составляет 34%. Вклады каждого из

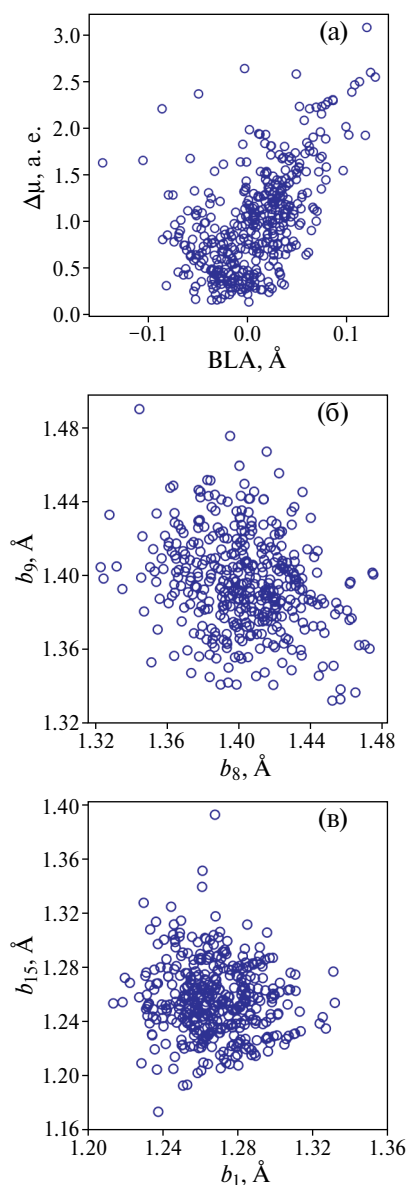


Рис. 3. Взаимосвязь свойств BLA и $\Delta\mu$ (а); длин мостиковых связей b_8 и b_9 (б), и связей углерод–кислород b_1 и b_{15} (в). Значения рассчитаны для набора кадров из КМ/ММ МД-траекторий.

остальных геометрических параметров составляют 3–7%.

При редуцировании количества параметров до двух наиболее важных — длин мостиковых связей (b_8 и b_9), ошибка определения увеличилась до 0.14 а.е. и $R^2 = 0.90$, при этом вклады обоих параметров были практически равные. Далее, в качестве параметра для обучения модели была также добавлена длина связи C_5-C_7 (b_6), примыкающей к мостику и находящейся в фенильном фрагменте. Это привело к уменьшению ошибки до 0.13 а.е. с

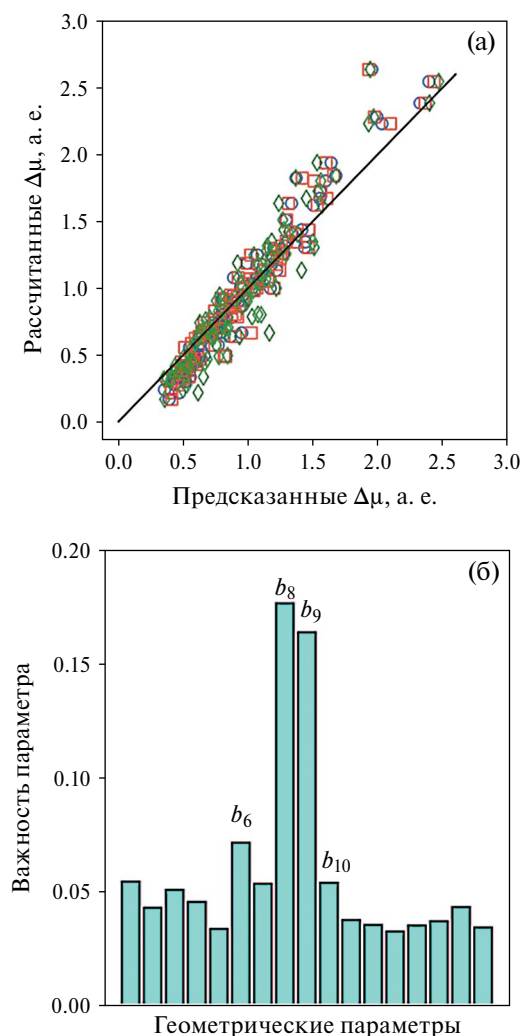


Рис. 4. Предсказанные и рассчитанные значения изменения дипольного момента при возбуждении (а) для моделей случайных лесов, построенных с использованием 17 (кругами), 4 (квадратами) и 2 геометрических параметров (ромбами). Значимость геометрических параметров в модели с 17 параметрами (б). Слева направо: длины связей по порядку $b_1 - b_{15}$ и двугранные углы d_1 и d_2 . Наиболее важные параметры имеют подпись.

$R^2 = 0.93$. Затем был добавлен четвертый по значимости параметр — длина связи, примыкающей к мостиковому фрагменту со стороны имидазолонного кольца C_9-C_{12} (b_{10}). Добавление этой связи позволило уменьшить среднюю абсолютную ошибку до 0.11 а.е. и $R^2 = 0.94$, что совпадает со значениями, получаемыми в модели с максимальным набором параметров.

Таким образом, для определения значений $\Delta\mu$ из молекулярно-динамической траектории недостаточно применять регрессионный анализ

и использовать BLA в качестве единственного дескриптора, поскольку другие геометрические параметры также вносят значительный вклад в характеристики электронного возбуждения. Оптимальным набором дескрипторов являются длины связей C₇–C₈, C₈–C₉, C₅–C₇, C₉–C₁₂.

ЗАКЛЮЧЕНИЕ

В работе показано, что изменение дипольного момента при электронном возбуждении в хромофоре зеленого флуоресцентного белка может описываться его геометрическими параметрами. Однако в отличие от минимумов на поверхности потенциальной энергии, при анализе МД-траектории недостаточно использовать две геометрические характеристики – длины мостиковых связей. По всей вероятности, это связано с колебаниями длин связей в результате эволюции системы и, как следствие, с отсутствием корреляции между длинами мостиковых связей. Для наиболее точного описания системы требуется добавление, по крайней мере, еще двух длин связей, соседствующих с мостиковыми связями.

БЛАГОДАРНОСТИ

Работа выполнена при финансовой поддержке Научно-образовательной школы МГУ “Мозг, когнитивные системы, искусственный интеллект” (проект 23-Ш03-04) с использованием оборудования Центра коллективного пользования сверхвысокопроизводительными вычислительными ресурсами МГУ им. М. В. Ломоносова.

СПИСОК ЛИТЕРАТУРЫ

1. *Enterina J.R., Wu L., Campbell R.E.* // *Curr. Opin. Chem. Biol.* 2015. V. 27. P. 10.
<https://doi.org/10.1016/j.cbpa.2015.05.001>
2. *Shinoda H., Shannon M., Nagai T.* // *Int. J. Mol. Sci.* 2018. V. 19. P. 1548.
<https://doi.org/10.3390/ijms19061548>
3. *Day R.N., Davidson M.W.* // *Chem. Soc. Rev.* 2009. V. 38. P. 2887.
<https://doi.org/10.1039/b901966a>
4. *Willig K.I., Wegner W., Müller A. et al.* // *Cell Rep.* 2021. V. 35. P. 109192.
<https://doi.org/10.1016/j.celrep.2021.109192>
5. *Lippincott-Schwartz J., Patterson G.H.* // *Trends Cell Biol.* 2009. V. 19. P. 555.
<https://doi.org/10.1016/j.tcb.2009.09.003>
6. *Tantama M., Hung Y.P., Yellen G.* // *J. Am. Chem. Soc.* 2011. V. 133. P. 10034.
<https://doi.org/10.1021/ja202902d>
7. *Ibraheem A., Campbell R.E.* // *Curr. Opin. Chem. Biol.* 2010. V. 14. P. 30.
<https://doi.org/10.1016/j.cbpa.2009.09.033>
8. *Kollenda S., Kopp M., Wens J. et al.* // *Acta Biomater.* 2020. V. 111. P. 406.
<https://doi.org/10.1016/j.actbio.2020.05.014>
9. *Tsien R.Y.* // *Annu. Rev. Biochem.* 1998. V. 67. P. 509.
<https://doi.org/10.1146/annurev.biochem.67.1.509>
10. *Rodriguez E.A., Campbell R.E., Lin J.Y. et al.* // *Trends Biochem. Sci.* 2017. V. 42. P. 111.
<https://doi.org/10.1016/j.tibs.2016.09.010>
11. *Lin C.-Y., Romei M.G., Oltrogge L.M. et al.* // *J. Am. Chem. Soc.* 2019. V. 141. P. 15250.
<https://doi.org/10.1021/jacs.9b07152>
12. *Khrenova M.G., Mulashkin F.D., Nemukhin A.V.* // *J. Chem. Inf. Model.* 2021. V. 61. P. 5125.
<https://doi.org/10.1021/acs.jcim.1c00981>
13. *Drobizhev M., Tillo S., Makarov N.S. et al.* // *J. Phys. Chem. B* 2009. V. 113. P. 12860.
<https://doi.org/10.1021/jp907085p>
14. *Bublitz G., King B.A., Boxer S.G.* // *J. Am. Chem. Soc.* 1998. V. 120. P. 9371.
<https://doi.org/10.1021/ja981606e>
15. *Drobizhev M., Makarov N.S., Tillo S.E. et al.* // *J. Phys. Chem. B* 2012. V. 116. P. 1736.
<https://doi.org/10.1021/jp211020k>
16. *Drobizhev M., Makarov N.S., Tillo S.E. et al.* // *Nat. Methods* 2011. V. 8. P. 393.
<https://doi.org/10.1038/nmeth.1596>
17. *Drobizhev M., Callis P.R., Nifosi R. et al.* // *Sci. Rep.* 2015. V. 5. P. 13223.
<https://doi.org/10.1038/srep13223>
18. *Khrenova M.G., Nemukhin A.V., Tsirelson V.G.* // *Chem. Phys.* 2019. V. 522. P. 32.
<https://doi.org/10.1016/j.chemphys.2019.02.010>
19. *Khrenova M.G., Mulashkin F.D., Bulavko E.S. et al.* // *J. Chem. Inf. Model.* 2020. V. 60. P. 6288.
<https://doi.org/10.1021/acs.jcim.0c01028>
20. *Nifosi R., Mennucci B., Filippi C.* // *Phys. Chem. Chem. Phys.* 2019. V. 21. P. 18988.
<https://doi.org/10.1039/C9CP03722E>
21. *De Meulenaere E., Nguyen Bich N., de Wergifosse M. et al.* // *J. Am. Chem. Soc.* 2013. V. 135. P. 4061.
<https://doi.org/10.1021/ja400098b>
22. *Spiess E., Bestvater F., Heckel-Pompey A. et al.* // *J. Microsc.* 2005. V. 217. P. 200.
<https://doi.org/10.1111/j.1365-2818.2005.01437.x>
23. *Best R.B., Zhu X., Shim J. et al.* // *J. Chem. Theory Comput.* 2012. V. 8. P. 3257.
<https://doi.org/10.1021/ct300400x>
24. *Denning E.J., Priyakumar U.D., Nilsson L. et al.* // *J. Comput. Chem.* 2011. V. 32. P. 1929.
<https://doi.org/10.1002/jcc.21777>
25. *Jorgensen W.L., Chandrasekhar J., Madura J.D. et al.* // *J. Chem. Phys.* 1983. V. 79. P. 926.
<https://doi.org/10.1063/1.445869>
26. *Phillips J.C., Hardy D.J., Maia J.D.C. et al.* // *Ibid.* 2020. V. 153. P. 044130.
<https://doi.org/10.1063/5.0014475>

27. *Adamo C., Barone V.* // J. Chem. Phys. 1999. V. 110. P. 6158.
<https://doi.org/10.1063/1.478522>
28. *Seritan S., Bannwarth C., Fales B.S. et al.* // WIREs Comput. Mol. Sci. 2021. V. 11. P. e1494.
<https://doi.org/10.1002/wcms.1494>
29. *Melo M.C.R., Bernardi R.C., Rudack T. et al.* // Nat. Methods 2018. V. 15. P. 351.
<https://doi.org/10.1038/nmeth.4638>
30. *Chai J.-D., Head-Gordon M.* // Phys. Chem. Chem. Phys. 2008. V. 10. P. 6615.
<https://doi.org/10.1039/b810189b>
31. *Neese, F.* // Wiley Interdiscip. Rev. Comput. Mol. Sci. 2012. V. 2. P. 73–78,
<https://doi.org/10.1002/wcms.81>.