

УДК 577.332+577.322+539.199+577.112.82

МЕТОД РОЗРАХУНКУ ПАРЦІАЛЬНОГО ОБ'ЄМУ ГЛОБУЛЯРНОГО БІЛКА НА ОСНОВІ ЙОГО АТОМНИХ КООРДИНАТ

І.Є. Щечкін, Т.О. Гуца

*Інститут біоорганічної хімії та нафтохімії НАН України, вул. Мурманська 1, Київ-94 МСП-660, 02660,
e-mail: shch406@bpci.kiev.ua*

Надійшла до редакції 14 лютого 2013 року

Прийнята 11 березня 2013 року

Запропоновано новий метод обчислення парціального об'єму глобулярного білка, в якому простір, що займає молекула, розглядається складеним з конусів, що мають загальну вершину, а основи котрих відповідають поверхневим атомам і складають поверхню молекули. Об'єм молекули приймається рівним сумі об'ємів цих конусів. Об'єм кожного конуса розраховується через площу основи, яка, в свою чергу, обчислюється виходячи з кількості вільного простору навколо відповідного атома. Розрахунки об'ємів декількох білкових молекул, проведені за допомогою запропонованого метода, показали добре узгодження з експериментальними даними.

КЛЮЧОВІ СЛОВА: парціальний об'єм, розчин білка, молекулярна поверхня, теорія гідрофобних взаємодій, конічна клітина, міжатомний простір, параметри підгонки.

МЕТОД РАСЧЕТА ПАРЦИАЛЬНОГО ОБЪЕМА ГЛОБУЛЯРНОГО БЕЛКА НА ОСНОВЕ ЕГО АТОМНЫХ КООРДИНАТ

И.Е. Щечкин Т.О.Гуца

Інститут біоорганічної хімії та нафтохімії НАН України, вул. Мурманська 1, Київ-94 МСП-660, 02660
Предложен новый метод вычисления парциального объема глобулярного белка, в котором пространство, занимаемое молекулой, рассматривается состоящим из конусов, имеющих общую вершину, основания которых соответствуют поверхностным атомам и составляют поверхность молекулы. Объем молекулы принимается равным сумме объемов конусов. Объем каждого конуса рассчитывается через площадь основания, которая, в свою очередь, вычисляется исходя из количества свободного пространства вокруг соответствующего атома. Расчеты объемов нескольких белковых молекул, проведенные при помощи предложенного метода, показали хорошее согласование с экспериментальными данными.

КЛЮЧЕВЫЕ СЛОВА: парциальный объем, раствор белка, молекулярная поверхность, теория гидрофобных взаимодействий, коническая ячейка, междуатомное пространство, параметры подгонки.

METHOD TO EVALUATE PARTIAL VOLUME OF A GLOBULAR PROTEIN BASED ON ITS ATOMIC COORDINATES

I.E. Shchekhin T.O. Hushcha

Institute of Bioorganic Chemistry and Petrochemistry of NAS of Ukraine, Murmansk 1, Kiev-94, 02660

A new technique is proposed to evaluate partial molecular volume of a globular protein where the bulk of the molecule is treated as consisting of cones with common vertex, whose bases correspond to surface atoms and constitute the surface of the molecule. Volume of molecule is assumed to be sum of the volumes of the cones. Volume of each cone is evaluated through the square of its base which, in turn, is calculated from amount of clear space around corresponding atom. Evaluating of volumes of some protein molecules using the method shows good agreement with experimental data.

KEY WORDS: partial volume, protein solution, molecular surface, theory of hydrophobic interactions, conic scaffold, interatomic space, fitting parameters.

Глобулярні білки, що існують в біологічних рідинах у вигляді індивідуальних молекул або малих кластерів, задіяні в цілому спектрі біохімічних процесів як ферменти, антитіла імунної системи, транспортери фізіологічно активних сполук. Спільна риса цих процесів полягає в асоціації ліганду з білком. Пошук ефективних

засобів контролю функції білка залежить від визначення термодинамічного профілю його взаємодії з лігандом, зокрема, об'ємного ефекту, який містить інформацію стосовно структурних змін білка.

Об'ємний ефект реакції можна визначити як різницю між парціальними об'ємами реагентів і продуктів [1]. Зміни парціального об'єму білка в залежності від температури, рН, складу розчинника і зв'язування лігандів вивчалися в основному експериментально протягом багатьох років [2-8]. Багата інформація, отримана в останні роки для послідовностей білків та їх тривимірної структури, створила можливості для розрахунку *in silico* парціального об'єму білка. Серед строгих теоретичних підходів найкращі результати показують способи обчислення термодинамічних величин на основі парних кореляційних функцій густини частинок розчину [1]. Однак складність розрахунків і недостатня кількісна відповідність їх результатів експерименту все ще обмежують застосування строгих методів для визначення парціального об'єму білка. Спрощені підходи, описані в літературі, дозволяють визначити різні види геометричного об'єму на основі декартових координат та радіусів атомів білка, і передбачають додаткові процедури для реконструкції парціального об'єму з використанням експериментальних даних [2, 9-12] або без них [13, 14].

Мета цієї статті – створити в межах спрощеного підходу єдиний алгоритм для розрахунку як парціального об'єму білка, так і складових парціального об'єму, важливих для характеристики структурних переходів білка і зв'язування ним ліганду.

ПОСТАНОВКА ЗАДАЧІ

Загальноприйнята стратегія інтерпретації парціального об'єму білка ґрунтується на розбитті його на внутрішні білкові складові та вклади, які пов'язані із сольватацією білка:

$$V_p = V_M + \Delta V_{sol}, \quad (1)$$

де V_p – парціальний об'єм білка; V_M – внутрішній об'єм молекули, що відповідає області простору, який є недосяжним для молекул води; а ΔV_{sol} – зміна об'єму розчинника внаслідок присутності білка, зумовлена зміною щільності розчинника, зв'язаного з поверхнею білка в порівнянні з вільним станом. Внутрішній об'єм V_M молекули у виразі (1), в свою чергу, розбивається на дві складові:

$$V_M = V_{vdW} + V_{free} \quad (2)$$

де V_{vdW} – сума ван-дер-ваальсових об'ємів всіх атомів, V_{free} – об'єм міжатомного простору. Процедура розбиття, представленого співвідношеннями (1) і (2), є неоднозначною внаслідок невизначеності поняття внутрішнього об'єму і існування різних способів його оцінки.

Серед описаних в літературі підходів до розрахунку молекулярного об'єму, виходячи з атомних координат білка, можна виділити дві групи методів. В першій групі об'єм обчислюється безпосередньо зондуванням простору, в якому знаходиться молекула. Такий підхід був вперше запропонований в роботі [10] для приблизної оцінки середньоквадратичних флуктуацій $\langle (\Delta V_M)^2 \rangle$ об'єму молекули білка без моделювання його молекулярної динаміки. В роботі [9] для розрахунку об'єму V_M в просторі будуються кубічні ґрати, що повністю вміщують молекулу. Об'єм V_{vdW} обчислюється як кількість таких вузлів, які знаходяться у ван-дер-ваальсовій сфері хоча б одного атома, помножена на об'єм елементарного куба ґрат. Для того, щоб врахувати додаток V_{free} і отримати повний об'єм молекули, цей розрахунок проводять з

використанням збільшених значень ван-дер-ваальсових радіусів атомів. Коефіцієнт збільшення підбирають емпірично. Аналогічний принцип застосовується при розрахунку V_M і V_p на основі об'єму пересічних сфер, а також на основі інтерпретації об'єму Вороного [13, 14]. Всі ці методи є простими для обчислювання, але їх недолік полягає в низькій розмірності простору параметризації – варіюється лише коефіцієнт збільшення атомних радіусів, що знижує якість апроксимації.

В другій групі методів спочатку визначається молекулярна поверхня, що виникає при «обкатуванні» ван-дер-ваальної оболонки молекули сферою із радіусом, який дорівнює радіусу молекули води. Ця поверхня розбивається на невеличкі фрагменти, після чого об'єм молекули обчислюється як сума об'ємів пірамід за формулою:

$$V_M = \frac{1}{3} \sum_i (r_i \cdot \Delta S_i), \quad (3)$$

де r_i – радіус-вектор від довільної точки простору до елемента поверхні ΔS_i , а сума береться по всім фрагментам молекулярної поверхні [2, 11, 12, 15]. Основна проблема при використанні формули (3) полягає у складності точної побудови явного образу молекулярної поверхні. При визначенні поверхні її елементи ΔS_i часто задають у вигляді геометричних фігур, зокрема, трикутників [15]. Існують також інші способи визначення молекулярної поверхні, такі як метод параметричного наближення сплайнами [16, 17], метод мінімальної молекулярної поверхні [18], а також найбільш сучасний метод множин рівня [19-21]. Парціальний об'єм розраховується за формулою (1) за спеціальним алгоритмом. При цьому доданок ΔV_{sol} часто додатково розбивається на дві складові: V_T , пов'язану з неідеальною упаковкою розчинника поблизу білка внаслідок їх теплового руху, та V_I , пов'язану із зміною об'єму розчинника внаслідок взаємодії з білком. Розрахунок кожного із внесків V_T і V_I проводиться на основі наближених емпіричних формул, які вимагають підбору параметрів [2]. В результаті, повна процедура обчислення об'єму V_p є достатньо громіздкою, і не обходиться без використання емпіричного тюнінгу.

Однак насправді для обчислення об'єму за формулою (3) явний вид поверхні не має значення, а потрібні лише векторні її фрагменти ΔS_i , які можна визначити так, як це робиться в сучасних ефективних алгоритмах обчислення енергії гідрофобних взаємодій [22, 23]. В цих алгоритмах поверхню не визначають явним чином, а вводять абстрактно, як сукупність деяких фрагментів, кожен з яких має такі атрибути, як точка розташування у просторі, вектор нормалі, та площа. Обчислення цих атрибутів параметризовано, і параметри підбираються так, щоб задовольнити відповідні експериментальні дані. Такий спосіб завдання молекулярної поверхні одразу неявно враховує всі складові парціального об'єму у формулі (1), та уявляється більш простим і в той же час більш гнучким ніж підхід, в якому молекулярна поверхня задається явно. До того ж такий підхід дозволяє обчислювати різні види молекулярного об'єму за допомогою єдиного алгоритму, змінюючи лише значення параметрів.

РЕЗУЛЬТАТИ Й ОБГОВОРЕННЯ

Ми використали концепцію неявно заданої молекулярної поверхні для обчислення молекулярного об'єму за формулою (3) на основі алгоритму описаному в роботі [22] і реалізованого в нашому програмному комплексі *FlexiChain*.

Так само, як і в роботах [22, 23], введемо фрагменти молекулярної поверхні такими, що відповідають частинам вільного (тобто незакритого іншими атомами молекули) простору навколо окремого атома у об'ємному куті (конусі), орієнтованому у

певному напрямку. Вектор цього напрямку є вектором нормалі до відповідного фрагменту, а точкою розташування фрагменту є центр атома, або точка на певній відстані від нього, як зображено на Рис.1. Для кожного атома введені чотири вектори напрямків, що починаються у центрі атома та є орієнтованими у просторі у тетраедральній симетрії. Таким чином, кожному атому відповідають чотири фрагменти молекулярної поверхні. Площа ΔS_{ik} фрагмента поверхні, асоційованої з напрямком k відносно атома i , обчислюється пропорційною кількості вільного простору A_{ik} у конусі, що відповідає даному напрямку:

$$\Delta S_{ik} = C_i A_{ik}, \quad (4)$$

де C_i - коефіцієнт пропорційності. Кількість вільного простору A_{ik} обчислюється через щільності сусідніх атомів за формулою:

$$A_{i,k} = \frac{\rho_i^* - \rho_{i,k}^*}{\rho_i^*} \pi (r_{H_2O} + r_i)^2, \quad (5)$$

де r_i – ван-дер-ваальсів радіус атома i , ρ_i^* – максимальне значення щільності сусідів, при якому в даному конусі ще є вільний простір. Якщо $\rho_i > \rho_i^*$, напрямок вважається «неповерхневим» і відповідна площа приймається рівною нулю.

Щільність сусідів атома i у конусі напрямку k , в свою чергу, обчислюється за формулою:

$$\rho_{i,k} = \sum_j^{m_{i,k}} G_{ij}, \quad (6)$$

$$G_{ij} = \exp \left[-\alpha_i \frac{d_{ij}^2}{(r_{H_2O} + r_j)^2} \right], \quad (7)$$

де $\rho_{i,k}$ - щільність сусідів i -го атома по k -му напрямку; $m_{i,k}$ – кількість сусідніх атомів i -го атома, центри яких належать k -му конусу; α_i – константа, що залежить від типу i -го атома, d_{ij} – відстань між центрами атомів i та j ; r_j – ван-дер-ваальсов радіус j -го сусіда атома i , r_{H_2O} – умовний радіус молекули води (1,4 Å). Сусідніми до i -го атома вважаються такі атоми, центри яких лежать від атома i не далі деякої певної граничної відстані d_{cut} . Величини C_i , ρ_i^* , r_{H_2O} , α_i та d_{cut} – параметри, що підлягають налагодженню.

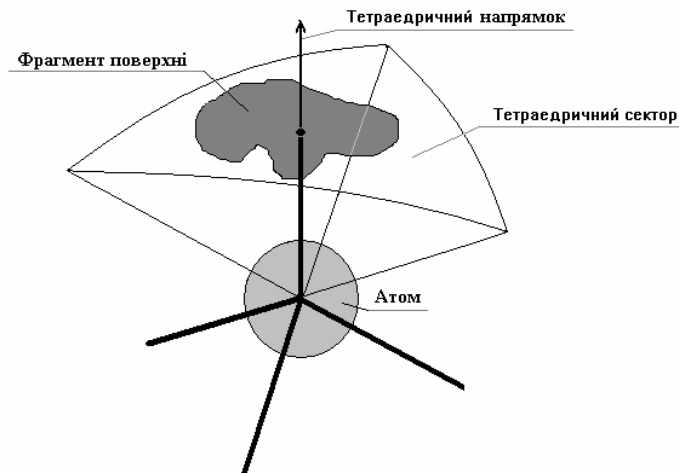


Рис.1 – Зображення фрагмента неявно заданої поверхні.

Ми провели розрахунки для серії білкових молекул, налагоджуючи чисельні параметри цього алгоритму так, щоб обчислені значення молекулярних об'ємів були найбільш близькими до значень V_p , експериментально отриманих в роботі [3]. В розрахунках були використані кристалографічні координати атомів, що містяться в PDB файлах банку даних білкових структур *Brookhaven Protein Data Bank*. Щоб досягти задовільної відповідності розрахункових результатів експериментальним даним, виявилось достатнім підібрати нові значення параметрів C_i та d_{cut} , а решту залишити такими, як в роботах [22, 23]. Оптимальні значення цих параметрів: $C_i = 2,596$ для всіх атомів, а d_{cut} - удвічі більше, ніж сума атомних радіусів сусідніх атомів. Крім того, найкраща відповідність розрахунків експериментальним даним була досягнута при зміні орієнтації векторів напрямків на протилежну порівняно з їхньою орієнтацією у вихідному алгоритмі, а замість суми радіуса атома та радіуса молекули води в формулах (4) та (5) використовувався радіус атома, оскільки ми намагались обчислювати об'єм, що обмежується молекулярною поверхнею, а не поверхнею, що доступна розчиннику, як у роботах [22, 23].

В Табл. 1 представлені, разом із відповідними молекулярними масами M та кодами PDB файлів, обчислені нами значення парціального питомого об'єму V_p для серії глобулярних білків у водному розчині. Для порівняння в Табл. 1 наведені результати розрахунків V_p за іншим методом [2], а також експериментальні значення V_p , отримані в роботі [2].

З Табл. 1 видно, що розрахунки V_p за обома методами в цілому добре узгоджуються з експериментом. Розраховані значення V_p залежать від якості рентгеноструктурних даних, на яких ґрунтуються розрахунки. При відсутності в PDB файлі білка координат значної кількості його атомів, такої, що їх сумарна маса $M \geq 500$ кДа, розбіжності між розрахованими і експериментальними значеннями V_p зростають (див. колонку 3 Табл. 1). При цьому середнє відхилення розрахованих від експериментальними значень V_p для нашого метода ($\pm 0,010$) см³/г вдвічі менше, в порівнянні з методом [2] ($\pm 0,020$) см³/г. Крім того, відхилення від експерименту при розрахунках за нашим методом краще корелює із кількістю невизначених важких атомів та цілих залишків у кристалічній структурі білків (PDB файли). Таким чином, розроблений метод цілком придатний для розрахунку парціальних об'ємів різних глобулярних білків у водному розчині за даними їх кристалічної структури.

Таблиця 1

Парціальний питомий об'єм V_p глобулярних білків у водних розчинах. 25 °С.

Білок	PDB файл	M , кДа*	Парціальний питомий об'єм V_p , см ³ /г		
			Розрахований в роботі [2]	Визначений в експерименті [2]	Розрахований нами
Трипсин	2PTN	23,0 (22,487)	0,745	0,718	0,725
Рибонуклеаза А	5RSA	13,6 (13,668)	0,721	0,702	0,693

Білок	PDB файл	M, кДа*	Парціальний питомий об'єм V_p , см ³ /г		
			Розрахований в роботі [2]	Визначений в експерименті [2]	Розрахований нами
α -Хімотрипсिनоген А	2CGA	25,7 (25,642)	0,753	0,727	0,727
Трипсिनоген	1TGN	23,0 (22,224)	0,741	0,721	0,740
α -Лактальбумін	1ALC	14,3 (13,801)	0,716	0,711	0,702
Лізоцим	4LYZ	14,3 (14,032)	0,708	0,699	0,698
Цитохром С	5CYT	12,4 (11,160)	0,706	0,735	0,744

* В дужках показані значення M , розраховані із PDB файлів.

ВИСНОВКИ

Створено простий і ефективний метод розрахунку парціального об'єму білка, що ґрунтується на концепції неявно заданої молекулярної поверхні з використанням даних про атомні координати білка. Властивості поверхні параметризуються, а значення параметрів підбираються шляхом підгонки розрахованих значень обмеженого поверхнею об'єму до експериментальних даних. Перевагою запропонованого способу в порівнянні з іншими емпіричними методами є можливість визначати не лише парціальний об'єм білка, а і його складові, такі як ван-дер-ваальсовий об'єм, внутрішній об'єм, а також вклади, пов'язані із сольватацією білка, за допомогою єдиного алгоритму, відповідно змінюючи значення параметрів підгонки. Апробація алгоритму на серії білкових молекул показала, що при визначенні парціального об'єму достатньо підібрати лише два параметра. При цьому кореляція розрахованих і отриманих експериментально значень парціального об'єму виявилась значно вищою, в порівнянні з існуючими методами. Розроблений розрахунковий метод представляє хорошу перспективу для характеристики молекулярної картини об'ємного ефекту в реакціях білків.

СПИСОК ЛІТЕРАТУРИ

1. Imai T. Molecular theory of partial molar volume and its applications to biomolecular systems / T. Imai // Condensed Matter Physics – 2007. – V. 10. – № 3(51). – P. 343–361.
2. The hydration of globular proteins as derived from volume and compressibility measurements: Cross correlating thermodynamic and structural data / T. V. Chalikian, M. Totrov, R. Abagyan, K. J. Breslauer // J. Mol. Biol.– 1996. – V. 260. – P. 588–603.
3. Gekko K. Compressibility – structure relationship of globular proteins / K. Gekko, Ya. Hasegava // Biochemistry. – 1987. – V. 25. – P. 6563–6571.

4. Poklar N. Acid- and base-induced conformational transitions of equinatoxin II / N. Poklar, J. Valker, G. Anderluh, P. Macek, T. V. Chalikian // *Biophys. Chem.* – 2001. – V. 90(2). – P. 103–121.
5. Filfil R. Binding of bovine pancreatic trypsin inhibitor to trypsinogen: spectroscopic and volumetric studies / R. Filfil, A. Rataivosi, T. V. Chalikian // *Biochemistry.* – 2004. – V. 43(5). – P. 1315–1322.
6. Filfil R. Volumetric and spectroscopic characterizations of glucose-hexokinase association / R. Filfil, T. V. Chalikian // *FEBS Lett.* – 2003. – 554(3). – P. 351–356.
7. Chalikian T. V. How large are the volume changes accompanying protein transitions and binding ? / T. V. Chalikian, R. Filfil // *Biophysical Chemistry.* – 2003. – V. 104. – P. 489–499.
8. The native and the heat-induced denatured states of alpha-chymotrypsinogen A: thermodynamic and spectroscopic studies / T. V. Chalikian, J. Valker, D. Anafi, K. J. Breslauer // *J Mol Biol.* – 1998. – V. 274(2). – P. 237-252.
9. Dadarlat V. M. Insights into protein compressibility from molecular dynamics simulations / V. M. Dadarlat, K. B. Post // *J. Phys. Chem. B.* – 2001. – V. 105. – №3. – P. 715–724.
10. Lee B. Calculation of volume fluctuation for globular protein models / B. Lee // *Proc. Natl. Acad. Sci. USA.* – 1983. – V. 80. – P. 622–626.
11. Richards F. M. Areas, volumes, packing, and protein structure / F. M. Richards // *Annu. Rev. Biophys. Bioeng.* – 1977. – V. 6. – P. 151–176.
12. Conolly M. L. Analytical molecular surface calculation / M. L. Conolly // *J. Appl. Crystallog.* – 1983. – V. 16. – P. 548–558.
13. Paci E. Intrinsic compressibility and volume compression in solvated proteins by molecular dynamics simulation / E. Paci, M. Marchi. // *Proc. Natl. Acad. Sci. USA.* – 1996. – V. 93. – P. 11609–11614.
14. Ким А. В. Исследование структурных и термодинамических особенностей гидратной оболочки амфифильной молекулы C₈E₆ / А. В. Ким, Н. Н. Медведев, А. Гайгер // Структура и динамика молекулярных систем. – 2011. – V. 10 А. – С. 36–42. / Kim A. V. Issledovanie strukturnyh i termodinamicheskikh osobennostej gidratnoj obolochki amffil'noj molekuly S8E6 / A. V. Kim, N. N. Medvedev, A. Gajger // *Struktura i dinamika molekulyarnyh sistem.* – 2011. – V. 10 А. – S. 36–42./
15. Totrov M. Contour-buildup algorithm to calculate the analytical molecular surface of proteins / M. Totrov, R. Abagyan // *J. Struct. Biol.* – 1996. – V. 116. – P. 138–143.
16. Dynamic maintenance and visualization of molecular surfaces / C. Bajaj, V. Pascucci, A. Shamir [et al.] // *Discrete Applied Mathematics.* – 2003. – V. 127. – P. 23–51.
17. Zhao W. An Algebraic Spline Model of Molecular Surfaces for Energetic Computations / W. Zhao, Ch. Bajaj, G. Xu // *IEEE/ACM Trans Comput Biol Bioinform.* – 2011. – V. 8 (6). – P. 1458–1467.
18. Bates P. Minimal molecular surfaces and their applications / P. Bates, G. W. Wei, S. Zhao // *J. Comput. Chem.* – 2008. – 29. – P. 380–391.
19. Can T. Efficient molecular surface generation using level-set methods / T. Can, C.-I. Chen, Y.-F. Wang // *Journal of Molecular Graphics and Modelling.* – 2006. – V. 25. – P. 442–454.
20. Application of the level-set method to the implicit solvation of nonpolar molecules / L.-T. Cheng, J. Dzubiella, J. A. McCammon, B. Li // *J. Chem. Phys.* – 2007. – V. 127 (8). – P. 084503-1–084503-10.
21. Qing Pan. Model the Solvent-Excluded Surface of 3D Protein Molecular Structures Using Geometric PDE-Based Level-Set Method / Qing Pan, Xue-Cheng Tai // *Commun. Comput. Phys.* – 2009. – V. 6. – № 4. – P. 777–792.
22. Weiser J. Approximate Solvent-Accessible Surface Areas from Tetrahedrally Directed Neighbor Densities / J. Weiser, P. S. Shenkin, W. C. Still // *Biopolymers.* – 1999. – V. 50. – P. 373–380.
23. Weiser J. Fast, approximate algorithm for detection of solvent-inaccessible atoms / J. Weiser, P. S. Shenkin, W. C. Still // *J. Comput. Chem.* – 1999. – V. 20. – P. 586–596.