

生物分子模拟中的静电计算

彭 波, 李翰林, 卢本卓

(科学与工程计算国家重点实验室中国科学院 数学与系统科学研究院, 北京 100190)

摘 要: 在分子尺度上对生物系统的研究主要通过模拟生物分子间的相互作用来完成. 在众多的分子间相互作用中, 静电相互作用及其重要, 对它的定量计算是了解生物分子的溶剂化效应、生物分子折叠和酶催化作用等的一个中心课题. 本综述着重介绍用于计算溶剂化效应的泊松-玻尔兹曼模型, 包括模型的理论基础、发展历程、数值求解方法、及在分子生物学研究中的应用等. 同时, 也介绍泊松-玻尔兹曼模型的一种近似模型, 广义玻恩模型.

关键词: 泊松-玻尔兹曼方程; 广义玻恩模型; 溶剂化效应; 边界元; 有限元

中图分类号: O242; O244; O561.4

文献标志码: R

1 蛋白质静电相互作用的重要性及研究状况

1.1 蛋白质静电相互作用的重要性

在我们周围, 通过量子力学与统计力学原理, 电磁力维持着生物体的结构和驱动生物分子的化学反应. 在化学、生物化学和生物分子学等领域中, 静电作用力作为一种长程作用力, 在研究生物分子中的相互作用时起着重要的作用. 目前, 在生物、计算机科学、计算数学、物理与化学和生物分子学等交叉学科中, 生物分子的静电性质得到了广泛的研究. 很久以来, 科学家们一直采用理论方法分析大分子中的结构动力学, 为生物学实验提供重要的补充. 然而, 当试图用物理化学原理来解释蛋白质分子折叠途径、生物酶分子如何发挥高效的催化功能、生物分子间如何通过特异性相互作用进行识别等问题时, 试图对生物分子的序列、结构或功能进行定量预测或理性设计时, 却发现我们面临着巨大的困难与技术鸿沟. 生物大分子由成千上万的原子所组成, 其结构和动力学性质通过大量的范德华、疏水、氢键、盐键等物理化学意义上的弱相互作用来维系或支配. 这些弱相互作用在微观上缺乏对称性, 使得很多经典的理论分析工具在生物大分子研究中失去了用武之地. 最近计算机分子模拟技术的发展, 为填平这道鸿沟提供了很大的机会. 在多维核磁共振实验和 X 射线晶体衍射实验数据的结构修正中, 分子动力学模拟已经成为一种常规手段. 近年来, 人工蛋白序列的设计、人造酶的设计等与定向进化等实验技术的结合, 在生物分子工程领域取得了一系列影响深远的成果. 分子动力学模拟在生物大分子中的成功应用, 至少依赖于两个方面: ①模拟中能量模型能够准确地刻画大生物分子上的折叠、相互作用和动力学变化相关的分子间相互作用; ②分子动力学或蒙特卡洛等模拟技术能够对大生物分子的相空间进行合理地采样. 后者与目前计算机模拟的时间尺度密切相关, 现有计算机的能力仅限于模拟大生物分子在几十到数百纳秒量级的时间演化, 而相当多的生物学过程包括蛋白质折叠、酶催化功能的运动等在毫秒或更长的时间尺度上发生. 为了研究这些生物学过程, 急待发展新的模拟技术, 以适用于生物分子工程的计算机模拟、实验数据的理论解释、大生物分子的装配问题和蛋白质热力学的理论研究等^[1-3].

蛋白质是构成生物体的基础物质之一, 是构成细胞的基本有机物, 是生命活动的主要承担者. 蛋白质的重要性很早就被认识到了. 早在 1838 年, 当荷兰化学家 Mulder 提出蛋白质这个名称时, 他就明确指出: 在植物和动物中存在着这样一种物质, 毫无疑问它是生命体中已知的最重要的物质, 如果没有它, 在我们这个地球上生命则是不可能存在的. 蛋白质约占细胞干重的 50% 以上. 现在, 人们虽然还远未了解生命现象的全部, 但那些已经被人类所揭示的生命活动, 却无一不与蛋白质密切相关. 生物体中的每一个细胞和所有重要组成部分都有蛋白质参与. 蛋白质分子通过和其它生物分子发生相互作用而发挥其生物学功能. 实验研究表

明,蛋白质分子的生理功能是通过复杂多样的蛋白质—蛋白质界面而发生相互作用.这些复杂多变的界面对应了多姿多彩、各具特色的蛋白质相互作用,有着复杂的物理化学性质和结构特征.在分析生物分子的生理功能时,静电作用是一个很重要的因素,它支配着生物分子中粒子的移动和相变.事实上,大量的生命活动都是由蛋白质之间的相互作用来调节完成的.例如,转录起始因子之间的相互作用调控着生物体内基因的表达;磷酸化激酶通过与受体相互作用控制细胞间的信号传输;抗体蛋白通过与各种抗原相互作用保护机体不受伤害;各种构成细胞骨架的蛋白之间的相互作用使白细胞得以变形移动,对付入侵;肌动蛋白和肌球蛋白之间的相互作用使肌肉收缩,心脏跳动等.随着生物化学和分子生物学的发展,人们不断认识到诸如胚胎发育、生长、分化和繁殖等都是蛋白质参与作用并受到蛋白质的控制和影响,一种生物的性状实际上是由它们自身产生的各种不同蛋白质相互作用所表现出来的总的结果^[4].

1924年,Linderstrøm-Lang等人对蛋白质分子静电力学做了开创性的工作^[5].随后一些快速计算大生物分子之间的相互作用的方法陆续出现,包括蛋白质—蛋白质之间相互作用的建模等等.近年来,伴随着结构基因组研究的开展,越来越多的蛋白质结构被测定出来,其中包括了大量的蛋白质复合物.迄今为止,除去序列一致性90%以上的结构,蛋白质结构数据库(protein data bank, PDB)中收录的结构中包含了约101 046个可能的蛋白质复合物(生物单元内>2条肽链,2014年6月18日查询).同时,PDB里收录的蛋白质结构数量随时间而快速地增加.现在,计算机模拟、化学材料合成、生物分子结构的测试等技术的结合构成了后基因组时代新药开发的新策略.这些新的技术有望能成功地指导蛋白质之间相互作用的设计、生产出全新的生物调控元件;也推动了蛋白质分子之间对接技术的进步;加快蛋白质之间相互作用的新药开发,为人类疾病的治疗贡献力量.

静电相互作用在研究生物分子结构与稳定性、酶的催化活性、生物分子识别和反应速率以及粒子通道选择性等方面发挥了举足轻重的作用.事实上,生物大分子中的离子基团和极性基团,蛋白质分子中有超过20%的氨基酸在生理条件下成电离态;有超过25%的蛋白质分子其侧链成极性基团存在.绝大多数的生物膜也需要有少量的磷脂基团来维持它的生理功能.

1.1.1 蛋白质分子结构的稳定性

蛋白质多肽链在空间中的走向是决定蛋白质构象的基础.多肽链主链构象是指主链上的原子在空间上的排列,不涉及侧链基团的原子,表示了蛋白质的二级结构.一种蛋白质的多肽链主链实际上是由连续的刚性肽基构成的,或者可以看作是由重复的 $-N-C_{\alpha}-C-$ 单位借肽键连接而成的线性结构.

螺旋结构是蛋白质二级结构重要的结构要素.如果一条多肽链绕它的每个 C_{α} 原子以相同的角度旋转,将会形成某种螺旋结构. α -螺旋是一种既有允许的构象角,又有最能有效形成氢键的多肽主链构象,是多肽链的一种特殊的刚性排列.在 α -螺旋中,氢键处在最适距离(0.28nm),是一种强有力的氢键.由于主链上所有的羰基和酰胺基都参与了这种氢键的形成,所有 α -螺旋是一种稳定的结构.在肽键的共振结构中,羰基氧带部分负电荷,而酰胺氮具有部分正电荷,从而形成了一种电偶极.由于所有的肽键都具有这种电偶极,而且它们会通过螺旋的氢键而衔接,导致净偶极沿螺旋延伸,并随螺旋长度而加强,从而形成螺旋的偶极.由于螺旋末端的4个残基不能完全参与螺旋氢键的形成,螺旋偶极的正电荷和负电荷实际上分别属于接近螺旋末端的氨基和羧基.基于这样的理由,带负电荷的氨基酸残基往往会在螺旋接近氨基末端的部位找到,它们可与螺旋偶极的正电荷相互作用,起到稳定螺旋的作用.若带正电荷的氨基酸残基出现在同样的部位,则会致螺旋去稳定.同样地,相对应的情况也会出现在羧基末端.

β -结构是由多个平行排列在一起的 β -链或肽段形成的一种多肽链主链折叠结构.与 α -螺旋不同的是,在 β -折叠片中,氢键出现在相邻多肽链之间或相邻肽段之间,而不是像 α -螺旋一样出现在直向的同一条链中. β -折叠片是蛋白质中普遍存在的主链结构要素.

当蛋白质多肽链未折叠时,它处在最不稳定的高能量状态,而当它折叠成天然构象后,它则处于热力学最稳定的低能量水平状态.蛋白质多肽链在生理条件下折叠成特定的构象是热力学上一种有利的过程,这告诉我们,在折叠中,总的自由能变化必定是负的^[6-7].自由能的变化

$$\Delta G = \Delta H - T\Delta S, \quad (1)$$

式中, ΔH 是焓的变化,大致相当于总的键能和之差, ΔS 是熵的变化,是系统有序度的量度.对于某一给定的

转变来说,负的 ΔH 表明,系统从较弱的键合到较强的键合作用转变. 正的 ΔS 表明,系统从较有序到较无序的状态转变. 但蛋白质折叠过程涉及从无序到有序状态的转变,因此该过程必然涉及混乱度的降低,即熵的减少. 自由能变化公式表明, ΔS 的负值将会产生正的 ΔG ,构象熵的变化对折叠做功. 在研究蛋白质折叠时,我们不能忽视水的作用,因为在讨论蛋白质折叠所涉及的物理化学因素时,必须考虑多肽链与水的相互作用. 此外,水溶液中的其它因素,例如小分子的阴离子和阳离子,在决定蛋白质结构的稳定性方面也能起重要的作用. 实际上,一个蛋白质折叠所成的特定结构是各种作用力相互抵消、精巧平衡的结果. 这里,我们将讨论蛋白质折叠的过程以及各种作用力的性质和对 ΔG 的负值的贡献.

1.1.2 推动蛋白质特定构象形成和稳定的作用力

1) 疏水作用

在天然蛋白质结构中,由于非极性侧链基团具有避开水的倾向,大多数在分子内部聚集,不与水溶剂接触,形成胶束状的构象. 因此,非极性基团的疏水作用是蛋白质结构的重要决定因素^[9].

蛋白质折叠时负值 ΔG 的产生涉及多肽链和溶剂水的焓变与熵变

$$\Delta G = \Delta H_p + \Delta H_s - T\Delta S_p - T\Delta S_s, \quad (2)$$

疏水作用源于水作为溶剂的特殊性质. 表1的热力学数据为疏水作用的产生提供了重要依据,因为碳氢化合物从水中转移到非极性溶剂中与蛋白质折叠时非极性侧链从蛋白质外部转移到内部是相似的. 碳氢化合物从水溶液中转移非极性溶剂中的自由能变化(ΔG)表明,这样的转移过程是一个自发的过程. 脂肪族化合物的转移过程是吸热的,即 ΔH 为正值,但是(ΔG)是负值,表明这种转移过程是熵增加推动的. 同样,可以认为蛋白质的非极性基团从水溶液环境中转移到蛋白质非极性内部也是熵的增加推动的. 当非极性分子或非极性侧链被水溶剂化时,熵会减少. 因为,非极性物质与水接触引起水分子在它们周围形成像笼状的结构,这种有序化使该系统水的混乱度降低,并伴随熵的减少.

蛋白质通常含有相当数量的疏水侧链基团(如 Leu、Ile、Phe 等的侧链). 当多肽链处在未折叠状态时,这些疏水基团与水接触,极性水分子诱导非极性基团产生的偶极可以与水分子产生弱的静电相互作用,引起周围水的结构有序化,于是便产生了不利的自由能变化. 但是,疏水侧链基团具有避开水的倾向,当它们减少与水接触时,它们中的大多数就会从水中转移到蛋白质分子内部. 此时,它们所占据的空间的表面积小于它们各自单独占用空间表面积之和,导致非极性基团表面水的有序化程度总量有很大的降低. 任何系统的稳定都必须在能量变化上是一个负值,熵的增加是产生负的自由能变化的主要因素. 尽管非极性基团内聚给出的 $-T\Delta S_p$ 是一个较大正值,而溶剂水分子的混乱度增加给出的 $-T\Delta S_s$ 是一个很大的负值,两者抵消后得到总的 $-T\Delta S$ 是一个很显著的负值,这对折叠时自由能的变化将贡献较显著的负值. 蛋白质折叠过程中多肽链的焓变与溶剂的焓变相互抵消后,总焓变是很小的,对折叠不起支配作用. 从折叠时的极性基团的焓变与熵变来看,它们给出的 ΔG 是一个很小的负值,对折叠的贡献是很小的. 因此,非极性基团在蛋白质内部聚集所导致的溶剂熵增加对蛋白质的折叠和结构稳定的贡献是最大的.

根据对天然蛋白质空间构象中的单个侧链基团的定位详细分析,表明它们都是定位在预期会对天然结构的稳定性作出贡献的部位,非极性侧链基团被包埋在蛋白质结构的内部,而极性侧链残基则位于蛋白质结构的表面,并与环境水分子结合. 当极性基团,尤其是主链上的羰基和酰胺基不与水结合时,它们总是强烈地彼此结合,其结果往往产生连续的重复结构,如 α -螺旋和 β -折叠片.

表1 碳氢化合物从水转移到非极性溶剂中的热力学变化(25 °C)^[11]

Table 1 Thermodynamic changes in transferring hydrocarbon from water to nonpolar solvent

过程	$\Delta H/$ (kJ · mol ⁻¹)	$-T\Delta S/$ (kJ · mol ⁻¹)	$\Delta G/$ (kJ · mol ⁻¹)
CH ₄ 从水转移到苯中	11.7	-22.6	-10.9
CH ₄ 从水转移到四氯化碳中	10.5	-22.6	-12.1
C ₂ H ₆ 从水转移到苯中	9.2	-25.1	-15.9
C ₂ H ₄ 从水转移到苯中	6.7	-18.8	-12.1
C ₂ H ₂ 从水转移到苯中	0.8	-8.8	-8.0
水中的苯转变成液态苯(18 °C)	0	-17.2	-17.2
水中的甲苯转变成液态甲苯(18 °C)	0	-20.0	-20.0

2) 氢键

氢键(D—H—A)是一种由弱酸性的供体基团(D—H)和一个具有孤对电子的原子(A)之间形成的最显著的静电作用力. 在生物学系统中, D 和 A 两者都是高电负性的 N、O 以及 S 原子. 氢键的结合能大约是 ($-12 \sim -30$) $\text{kJ} \cdot \text{mol}^{-1}$; 与范德华力相比, 它有较强的方向性. D—A 间的距离正常情况下为(0.27 ~ 0.31) nm. 蛋白质具有众多的氢键供体和受体, 包括主链的羰基和酰胺基以及极性侧链基团^[10].

由于蛋白质分子内部的氢键结合基团几乎都有可能形成氢键, 因此氢键对蛋白质的结构有很大的影响. 但是, 当蛋白质处在未折叠状态时, 它的所有能形成氢键的基团都与水形成了氢键, 因此内部氢键不可能为天然蛋白质折叠提供比它处在未折叠时形成的氢键更多的自由能. 但蛋白质内部氢键处在一个疏水环境下, 几乎没有水分子与其竞争, 因而内部氢键比分子表面的或近表面的氢键更加稳定; 同时, 蛋白质内部的氢键也为它的天然折叠格局提供了结构基础. 如果蛋白质以阻止它的某些内部氢键形成的方式折叠, 那么就会损失一部分自由能, 其构象就会比所有氢键完全形成时的构象的稳定性低得多. 因此, 蛋白质氢键是蛋白质折叠和结构稳定的协同者或促进者. 在很多情况下, 多肽链的极性基团与水分子之间形成的氢键和静电作用力产生有利的焓(负值). 通常, 焓对系统获得有利的水化作用比熵更重要.

3) 静电相互作用

蛋白质分子的相反电荷基团的结合成为盐键、离子键或离子对. 当典型离子对(如谷氨酸残基的侧链羧基和赖氨酸残基的侧链氨基)的电荷中心处在介电常数为 4 的介质中被分开 0.4 nm 时的能量大约是 $86 \text{ kJ} \cdot \text{mol}^{-1}$. 但是, 两个被分开离子的溶剂化自由能(自由离子在水溶液中是高度溶剂化的)与它们处在非溶剂化(处在蛋白质分子内部)状态所形成的离子对的能量几乎是相等的. 因此, 尽管离子对形成的自由能比较高, 但它们对蛋白质天然结构的稳定性贡献不大. 从在蛋白质的疏水内部很少存在非溶剂化的离子对可以看出来. 然而, 由于水分子的偶极性质, 暴露在水溶剂的电荷离子与水分子的相互作用对蛋白质的结构起着稳定的作用.

在电中性分子之间的非共价结合称为范德华力或范德华相互作用, 产生于永久的或诱导的偶极之间的静电相互作用, 与那些非键合的邻近原子间的各种相互作用有关. 永久偶极的相互作用是蛋白质的重要结构因素. 因为蛋白质的许多基团, 例如肽主链上的羰基和酰胺基存在永久的偶极矩. 这些相互作用比离子对的电荷间的相互作用要弱得多. 但是, 在 α -螺旋和 β -折叠中, 多肽主链上的羰基和酰胺基都指向同一方向, 它们的相互作用是可以联合和相加的. 因此, 偶极与偶极的相互作用也显著地影响着蛋白质的折叠.

范德华力实际上包含静电吸引和静电排斥两种相反的作用力. 上面偶极间的相互作用是静电吸引. 但是, 当非键合原子间彼此太接近, 就会产生排斥力. 这种排斥作用是原子间电子云重叠的结果.

4) 二硫键

二硫键也是维持某些蛋白质空间结构稳定的重要因素, 它是由两个半胱氨酸的侧链—SH 氧化形成的. 但处于还原状态的细胞质会极大地降低细胞内二硫键的稳定性. 事实上, 几乎所有含有二硫键的蛋白质(如核糖核酸酶 A 和胰岛素)都分泌到氧化程度高的细胞外的特定部位, 在那里, 蛋白质中的二硫键才能有效地发挥稳定其结构的作用.

1.1.3 酶分子的催化反应

酶是生物细胞产生、以蛋白质为主要成分、能加快反应速度、且具有催化专一性的生物催化剂. 生物体内绝大多数反应都是由各种不同的酶催化的. 人们对蛋白质的研究, 大多数是以酶为对象的. 这是因为生命现象的本质是它能进行新陈代谢, 与自然界进行物质和能量交换, 而这些基本过程都是在酶的参与下完成的. 按照热力学的观点, 一个化学反应并不是从反应物直接向产物的方向进行的. 反应的发生, 即反应物向产物的转变, 取决于过渡态的能量状况. 只有当反应物在空间定向上有利于反应, 而且反应物达到过渡态所需的能量时, 反应才能发生. 反应物从基态转变成过渡态的能量差称为活化能(ΔG^\ddagger). 化学反应的速度常数 k 与活化能之间的关系根据转换态理论由阿伦尼乌斯方程(Arrhenius equation)给出^[11]

$$k = \frac{k_B T}{h} e^{-\Delta G^\ddagger / RT}, \quad (3)$$

这里 h 是普朗克常数(Planck's constant). 每一个化学反应都有 ΔG^\ddagger , 但活化能是可以改变的, 它取决于催化

剂是否存在和催化剂的类型. 过渡态或者活化态是一种短暂的分子结构状态,它既不是反应物,也不是产物,是反应途径中一种最不稳定、具有最高自由能的分子形式. 根据过渡态理论,催化剂的作用在于以某种途径降低反应活化能而加快化学反应速度. 典型的酶促反应速度与无催化剂存在时相比,反应速度要高出 $10^6 \sim 10^{12}$ 倍,也比非生物催化剂存在下的反应高出几个数量级. 从经典的观点来看,静电相互作用比较适合描述酶分子催化作用的几个方面. Warshel 等人在 1998 年提出静电相互作用降低反应活化能^[12].

酶的高效性和专一性是同一事物的两个方面,两者是统一的. 两个基本而又相互关联的原理为酶的作用机制提供了普遍性解释:①酶的大部分催化效力最终来自酶和它底物之间产生的多种弱作用力和相互作用所释放的自由能. 这种自由能既贡献于酶作用的专一性,又贡献于它的催化效力. ②在反应的过渡态中,使这样的弱相互作用处于最佳状态.

根据对底物和酶功能基团之间的化学反应研究,酶的催化功能基团(活性部位特定氨基酸的侧链基团、辅酶或者金属离子)能与底物发生瞬间的相互作用,使底物激活. 在很多情况下,这些功能基团通过提供一种较低能量反应途径降低活化能,从而使反应加速. 但催化功能基团不是酶催化作用唯一的贡献者. 降低活化能所需的能量通常来自底物和酶之间的弱的、非共价的相互作用. 酶和底物相互作用产生的自由能叫做结合能. 这种结合能的重要性已超出了单纯的酶-底物相互作用的范围,它是酶用来降低反应活化能所需自由能的主要来源. 只有正确的底物才能参加到与酶的完全的相互作用中去,才能使结合能达到最大化. 这种完全的互补的相互作用只有当底物处在它的过渡态时才能形成,因此,只有当酶与底物间的相互作用促使底物转变成过渡态时,最大结合能的释放才会出现. 最稳定的相互作用(最大的结合能)发生在酶和底物的过渡态之间,但同时过渡态也是最不稳定的反应中间态,它存在的时间太短,以致很快分解为底物或产物. 大多数相互作用涉及酶和它底物的非反应部位间的基团. 这表明,酶和它底物的非反应部位间的相互作用也起到了提供结合能的作用. 这种能量的补偿是使结合能达到最大的重要原因.

底物(S)的紧密结合加深了酶(E)-底物复合物的能量陷阱,实际上降低了反应速度. 升高 ES 复合物的能量水平将增高酶促反应速度. ES 复合物的能量水平是通过两种途径达到的:①由 S 和 E 的结合所造成的熵的减少;②ES 复合物由于张力、变形、去溶剂化以及其它的效应引起的去稳定作用. ES 复合物是高度组织化的统一体,而溶液中的 E 和 S 是处在无序、高熵状态. 底物进入到酶的活性部位引起反应基团及其它相关基团与底物一起进入较合适的位置. 这种从无序到相对有序的转变必然导致熵的减少. 此外,酶和底物两者在三维方向上的平移运动能和旋转能在酶同底物结合后将会有一定程度的损失. 去稳定化作用涉及底物分子的荷电基团同酶活性部位的结合所产生的去溶剂化作用. 在水中,荷电基团是很稳定的. 当底物的荷电基团从水中移入到酶的活性部位时,往往会导致它们具有某种程度的去溶剂化,因为酶的活性部位相对是疏水的,从而使得这些荷电基团变得不稳定而表现出较大的反应性.

1.1.4 生物分子识别

结构互补性是生物分子间识别的手段. 生命复杂而高度组织化的形式取决于生物分子彼此识别和相互作用的能力. 如果一种分子的结构与另一种分子的结构是互补的,如酶与它的专一性底物、激素与它的受体、抗原与抗体等,那么这种分子之间的相互作用就能准确地实现. 它们的这种关系如同锁和钥匙之间的关系. 结构互补性是生物分子识别的基本要素,为认识生物系统的功能特征提供了思路. 从大分子到细胞,所有生物系统都是经由基于结构互补性的特殊分子识别机制运转的:酶对它专一性底物的识别、DNA 两条互补链的识别、精子与卵子的结合. 所有这些过程都涉及分子间的结构互补性^[13-14].

由结构互补性所发生的生物分子间的识别是由前面介绍的非共价作用力介导的. 由于这些非共价作用力很弱,在生理条件下这些过程是很容易可逆的. 因此,生物分子的相互作用存在瞬息变化的潜力,生物分子刚性而静态的状态(这种状态会使细胞活性丧失)也因此不会形成. 所有经由结构互补性介导的生命过程都有着动态的相互作用,由互补分子间的特殊识别所发动,最终导致特有生理活性的产生. 生物分子的功能就是根据结构互补性和弱的化学相互作用的机制来实现的.

结构互补性原理可以向更高层次的相互作用延伸,这对于生命状态的确立是必不可少的. 例如,超分子复合物的形成是由于复合物中的大分子成员间的识别和相互作用所致,是由各成员间形成的弱的作用力决定的. 如果能形成有效数量的弱的非共价作用力,如同在结构上彼此互补的大分子那样,更大的结构将会自

动装配. 分子的非极性倾向能启动超分子装配体的形成. 复杂的亚细胞结构实际上也是某种装配过程而自发形成的, 这种装配过程也是通过结构互补性所积累的弱相互作用力推动的.

1.2 蛋白质静电相互作用的研究现状

人体是个复杂的生命机器, 每时每刻都在进行复杂的生理生化演变. 生物体内发生的反应, 主要发生在水溶液中, 紧密地调节着一些重要离子的浓度, 如钠离子, 氯离子, 钾离子等等, 这些溶剂分子强烈地影响着分子的结构和功能, 以及蛋白质的折叠稳定性和选择特异性等. 因此, 发展一种理论模型来研究这些生理现象的溶剂化效应势在必行, 也因此出现了大量的计算方法和算法来研究蛋白质之间的相互作用. 比较典型的有分子动力学模拟、布朗运动模拟、蛋白质分子 pKa 计算、蛋白质设计算法、蛋白质药物设计和蛋白质对接算法等等^[15-17]. 分子动力学方法, 需要考虑溶剂中巨大数量的水分子和电解质粒子, 很细致地刻画了溶剂化过程. 从计算的角度来看其模拟是个很漫长的过程. 实际上, 我们比较感兴趣的是这些过程的统计平均效应, 而隐式溶剂模型将溶剂分子的溶剂化效应隐式化处理, 大大地降低了模拟时间, 达到了很高的计算效率. 然而, 这种简化过程也意味着隐式化溶剂方法不是万能的, 在某些应用中可能会与实验结果发生较大的偏差, 因此在提高隐式化模型方面也出现了大量的优秀工作.

目前已经有各种不同的隐式溶剂化模型, 它们着重从计算精度和计算时间等方面来考虑. 这些模型一般是通过物理分析而得到的, 在一定程度上它们是一种合理的近似. 相比之下, 在计算物理中的其它领域, 如计算流体力学、计算电磁学等, 纳维叶-斯托克斯(Navier-Stokes)方程和麦克斯韦(Maxwell)方程是人们熟知的更加严格方程. 对于这些方程来说, 目前已经有了很多理论分析和实验结果来研究它们在计算精度与计算时间之间的关系. 它们的发展要比隐式溶剂化模型早很多, 数值理论分析相对来说也已经相当成熟. 因此, 给我们隐式溶剂模型带来了新的挑战, 当我们用隐式溶剂化模型来模拟溶液现象或提出一种新的模型时候, 需要着重考虑它的实用性与合理性.

计算代价几乎总是一个严格的限制因素. 这种限制作用不断地促进这些模型和数值算法之间的交互与回馈作用, 产生了一些新的模型和数值计算方法. 总的来说, 新模型或新算法的提出要比以前的方法在数值计算方面更快. 介电系数的连续化即是一种很好的改进方法, 在交界面厚度趋于零的极限情况下便可以恢复成原来的经典两介质模型. 据作者所知, 虽然一些研究者们对这种模型已经有了计算方法甚至是程序包, 然而这方面严格的理论分析工作少之又少, 仅仅是从数值结果方面对他们的模型进行分析, 恰恰也是这些数值结果显现出这些近似模型的强健性. 对于泊松理论, 人们可以引入非线性介质影响, 但不会失去计算可行性. 新的计算平台能有效地加快模拟过程, 特别是图形处理器(GPU). 在这些模型的求解过程中, 带动和引发了一些新的快速计算方法, 如快速多极子方法, 并行计算和多重网格等^[18-19].

的确, 最近也涌现出了大量的关于蛋白质溶液模拟的研究性文章与综述^[20-27,130], 它们或者是从计算数学的角度来探讨其算法的高效性, 或是从物理学的观点对模型进行修正, 扩宽其应用范围. 这里我们忽略掉了所有的具体细节过程, 但参考文献中完整的分析过程相信能给读者以收获.

2 蛋白质静电相互作用的计算方法与应用

2.1 Poisson-Boltzmann (PB) 方程

2.1.1 PB 方程的研究历史与现状

泊松-玻尔兹曼方程(Poisson-Boltzmann Equation)是用来计算电解质溶液中的静电相互作用的偏微分方程. 该方程的雏形最早出现于 Gouy-Chapman 模型中的双电层(Double Layer)理论^[28]. 在 Guoy-Chapman 模型中, 生物分子表面附近的离子用玻尔兹曼分布来描述. 泊松-玻尔兹曼方程将溶液中的水简化为均匀电介质, 这种隐式溶剂化(Implicit Solvent)的处理方法极大地提高了生物大分子溶液体系中的模拟和计算效率. 目前, 泊松-玻尔兹曼方程被广泛地应用于电解质溶液体系性质的计算和分子动力学模拟中, 特别是生物体系中各种生物大分子在溶液中的溶剂化效应(Solvation)的计算中.

泊松-玻尔兹曼方程是二阶非线性椭圆偏微分方程, 当离子溶液的电势能绝对值较小时, 可以把泊松-玻尔兹曼方程中的指数项展开到一阶, 得到线性的德拜-休克尔(Debye-Hückel)方程^[29]. 在稀溶液中, 德拜-休克尔方程对于泊松-玻尔兹曼方程是很好地近似. 除了在一些特定的简化体系中能求得它的解析解外, 一

般采用数值求解方法,例如有限差分、有限元或者边界元,以及一些其他随机方法,同时也涌现出了大量的求解泊松-玻尔兹曼方程的软件包. 这些软件包中,有的已经有了并行实现,真正地成为了分子尺度上模拟电解液的一种有效方法.

泊松-玻尔兹曼方程采用平均场近似,当溶液中出现一定浓度高价离子导致离子间相互作用和关联增强,泊松-玻尔兹曼方程的解将无法解释一些由关联所产生的现象,比如带相同电荷的物体在高价盐溶液中相互吸引,以及带电胶体在高价盐溶液中的电泳呈现电荷反转,这些现象必须考虑离子间的关联才能得到合理解释^[30-31]. 近年来,通过在泊松-玻尔兹曼方程中加入一些被平均场忽略掉的项来修正泊松-玻尔兹曼方程,获得了比较理想的结果^[32-33].

2.1.2 PB 方程的导出和适用范围

库仑定律 经典电磁学中的第一个定量定律,库仑定律,真空中两静止的点电荷之间的相互作用力,与它们电荷量的乘积成正比,距离的平方成反比,作用力的方向在它们的连线上. 其数学表达式

$$\mathbf{F} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r^2} \mathbf{r}_0, \quad (4)$$

其中 r 为两者之间的距离, \mathbf{r}_0 为 q_1 到 q_2 的方向矢径, $\epsilon_0 \sim 8.854\,187\,8 \times 10^{-12} \text{C}^2 \cdot (\text{J} \cdot \text{m})^{-1}$ 表示真空中的介电常数.

经典静电力学中一个最基本的问题是:给定空间中的电荷分布,如何计算空间中任意一点的静电势以及整个体系的静电自由能. 考虑一种比较简单的情形,在均匀电介质中分布着 N 个点电荷,它们在空间中的点 \mathbf{r} 处的静电势 $\phi(\mathbf{r})$ 可以根据库仑定律(4)得到

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon\epsilon_0} \sum_i \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|}, \quad (5)$$

其中 ϕ 为静电势, q_i 和 \mathbf{r}_i 分别表示点电荷的带电量 and 空间坐标 ($i = 1, \dots, N$), ϵ 为介质的相对介电系数.

同样地,根据库仑定律,整个体系的静电自由能可以通过下面的公式计算

$$\Delta G = \frac{1}{8\pi\epsilon\epsilon_0} \sum_{j \neq i} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|}, \quad (6)$$

式(6)在全原子模型中经常被用于计算体系的静电自由能.

泊松方程 由于生物分子-溶液系统(含有生物大分子的电解质溶液,见图1)往往含有巨大数量的带电粒子,计算该系统的静电相互作用是一件极其耗时的工作. 隐式溶剂方法是减少该计算耗时的一种有效方法,该方法将溶剂当作均匀电介质,隐式地表示溶剂中的电荷分布. 在隐式溶剂模型中,静电势 $\phi(\mathbf{r})$ 满足如下的泊松方程

$$-\nabla \cdot (\epsilon\epsilon_0(\mathbf{r}) \nabla \phi(\mathbf{r})) - \rho(\mathbf{r}) = 0, \quad (7)$$

其中 $\rho(\mathbf{r})$ 为电荷密度分布函数, $\epsilon(\mathbf{r})$ 为相对介电系数,该物理量用于衡量电介质的极化效应强度. 一般而言,介质的介电系数并不是常数,而与体系的电荷分布密度、电磁场、温度等物理量有着极为复杂的依赖关系. 在连续化模型中,通常将生物大分子和溶剂近似为均匀的电介质,而将其相对介电系数取做不同的常数,生物大分子的相对介电系数通常取值在 1 到 20 之间,而溶剂的相对介电系数通常取 80 左右^[34].

玻尔兹曼分布定律 为了从式(7)来计算生物分子-溶液系统的静电属性,需要定量地描述电荷密度分布函数 $\rho(\mathbf{r})$. 生物大分子中的带电粒子(原子)由于受到化学键的束缚,它们不容易受到周围静电场和溶液中带电粒子的影响,其电荷分布通常较为固定. 通常称这些电荷为固定电荷,其电荷密度分布函数 $\rho^f(\mathbf{r})$ 可以通过一系列 Dirac-Delta 函数的和来表示

$$\rho^f(\mathbf{r}) = \sum_{i=1}^M q_i \delta(\mathbf{r} - \mathbf{r}_i) = \sum_{i=1}^M z_i e_c \delta(\mathbf{r} - \mathbf{r}_i), \quad (8)$$

其中 $\delta(\cdot)$ 为 Dirac-Delta 函数, q_i 和 \mathbf{r}_i 分别表示各个固定电荷的带电量 and 空间坐标 ($i = 1, \dots, M$), z_i 表示分子中原子的电荷数,为无量纲参数,而 e_c 表示电子电量. 相比生物分子,溶液中的带电粒子 (Na^+ , Cl^- , K^+ , Mg^{2+} 等) 由于受到周围静电场影响以及粒子的碰撞作用,将会频繁地改变它们的位置,因此定量描述溶液中的电荷密度分布是非常困难的. 为了定量描述溶液中粒子的分布,物理学家们用概率的方法—玻尔兹曼分布

来描述生物分子附近带电粒子的分布情况:

$$c(\mathbf{r}) = c^{\text{bulk}} \exp\left(-\frac{q\phi(\mathbf{r})}{k_B T}\right), \quad (9)$$

其中 $c(\cdot)$ 和 $\phi(\cdot)$ 表示离子浓度和静电势, c^{bulk} 为宏观溶液的离子浓度, q 为离子的带电量, k_B 为玻尔兹曼常数 ($k_B \sim 1.38 \times 10^{-23} \text{J} \cdot \text{K}^{-1}$), T 为热力学温度.

因此,对于含有 K 种粒子的电解质溶液,其离子电荷密度分布函数 $\rho^{\text{ion}}(\mathbf{r})$ 可以表示为

$$\rho^{\text{ion}}(\mathbf{r}) = \sum_{i=1}^K q_i c_i(\mathbf{r}) = \sum_{i=1}^K q_i c_i^{\text{bulk}} \exp\left(-\frac{q_i \phi(\mathbf{r})}{k_B T}\right). \quad (10)$$

泊松-玻尔兹曼方程 将式(9)、(10)与式(7)相结合,就得到了描述生物分子-溶液系统静电相互作用的泊松-玻尔兹曼(PB)方程

$$-\nabla \cdot (\varepsilon \varepsilon_0(\mathbf{r}) \nabla \phi(\mathbf{r})) - \rho^f - \lambda \sum_{i=1}^K q_i c_i^{\text{bulk}} \exp\left(-\frac{q_i \phi(\mathbf{r})}{k_B T}\right) = 0, \quad (11)$$

其中 λ 在生物分子区域 Ω_p 取 0, 在水溶液 Ω_w 中取 1 见图 1.

特别地,在溶液中只有正负两种离子 ($\pm 1e_c$)、浓度比为 1:1 的时候,式(11)可以简化为

$$-\nabla \cdot (\varepsilon \varepsilon_0(\mathbf{r}) \nabla \phi(\mathbf{r})) = \rho^f - 2\lambda c^{\text{bulk}} e_c \sinh\left(\frac{e_c}{k_B T} \phi(\mathbf{r})\right). \quad (12)$$

为了方便计算,取 $u(\mathbf{r}) = e_c \beta \phi(\mathbf{r})$ 将式(12)无量纲化,这里 $\beta = 1/(k_B T)$, 得到无量纲形式的 PB 方程

$$-\nabla \cdot (\varepsilon(\mathbf{r}) \nabla u(\mathbf{r})) = \frac{e_c^2}{\varepsilon_0 k_B T} \sum_{i=1}^M z_i \delta(\mathbf{r} - \mathbf{r}_i) - \kappa^2 \sinh(u(\mathbf{r})). \quad (13)$$

这里 κ 在生物分子区域中取为 0, 而在溶剂中取作如下形式

$$\kappa^2 = \frac{2c^{\text{bulk}} e_c^2}{\varepsilon_0 k_B T} = \frac{2I e_c^2}{\varepsilon_0 k_B T}, \quad (14)$$

其中离子强度 $I = \frac{1}{2} \sum_{i=1}^K c^{\text{bulk}} z_i^2$, z_i 为第 i 种粒子的化合价. 溶

液中离子强度 I 对于蛋白质溶液中的静电作用有显著的影响,通过改变溶液中离子强度,可以增加或者减少溶液中电荷之间的静电吸引与排斥,因此为实验提供了一个途径来研究静电相互作用的重要性.

在 PB 方程中, ρ^{ion} 关于电势是非线性的,因此 PB 方程是非线性偏微分方程. 然而,在一定条件下可以对 PB 方程中的非线性项作线性近似, $\sinh[u(\mathbf{r})] \sim u(\mathbf{r})$, 得到下面的线性泊松-玻尔兹曼(LPB)方程

$$-\nabla \cdot (\varepsilon(\mathbf{r}) \nabla u(\mathbf{r})) = \frac{e_c^2}{\varepsilon_0 k_B T} \sum_{i=1}^M z_i \delta(\mathbf{r} - \mathbf{r}_i) - \kappa^2 u(\mathbf{r}). \quad (15)$$

PB 方程的边界条件 在一些连续模型中,通常将生物分子-溶液体系划分为三部分:溶质区域(生物大分子),溶剂区域,以及溶质与溶剂的隔离层. 隔离层由溶液构成,不同的是其中的离子受到一些生物分子的静电作用影响,无法自由移动(也可以看作是没有游离的粒子). 一个更为流行的是两介质模型(图 1),即将隔离层直接作为溶剂的一部分,整个体系由溶质区域与溶剂区域两部分构成. 本文仅考虑这种两介质模型. 在两个区域的交界面,介电系数急剧地跳跃.

根据经典静电力学,静电势函数 $\phi(\mathbf{r})$ 应满足下面两个条件

1) 无穷远条件,即

$$\lim_{r \rightarrow \infty} \phi = 0. \quad (16)$$

2) 它本身和电位移在法方向上的值 $\left(\varepsilon \frac{\partial \phi}{\partial n}\right)$ 在整个区域保持连续^[35]. 特别地,在生物分子与溶液的接触

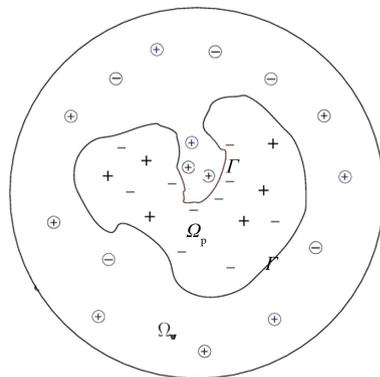


图 1 蛋白质分子-溶液模型(Ω_p 表示蛋白区域, Ω_w 表示溶液,他们相应的介电系数为 ε_p 和 ε_w .)

Fig. 1 A schematic view of a solvated protein molecule

(Ω_p denotes solute region, Ω_w denotes solvent region.

Corresponding dielectric constants are ε_p and ε_w , respectively.)

表面,满足下面的跳跃性条件:

$$\begin{aligned} [u] &= 0, \\ \left[\frac{\partial u}{\partial n} \right] &= 0. \end{aligned} \quad (17)$$

这里符号 $[\cdot]$ 表示从蛋白质分子到溶液界面上函数值的跳量.边界条件(16)和(17)一般用于边界元方法求解PB方程中,在其它的一些方法如有限差分 and 有限元计算中,通常需要附加Dirichlet边界条件来求解^[22].

接下来,我们描述在运用经典的静电学理论于蛋白质分子三维结构时的一些相关要素.实验研究表明,蛋白质分子中的电子向电负性大的原子集中的趋势导致这些原子的负电荷过剩,而那些失去电子的原子则将带部分正电荷,在这些原子对之间会形成永久性的电偶极子.在静电计算中,这些电偶极子是通过分子力场(实验得到或*ab initio*计算^[36])得到的部分电荷.蛋白质分子中的这些电偶极子可与其它偶极子、诱导偶极子及其周围水分子发生相互作用.这些相互作用非常复杂且蛋白质分子中的永久偶极子会发生位置的改变(也称重排),然而,其改变程度受限于蛋白质分子中的共价键(氢键或范德华作用)及空间位阻的影响,使得这些电偶极子在蛋白质分子侧链上排成特定的几何形状,在蛋白质分子骨架上形成固定的结构.因此,如果知道蛋白质分子的三维结构,我们可以对蛋白质分子中的原子和共价键赋予电荷分布,将它们用于PB方程.而配体和离子基团的引入导致的蛋白质结构微扰,可以用分子动力学模拟方法或者直接从蛋白质三维结构来估计偶极子的重排度.因此,如果蛋白质的三维结构表达合理,所有电极化效应都将隐式地通过蛋白质溶液的介电系数反映出来.

相比之下,由水分子形成的偶极子则有不同的行为,其行为依赖于它们在溶液中的位置.我们称远离蛋白质分子的水偶极子为溶剂水,比起那些在蛋白质分子界面上的水分子和蛋白质中的永久偶极子,它们的自由度大得多.因此,将溶剂水分子的复杂平移与旋转运动隐式化,在PB模型中将它当成一种具有较大介电系数的电介质来刻画.它的介电系数可以通过实验精确地测量,在0.1 MPa大气压25 °C下,其值为78.46^[37].而蛋白质分子界面上的水分子受到周围环境的静电场影响,它的重排则受限于蛋白质分子中原子形成的氢键网和界面上其它分子的作用.近年来,一些研究表明应将蛋白质表面上的水分子当作蛋白质分子的一部分^[38].蛋白质分子界面上的这些水分子,在蛋白质原子相互作用中发挥了桥梁的作用,在酶分子的催化作用和与配体的结合等方面起到了极其重要的作用^[39].虽然这些理论研究让我们知道它所发挥的重要作用,然而却也是极其有限的.为了简单化,在PB模型中,蛋白质分子界面上的水分子通常还是被当成溶剂水分子.蛋白质分子中的诱导偶极子(称为极化),来源于局部静电场使得分子中的原子核周围电子云位置发生偏移,可以用下面式子来计算它的偶极矩

$$(\mu_i)^{n+1} = \chi E(\mathbf{r}_i)^n, \quad (18)$$

式中 χ 是分子的极化常数, $E(\mathbf{r}_i)$ 是总的静电场强,包括了蛋白质分子中的部分电荷、溶液与蛋白质中的离子和蛋白质中其它原子电子云的重排等所产生的总电场强度.因此,所有诱导偶极子的偶极矩 μ_i 可以利用式(18)通过迭代的方法得到.在多粒子体系中,这种迭代过程(有时称为极化灾难),成为在原子尺度上精确计算蛋白质中的静电相互作用的一个主要障碍.为了避免这种极化灾难,可以在PB模型中引入阻尼参数.进一步,为了避免在方程(15)引入极化电荷,通常将蛋白质分子中的介电系数 ϵ_p 取为2以表示所有诱导偶极子和永久偶极子等其它因素的影响,而水分子产生的诱导偶极子则用介电系数 $\epsilon_w = 80$ 来代替.

上述边界条件(16),(17)与PB方程一起构成了一个封闭系统.对于给定的生物分子-溶液系统(生物大分子的结构、溶液的离子种类和离子浓度、介电系数等等)和外界条件(温度等等),PB方程的解存在且唯一.

对于PB方程,无法求出它的解析解,而只能进行数值求解.仅仅在某些特殊情况下,可以求出线性PB方程的格林函数,然后由格林表示定理得到它的解析解.下面简单地介绍几种特殊情况下的线性PB方程的解析解.

均匀带电球壳 考虑单价溶液中半径为 R 的带电球壳,其所带电荷总量为 q 且均匀分布在小球表面,小球的介电系数为 ϵ_p ,溶液的介电常数为 ϵ_w .以小球的中心为坐标原点,由于电荷均匀分布在球表面,因此球壳表面电荷面密度 $\sigma = q/(4\pi R^2)$.利用小球的对称性,通过球坐标变换,线性PB方程可以转化为如下的常

微分方程

$$\begin{aligned} -\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d}{dr} u_1 \right) &= 0, \quad r \in \Omega_p, \\ -\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d}{dr} u_2 \right) + \bar{\kappa}^2 u_2 &= 0, \quad r \in \Omega_w, \end{aligned} \quad (19)$$

这里 $\bar{\kappa}^2 = \kappa^2 / \varepsilon_w$, 而由于小球的面电荷密度为 σ , 因此当 $r = R$ 时,

$$u_1(r) = u_2(r), \quad \varepsilon_p \frac{du_1}{dr} - \varepsilon_w \frac{du_2}{dr} = \sigma = -\frac{q}{4\pi R^2}. \quad (20)$$

结合式(19), (20), 解得

$$\begin{aligned} u(r) &= \frac{q}{4\pi\varepsilon_w R} \frac{1}{1 + \bar{\kappa}R}, \quad r \in \Omega_p, \\ u(r) &= \frac{q}{4\pi\varepsilon_w R} \frac{e^{-\bar{\kappa}(R-r)}}{(1 + \bar{\kappa}R)r}, \quad r \in \Omega_w. \end{aligned} \quad (21)$$

带电小球理想情况下, 视分子为点电荷, 这时边界条件将变为连续条件

$$u_1(R) = u_2(R), \quad \varepsilon_p \frac{du_1}{dr} = \varepsilon_w \frac{du_2}{dr}. \quad (22)$$

解析解可以按照上面的方法来进行求解. 这里采用另外一种方法: 通过方程(29)来求解. 由于对称性, 我们可以将积分号下的未知函数提到积分号前, 直接计算可得

$$\begin{aligned} \oint_r G_{pt} dS_t &= R \oint_r \frac{\partial G_{pt}}{\partial n} dS_t = -\frac{1}{2}, \\ \oint_r F_{pt} dS_t &= \frac{1}{2\bar{\kappa}} (1 - \exp(-2\bar{\kappa}R)), \\ \oint_r \frac{\partial F_{pt}}{\partial n} dS_t &= \frac{1}{2\bar{\kappa}R} (1 + \bar{\kappa}R) \exp(-2\bar{\kappa}R). \end{aligned} \quad (23)$$

经过简单的代数运算, 得到整个系统的静电势

$$\begin{aligned} u(r) &= \frac{q}{4\pi\varepsilon_p r} - \frac{q}{4\pi\varepsilon_p R} + \frac{q}{4\pi\varepsilon_w R(1 + \bar{\kappa}R)}, \quad r < R, \\ u(r) &= \frac{q}{4\pi\varepsilon_w (1 + \bar{\kappa}R)} \frac{\exp(-\bar{\kappa}(r - R))}{r}, \quad r > R. \end{aligned} \quad (24)$$

溶剂完全渗透 考虑一个细长的杆状分子放入到一个完全渗透的离子溶液中, 在这种情况下整个溶液里面的电势可以用下式来描述,

$$-\nabla^2 u(\mathbf{r}) + \bar{\kappa}^2 u(\mathbf{r}) = \frac{1}{\varepsilon_w} \sum_{i=1}^N q_i \delta(\mathbf{r} - \mathbf{r}_i).$$

由于三维自由空间中的 Helmholtz 方程 $-\nabla^2 G(\mathbf{r}) + \bar{\kappa}^2 G(\mathbf{r}) = \delta(\mathbf{r} - \mathbf{r}_i)$ 的格林函数为

$$G_1(\mathbf{r}, \mathbf{r}_i) = \frac{1}{4\pi} \frac{e^{\bar{\kappa}|\mathbf{r} - \mathbf{r}_i|}}{|\mathbf{r} - \mathbf{r}_i|}, \quad G_2(\mathbf{r}, \mathbf{r}_i) = \frac{1}{4\pi} \frac{e^{-\bar{\kappa}|\mathbf{r} - \mathbf{r}_i|}}{|\mathbf{r} - \mathbf{r}_i|}.$$

结合无穷远处的边界条件, 因此单个粒子情况下的解

$$u^{(i)}(\mathbf{r}) = \frac{q_i}{4\pi\varepsilon_w} \frac{e^{-\bar{\kappa}|\mathbf{r} - \mathbf{r}_i|}}{|\mathbf{r} - \mathbf{r}_i|}. \quad (25)$$

根据线性叠加原理得到细长杆状分子溶液的静电势

$$u(\mathbf{r}) = \sum_{i=1}^N u^{(i)}(\mathbf{r}) = \sum_{i=1}^N \frac{q_i}{4\pi\varepsilon_w} \frac{e^{-\bar{\kappa}|\mathbf{r} - \mathbf{r}_i|}}{|\mathbf{r} - \mathbf{r}_i|}. \quad (26)$$

以上仅列举了三种简单情形, 更多复杂情况 LPB 方程解析解的研究可以参考文献[40].

2.1.3 PB 方程的求解方法

通常情况下很难求出 PB 方程的解析解,而只能通过数值方法求其近似解.这里将系统地介绍 PB 方程的一些常用数值解法及其最新进展,包括有限差分方法、有限元方法和边界元方法等等.

2.1.3.1 有限差分方法

有限差分方法基本步骤是首先方程的求解区域进行剖分,然后对方程中的导数项用网格点上的函数值来近似,得到关于网格点上未知函数的线性代数方程组.对于一些区域比较规则的偏微分方程,可以利用傅里叶方法来研究它的收敛性及稳定性.

应用有限差分方法求解 PB 方程面临的首要问题是计算精度问题.由于 PB 方程中含有带奇异性的 Dirac-Delta 函数,在进行网格剖分和对方程进行离散时极易产生误差,这些误差会对差分格式的收敛性和稳定性产生很大影响.此外,由于介电系数在生物分子表面的间断性,以及连续性条件的存在,如何定义分子表面和如何进行区域剖分也会对数值解的精度产生较大的影响.

直到 20 世纪 90 年代,Dirac-Delta 函数引发的离散格式精度问题得以解决,一些基于有限差分方法的解法器被陆续提出,包括 DelPhi^[41], APBS^[42], GRASP^[43], MEAD^[44], UHBD^[45], PBSA^[46], ZAP^[47] 和 CHARMM^[48] 等等.

近年来,结构化网格上椭圆界面问题的数值方法的快速发展,使得如何处理 PB 方程的界面连续条件的问题在一定程度上得到解决. Wang 等人通过在界面两侧对方程进行局部泰勒展开对界面附近的差分格式进行修正,得到一种新的差分格式^[49-50].这种格式的优点在于满足极值原理,因此离散得到的刚度矩阵对称正定,可以采用预条件共轭梯度、多重网格等方法进行快速求解.但需要一个从不规则分子区域到参考网格的贴体网格映射,因此目前该方法只适用于一些表面充分光滑的简单生物分子. MIB (Matched Interface and Boundary) 方法利用界面条件对区域进行光滑延拓,然后利用中心差分格式,通过反复利用界面条件得到了高阶精度的差分格式^[51]. MIB 方法的优点是具有高阶的精度和收敛速度,数值试验显示这类方法具有二阶收敛速度^[52-53],而如果将电势进行正则分解^[49]以消去 PB 方程中 Dirac-Delta 函数引发的奇异性, MIB 方法的精度和收敛速度可以达到更高^[54]. MIB 方法的缺点在于将 PB 方程离散所得的刚度矩阵是不定的和反对称的,这导致包括共轭梯度算法在内的一些常用方法很难发挥功效.此外, IIM (Immersed Interface Method) 方法也是一种有效的差分方法,该方法在远离蛋白质分子界面的网格点上采用标准的差分格式,而界面附件则通过界面连续性条件来修正其差分格式^[55]. IIM 方法的优点在于具有较高的精度^[56],并且可以用于求解非线性 PB 方程^[57].另外有一种值得借鉴的处理界面问题的方法是求解稳态抛物方程快速时间演化方法,该方法通过将界面上不连续的系数光滑化,使得扩散作用和原 PB 中的变系数作用可以分解成新的扩散作用及常系数和一个光滑的对流作用^[58].

2.1.3.2 有限元方法

有限元方法是一种基于变分原理的微分方程数值求解方法,被广泛应用于拉普拉斯方程和泊松方程的求解中.有限元方法的求解通常包含三个步骤:单元剖分、有限元离散和求解离散变分问题.

比较早期的有限元求解生物分子 PB 方程有 Holst 和 Harvey 等人的工作^[59-60], Bowen, Cortis, Friesner, Baker, Holst 和 Lu 等人之后做了进一步的工作^[61-68].由于 PB 方程中广义函数 (Dirac-Delta 函数) 的奇异性, Chen 等人详细地讨论了 PB 方程 Galerkin 有限元方法解的先验估计与收敛性分析理论^[69]. Chern 等人提出了解的正则格式分离方法,将解的正则部分和奇异部分分开处理^[49],这一分解方法已被应用到之后 Lu 和 Zhou 等人有限元方法中^[22,70].利用有限元方法结合相应的界面条件求解正则部分,它的计算精度有了本质上的改进. Cortis 等人利用伽辽金方法改进了有限元方法,但他们没能处理好非线性 PB 方程和分子界面上的条件^[63], Holst 和 Baker 等人用自适应有限元方法,采用 Newton-AMG 方法对离散系统进行求解^[65-66,71]. Xie 等人提出了一种新的有限元离散格式,很好地处理了界面条件^[72],最近他们又提出了对方程进行正则化处理的新格式^[73],得到了较高的计算精度. Shestakov 等人采用 pseudo-transient continuation 方法结合 Newton-Krylov 子空间迭代方法求解 PB 方程,然而这种方法目前尚未被用于具有复杂界面的生物分子系统的计算中,而仅仅被用于具有简单几何结构的胶粒的计算^[74].

2.1.3.3 有限差分法和有限元法中的快速方法

应用有限差分方法求解 PB 方程还面临求解速度问题,生物大分子的复杂性和巨大规模对求解器的求

解速度提出了更高的要求. 近年来, 在发展 PB 方程快速求解器方面也有很多开创性的进展. 比较常用的有 AMG、Inexact-Newton-Algebraic MultiGrid 和牛顿方法^[59,71,75]. 由于牛顿法求解非线性问题时的迭代次数不依赖于网格尺寸, 牛顿法求解 PB 方程的时间复杂度实际上等于迭代过程中求解线性系统的时间复杂度^[76]. 因此, 牛顿法求解 PB 方程的空间复杂度和时间复杂度是线性的^[59,71,75]. 从理论上讲, 将牛顿法与 AMG 方法结合可以达到最优的时间复杂度与空间复杂度, 然而这类方法非常难以实现. 一些比较容易实现的算法如预条件共轭梯度法 (PCG)、经典的静迭代法 (SOR, Gauss-Seidel, Jacobi)、MICCG (Modified Incomplete Cholesky Factorization Based PCG) 方法等等, 可以在 $O(N^{1.25})$ 到 $O(N^2)$ 时间内获得方程组的解, 仅有少数求解器可以达到线性时间复杂度^[77].

应用有限差分方法求解 PB 方程的另一个问题是计算规模问题. 对于 PB 方程而言, 计算区域的网格单元尺寸一般取 0.2 \AA 到 1 \AA 之间, 对于一些尺寸达到上千 \AA 的生物大分子, PB 方程的自由度很容易就达到 $O(1\ 000^3)$. 解决该方法包括自适应网格加密方法和 Telescoping Technique 等等. 自适应网格加密方法首先在粗网格上进行计算, 然后在迭代过程中对解变化较大的区域进行局部加密. Telescoping Technique 首先在粗网格上得到一个近似解, 通过对粗网格上的结果进行插值方法得到细网格上的边界条件, 然后在局部用细网格进行求解^[77]. 更多这方面的工作可以参考文献^[78].

在自适应有限元方面, Holst 等人提出了一种求解 PB 方程的自适应有限元方法^[65-66,69], 该方法采用 Newton-AMG 方法或者 Inexact-Newton-AMG 方法对离散有限元系统进行求解^[59,71,76], 具有线性的时间复杂度和线性的空间复杂度^[79]. 最近, Olson 等人提出了一种求解 PB 方程的混合自适应有限元方法 (Weighted Adaptive Least-Squares FEM, the First-Order System Least-Square FEM, FOSLS FEM)^[80-81], 该方法采用传统伽辽金方法和混合型伽辽金方法 (Mixed Galerkin FEM) 对 PB 方程进行有限元离散, 通过引入残量泛函并极小化该泛函来得到后验误差估计并进行自适应网格加密. 这种有限元方法的优点在于可以使得 PB 方程的解和解的梯度具有较高精度, 此外该方法具有较高的收敛速度.

此外, Lu 等人利用基于 PNP (Poisson-Nernst-Planck) 方程的时间演化方程获得 PB 的结果^[82-83]:

$$\begin{aligned} \frac{\partial c_i(\mathbf{r}, t)}{\partial t} &= \nabla \cdot (D_i(\mathbf{r}) \nabla c_i(\mathbf{r}, t) + \beta D_i(\mathbf{r}) c_i(\mathbf{r}, t) \nabla \phi(\mathbf{r}, t)), \quad \mathbf{r} \in \Omega_s, \quad i = 1, \dots, K; \\ \nabla \cdot (\varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r}, t)) &= -\rho^f(\mathbf{r}) - \sum_i q_i c_i(\mathbf{r}, t), \quad \mathbf{r} \in \Omega, \quad i = 1, \dots, K, \end{aligned} \quad (27)$$

其中 c_i 和 $D_i(\mathbf{r})$ 分别表示第 i 种离子的浓度和扩散系数, K 为溶液中离子的种类, $\beta = 1/(k_B T)$. 求解 PNP 方程的方法可以参看文献^[70,83-84]. PB 方程可以看做 PNP 方程的平衡态 (即处处流等于零) 解.

$$\begin{aligned} \nabla \cdot (D_i(\mathbf{r}) \nabla c_i(\mathbf{r}, t) + \beta D_i(\mathbf{r}) c_i(\mathbf{r}, t) \nabla \phi(\mathbf{r}, t)) &= 0, \quad \mathbf{r} \in \Omega_s, \quad i = 1, \dots, k; \\ \nabla \cdot (\varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r}, t)) - \rho^f(\mathbf{r}) - \sum_i q_i c_i(\mathbf{r}, t) &= 0, \quad \mathbf{r} \in \Omega, \quad i = 1, \dots, k. \end{aligned} \quad (28)$$

另外值得一提的是, 最近随机方法也广泛应用于一些偏微分方程的边值问题中. 由于统计采样方法固有的特性, 随机方法不要求求解区域的性质, 可以处理任意复杂的几何区域. 因此, 现在蒙特卡洛随机方法也是受欢迎的一类求解 PB 方程的方法. 这种随机方法利用拉普拉斯方程的随机行走方法和积分中值公式结合 PB 方程的边界条件, 根据轨迹的存活概率模拟出一系列的轨迹以获得 PB 方程在某些点上的解, 它显著的优点在于其消耗的计算机内存非常小^[85-88].

2.1.3.4 生物分子网格生成方法

在结构生物学和结构生物信息学中, 分子表面网格广泛应用于可视化和几何分析计算. 随着生物系统的数学建模和数值模拟的发展, 对分子表面网格有了一些新的需求, 例如合格性、稳定性、有效性等等, 特别是对于上面提到的隐式溶剂化模型. 在计算模拟中, 需要频繁产生网格或为大生物分子体系产生网格时, 高效是十分必要的. 它要求网格产生方法是稳定的, 且在计算机许可条件下, 能为任意尺度大小的分子系统产生网格. 网格质量涉及网格光滑 (避免锋利固体角等) 和拓扑正确性 (避免孤立的顶点、相交单元等), 其对于数值方法在求解偏微分方程的时候是至关重要的, 能影响数值算法的收敛性及结果的合理性. 网格剖分技术已经有几十年的发展历史了. 到目前为止, 结构化网格技术发展得相对比较成熟, 而非结构化网格技术由于起步较晚, 实现比较困难等方面的原因, 现在正在处于逐渐走向成熟的阶段.

贴体网格与非贴体网格 在基于有限差分的 PB 求解器里,需要对求解域进行网格剖分,通常采用结构化的网格来加速有限差分法的计算.而由于生物分子固有的特性:复杂外形,产生生物分子的贴体网格 (body-fitted mesh) 是及其复杂的过程,因此,一般采用结构化的非贴体网格 (body-unfitted mesh) 来离散求解域.目前结构化网格生成技术已有大量的文献.一般说来,它生成速度快、网格质量好、数据结构简单.然而它们给 PB 方程的求解在计算上带来难以估计的人为误差.为了降低非贴体网格带来的人为误差,目前另一种有效的方法是采用非结构化的贴体网格.非结构化网格没有规则的拓扑结构,网格节点的分布是随意的,因此具有很强的灵活性.在 20 世纪 90 年代时,非结构化网格的文献达到了它的高峰时期.随着人们对求解区域的复杂性的不断提高,对非结构化网格生成技术的要求越来越高.从现在的文献调查的情况来看,非结构化网格生成技术中只有平面三角形的自动生成技术比较成熟(边界的恢复问题仍然是一个难题,现在正在广泛讨论),平面四边形网格的生成技术正在走向成熟.而空间任意曲面的三角形、四边形网格的生成,三维任意几何形状实体的四面体网格和六面体网格的生成技术还远远没有达到成熟,需要解决的问题还非常多.非结构化网格技术,一般有 Grid Generation Technology(应用于差分法)和 Mesh Generation Technology(应用于有限元和边界元).对平面三角形网格生成方法,比较成熟的是基于 Delaunay 准则的一类网格剖分方法 (Bowyer-Watson Algorithm and Watson's Algorithm) 和波前法 (Advancing Front Triangulation) 的网格生成方法.另外还有一种基于梯度网格尺寸的三角形网格生成方法,这一方法现在还在发展当中.基于 Delannay 准则的网格生成方法的优点是速度快,网格的尺寸比较容易控制.缺点是对边界的恢复比较困难,很可能造成网格生成的失败,对这个问题的解决方法现在正在讨论之中.波前法 (Advancing Front Triangulation) 的优点是对区域边界拟合的比较好,所以在流体力学等对区域边界要求比较高的情况下,常常采用这种方法.它的缺点是对区域内部的网格生成的质量比较差,生成的速度比较慢.

三维实体的四面体和六面体网格生成方法现在还远远没有达到成熟.部分四面体网格生成器虽然已经达到了使用的阶段,但是对任意几何体的剖分仍然没有解决,现在的解决方法就是采用分区处理的办法,将复杂的几何区域划分为若干个简单的几何区域然后分别剖分再合成.对凹区的处理更是如此.六面体的网格生成技术主要采用的是间接方法,即由四面体网格剖分作为基础,然后生成六面体.这种方法生成的速度比较快,但是生成的网格很难达到完全的六面体,会剩下部分的四面体,四面体和六面体之间需要金字塔形的网格来连接.现在还没有看到比较成熟的直接生成六面体的网格生成方法.

贴体表面网格的生成 为了描述分子结构的形状,人们提出了各种分子表面定义,包括范德瓦尔斯表面 (van der Waals surface), 溶剂可接触表面 (solvent-accessible surface)、溶剂排斥表面 (solvent-excluded surface), 分子皮肤表面 (molecular skin surface), 最小分子表面 (minimal molecular surface) 和高斯表面 (Gaussian surface) 等.范德瓦尔斯表面被定义为球形原子的结合体的暴露表面,这些球的半径为分子中每个原子的范德瓦尔斯半径.溶剂可接触表面和溶剂排斥表面分别被定义为探针在范德瓦尔斯表面滚动时轨迹的中心和内边界.最小分子表面被定义为自由能最小化得到的表面.不同于这些定义,高斯表面被定义为一些高斯核函数和的水平集.通过选取合适的参数,高斯表面能很好地逼近范德瓦尔斯表面,溶剂可接触表面和溶剂排斥表面^[89-90].高斯表面被广泛应用于计算生物学的许多问题,例如对接问题、分子形状对比、计算溶剂可接触表面的面积以及广义玻恩模型中.随着各种分子表面定义的提出,许多致力于分子表面计算的工作出现了.1983年,Connolly等人提出了一系列算法来计算和分析了溶剂可接触表面^[91].1997年,Vorobjev等人提出了SIMS算法,该方法计算平稳不变分子点表面,能去掉自相交的部分和光滑溶剂排斥表面的奇异区域^[92].Sanner等人开发了一个基于 α 形的工具MSMS,为溶剂排斥表面产生三角形网格^[93].Ryua等人推广了 α 形方法,提出了 β 形的方法来产生三角形网格^[94].最近,Yu等人使用marching cube method为生物分子结构产生网格^[95].Can等人利用水平集的方法开发了LSMS用于生成溶剂排斥表面^[96].Cheng等人使用球的约束集合为分子皮肤表面生成网格^[97].此外,在计算结构生物学或结构生物信息学中,大多数的这些方法,如GRASP,MSMS,LSMS主要用于分子可视化和几何分析.对于一些数值模拟方法而言,例如有限元或边界元方法,通常需要采取一些其它方法来提高网格质量^[22].最近,Chen与Lu等人用表面trace的方法开发的高斯表面生成程序包TMSmesh,成功地为一百多万原子的大生物分子产生表面网格并用于BEM计算^[98-99].

体网格的生成 一般来说,一个产生体网格的合理策略可分为以下两步:首先产生一个分子表面的一致网格,然后基于表面网格去生成体网格. Lu 研究组发展了这样一个产生体网格的工具链^[22,67,100]. 该工具链主要包括三部分:表面网格的生成,网格质量的提高,体网格的产生. 首先,用 TMSmesh 为高斯表面进行三角行网格剖分,它能避免自相交,保持分子表面的拓扑结构^[101]. 在第二步,首先使用软件 ISO2mesh^[102]去简化表面网格,如果存在自相交的三角形,再使用 TransforMesh^[103]去掉那些自相交的三角形. 最后,使用 TetGen^[104]为上面得到的表面网格产生体网格.

2.1.3.5 线性 PB 方程的边界元求解方法

边界元方法是一种求解线性偏微分方程的数值方法,其基本思想是:将原方程转化为边界积分方程,再结合边界条件、界面条件等解出未知函数在边界上的值,从而得到它的整体解.

对于线性 PB 方程,通过格林第二公式可以转换成下面的边界积分方程^[105]

$$\begin{aligned}\alpha_p f_p &= \oint_{\Gamma} [\varepsilon G_{pt} h_t - \frac{\partial G_{pt}}{\partial \mathbf{n}} f_t] dS_t + \frac{1}{\varepsilon_w} \sum_k q_k G_{pk}, \quad p \in \Gamma, \\ (1 - \alpha_p) f_p &= \oint_{\Gamma} [-F_{pt} h_t + \frac{\partial F_{pt}}{\partial \mathbf{n}} f_t] dS_t, \quad p \in \Gamma,\end{aligned}\quad (29)$$

其中 $f = u^{\text{ext}}$, $h = \frac{\partial u^{\text{ext}}}{\partial \mathbf{n}}$ 为未知量, \mathbf{n} 是表面上点 t 的单位外法向量, $\varepsilon = \varepsilon_p / \varepsilon_w$, $G_{pt} = 1 / (4\pi r_{pt})$, $F_{pt} = \exp(-\kappa r_{pt}) / (4\pi r_{pt})$, r_{pt} 表示 p, t 两点之间距离. 参数 α_p 依赖于点 p 的几何特征,一般的可以取为 $A_p / (4\pi)$, A_p 为点 p 所张成的立体角. 对于光滑的几何表面, $\alpha_p = 1/2$. 在分子表面足够光滑时,对公式(2) - (29)和它的导数线性组合,便得到下面的 dBIEs 公式^[106]

$$\begin{aligned}\left(\frac{1}{2\varepsilon} + \frac{1}{2}\right) f_p &= \oint_{\Gamma} \left[(G_{pt} - F_{pt}) h_t - \left(\frac{1}{\varepsilon} \frac{\partial G_{pt}}{\partial \mathbf{n}} - \frac{\partial F_{pt}}{\partial \mathbf{n}} \right) f_t \right] dS + \frac{1}{\varepsilon_w} \sum_k q_k G_{pk}, \\ \left(\frac{1}{2} + \frac{1}{2\varepsilon}\right) h_p &= \oint_{\Gamma} \left[\left(\frac{\partial G_{pt}}{\partial \mathbf{n}_0} - \frac{1}{\varepsilon} \frac{\partial F_{pt}}{\partial \mathbf{n}_0} \right) h_t - \frac{1}{\varepsilon} \left(\frac{\partial^2 G_{pt}}{\partial \mathbf{n}_0 \partial \mathbf{n}} - \frac{\partial^2 F_{pt}}{\partial \mathbf{n}_0 \partial \mathbf{n}} \right) f_t \right] dS + \frac{1}{\varepsilon_w} \sum_k q_k \frac{\partial G_{pk}}{\partial \mathbf{n}_0}.\end{aligned}\quad (30)$$

对于非线性 PB 方程,解的积分表示需要用到它的体积分,降低了 BEMs 方法的效率. 文献[107]较详细的讨论了边界元方法与有限差分方法在求解 PB 方程时候的性能,认为边界元方法即使在应用了快速多极子方法加速积分,其求解速度仍然要慢于有限差分. 如果将(29)看作是直接边界元法,文献[108]则提出了一种间接边界元法:将分子表面电势求解转化为分子表面上电荷分布的求解.

通常可以采用伽辽金方法或者配点法对边界积分方程进行离散,对于 PB 方程,后者成为一种常用的选择. 一般来说,采用边界元方法求解 PB 方程的优点在于:①相比体离散方法,边界元方法降低了一维,从而减少了方程自由度;②边界元方法可以很好地处理复杂的几何形状和 PB 方程的边界条件;③采用边界元方法对 PB 方程进行离散可以得到一个条件数较好的线性方程组(见式(30)),因而无需采用预条件子求解;④它能很好地处理 PB 方程中的奇异项. 而它的缺点在于:①相比体离散方法来说,尽管边界元法离散的自由度 N 有所减少,但它所得到的刚度矩阵通常是稠密的,因此边界元方法通常具有极高的空间复杂度和时间复杂度(分别为 $O(N^2)$ 和 $O(N^3)$),这极大限制其在生物分子模拟中的应用,需要有效地降低边界元方法在应用于 PB 方程时的计算复杂度. ②由于无法得到非线性 PB 方程的格林函数,边界元方法仅适用于求解线性 PB 方程. ③计算边界单元上的积分较为耗时;④奇异积分的计算会影响解的精度与稳定性. 此外,对于一些生物分子系统中出现的溶剂空腔和隔离层,边界元法需要额外的计算量.

在最近几十年间,科学家在改进方法上和优化算法方面做了很多工作,主要在加快求解速度方面. 文献[105]通过采用迭代法求解线性方程组、不完全计算矩阵系数,大大地节省了内存与 CPU 时间. Krylov 子空间迭代方法在求解线性代数方程组解是一个很好的选择, Lu 等人实现了多种 Krylov 子空间迭代求解方法^[109]. 在采用 Krylov 子空间迭代求解方程(30)的时候需要反复计算右端的积分,它在离散情况下即为矩阵乘向量运算,在整个求解过程中是最耗时的. 目前的加速计算它的一些方法有快速多极子方法(FMM)^[110], 快速傅里叶变换(FFT)^[111-112], 面板聚类方法(Panel Clustering Method)^[113], 小波压缩方法(Wavelet Compression Method)^[114]等等,在边界元求解 PB 方程中前两种方法较为常用.

快速多极子方法的基本思想是采用八叉树结构进行空间管理,从而达到减少计算量和加快计算速度的目的,早期的快速多极子方法用于计算库仑作用,计算复杂度为 $O(N \log N)$,因此应用它可以加速泊松方程的求解^[115-116]. Bordner 和 Huber 等人采用自适应多极子展开的方法,将它应用到边界元方法求解泊松方程中^[117]. 随后, Boschitsch 等人将快速多极子应用到了屏蔽库仑作用的计算且成功地将它应用于 BEM 方法中^[118]. 最近,一种新的快速多极子方法将计算时间复杂度降至 $O(N)$ ^[110]. 比起早期的 FMM,新方法用对角变换进行平面波展开,大大地降低了运算量. 这种方法应用在 BEM 方法中出现在文献[109,119],其总体性能(包括 Node-Patch 的应用)得到了很大的提高. 使用 FMM 这个框架能延伸到任意体系生物分子^[18],克服了在求解 PB 方程时候的计算瓶颈,使得它在生物计算领域成为一个很有效的快速计算静电的工具.

相比之下,通常的 FFT 能在 $O(N \log N)$ 的时间内计算式(30)中的积分. White 等人在他们的边界元 PB 求解器中采用了预校正的快速傅里叶变换方法(Pre-corrected FFT)^[112]. 最近他们又提出了一种 FFTSVD 方法,该方法利用八叉树结构将原空间进行分解,利用抽样方法计算源项和线性响应主成分的低秩近似,在每一个尺度里面,其长程相互作用应用 FFT 计算^[120-121]. 其实验结果显示 FFTSVD 的性能要优于预校正的快速傅里叶变换. 此外还有将 FFT 与快速多极子相结合的方法^[122-123],其主要思想是利用小波基函数使得伽辽金矩阵的元素很小以致可以忽略,其运算复杂度接近于线性复杂度^[124].

此外,也可以通过对网格重新组织来加快边界元方法的求解速度. 例如可以通过合并相邻单元来降低自由度,以加快求解速度,但是这种方法通常会造成精度损失^[125]. Node-Patch 方法很像有限元方法中的 Voronoi 网格,通过简单地重新安排每个 Patch,可以达到线性边界元方法的精度而同时又能保持常数边界元方法的运算量^[105]. 另外,由于高阶基函数方法可以使得算法具有高的收敛阶,因此高阶方法可以采用较少的自由度来达到与低阶方法相同的精度^[111,126]. White 等人用高阶边界元方法在一些小分子的计算,得到了很好的收敛阶^[127]. Bardhan 等人采用曲面元离散分子表面,得到了很好的计算精度,然而他们没有系统地比较这种曲面元方法的整体性能,可能是由于产生这种曲面元以及在曲面元上进行数值积分需要消耗更多的 CPU 时间^[128]. 最后,好的计算格式可以提高数值计算的稳定性. Grandison 等人将总的电势分解为库仑电势和界面上极化电荷的产生的反应电势进行计算,其结果显示了此方法具有很好的稳定性^[129],但整体计算时间仍然较长.

2.1.3.6 超越 PB 方程

PB 方程形式简洁且便于计算,在弱耦合(低表面电荷密度,低浓度的单价离子溶液,高温)条件下,PB 方程可以给出与实验较为一致的结果,是模拟生物分子-溶液系统静电相互作用的有效手段,因而被广泛应用于很多研究领域. 尽管如此,传统 PB 模型仍然存在局限性:首先,传统 PB 模型将带电粒子视为点电荷,忽略了粒子的尺寸效应;其次,传统 PB 模型将蛋白质和溶液视为连续均匀的电介质,忽略了粒子的极化效应^[130];此外,传统 PB 模型基于平均场假设,因而无法反应系统的涨落^[30-131],且忽略了粒子相互关联(ionic correlation)效应. 这些被传统模型所忽略的效应被认为在溶液分层、电荷反转、同性电荷相吸、离子通道选择性等一系列生物进程中具有极其重要的作用^[132-135].

近年来,为了克服传统 PB 模型的局限性,一些新模型被陆续提出. 通过引入被传统 PB 模型所忽略的效应,这些新模型可以更精确地描述离子溶液体系的溶剂化效应,并有助于解释一些传统 PB 模型无法解释的物理现象.

一些研究致力于通过能量变分方法得到加入粒子尺寸效应的修正 PB 模型,这些修正模型主要包括 MPB(Modified Poisson-Boltzmann)模型^[136], SMPB(Size Modified Poisson-Boltzmann)模型^[33],以及 SMPNP(Size Modified Poisson-Nernst-Planck)模型^[84,20]等等. MPB 模型只适用于溶液中的粒子具有相同尺寸的情形. SMPB 模型的适用范围有所放宽,该模型可适用于溶液中的粒子具有两种不同的尺寸的情形. SMPNP 模型则完全克服了粒子种类和粒子尺寸的约束,该模型可适用于溶液中含有任意多种不同尺寸的粒子的情形;另外,SMPNP 方程作为含时方程,不仅可以模拟蛋白质-溶液体系的平衡态,还可以用于模拟体系的非平衡动态过程以及溶液离子的输运过程.

SMPNP 方程:

$$\frac{\partial c_i(\mathbf{r}, t)}{\partial t} = -\nabla \cdot D_i(\mathbf{r}) [c_i(\mathbf{r}, t) + \frac{k_i c_i(\mathbf{r}, t) \sum_l a_l^3 c_l(\mathbf{r}, t)}{1 - \sum_l a_l^3 c_l(\mathbf{r}, t)} + \beta c_i(\mathbf{r}, t) q_i \phi(\mathbf{r})],$$

$$\mathbf{r} \in \Omega_w, \quad i = 1, \dots, K.$$

$$-\nabla \cdot (\varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r}, t)) = \rho^f(\mathbf{r}) + \sum_i q_i c_i(\mathbf{r}, t), \quad \mathbf{r} \in \Omega.$$

一些研究致力于修正 PB 模型中的介电系数, 主要包括 DPB (Dipolar Poisson-Boltzmann) 模型^[130], NE (Nonlocal Electrostatic) 模型^[137], VDPB (Variable Dielectric Poisson-Boltzmann) 模型^[21] 等等. DPB 模型主要考虑水分子极化效应对介电系数的影响, 采用统计场论方法得到了一种介电系数依赖于电场强度和水分子偶极矩的变系数 PB 方程. NE 模型基于溶液模拟中的全局静电力学方法, 得到了一种介电系数依赖于电势在整个空间的分布的变系数 PB 方程. VDPB 模型主要考虑粒子的体积排斥效应对溶液介电系数的影响, 得到了一种介电系数依赖于离子浓度和离子尺寸的变系数 PB 方程.

DPB 方程

$$-\nabla \cdot \left((\varepsilon_0 + \frac{c_d p_0}{|\phi(\mathbf{r})|} G(\beta p_0 |\phi(\mathbf{r})|)) \nabla \phi(\mathbf{r}) \right) = \rho^f(\mathbf{r}) + \sum_i q_i c_i e^{-\beta q_i \phi(\mathbf{r})}, \quad \mathbf{r} \in \Omega.$$

NE 方程

$$-\nabla \cdot (\varepsilon_0 \int_{\Omega} \mathbf{r}, \mathbf{r}') \nabla \phi(\mathbf{r}') d\mathbf{r}') = \rho^f(\mathbf{r}) + \sum_i q_i c_i e^{-\beta q_i \phi(\mathbf{r})}, \quad \mathbf{r} \in \Omega.$$

VDPB 方程

$$-\nabla \cdot (\varepsilon(\phi(\mathbf{r})) \nabla \phi(\mathbf{r})) = \rho^f(\mathbf{r}) + \lambda \sum_{i=1}^K q_i c_i e^{-\beta q_i \phi(\mathbf{r})}, \quad \mathbf{r} \in \Omega.$$

$$\varepsilon(\phi(\mathbf{r})) = \begin{cases} \varepsilon_p, & \mathbf{r} \in \Omega_p, \\ \max \{ \varepsilon_p, \varepsilon_w - (\varepsilon_w - \varepsilon_p) \sum_{i=1}^K a_i^3 c_i e^{-\beta q_i \phi(\mathbf{r})} \}, & \mathbf{r} \in \Omega_w. \end{cases}$$

以上方程中的变量及参数的详细说明参考相关文献.

此外, 由于平均场理论的局限性, 在强耦合条件下, 传统 PB 方程的结果与实验结果存在较大偏差. 因此, 一些研究致力于发展适用于强耦合条件下的修正 PB 模型, 这方面的工作可以参考文献[138].

2.1.4 并行计算

一般来说, 串行程序局限于较小的生物分子系统. 对生物大分子溶液来说, 即使是很快的解法器, 在很精细的网格上也需要很长的计算时间. 因此, 加速原来的串行解法器来研究大分子系统的静电问题是一个亟待解决且艰巨的任务. 好在最近新一代计算机的发展, 为实现这个目的提供了重要的基础. 逐个介绍各种类型计算机的并行架构超出了本综述范畴, 这里主要介绍目前的一些主流并行程序在生物计算方面的应用, 同时, 它也受到越来越多人的关注.

CUDA 程序 CUDA (Computer Unified Device Architecture) 由 NVIDIA 公司开发, 在 GPU (Graphics Processing Units) 上计算的并行程序^[139]. 工具集的核心是一个 C 语言编译器, CUDA 3.0 已经开始支持 C++ 和 FORTRAN. CUDA 工具包包括 nvcc C 语言编译器 (适用于 GPU 的 CUDA FFT 和 BLAS 库); 分析器 (适用于 GPU 的 gdb 调试器和 CUDA 运行时驱动程序) 等等. CUDA 开发者软件包 SDK 提供了大量的范例, 包括矩阵乘法、矩阵转置、OpenGL 和 Direct 3D 图形互操作、BLAS 和 FFT 库的使用等等. CUDA 核心有三个重要抽象概念: 线程组层次结构、共享存储器、屏障同步 (Barrier Synchronization).

CUDA 体系结构的组成包含三个部分: 开发库、运行期环境和驱动. CUDA 1.1 版提供了两个标准的数学运算库 CUFFT (离散快速傅立叶变换) 和 CUBLAS (离散线性代数计算) 的实现. 这两个数学运算库所解决的是典型的大规模的并行计算问题, 也是在密集数据计算中非常常见的计算类型. 开发人员在开发库的基础上可以快速、方便的建立起自己的计算应用. 运行期环境提供了应用开发接口和运行期组件, 包括基本数据类型的定义和各类计算、类型转换、内存管理、设备访问和执行调度等函数. 基于 CUDA 开发的程序代

码在实际执行中分为两种,一种是运行在 CPU 上的宿主代码(Host Code),一种是运行在 GPU 上的设备代码(Device Code)。不同类型的代码由于其运行的物理位置不同,能够访问到的资源不同,因此对应的运行期组件也分为公共组件、宿主组件和设备组件三个部分,基本上囊括了所有在 GPGPU 开发中所需要的功能和能够使用到的资源接口,开发人员可以通过运行期环境的编程接口实现各种类型的计算。由于目前存在着多种 GPU 版本的 Nvidia 显卡,不同版本的 GPU 之间都有不同的差异,因此驱动部分基本上可以理解为是 CUDA-Enable 的 GPU 的设备抽象层,提供硬件设备的抽象访问接口。

OpenMP 程序 OpenMP(Open Multi-processing)是由 OpenMP Architecture Review Board 牵头提出的,并已被广泛接受的,用于共享内存并行系统的多线程程序设计的一套指导性的编译处理方案(Compiler Directive)。OpenMP 支持的编程语言包括 C、C++ 和 FORTRAN^[140],而支持 OpenMP 的编译器包括 Sun Compiler, GNU Compiler 和 Intel Compiler 等。OpenMP 提供了对并行算法的高层的抽象描述,程序员通过在源代码中加入专用的 pragma 来指明自己的意图,由此编译器可以自动将程序进行并行化,并在必要之处加入同步互斥以及通信。当选择忽略这些 pragma,或者编译器不支持 OpenMP 时,程序又可退化为通常的串行程序,代码仍然可以正常运作,只是不能利用多线程来加速程序执行。OpenMP 提供的这种对于并行描述的高层抽象降低了并行编程的难度和复杂度,这样程序员可以把更多的精力投入到并行算法本身,而非其具体实现细节。对基于数据分集的多线程程序设计,OpenMP 是一个很好的选择。同时,使用 OpenMP 也提供了更强的灵活性,可以较容易地适应不同的并行系统配置。线程粒度和负载平衡等是传统多线程程序设计中的难题,而 OpenMP 库从程序员手中接管了部分这两方面的工作。作为高层抽象,OpenMP 并不适合需要复杂的线程间同步和互斥的场合。

MPI 程序 多线程是一种便捷的模型,其中每个线程都可以访问其它线程的存储空间。因此,这种模型只能在共享存储系统之间移植。一般来讲,并行机不一定在各处理器之间共享存储,当面向非共享存储系统开发并行程序时,程序的各部分之间通过来回传递消息的方式通信。要使得消息传递方式可移植,就需要采用标准的消息传递库 MPI(Message Passing Interface)^[141]。与 OpenMP 并行程序不同,MPI 是一种基于消息传递的并行编程技术。消息传递接口是一种编程接口标准,而不是一种具体的编程语言。简而言之,MPI 标准定义了一组具有可移植性的编程接口。各个厂商或组织遵循这些标准实现自己的 MPI 软件包,典型的实现包括开源代码 MPICH、LAM MPI 以及不开放源代码的 Intel MPI。MPI 程序不仅可以适用多线程的方式并行运算还可以让程序以多进程的方式执行。这样做的好处是完成某一计算任务的不同进程可以运行在不同处理器上(不只是处理器的不同核上),甚至是不同的结点计算机上,方便分布式计算系统的构建。在多核上使用 MPI 可以采用两种方式,一种是在多核平台上开发传统的多进程 MPI 并行程序,一个核执行一个 MPI 进程。另外一种方式是采用 MPI + OpenMP 的方法,在结点内采用多线程方式,结点间采用 MPI 多进程方式。

Cilk Plus 程序 Cilk 多线程编程技术最早由 MIT 开发,是一个基于 GCC 编译器的开源项目,之后被 Intel 公司收购,整合进 Intel 编译器中。Cilk 为 Intel 编译器的语言扩展,通过提供一些关键字、环境变量、函数库等为 C/C++ 语言增加了细粒度任务支持,使其为新的和现有的软件增加并行性来充分发掘多处理器能力变得更加容易^[142]。Cilk 技术的设计特别适合但不限于 Divide 和 Conquer 的算法。它将问题分解成可以独立完成的子问题(任务),然后再将这些执行结果合并起来。

并行的 PB 解法器 虽然上面介绍的这些并行程序库很容易将一个串行解法器实现并行,然而在各个处理器上合理调度工作任务也不乏技巧。

据我们所知,一些广为使用的 PB 求解器,已经有了不同的并行程序实现,如 APBS, PBSA, DelPhi, AFMPB。这里,将简单描述它们中的并行技术以及最近发展的新技术以加速 PB 方程的求解^[143]。

Band 和 Holst 等人发展的偏微分方程有限元求解器(Adaptive Multilevel Finite Element Package),首先在每个处理器上求出粗网格上的近似解,然后将整个区域剖分成一些小区域,将小区域上的计算分派给处理器,将粗网格上的解当作初始解在细网格上进行迭代求解。APBS 延伸了 Band 和 Holst 等人的并行技术,将“Parallel Focusing Method”与“Electrostatic Focusing Method”结合^[64]。在“Parallel Focusing Method”中,将整个网格剖分成一些子网格(它们之间可以有重叠),每个子网格交给一个处理器来计算。粗网格上的解用来给予网格上定义边界条件,以减少了处理器之间的通信。Luo 等人则在 FDPB 中采用 Coarse-Grained

Distributive Method, 结合“Electrostatic Focusing Method”, 将细网格剖分成多个独立的 FDPB 进行计算^[144]. 此外, DelPhi 将分子表面网格、静电势和静电自由能的计算分别采用特定的并行技术来实现在多处理器上的并行计算^[145]. 最近, Hwang 等人在他们的有限元离散里面引入了一种全新地并行的 NKS (Newton-Krylov-Schwarz) 算法^[143]. NKS 算法利用 INB (Inexact Newton Method with Backtracking) 方法, 可以对非线性 PB 方程进行求解. 在每一个牛顿迭代步, 利用 Krylov 子空间方法求解离散得到的雅可比线性系统, 结合区域分解方法^[143] 寻求预条件子来加快它的收敛性. 用自适应网格加密的并行技术^[146] 来加快 PB 方程的求解也正在发展中. Lu 等人将并行自适应有限元软件平台 PHG^[147] 应用于 PB 方程, 自适应地产生复杂分子区域的有限元网格, 且成功地应用于 gA 离子通道的计算中^[67-68].

在边界元领域, 不同的并行实现出现在分子静电计算软件包. Yokota 等人将 FMM 的并行用 GPU 来实现, 其结果显示在 512 GPU 的时候达到了 0.5 的并行效率^[19, 148]. Geng 等人用通过 MPI 库实现了高阶边界元方法^[149] 和用 GPU 对直接边界元方法^[150] 及对树结构算法进行了加速计算 (TABI)^[151]. Barba 等人在 GPU 上用树结构算法加速边界元计算 (PyGBe)^[152], 很好地处理了生物系统中的溶剂空洞现象. 像最近的 Lu 等人把 FMM 计算过程用 DAG (Directed Acyclic Graph) 表示, 借助图论里面的 NP 问题近似优化方法将 FMM 计算动态地分配给处理器以实现并行, 各核之间的通信降到近似最低. 他们最近还用 Intel 编译器中的 Cilk 库来并行 AFMPB 软件包并得到了很好的并行效率, 且利用动态分配内存的方法实现了超大分子的静电计算^[18].

2.1.5 静电计算结果的可视化

当我们通过计算得到数值结果时, 可以借助可视化工具将数据转化为图片和视频等多媒体文件. 这些多媒体文件对于科学研究可以起到很好的辅助作用, 它们有助于研究者从整体上、从更直观的角度了解和掌握这些计算结果, 同时也有助于各个领域的研究者之间的学术交流. 尤其是对于生物模拟这种大规模问题, 其计算结果往往包含大量的数据和信息. 因此, 发展适用于生物分子模拟的可视化平台是很有必要的. 目前国际上分子模拟领域常用的可视化软件包括: PyMOL^[153], VMD^[154], Chimera^[155], GRASP^[156] 等等. 其中 PyMOL 采用 Python 编写而成, 具有非常强大的图片制作功能. VMD 采用了 Tk/Tcl 脚本语言, 具有操作简便、界面友好等特点. PyMOL 和 VMD 都加入了很多由研究者根据工作需要自行编写的脚本, 大大地方便了其使用. Chimera 通常被用于分子对接和药物设计领域, 而 GRASP 通常被用于分子表面的可视化. 这些软件的缺点是不能处理一般非结构化网格数据及基于它获得的模拟数据结果.

最近, Bai 等人开发了一款针对三维生物分子的可视化软件 VCMM^[157]. VCMM 提供了生物分子模型、计算网格和数值结果于一体的可视化功能, 并有多种小工具包括分子力场、网格质量分析、各种采样和体绘制等功能. 在软件内部, 运用了分子结构区域划分的方法和多种其他的采样方法, 加快了大型分子数据的渲染速度. 此外, VCMM 拥有简单易用的界面和广泛的数据格式支持, 包括 PDB、PQR、OFF、MESH、Medit、VTK 等常用的生物分子结构、网格结构以及数值计算结果等数据格式.

2.2 广义 Born 模型

广义玻恩模型 (Generalized Born Model, GB) 是 PB 模型的一种近似方法, 该模型被广泛用于生物分子中的长程相互作用与溶剂化效应的计算. 相比经典的 PB 模型, GB 模型采用半解析的计算方法快速地计算生物分子的溶剂化能. 广义玻恩模型目前被广泛应用于分子动力学研究^[158-159]. 这里主要介绍 GB 模型的一些计算方法和它在蛋白质溶液系统模拟中的一些应用.

2.2.1 Born 模型及广义 Born 模型的计算方法

我们从泊松方程(7)的格林函数 $G(\mathbf{r}_i, \mathbf{r}_j)$ 出发来推导玻恩公式 (Born Formula). 为了方便起见, 这一章中我们省略真空中的介电常数 ϵ_0 . 点电荷泊松方程的格林函数 $G(\mathbf{r}_i, \mathbf{r}_j)$ 满足如下方程

$$\nabla \cdot (\epsilon(\mathbf{r}) \nabla G(\mathbf{r}_i, \mathbf{r}_j)) = -\delta(\mathbf{r}_i - \mathbf{r}_j). \quad (36)$$

这里考虑简单的两介质模型, 在生物分子内, 格林函数 $G(\mathbf{r}_i, \mathbf{r}_j)$ 可以表示为

$$G(\mathbf{r}_i, \mathbf{r}_j) = \frac{1}{4\pi\epsilon_p |\mathbf{r}_i - \mathbf{r}_j|} + G'(\mathbf{r}_i, \mathbf{r}_j). \quad (37)$$

这里的 $G'(\mathbf{r}_i, \mathbf{r}_j)$ 反映了生物分子表面极化电荷所产生的反应电场. 一般地, 当介电表面存在时 $G'(\mathbf{r}_i, \mathbf{r}_j) \neq$

0. 因此,分子的溶剂化能可以写为

$$\Delta G_{\text{el}} = \frac{1}{2} \sum_{i,j} G'(\mathbf{r}_i, \mathbf{r}_j) q_i q_j. \quad (38)$$

对任意形状、任意电荷分布的生物分子,计算 $G'(\mathbf{r}_i, \mathbf{r}_j)$ 与求解泊松方程(7) 同样困难. 文[160] 提供了在球形分子情况下泊松方程的解析解,得到了自作用项 $G'(\mathbf{r}_i, \mathbf{r}_i)$ 和交叉项 $G'(\mathbf{r}_i, \mathbf{r}_j)$ 的表达式^[161]

$$G'(\mathbf{r}_i, \mathbf{r}_i) = -\frac{1}{4\pi} \left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w} \right) \frac{1}{R} \sum_{l=0}^{\infty} \frac{t_{ii}^l}{1 + (l/(l+1))\alpha},$$

$$G'(\mathbf{r}_i, \mathbf{r}_j) = -\frac{1}{4\pi} \left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w} \right) \frac{1}{R} \sum_{l=0}^{\infty} \frac{t_{ij}^l P_l(\cos\theta)}{1 + (l/(l+1))\alpha}, \quad (39)$$

其中 R 是分子的半径, r_i 是体系中第 i 个原子的质心到分子中心的距离, $t_{ij} = r_i r_j / R^2$, θ 是 \mathbf{r}_i 与 \mathbf{r}_j 的夹角, $\alpha = \varepsilon_p / \varepsilon_w$.

在实际计算中,由于大量的原子分布在生物分子表面附近,因此 t_{ij} 接近于 1, 导致式(39) 的收敛非常缓慢. 由于 $\varepsilon_w \gg \varepsilon_p \geq 1$, 我们设想 ε_w 趋于无穷大, 因而 $\alpha = 0$. 经过一些代数运算, 可以得到

$$G'(\mathbf{r}_i, \mathbf{r}_i) = -\frac{1}{4\pi\varepsilon_p} \frac{1}{R - r_i^2/R}. \quad (40)$$

$$G'(\mathbf{r}_i, \mathbf{r}_j) = -\frac{1}{4\pi\varepsilon_p} \frac{1}{\sqrt{r_{ij}^2 + (R - r_i^2/R)(R - r_j^2/R)}}. \quad (41)$$

结合式(40), (41) 和式(38), 便得到溶剂化能 ΔG_{el} 的计算公式. 然而我们还需要确定式中的一些参数, 如生物分子的半径 R , 原子到分子中心的距离 r_i . 其实, 方程(40) 和(41) 为广义玻恩模型提供了一个奠基石. 由于自作用项(40) 和交叉项(41) 都只依赖于 $(R - r_i^2/R)$, 我们将这个量定义为原子的等效玻恩半径, 记为 R_i . 因此, 如果求得生物分子中所有原子的 $G'(\mathbf{r}_i, \mathbf{r}_i)$ 或 $\Delta G_{ii}^{\text{el}}$, 便可由式(41) 求得所有的交叉项 $G'(\mathbf{r}_i, \mathbf{r}_j)$.

在实际计算中, 一般采用下面玻恩公式来计算生物分子的溶剂化能, 它作为一种半解析计算的方法其计算速度非常之快, 已经广泛地应用于一些生物模拟方面的计算:

$$R_i = -\frac{1}{8\pi} \left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w} \right) \frac{q_i^2}{\Delta G_{ii}^{\text{el}}},$$

$$\Delta G_{\text{el}} = \frac{1}{2} \sum_{i,j} F(\mathbf{r}_i, \mathbf{r}_j) q_i q_j \approx -\frac{1}{8\pi} \left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w} \right) \sum_{i,j} \frac{q_i q_j}{\sqrt{r_{ij}^2 + R_i R_j \exp\left(-\gamma \frac{r_{ij}^2}{R_i R_j}\right)}}. \quad (42)$$

Still 等人^[162] 取 $\gamma = 1/4$, 是一种常用的形式, 还有其它的选择如 $1/2, 1/10$ 等. 虽然当 $\gamma \neq 0$ 时, 方程(42) 已经不再是泊松方程的解析解, 然而在一些非球形状的蛋白分子中, 它却可以比较精确地用来估算溶剂化自由能^[163], 利用方程(42) 来计算溶剂化自由能, 需要首先计算每个原子的等效半径 R_i . 实际上, 方程(42) 即是在单原子情况下的玻恩公式, 而广义玻恩模型计算溶剂化自由能可以追溯到大概 50 年前, 虽然当时“Generalized Born”还未曾出现在文献中^[164]. 可以说, 玻恩公式不仅公式简单, 还提供了一种解析的计算相互作用力的有效方法. 然而, 两个问题也随之出现了:

- 1) 如何快速计算每个原子的有效半径;
- 2) 玻恩公式(42) 能否合理地估算实际分子的溶剂化自由能.

在文[165] 利用数值的方法求解泊松方程得到每个原子的等效半径, 从而给予了肯定答案. 然而, 他们用的数值方法求解每个原子的等效半径是非常耗时的. 因此广义玻恩模型中, 计算每个原子的等效半径成了 GB 模型中至关重要的任务. 实际上, 现在几乎所有的 GB 模型, 都采用方程(42) 来计算溶剂化效应, 而它们之间的差异则反映在原子等效半径的计算上, 下面我们主要介绍原子的等效半径的一些计算方法. 一般来说, 通常可以采用体积分的方法^[166] 或面积分的方法^[167] 来计算(42). 然而我们却面临着两方面的挑战

- 1) 用简单的公式来计算 $\Delta G_{ii}^{\text{el}}$;

2) 一个好的数值积分方法来计算上面提到的积分公式.

David 和 Jackson 等人采用库仑近似 (Coulomb Field Approximation, CFA) 方法^[168] 和能量密度积分方法得到原子的等效玻恩半径公式

$$\alpha_4 = R_i^{-1} = \frac{1}{4\pi} \int_{\text{ext}} \frac{1}{|\mathbf{r} - \mathbf{r}_i|^4} dV = \rho_i^{-1} - \frac{1}{4\pi} \int_{r > \rho_i} |\mathbf{r}|^{-4} dV, \quad (43)$$

其中, ρ_i 表示原子 i 的范德华半径, 方程(43) 中的积分区域为溶液中除去原子 i 所剩下的部分. 值得注意的是, 库仑近似只对点电荷分布在对称的球形分子情况下成立. 对于实际分子, 库仑近似得到的等效玻恩半径往往会偏大^[169]. 此外, 通过参数化方法也可以提高等效玻恩半径的精度, 如下面的计算公式:

$$\alpha_N = \left(\frac{1}{4\pi} (N-3) \int_{\text{ext}} \frac{1}{|\mathbf{r} - \mathbf{r}_i|^N} dV \right)^{\frac{1}{N-3}} = \left(\rho_i^{3-N} - \frac{1}{4\pi} (N-3) \int_{r > \rho_i} |\mathbf{r}|^{-N} dV \right)^{\frac{1}{N-3}}. \quad (44)$$

从 α_N 的线性组合出发可以改进库仑近似, 同时提高广义玻恩模型的精度, 如 CHARMM 中的 ‘GBMV’ 利用 α_4 和 α_5 的线性组合来计算原子的等效玻恩半径, 随后, ‘GBMV2’ 用 $R^{-1} = (1 - \sqrt{2}/2)\alpha_4 + \sqrt{2}/2\alpha_5$ 计算原子的等效玻恩半径, 提高了玻恩半径的计算精度^[169]. 虽然在方程(44) 中, 当 $N \neq 4$ 也许纯粹是一种经验性公式, 然而文献[170] 却指出在 $N = 6$ 时可以导出精确的等效半径计算公式, 甚至都不要求原子像库仑近似那样成对称分布.

此外, 我们可以类比点电荷分子情况下的泊松方程那样来推导一般溶液的等效玻恩半径公式. 然而这时情况会变得非常复杂, 为推导带来困难. 实际上, 人们更多的是采用经验公式与半解析的方法, 如下面的方法来处理溶液中存在的离子^[171]:

$$\Delta G_{\text{el}} = -\frac{1}{8\pi} \sum_{i,j} \left(\frac{1}{\varepsilon_p} - \frac{\exp(-0.73\kappa f^{\text{GB}})}{\varepsilon_w} \right) \frac{q_i q_j}{f^{\text{GB}}}, \quad (45)$$

这里 $f^{\text{GB}} = \sqrt{r_{ij}^2 + R_i R_j \exp\left(-\gamma \frac{r_{ij}^2}{R_i R_j}\right)}$. 其实, 我们可以从单个原子的 PB 方程的解的形式 $\phi_i \sim \frac{\exp(-\kappa r)}{\varepsilon_w} \frac{q_i}{r}$

出发, 将方程(42) 中 $\left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w}\right) \rightarrow \left(\frac{1}{\varepsilon_p} - \frac{\exp(-\kappa f^{\text{GB}})}{\varepsilon_w}\right)$. 式中的常数 0.73 是通过与 PB 方程的预测结果进行对比而得到的拟合参数.

2.2.2 广义 Born 模型的应用

蛋白质折叠与设计 利用隐式溶剂模型来了解生物分子系统的构象转变具有广泛的应用前景, 特别是广义 Born 模型. 像研究一些小的蛋白质分子 (几十个残基) 的折叠过程^[172-173] 和一些较大分子的折叠过程^[174], 在这些模拟中, 用隐式溶剂的方法来计算它们在折叠过程中的能量地形. 广义 Born 模型也能用来探测环境温度、摩擦力和随机力对蛋白质折叠的影响. Lopes 和 Felts 等人在他们的蛋白质设计实验中, 利用广义 Born 模型来探测点突变对蛋白质稳定性改变的影响^[175-176].

大分子上的大尺度运动 Amaro 等人用广义 Born 模型研究了禽流感病毒的构象变化^[177]. 作为最近一个新的研究方向 (DNA - 蛋白质的结合^[178]), 广义 Born 模型也大有用武之地. Ruscio 等人研究了蛋白质 - DNA 的动力学性质^[179].

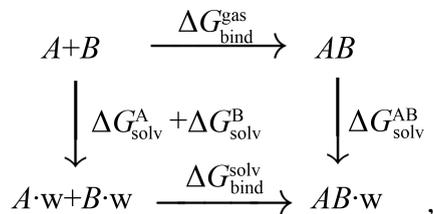
2.3 静电相互作用的若干应用

近年来, 分子模拟技术得以广泛应用, 在对生物分子的结构、功能和特性的研究中, 计算机模拟技术逐渐占据主导地位. 相比全原子模型, 以 PB 模型和 GB 模型为代表的隐式溶剂模型可以极大降低计算量. 隐式溶剂模型由于其简便性和高效性而被广泛应用于分子模拟. 本章将着重介绍 PB 模型和 GB 模型在生物计算和分子模拟中的应用.

2.3.1 静电相互作用在自由能计算中的应用

最近的现代药物设计中, 人们最感兴趣的是计算蛋白质与配体、核酸与配体之间的结合自由能. 随着现代计算机技术的发展, 现在已经能够快速计算蛋白质折叠、蛋白 - 配体结合、蛋白 - 蛋白结合等生物反应中的自由能变化. 通过分子动力学模拟采样, 计算蛋白质复合物结合自由能的方法主要有 MM-PBSA (Molecular Mechanics Poisson-Boltzmann Surface Area) 和 MM-GBSA (Molecular Mechanics Generalized Born Surface Area).

它们可以计算两个蛋白质的结合自由能,也可以计算复合物界面残基突变造成的自由能差.而通过自由能微扰和利用 Jarzynski 公式进行非平衡态模拟,原则上也可以计算蛋白质复合物的结合自由能,但由于通常蛋白质相互作用体系庞大,需要的采样计算量远远大于 MM-PB(GB)SA,因此它们的应用受到了一定程度的限制.在 MM-PB(GB)SA 中,需要对蛋白质复合物进行一段分子动力学模拟,利用模拟中获得的蛋白质构象来计算结合自由能.它采用了下面的热力学循环^[180]:



根据这一循环,蛋白质的结合自由能可表示如下

$$\Delta G_{\text{bind}}^{\text{solv}} = \Delta G_{\text{bind}}^{\text{gas}} + \Delta G_{\text{solv}}^{AB} - (\Delta G_{\text{solv}}^A + \Delta G_{\text{solv}}^B) = (\Delta H_{\text{bind}}^{\text{gas}} - T\Delta S) + \Delta G_{\text{solv}}^{AB} - (\Delta G_{\text{solv}}^A + \Delta G_{\text{solv}}^B), \quad (46)$$

其中, $\Delta H_{\text{bind}}^{\text{gas}}$ 表示气态复合物形成前后的焓变,包括了 A 与 B 间的静电和范德华作用能; ΔS 表示复合物形成前后的熵变,包括平动熵、转动熵、振动熵和构象熵; ΔG_{solv} 表示复合物 AB 或蛋白质 A, B 从气态到溶液中的自由能变化,也就是溶剂化自由能,它可以分解为静电项和非极性项

$$\Delta G_{\text{solv}} = \Delta G_{\text{elec, solv}} + \Delta G_{\text{np, solv}}, \quad (47)$$

这里 $\Delta G_{\text{elec, solv}}$ 可通过解 Poisson-Boltzmann 方程获得(MM-PBSA),或者利用广义波恩模型获得(MM-GBSA).这方面的研究可以参考文献[182-187].

2.3.2 静电相互作用在蛋白质相互作用中的应用

蛋白-蛋白和蛋白-配体相互作用 PB模型广泛应用于分析蛋白-蛋白结合过程,它们的静电结合自由能定义为

$$\Delta G_{AB} = G_{AB} - (G_A + G_B), \quad (48)$$

这里, $G_X = \frac{1}{2} \sum_i q_i \phi(r_i)$, 其中 X 表示分子或者复合物.然而早期的一些 PB 计算结果通常显示两分子结合自由能 ΔG_{AB} 为正,即不利于它们的结合.主要是由于一些带电基团和极性基团的去溶剂化效应不足以弥补蛋白质分子界面上的电荷与电荷之间相互作用^[188-190]. 这些不利的因素主要是用溶剂排斥表面作为蛋白质分子-溶剂的分界面计算而得到的.

芽孢杆菌 RNA 酶,细胞外解淀粉芽孢杆菌的核糖核酸酶,当前对于研究及发展抗癌药物有很重要的指导价值^[191-192]. 由于很容易获得它们的单体与复合物的高分辨率结构^[193], 研究它与芽孢杆菌 RNA 酶抑制剂的相互作用是一个很好的例子. 芽孢杆菌 RNA 酶抑制剂通过盐桥、氢键和芽孢杆菌 RNA 酶紧密结合. 它们的反应界面主要是一些极性残基、带电残基和水分子(不与溶剂接触). 这些残基之间高度地静电补偿和静电吸引有助于它们复合物的稳定. 然而这些残基的去溶剂化则给复合物的结合带来了不稳定的影响,换句话说,蛋白质分子表面带电残基的溶剂化效应,以及复合物形成将这些残基转化成复合物分子内部(因此它们具有低介电系数),这些作用将反抗它与酶分子的结合. Fersht 等人运用停留光谱法、等温滴定量热法和双向突变周期分析等方法测量了这些残基对于复合物形成时的贡献作用^[194-195]. Dong 等人^[196]和 Wang 等人^[197]运用 PB 模型对它们进行定量计算.

Steitz 等人最早关于蛋白质分子-核酸相互作用的例子(大肠杆菌 DNA 聚合酶片段),他们发现 C-末端结构域深裂中的组氨酸,是 DNA 底物上的活性区,能改变它们的状态与 DNA 结合,对蛋白质-DNA 结合能有很重要的贡献^[198]. 众所周知, DNA 中的磷酸基与 RNA 在蛋白质分子的活性位点上会形成盐桥. 然而, Qin 和 Zhou 等人提供了大量的计算结果显示蛋白质-核酸的结合能为正的,但他们采用溶剂排斥表面计算,而当用范德瓦尔斯表面进行计算时结合能则变为负的^[199]. Szklarczyk 等人计算了三种形式的 eIF4E 蛋白质的结合自由能,指出当加入非极性和相变作用时,结合自由能与实验结果很吻合^[200].

配体-受体结合速率常数 PB模型结合布朗运动(BD)模拟可以计算配体-受体的结合速率. 为了实现结合,配体须扩散输送到受体表面^[201]. 扩散过程中遇到受体的速率受几个因素的影响,主要是反应物上

复杂电荷分布产生的长程库仑力. 由于多数分子的复杂外形结构和结合模式各不相同, 结合速率通常不能解析计算得到, 因此采用数值模拟的方法来计算得到. 布朗运动轨迹模拟方法便是一个很好的数值模拟方法. 为了采用它来计算结合速率, 首先为配体分子生成很多可能的轨迹, 其初始时候位于一个半径为 r_b , 中心为受体的坐标中心的球面上. 球的半径应该充分大以致受体与配体之间的静电作用力大致成中心对称. 每条轨迹将达到一个相遇准则或达到一个半径为 r_q 的球面上而退出模拟. 配体扩散遇到受体的速率 k_d , 采用下面式子计算^[202]

$$k_d = k_b \frac{\beta_b}{1 - (1 - \beta_b) k_b / k_q}, \quad (49)$$

这里^[203]

$$k_{b(q)} = 4\pi \left[\int_{r_b(q)} \frac{\exp(U(r)/(k_B T))}{r^2 D_w} dr \right]^{-1}, \quad (50)$$

β_b 是配体最终遇到受体轨迹的百分比, $U(r)$ 是分子间作用势, D_w 是相对平移扩散系数. 目前, 最重要的因素是配体最终遇到受体轨迹的百分比, $U(r)$ 是分子间作用势, D_w 是相对平移扩散系数. 目前, 最重要的因素是配体的轨迹是在受体激发的静电场中模拟所产生的, 而静电场可以通过求解 PB 方程计算得到, 配体通常当成点电荷. 分子结合的布朗运动模拟方法假定配体的运动受到分子间的静电相互作用力与溶剂分子碰撞产生的随机作用力影响.

BD-PB 方法的一个应用例子是铜锌超氧化物歧化酶催化 O_2^- 自由基的破坏,



虽然铜锌超氧化物歧化酶在生理 pH 条件下带负静电荷, 但它活性位点区主要是带正电的. Sharp 等人^[204] 和 Sines 等人^[205] 对牛红细胞与铜锌超氧化物歧化酶做了一系列的布朗运动模拟, 旨在分析 O_2^- 对超氧化物歧化酶的活性位点的静电转向. 他们考虑离子强度与结合速率的相关性以及选定的带电残基的突变对结合速率的影响. 同时采用实验模拟的方法得到了实验观测数据, 随着离子强度的增加, O_2^- Cu/Zn SOD 的反应速率下降^[206]. 确实, Getzoff 等人的实验数据推测了定点突变特征, 当维持定向结构时, 位点特异性突变将增加局部的正电荷, 导致它们有很快的反应速率, 因此反应速率与离子强度的依赖性增加, 符合布朗运动模拟结果^[207].

另一个用 BD-PB 模拟的例子是乙酰胆碱酯酶 (AChE), 它能水解神经递质乙酰胆碱, 产生胆碱和醋酸. 它主要存在于神经肌肉接头和中枢神经系统胆碱神经突触中, 它的活性能终止突触传递^[208]. AChE 与它天然的底物发生反应, 是一种非常高效的催化剂. 然而它的活性区位于附近窄且 $\sim 20 \text{ \AA}$ 深的口袋底部, 称活性位点峡谷. 文献[209]采用布朗运动模拟计算了乙酰胆碱酯酶与乙酰胆碱离子的扩散反应速率.

酶催化反应 - 蛋白酪氨酸磷酸酶 PTP PB 模型也可以应用到酶促反应静电方面的分析, 预测酶分子在发生催化功能时其残基的电离状态. 因为我们在隔离态与束缚态研究它们的反应机理时候, 必须考虑酶分子和它的底物上的官能团的电离状态.

蛋白酪氨酸磷酸酶 (PTPases) 的重要性的功能的鉴定是通过酶分子动力学, 位点定向突变和结构测定等方面体现出来的. 去磷酸化时首先需要半胱氨酸离子 (失去硫元素) 对亲核磷酸盐在活性区进行碰撞形成磷酸化酶中间产物, 而活性位点上的天冬氨酸残基须以酸的身份捐赠氢离子给酪氨酸基, 其底物上的磷酸基在反应中被认为是完全电离态. Dillet 等人推断底物是一个双质子化的磷酸基团^[210], 而 Kolmodin 和 Åqvist 用了一个完全去质子化的底物和半胱氨酸酶之间存在着强烈的静电排斥力作为例子和它反应的结合自由能来反对 Dillet 等人的观点^[211]. 此外, 在解释 pH 对蛋白酪氨酸磷酸酶活性的影响时也存在歧义, 这是一种钟形依赖关系, 具有最佳的 pH 值在 5 到 6 之间, 大多数的实验研究和机理解释了这种关系.

蛋白质分子结合稳定性与环境 pH 之间的相关性 蛋白质分子结构的稳定性、配体与受体分子的结合是依赖于环境 pH 的, 因为相变和分子复合物的形成将会改变蛋白质分子中可滴定基团的质子化均衡, 进而改变离子基团的平均电荷及分子的静电自由能.

Schreiber 和 Fersht 等人测量了芽孢杆菌 RNA 酶和芽孢杆菌 RNA 酶抑制剂 (Barnase-Barstar) 复合物的结合解离速率与 pH 之间的依赖关系^[212]. 同样地, Dong 等人运用 PB 模型来计算它的反应速率. 他们采用下式来计算 pH 与结合速率之间的关系^[196]

$$K(pH) = K_0 \frac{1 + 10^{pK^{AB} - pH}}{1 + 10^{pK^A - pH}} \quad (52)$$

其中 pK^{AB} 和 pK^A 分别为 Barbase 在复合物与单体和 Barstar 的 pKa.

2.3.3 隐式溶剂分子模拟方法

分子动力学模拟一般可用于生物分子体系的结构和动力学性质的研究. 目前来说, 分子动力学模拟一般受限于模拟时间长度和体系自由度. 随机动力学(SD)模拟方法正是在这一思想指导下提出来的. 通常的随机动力学方法不考虑溶剂平均力势和溶剂摩擦力的含时性, 这限制了它的发展和运用. 溶剂平均力势反映了溶剂对生物分子的平均效应, 对生物分子的动力学性质有重要影响. Zauhar 和 Gilson 等人利用 PB 方程把溶剂化效应加到分子动力学模拟中^[213]. Wang 等人把由边界元法决定的水化力结合到随机动力学模拟中, 研究水化力和摩擦记忆性对系统相变和动力学行为的影响^[214-217]. Gilson 等人利用分子动力学模拟结合 PB 模型和 GB 模型研究了 HIV-1 蛋白酶的静电性质^[218]. 随后, PB 模型和 GB 模型结合分子动力学研究生物分子的动态性质和结构、稳定性和预测 pKa 值相继提出. Luo 等人比较了 TIP3P 显示溶剂模型与 PB 隐式溶剂模型在计算反应场能量时候的差异^[219]; Wei 等人用 MIB 方法求解 PB 方程进行分子动力学模拟^[220]. Juffer 等人用边界元方法求出了两个大分子体系的静电势与蒙特卡洛模拟结果进行了比较^[221]. 在这些方法的基础上, 一个重要的动力学方法, constant pH MD, 结合蒙特卡洛采样原理被提出, 很好地处理了生物分子构象变化、结构稳定性与溶液 pH 之间的依赖关系^[222].

2.3.4 其它

蛋白质-配体复合物的形成对电离态的影响 pKa 值, 反应溶液的酸度系数, 在生物化学中表示酸离解氢离子的能力, 其计算式为

$$pKa = -\log_{10} Ka, \quad (53)$$

其中 Ka 表示酸的离解常数. 对于酸 HA 的电离 $HA \rightleftharpoons H^+ + A^-$ 平衡式来说,

$$Ka = \frac{[H^+][A^-]}{[HA]}. \quad (54)$$

运用 PB 模型来预测 pKa 值需要计算蛋白质分子各个不同的质子态下的溶剂化能, 近年来涌现了大量的作者对它的关注, 其预测结果的精度依赖于 PB 模型计算精度^[223-224]. 由于广义玻恩模型计算的高效性, 目前它结合分子动力学模拟方法广泛地应用于计算 pKa 值^[225-226].

蛋白质分子和配体中的可滴定基团在它们复合物的形成中可以发生改变, 导致结合自由能也发生改变. 由于实验的复杂性, 系统地研究它们复合物形成时其质子化效应相当罕见. 事实上, 质子化状态的改变可以通过实验探测到, 如多维核磁共振光谱(NMR)^[227], 等温滴定量热法(ITC)^[228]. 多维核磁共振光谱实验可以提供可滴定基团在质子态下的详细微观信息, 而等温滴定量热法只测量整体的质子交互现象. 然而, 我们可以通过复合物形成中官能团等 pKa 值的改变比较详细地得到其微观性质. Czodrowski 等人运用 PB 模型研究了一系列酶抑制剂与丝氨酸蛋白酶、胰蛋白酶和凝血酶的结合反应过程^[229].

细胞膜的离子输运 细胞中离子跨膜输运是细胞生理学的一个关键过程. 离子通道是一种成孔蛋白, 能催化脂双层中疏水环境下的离子运动, 创建一个亲水性的环境, 因此离子很容易在它里面扩散^[230]. 通道大部分时间处于关闭状态以减少跨膜离子梯度的耗散作用, 在需要的时候成打开状态. 它的打开与关闭, 主要通过两种方式调节: 与配体的结合或跨膜电位的变化. 研究离子通道传统的方法主要为一些实验手段, 比如电压固定技术, 膜片钳技术, 穿孔钳技术等. Grabe 等人对钾离子通道的低分辨率结构运用 PB 模型求得通道的溶剂化能垒和膜电压对 S4 螺旋运动的影响^[231]. 最近, Tu 和 Lu 等人也搭建了用隐式溶剂方法研究三维离子通道内离子输运的完整有限元模拟软件平台^[67,100], 该方法采用了静电与扩散耦合的 PNP 模型.

参 考 文 献

- [1] Kolinski A. Multiscale approaches to protein modeling[M]. Beijing: Science Press, 2014.
- [2] Philips R. Physical biology of the cell[M]. Beijing: Science Press, 2012.
- [3] Gorham R D, Kieslich C A, Morikis D. Electrostatic clustering and free energy calculations provide a foundation for protein

- design and optimization[J]. *Ann Biomed Eng*, 2011, **39**(4):1252 – 63.
- [4] Bai H J, Lai L H. Protein-protein interactions; Interface analysis, binding free energy calculation and interaction design[J]. *Acta Phys Chim Sin*, 2010, **26**(7):1988 – 1997.
- [5] Linderstrøm-Lang. On the ionization of proteins[J]. *CRLC*, 1924, **15**:7.
- [6] Dobson C M. Principles of protein folding, misfolding and aggregation[J]. *Semin Cell Dev Biol*, 2004, **15**(1):3 – 16.
- [7] Hartl F U, Hayer H M. Converging concepts of protein folding in vitro and in vivo[J]. *Nat Struct Mol Biol*, 2009, **16**(6):574 – 81.
- [8] Irbäck A, Mohanty S. Folding thermodynamics of peptides[J]. *Biophys*, 2005, **88**(3):1560 – 1569.
- [9] Aurora R, Rose G D. Helix capping[J]. *Proteins*, 1998, **7**:21 – 38.
- [10] Donald J E, Kulp D W, DeGrado W F. Salt bridges: Geometrically specific, designable interactions[J]. *Proteins*, 2011, **79**(3):898 – 915.
- [11] Zhang C F. Principles of biochemistry[M]. Beijing:Higher Education Press, 2011:90 – 134.
- [12] Warshel A. Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites[J]. *Journal of Biological Chemistry*, 1988, **273**(42):27035 – 27038.
- [13] Sela-Culang I, Kunik V, Ofra Y. The structural basis of antibody-antigen recognition[J]. *Front Immunol*, 2013, **4**:302 – 314.
- [14] Lo C L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites[J]. *J Mol Biol*, 1999, **285**(5):2177 – 2198.
- [15] Roca M, Vardi-Kilshtain, Warshel A. Toward accurate screening in computer aided enzyme design[J]. *Biochemistry*, 2009, **48**(14):3046 – 3056.
- [16] Schreiber G, Haran G, Zhou H X. Fundamental aspects of protein-protein association kinetics[J]. *Chem Rev*, 2009, **109**(3):839 – 860.
- [17] Tjong H, Zhou H X. Accurate calculations of binding, folding, and transfer free energies by a scaled generalized Born method [J]. *J Chem Theory Comput*, 2008, **4**(10):1733 – 1744.
- [18] Peng B, Zhang B, Huang J F, Pitsianis N B, Sun X B, Lu B Z. Parallel AFMPB solver with automatic surface meshing for calculations of molecular solvation free energy[J]. *Computer Physics Communications*, 2014, accepted.
- [19] Yokota R, Bardhan J P, Knepley M G, Barba L A, Tsuyoshi H. Biomolecular electrostatics using a fast multipole BEM on up to 512 GPUs and a billion unknowns[J]. *Comp Phys Comm*, 2011, **182**(6):1271 – 1283.
- [20] Qiao Y, Tu B, Lu B Z. Ionic size effects to molecular solvation energy and to ion current across a channel resulted from the nonuniform size-modified PNP equations[J]. *J Chem Phys*, 2014, **140**(17):174102.
- [21] Li H L, Lu B Z. An ionic concentration and size dependent dielectric permittivity Poisson-Boltzmann model for biomolecular solvation studies[J]. *The Journal of Chemical Physics*, 2014, **141**:024115.
- [22] Lu B Z, Zhou Y C, Holst M J, McCammon J A. Recent progress in numerical methods for the Poisson-Boltzmann equation in biophysical applications[J]. *Commun Comput Phys*, 2008, **3**(5):973 – 1009.
- [23] Bardhan J P. Biomolecular electrostatics-I want your solvation[J]. *Computational Science & Discovery*, 2012, **5**:013001.
- [24] Kukić P, Nielsen J E. Electrostatics in proteins and protein-ligand complexes[J]. *Future Med Chem*, 2010, **2**(4):647 – 666.
- [25] Fogolari F, Brigo A, Molinari H. The Poisson-Boltzmann equation for biomolecular electrostatics: A tool for structural biology [J]. *J Mol Recognit*, 2002, **15**(6):377 – 392.
- [26] Koehl P. Electrostatics calculations: Latest methodological advances[J]. *Curr Opin Struct Biol*, 2006, **16**(2):142 – 151.
- [27] Ren P, Chun J, Thomas D, Schnieders M J, Marucho M, Zhang J, Baker N A. Biomolecular electrostatics and solvation: A computational perspective[J]. *Rev Biophys*, 2012, **45**(4):427 – 491.
- [28] David C G. The electrical double layer and the theory of electrocapillarity[J]. *Chem Rev*, 1947, **41**(3):441 – 501.
- [29] Debye P, Hückel E. Zur theorie der elektrolyte[J]. *Phys Zeitschr*, 1923, **24**:185 – 206.
- [30] Vlachy V. Ionic effects beyond Poisson-Boltzmann theory[J]. *Annu Rev Phys Chem*, 1999, **50**:145 – 165.
- [31] Collins K D. Why continuum electrostatics theories cannot explain biological structure, polyelectrolytes or ionic strength effects in ion-protein interactions[J]. *Biophys Chem*, 2012, **167**:43 – 59.
- [32] Grochowski P, Trylska J. Continuum molecular electrostatics, salt effects, and counterion binding-A review of the Poisson-

- Boltzmann theory and its modifications[J]. *Biopolymers*, 2008, **89**(2):93 – 113.
- [33] Chu V B, Bai Y, Lipfert J, Herschlag D, Doniach S. Evaluation of ion binding to DNA duplexes using a size-modified Poisson-Boltzmann theory[J]. *Biophys J*, 2007, **93**(9):3202 – 3209.
- [34] Schutz C N, Warshel A. What are the dielectric “constants” of proteins and how to validate electrostatic models? [J]. *Structure Function Genetics*, 2001, **44**(4):400 – 417.
- [35] Bernard F, Leitner A. Principles and techniques of applied mathematics[J]. *American Journal of Physics*, 1957, **25**(3):208.
- [36] Warshel A, Mitsunori K, Andrei V P. Polarizable force fields: History, test cases, and prospects[J]. *Journal of Chemical Theory and Computation*, 2007, **3**(6):2034 – 2045.
- [37] Uematsu M, Franck E U. Static dielectric constant of water and steam[J]. *Journal of Physical and Chemical Reference Data*, 1980, **9**(4):1291 – 1306.
- [38] Lebard D N, Matyushov D V. Redox entropy of plastocyanin: Developing a microscopic view of mesoscopic polar solvation[J]. *Journal of Chemical Physics*, 2008, **128**(15):106 – 155.
- [39] Li Z, Lazaridis T. Water at biomolecular binding interfaces[J]. *Physical Chemistry Chemical Physics*, 2007, **9**(5):573 – 581.
- [40] Niedermeier C, Schulten K. Molecular dynamics simulations in heterogeneous dielectrics and Debye-Hückel media—Application to the protein bovine pancreatic trypsin inhibitor[J]. *Molecular Simulation*, 1992, **8**(6):361 – 387.
- [41] Gilson M, Sharp K A, Honig B. Electrostatic interactions in proteins[J]. *Protein Structure, Folding and Design* 2, 1987, **2**:235 – 244.
- [42] Baker N A, Sept D, Joseph S, Holst M J, McCammon J A. Electrostatics of nanosystems: Application to microtubules and the ribosome[J]. *Proceedings of the National Academy of Sciences*, 2001, **98**(18):10037 – 10041.
- [43] Nicholls A, Bharadwaj R, Honig B. Grasp-graphical representation and analysis of surface-properties[J]. *Biophysical Journal*, 1993, **64**(2):166.
- [44] Bashford D. An object-oriented programming suite for electrostatic effects in biological molecules[J]. *Lecture Notes in Computer Science*, 1997, **1343**:233 – 240.
- [45] Luty B A, Ilin A, Antosiewicz, Gilson M K, McCammon J A. Electrostatics and diffusion of molecules in solution—simulations with the university of houston brownian dynamics program[J]. *Computer Physics Communications*, 1995, **91**(1/3):57 – 95.
- [46] Jun W, Qin C, Li Z L, Zhao H K, Luo R. Achieving energy conservation in Poisson-Boltzmann molecular dynamics: Accuracy and precision with finite-difference algorithms[J]. *Chemical Physics Letters*, 2009, **468**(4/6):112 – 118.
- [47] Prabhu N V, Panda M, Yang Q, Sharp K A. Explicit ion, implicit water solvation for molecular dynamics of nucleic acids and highly charged molecules[J]. *Journal of Computational Chemistry*, 2008, **29**(7):1113 – 1130.
- [48] Brooks B R, Bruccoleri R E, Olafson B D, States D J, Swaminathan S, Karplus M. Charmm: A program for macromolecular energy, minimization, and dynamics calculations[J]. *Journal of Computational Chemistry*, 1983, **4**(2):187 – 217.
- [49] Chern I L, Liu J G, Wang W C. Accurate evaluation of electrostatics for macromolecules in solution[J]. *Methods and Applications of Analysis*, 2003, **10**(2):309 – 328.
- [50] Wang W C. A jump condition capturing finite difference scheme for elliptic interface problems[J]. *SIAM Journal on Scientific Computing*, 2004, **25**(5):1479 – 1496.
- [51] Zhou Y C, G W, Wei G W. On the fictitious-domain and interpolation formulations of the matched interface and boundary (MIB) method[J]. *Journal of Computational Physics*, 2006, **219**(1):228 – 246.
- [52] Yu S N, Wei G W. Three dimensional matched interface and boundary (MIB) method for geometric singularities[J]. *Journal of Computational Physics*, 2007, **227**(1):602 – 632.
- [53] Yu S N, Geng W H, Wei G W. Treatment of geometric singularities in implicit solvent models[J]. *Journal of Computational Physics*, 2007, **126**(24):244108 – 244120.
- [54] Geng W H, Yu S, Wei G W. Treatment of charge singularities in implicit solvent models[J]. *Journal of Chemical Physics*, 2007, **127**(11):114106 – 114125.
- [55] Zhou Y C. Matched interface and boundary (MIB) method and its applications to implicit solvent modeling of biomolecules [D]. America:Michigan State University, 2006.
- [56] Wang C, Wang J, Cai Q, Li Z, Zhao H K, Luo R. Exploring accurate Poisson-Boltzmann methods for biomolecular

- simulations[J]. *Computational and Theoretical Chemistry*, 2013, **1024**:34 – 44.
- [57] Qiao Z, Li Z L, Tang T. A finite difference scheme for solving the nonlinear Poisson-Boltzmann equation modeling charged spheres[J]. *Journal of Computational Mathematics*, 2006, **24**(3):252 – 264.
- [58] Sayyed-ahmad A, Tuncay K, Ortoleva P J. Efficient solution technique for solving the Poisson-Boltzmann equation[J]. *Journal of Computational Chemistry*, 2004, **25**(8):1068 – 1074.
- [59] Holst M J. Multilevel methods for the Poisson-Boltzmann equation[D]. Urbana-Champaign:University of Illinois, 1993.
- [60] You T J, Harvey S C. Finite-element approach to the electrostatics of macromolecules with arbitrary geometries[J]. *J Comput Chem*, 1993, **14**:484-501.
- [61] Bowen W R, Sharif A O. Adaptive finite-element solution of the nonlinear Poisson-Boltzmann equation: A charged spherical particle at various distances from a charged cylindrical pore in a charged planar surface[J]. *Journal of Colloid and Interface Science*, 1997, **187**(2):363 – 374.
- [62] Cortis C M, Friesner R A. An automatic three-dimensional finite element mesh generation system for the Poisson-Boltzmann equation[J]. *J Comput Chem*, 1997, **18**:1570 – 1590.
- [63] Cortis C M, Friesner R A. Numerical simulation of the Poisson-Boltzmann equation using tetrahedral finite-element meshes [J]. *J Comput Chem*, 1997, **18**:1591 – 1608.
- [64] Baker N A, Sept D, Joseph S, Holst M J, McCammon J A. Electrostatics of nanosystems: Application to microtubules and the ribosome[J], *Proceedings of the National Academy of Sciences*, 2001, **98**:10037 – 10041.
- [65] Holst M J, Baker N A, Wang F. Adaptive multilevel finite element solution of the Poisson-Boltzmann equation I: Algorithms and examples[J]. *Journal of Computational Chemistry*, 2000, **21**(15):1319 – 1342.
- [66] Baker N A, Holst M J, Wang F. Adaptive multilevel finite element solution of the Poisson-Boltzmann equation II: Refinement schemes based on solvent accessible surfaces[J]. *Journal of Computational Chemistry*, 2000, **21**(15):1343 – 1352.
- [67] Tu B, Chen M X, Xie Y, Zhang L B, Eisenberg B, Lu B Z. A parallel finite element simulator for ion transport through three-dimensional ion channel systems[J]. *J Comp Chem*, 2013, **34**:2065 – 2078.
- [68] Xie Y, Tu B, Lu B Z, Zhang L B. Parallel finite element algorithms and adaptive mesh generation for solving nonlinear Poisson-Boltzmann equations[J]. *Journal of Software*, 2013, **24**:110 – 117.
- [69] Chen L, Holst M J, Xu J C. The finite element approximation of the nonlinear Poisson-Boltzmann equation[J]. *SIAM Journal on Numerical Analysis*, 2007, **45**(6):2298 – 2320.
- [70] Lu B Z, Holst M J, McCammon J A, Zhou Y C. Poisson-Nernst-Planck equations for simulating biomolecular diffusion-reaction processes I: Finite element solutions[J]. *Journal of Computational Physics*, 2010, **229**:6979 – 6994.
- [71] Holst M J, Saied F. Multigrid solution of the Poisson-Boltzmann equation[J]. *Journal of Computational Chemistry*, 1993, **14**(1):105 – 113.
- [72] Xie D X, Zhou S. A new minimization protocol for solving nonlinear Poisson-Boltzmann mortar finite element equation[J]. *BIT Numerical Mathematics*, 2007, **47**(4):853 – 871.
- [73] Xie D X. New solution decomposition and minimization schemes for Poisson-Boltzmann equation in calculation of biomolecular electrostatics[J]. *J Comp Phys*, 2014, **275**:294 – 309.
- [74] Shestakov A I, Milovich J L, Noy A. Solution of the nonlinear Poisson-Boltzmann equation using pseudo-transient continuation and the finite element method[J]. *Journal of Colloid and Interface Science*, 2002, **247**(1):62 – 79.
- [75] Holst M J, Vandewalle S. Schwarz methods: To symmetrize or not to symmetrize[J]. *SIAM Journal on Numerical Analysis*, 1997, **34**(2):699 – 722.
- [76] Holst M J, Saied F. Numerical solution of the nonlinear Poisson-Boltzmann equation: Developing more robust and efficient methods[J]. *Journal of Computational Chemistry*, 1995, **16**(3):337 – 364.
- [77] Luo R, David L, Gilson M K. Accelerated Poisson-Boltzmann calculations for static and dynamic systems[J]. *Journal of Computational Chemistry*, 2002, **23**(13):1244 – 1253.
- [78] Holst M J. Applications of domain decomposition and partition of unity methods in physics and geometry [J]. *Domain decomposition methods*, 2010: 63 – 78.
- [79] Aksoylu B, Holst M J. Optimality of multilevel preconditioners for local mesh refinement in three dimensions[J]. *SIAM Journal on Numerical Analysis*, 2006, **44**(3):1005 – 1025.
- [80] Bond S D, Chaudhry J H, Cyr E C, Olson L N. A first-order system least-squares finite element method for the Poisson-

- Boltzmann equation[J]. *Journal of Computational Chemistry*, 2010, **31**(8):1625 – 1635.
- [81] Chaudhry J H, Bond S D, Olson L N. A weighted adaptive least-squares finite element method for the Poisson-Boltzmann equation[J]. *Applied Mathematics and Computation*, 2012, **218**(9):4892 – 4902.
- [82] Lu B Z, Zhou Y C, Huber G A, Bond S D, Holst M J, McCammon J A. Electrodifffusion: A continuum modeling framework for biomolecular systems with realistic spatiotemporal resolution[J]. *J Chem Phys*, 2007, **127**(13):102 – 135.
- [83] Li B, Lu B Z, Wang Z M, McCammon J A. Solutions to a reduced Poisson-Nernst-Planck system and determination of reaction rates[J]. *Physica A*, 2010, **389**:1329 – 1345.
- [84] Lu B Z, Zhou Y C. Poisson-Nernst-Planck equations for simulating biomolecular diffusion-reaction processes II: Size effects on ionic distributions and diffusion-reaction rates[J]. *Biophysical Journal*, 2011, **100**:2475 – 2485.
- [85] Fenley M O, Mascagni M, McClain J, Silalahi R J, Simonov N A. Using correlated Monte Carlo sampling for efficiently solving the linearized Poisson-Boltzmann equation over a broad range of salt concentration[J]. *Journal of Chemical Theory and Computation*, 2010, **6**:300 – 314.
- [86] Mackoy T, Harris R C, Johnson J, Mascagni M, Fenley M O. Numerical optimization of a walk-on-spheres solver for the linear Poisson-Boltzmann equation[J]. *Communications in Computational Physics*, 2013, **13**:195 – 206.
- [87] Mascagni M, Simonov N A. Monte Carlo methods for calculating some physical properties of large molecules[J]. *SIAM Journal on Scientific Computing*, 2004, **26**:339 – 357.
- [88] Simonov N A, Mascagni M, Fenley M O. Monte Carlo-based linear Poisson-Boltzmann approach makes accurate salt-dependent solvation free energy predictions possible[J]. *Journal of Chemical Physics*, 2007, **127**(18):185105.
- [89] James F B. A generalization of algebraic surface drawing[J]. *ACM Transactions on Graphics*, 1982, **1**(3):235 – 256.
- [90] Duncan B S, Olson A J. Shape analysis of molecular surfaces[J]. *Biopolymers*, 1993, **33**(2):231 – 238.
- [91] Connolly M L. Analytical molecular surface calculation[J]. *J Appl Crystallogr*, 1983, **16**(5):548 – 558.
- [92] Vorobjev Y N, Hermans J. SIMS: Computation of a smooth invariant molecular surface[J]. *Biophysical Journal*, 1997, **73**(2):722 – 732.
- [93] Sanner M F, Olson A J, Spehner J C. Reduced surface: An efficient way to compute molecular surfaces[J]. *Biopolymers*, 1996, **38**:305 – 320.
- [94] Ryua J, Parka R, Kimb D S. Molecular surfaces on proteins via β shapes[J]. *Computer-Aided Design*, 2007, **39**(12):1042 – 1057.
- [95] Yu Z Y, Holst M J, McCammon J A. High-fidelity geometric modeling for biomedical applications[J]. *Finite Elements in Analysis and Design*, 2008, **44**(11):715 – 723.
- [96] Tolga Can, Chen C I, Wang Y F. Efficient molecular surface generation using level-set methods[J]. *J Mol Graphics Modell*, 2006, **25**(4):442 – 454.
- [97] Cheng H L, Shi X. Quality mesh generation for molecular skin surfaces using restricted union of balls[J]. *Comput Geom*, 2009, **42**(3):196 – 206.
- [98] Chen M X, Lu B Z. TMSmesh: A robust method for molecular surface mesh generation using a trace technique[J]. *J Chem Theory Comput*, 2011, **7**(1):203 – 212.
- [99] Chen M X, Tu B, Lu B Z. Triangulated manifold meshing method preserving molecular surface topology[J]. *Journal of Molecular Graphics and Modelling*, 2012, **38**:411 – 418.
- [100] Tu B, Bai S Y, Chen M X, Xie Y, Zhang L B, Lu B Z. A software platform for continuum modeling of ion channels based on unstructured mesh[J]. *Computational Science & Discovery*, 2014, **7**:014002.
- [101] Chen M X, Tu B, Lu B Z. Manifold meshing method preserving molecular surface topology[J]. *J Mol Graph Model*, 2012, **38**:411 – 418.
- [102] Fang Q Q, Boas D. Tetrahedral mesh generation from volumetric binary and gray-scale images[C]//IEEE International Symposium on Biomedical Imaging, 2009: 1142 – 1145.
- [103] Zaharescu A, Boyer E, Horaud R P. Transformesh: A topology-adaptive mesh-based approach to surface evolution[C]//The Eighth Asian Conference on Computer Vision, 2007: 166 – 175.
- [104] Si H. Tetgen: A quality tetrahedral mesh generator and a 3D delaunay triangulator[EB/OL]. <http://tetgen.berlios.de>.
- [105] Lu B Z, McCammon J A. Improved boundary element methods for Poisson-Boltzmann electrostatic potential and force calculations[J]. *Journal of Chemical Theory and Computation*, 2007, **3**(3):1134 – 1142.

- [106] Juffer A H, Botta F F, Vankeulen A M, Vanderploeg A, Berendsen J C. The electric-potential of a macromolecule in a solvent—a fundamental approach[J]. *J Comput Phys*, 1991, **97**(1):144–171.
- [107] Boschitsch A H, Fenley M O. Hybrid boundary element and finite difference method for solving the nonlinear Poisson-Boltzmann equation[J]. *Journal of Computational Chemistry*, 2004, **25**(7):935–955.
- [108] Zauhar R J, Morgan R S. A new method for computing the macromolecular electric potential[J]. *Journal of Molecular Biology*, 1985, **186**(4):815–820.
- [109] Lu B Z, Cheng X L, Huang J F, McCammon J A. Order N algorithm for computation of electrostatic interactions in biomolecular systems[J]. *Proceedings of the National Academy of Sciences*, 2006, **103**(51):19314–19319.
- [110] Greengard L F, Huang J F. A new version of the fast multipole method for screened coulomb interactions in three dimensions [J]. *Journal of Computational Physics*, 2002, **180**(2):642–658.
- [111] Bruno O P, Kunyansky L A. A Fast, high-order algorithm for the solution of surface scattering problems: Basic implementation, tests, and applications[J]. *Journal of Computational Physics*, 2001, **169**(1):80–110.
- [112] Kuo S S, Altman M D, Bardhan J P, Tidor B, White J K. Fast methods for simulation of biomolecule electrostatics[C]//ICCAD'02. IEEE/ACM International Conference on Computer-aided Design, 2012: 466–473.
- [113] Hackbusch W, Nowak Z P. On the fast matrix multiplication in the boundary element method by panel clustering[J]. *Numerische Mathematik*, 1989, **54**(4):463–491.
- [114] Tausch J, White J. Multiscale bases for the sparse representation of boundary integral operators on complex geometry[J]. *SIAM Journal on Scientific Computing*, 2003, **24**(5):1610–1629.
- [115] Bharadwaj R, Windemuth A, Sridharan S, Honig B, Nicholls A. The fast multipole boundary element method for molecular electrostatics an optimal approach for large systems[J]. *J Comput Chem*, 1995, **16**(7):898–913.
- [116] Zauhar R J, Varnek A. A fast and space-efficient boundary element method for computing electrostatic and hydration effects in large molecules[J]. *J Comput Chem*, 1996, **17**(7):864–877.
- [117] Bordner A J, Huber G A. Boundary element solution of the linear Poisson-Boltzmann equation and a multipole method for the rapid calculation of forces on macromolecules in solution[J]. *J Comput Chem*, 2003, **24**(3):353–367.
- [118] Boschitsch A H, Fenley M O, Zhou H X. Fast boundary element method for the linear Poisson-Boltzmann equation[J]. *J Phys Chem B*, 2002, **106**(10):2741–2754.
- [119] Lu B Z, Cheng X L, McCammon J A. “New-Version-Fast-Multipole Method” accelerated electrostatic calculations in biomolecular systems[J]. *J Comput Phys*, 2007, **226**(2):1348–1366.
- [120] Altman M D, Bardhan J P, Tidor B, White J K. FFTSV-D: A fast multiscale boundary element method solver suitable for BioMEMS and biomolecule simulation[J]. *IEEE Trans Comput-Aided Des Integr Circuits Syst*, 2006, **25**(2):274–284.
- [121] Ying L, Biros G, Zorin D. A kernel-independent adaptive fast multipole algorithm in two and three dimensions[J]. *J Comput Phys*, 2004, **196**(2):591–626.
- [122] Ong E T, Lim K M, Lee K H, Lee H P. A fast algorithm for three-dimensional potential fields calculation: Fast Fourier transform on multipoles[J]. *J Comput Phys*, 2003, **192**(1):244–261.
- [123] Ong E T, Lee K H, Lim K M. A fast algorithm for three-dimensional electrostatics analysis: Fast Fourier transform on multipoles (FFTM)[J]. *Int J Numer Methods Eng*, 2004, **61**(5):633–656.
- [124] Alpert B, Beylkin G, Coifman R, Rokhlin V. Wavelet-like bases for the fast solutions of second-kind integral equations[J]. *SIAM J Sci Comput*, 1993, **14**(1):159–184.
- [125] Totrov M, Abagyan R. Rapid boundary element solvation electrostatics calculations in folding simulations: Successful folding of a 23-residue peptide[J]. *Biopolymers*, 2001, **60**(2):124–133.
- [126] Ying L, Biros G, Zorin D. A high-order 3d boundary integral equation solver for elliptic PDEs in smooth domains[J]. *J Comput Phys*, 2006, **219**(1):247–275.
- [127] Kuo S H, White J. A spectrally accurate integral equation solver for molecular surface electrostatics[C]//ICCAD'06. IEEE/ACM international Conference on Computer-aided Design, 2006: 899–906.
- [128] Bardhan J P, Altman M D, Willis D J, Lippow S M, Tidor B, White J K. Numerical integration techniques for curved-element discretizations of molecule-solvent interfaces[J]. *Journal of Chemical Physics*, 2007, **127**(1):014701–014718.
- [129] Grandison S, Penfold R, Jean-Marc Vanden-Broecka. A rapid boundary integral equation technique for protein electrostatics [J]. *Journal of Computational Physics*, 2007, **224**(2):663–680.

- [130] Abrashkin A, Andelman D, Orland H. Dipolar Poisson-Boltzmann equation: Ions and dipoles close to charged interfaces[J]. *Phys Rev Lett*, 2007, **99**(1/4): 077801.
- [131] Netz RR, Orland H. Beyond Poisson-Boltzmann: Fluctuation effects and correlation functions[J]. *The European Physical Journal. E: Soft Matter*, 2000, **1**(2/3):203–214.
- [132] Allahyarov E, Gompper G, Lowen H. Attraction between DNA molecules mediated by multivalent ions[J]. *Phys Rev E*, 2004, **69**(4):1–13.
- [133] Tresset G, Cheng C D, Tan Y, Boulaire J, Lam Y M. Phospholipid-based artificial viruses assembled by multivalent cations [J]. *Biophysical Journal*, 2007, **93**(2):637–644.
- [134] Bhuiyan L B, Vlachy V, Outhwaite C W. Understanding polyelectrolyte solutions: Macroion condensation with emphasis on the presence of neutral co-solutes[J]. *International Reviews in Physical Chemistry*, 2002, **21**(1):1–36.
- [135] Lamperski S, Bhuiyan L B. Counterion layering at high surface charge in an electric double layer. Effect of local concentration approximation[J]. *Journal of Electroanalytical Chemistry*, 2003, **540**:79–87.
- [136] Borukhov I, Andelman D, Orland H. Steric effects in electrolytes: A modified Poisson-Boltzmann equation[J]. *Phys Rev Lett*, 1997, **79**(3):435–438.
- [137] Xie D X, Volkmer H W. A modified nonlocal continuum electrostatic model for protein in water and its analytical solutions for ionic Born models[J]. *Communications in Computational Physics*, 2013, **13**(1):174–194.
- [138] Naji A, Kanduc M, Forsman J, Podgornik R. Perspective: Coulomb fluids-Weak coupling, strong coupling, in between and beyond[J]. *The Journal of Chemical Physics*, 2013, **139**:150901.
- [139] Göddeke D, Strzodka R, Mohd Yusof, McCormick P, Buijssen S, Grajewski M, Tureka S. Exploring weak scalability for FEM calculations on a GPU-enhanced cluster[J]. *Parallel Comput*, 2007, **33**:685–699.
- [140] OpenMP Specification[EB/OL]. <http://openmp.org/wp/about-openmp>.
- [141] Message Passing Interface (MPI)[EB/OL]. <http://www.mcs.anl.gov/research/projects/mpi>.
- [142] Blumofe R D, Joerg C F, Kuszmaul B C, Leiserson C E, Randall K H, Zhou Y. Cilk: An efficient multithreaded runtime system[C] // *Proceedings of the Fifth ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP)*, 1995: 207–216.
- [143] Hwang F N, Cai S R, Shao Y L, Wu J S. Parallel Newton-Krylov-Schwarz algorithms for the three-dimensional Poisson-Boltzmann equation in numerical simulation of colloidal particle interactions[J]. *Computer Physics Communications*, 2010, **181**:1529–1537.
- [144] Hsieh M J, Luo R. Exploring a coarse-grained distributive strategy for finite-difference Poisson-Boltzmann calculations[J]. *Journal of Molecular Modeling*, 2011, **17**:1985–1996.
- [145] Li C, Li L, Zhang J, Alexov E. Highly efficient and exact method for parallelization of grid-based algorithms and its implementation in DelPhi[J]. *Journal of Computational Chemistry*, 2012, **33**:1960–1966.
- [146] Lian Y Y, Hsu K H, Shao Y L, Lee Y M, Jeng Y W, Wu J S. Parallel adaptive mesh-refining scheme on a three-dimensional unstructured tetrahedral mesh and its applications[J]. *Computer Physics Communications*, 2006, **175**:721–737.
- [147] Zhang L B. A parallel algorithm for adaptive local refinement of tetrahedral meshes using bisection[J]. *Numer Math Theory Methods Appl*, 2009, **2**:65–89.
- [148] Yokota R, Narumi T, Sakamaki R, Kameoka S, Obi S, Yasuoka K. Fast multipole methods on a cluster of GPUs for the meshless simulation of turbulence[J]. *Computer Physics Communications*, 2009, **180**(11):2066–2078.
- [149] Geng W H. Parallel higher-order boundary integral electrostatics computation on molecular surfaces with curved triangulation [J]. *J Comput Phys*, 2013, **241**:253–265.
- [150] Geng W H, Jacob F. A GPU-accelerated direct-sum boundary integral Poisson-Boltzmann solver[J]. *Comp Phys Comm*, 2103, **184**(6):1490–1496.
- [151] Geng W H, Krasny R. A treecode-accelerated boundary integral Poisson-Boltzmann solver for solvated biomolecules[J]. *J Comp Phys*, 2013, **247**:62–78.
- [152] Christopher D C, Bardhan J P, Barba L A. A biomolecular electrostatics solver using Python, GPUs and boundary elements that can handle solvent-filled cavities and Stern layers[J]. *Computer Physics Communications*, 2014, **185**(3):720–729.
- [153] Delano W L. The PyMOL molecular graphics system[EB/OL]. <http://www.pymol.org/>.

- [154] Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics[J]. J Mol Graphics, 1996, **14**(1):33–38.
- [155] Pettersen E F, Goddard T D, Huang C C, Couch G S, Greenblatt D M, Meng E C, Ferrin T E. UCSF Chimera-A visualization system for exploratory research and analysis[J]. J Comput Chem, 2004, **25**(13):1605–1612.
- [156] Nicholls A, Bharadwaj R, Honig B. Grasp: Graphical representation and analysis of surface properties[J]. J Biophys, 1995, **64**:166–167.
- [157] Bai S Y, Lu B Z. VCMM: A visual tool for continuum molecular modeling[J]. Journal of Molecular Graphics and Modelling, 2014, **50**:44–49.
- [158] Dominy B N, Brooks C L. Development of a generalized Born model parametrization for proteins and nucleic acids[J]. J Phys Chem B, 1999, **103**:3765–3773.
- [159] Nymeyer H, Garcia A E. Free in PMC simulation of the folding equilibrium of α -helical peptides: A comparison of the generalized Born approximation with explicit solvent[J]. Proc Natl Acad Sci USA, 2003, **100**:13934–13949.
- [160] Kirkwood J G. Theory of solution of molecules containing widely separated charges with special application to zwitterions[J]. J Chem Phys, 1934, **2**:351–361.
- [161] Sigalov G, Scheffel P, Onufriev A. Incorporating variable dielectric environments into the generalized Born model[J]. J Chem Phys, 2005, **122**(9):094511–094511.
- [162] Still W C, Tempczyk A, Hawley R C, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics[J]. J Am Chem Soc, 1990, **112**:6127–6129.
- [163] Onufriev A. The generalized Born model: Its foundation, application, and limitations[EB/OL]. <http://people.cs.vt.edu/onufriev/PUBLICATIONS/gbreview.pdf>.
- [164] Constanciel R, Contreras R. Self consistent field theory of solvent effects representation by continuum models: Introduction of desolvation contribution[J]. Theoretica Chimica Acta, 1984, **65**(1):1–11.
- [165] Onufriev A, Case D A, Bashford D. Effective Born radii in the generalized Born approximation: The importance of being perfect[J]. Journal of Computational Chemistry, 2002, **23**(14):1297–1304.
- [166] Onufriev A, Bashford D, Case D A. Exploring protein native states and large-scale conformational changes with a modified generalized Born model[J]. Proteins, 2004, **55**(2):383–394.
- [167] Romanov A N, Jabin S N, Martynov Y B, Sulimov A V, Grigoriev F V, Sulimov V B. Surface generalized Born method: A simple, fast, and precise implicit solvent model beyond the Coulomb approximation[J]. Journal of Physical Chemistry A, 2004, **108**(43):9323–9327.
- [168] Bashford D, Case A D. Generalized Born models of macromolecular solvation effects[J]. Annu Rev Phys Chem, 2000, **51**:129–152.
- [169] Lee M S, Salsbury F R, Brooks C L. Novel generalized Born methods[J]. J Chem Phys, 2002, **116**:10606–10614.
- [170] Grycuk T. Deficiency of the Coulomb-field approximation in the generalized Born model: An improved formula for Born radii evaluation[J]. Journal of Chemical Physics, 2003, **119**(9):4817–4826.
- [171] Srinivasan J, Trevathan M, Beroza P, Case D L. Application of a pairwise generalized Born model to proteins and nucleic acids: Inclusion of salt effects[J]. Theor Chem Acta, 1999, **101**:426–434.
- [172] Chen J, Im W, Brooks C L. Balancing solvation and intramolecular interactions: Toward a consistent generalized Born force field[J]. J Am Chem Soc, 2006, **128**(11):3728–3736.
- [173] Jang S, Kim E, Pak Y. All-atom level direct folding simulation of a $\beta\beta\alpha$ miniprotein[J]. The Journal of Chemical Physics, 2008, **128**(10):102–105.
- [174] Jagielska A, Scheraga H A. Influence of temperature, friction, and random forces on folding of the B-domain of staphylococcal protein A: All-atom molecular dynamics in implicit solvent[J]. Journal of Computational Chemistry, 2007, **28**(6):1068–1082.
- [175] Lopes A, Alexandrov A, Bathelt C, Archontis G, Simonson T. Computational sidechain placement and protein mutagenesis with implicit solvent models[J]. Proteins, 2007, **67**(4):853–867.
- [176] Felts A K, Gallicchio E, Chekmarev D, Paris K A, Friesner R A, Levy R M. Prediction of protein loop conformations using the AGBNP implicit solvent model and torsion angle sampling[J]. J Chem Theory Comput, 2008, **4**(5):855–868.
- [177] Amaro R E, Cheng X, Ivanov I, Xu D, Mccammon J A. Characterizing loop dynamics and ligand recognition in human and avian-type influenza neuraminidases via generalized Born molecular dynamics and end-point free energy calculations[J].

- Journal of the American Chemical Society, 2009, **131**(13):4702 – 4709.
- [178] Chocholousová J, Feig M. Implicit solvent simulations of DNA and DNA-protein complexes: Agreement with explicit solvent vs experiment[J]. J Phys Chem B, 2006, **110**(34):17240 – 17251.
- [179] Ruscio J Z, Onufriev A. A computational study of nucleosomal DNA flexibility[J]. Biophys J, 2006, **91**(11):4121 – 4132.
- [180] Kollman P A, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case D A, Cheatham T E. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models[J]. Acc Chem Res, 2000, **33**:889 – 897.
- [181] Wang J M, Hou T J, Xu X J. Recent advances in free energy calculations with a combination of molecular mechanics and continuum models[J]. Curr Comput Aided Drug Des, 2006, **2**:287 – 306.
- [182] Meirovitch H. Recent developments in methodologies for calculating the entropy and free energy of biological systems by computer simulation[J]. Curr Opin Struct Biol, 2007, **17**:181 – 186.
- [183] Boresch S, Tettinger F, Leitgeb M, Karplus M. Absolute binding free energies: A quantitative approach for their calculation [J]. J Phys Chem B, 2003, **107**:9535 – 9551.
- [184] Kuhn B, Gerber P, Schulz-Gasch, Stahl M. Validation and use of the MM-PBSA approach for drug discovery[J]. J Med Chem, 2005, **48**:4040 – 4048.
- [185] Yang T Y, Wu J C, Yan C L, Wang Y F, Luo R, Gonzales M B, Dalby K N, Ren P Y. Virtual screening using molecular simulations[J]. Proteins, 2011, **79**:1940 – 1951.
- [186] Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. The accuracy of binding free energy calculations based on molecular dynamics simulations[J]. J Chem Info Model, 2011, **51**:69 – 82.
- [187] Hou T, Wang J, Li Y, Wang W. Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking[J]. J Comput Chem, 2011, **32**:866 – 877.
- [188] Froloff N, Windemuth A, Honig B. On the calculation of binding free energies using continuum methods: Application to MHC class I protein-peptide interactions[J]. Protein Sci, 1997, **6**(6):1293 – 1301.
- [189] Swanson J M, Henchman R H, McCammon J A. Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy[J]. Biophys J, 2004, **86**:67 – 74.
- [190] Brock K, Talley K, Coley K, Kundrotas P, Alexov E. Optimization of electrostatic interactions in protein-protein complexes [J]. Biophys J, 2007, **93**:3340 – 3352.
- [191] Rutkoski T J, Raines R T. Evasion of ribonuclease inhibitor as a determinant of ribonuclease cytotoxicity[J]. Curr Pharm Biotechnol, 2008, **9**:185 – 199.
- [192] Ardel T, Ardel B, Drzynekiewicz Z. Ribonucleases as potential modalities in anticancer therapy[J]. Eur J Pharmacol, 2009, **625**(1/3):181 – 189.
- [193] Buckle A M, Schreiber G, Fersht A R. Protein-protein recognition: Crystal structural analysis of a barnase-barstar complex at 2.0 Å resolution[J]. Biochemistry, 1994, **33**(30):8878 – 8889.
- [194] Schreiber G, Fersht A R. Energetics of protein-protein interactions: Analysis of the barnase-barstar interface by single mutations and double mutant cycles[J]. J Mol Biol, 1995, **248**(2):478 – 486.
- [195] Frisch C, Schreiber G, Johnson C M, Fersht A R. Thermodynamics of the interaction of barnase and barstar: Changes in free energy versus changes in enthalpy on mutation[J]. J Mol Biol, 1997, **267**(3):696 – 706.
- [196] Dong F, Vijayakumar M, Zhou H X. Comparison of calculation and experiment implicates significant electrostatic contributions to the binding stability of barnase and barstar[J]. Biophys J, 2003, **85**(1): 49 – 60.
- [197] Wang T, Tomic S, Cabdoulle R R, Wade R C. How optimal are the binding energetics of barnase and barstar? [J]. Biophys J, 2004, **87**(3):1618 – 1630.
- [198] Warwicker J, Ollis D, Richards F M, Steitz T A. Electrostatic field of the large fragment of Escherichia coli DNA polymerase [J]. J Mol Biol, 1985,**186**: 645 – 649.
- [199] Qin S B, Zhou H X. Do electrostatic interactions destabilize protein-nucleic acid binding? [J]. Biopolymers, 2007, **86**:112 – 118.
- [200] Szklarczyk O, Zuberek J, Antosiewicz J M. Poisson-Boltzmann model analysis of binding mRNA cap analogues to the translation initiation factor eIF4E[J]. Biophys Chem, 2009, **140**:16 – 23.

- [201] Madura J D, Davis M E, Gilson M K, Wade R, Luty B A, McCammon J A. Biological applications of electrostatics calculations and Brownian dynamics simulations[J]. *Rev Comput Chem*, 1994, **5**:229 – 267.
- [202] Madura J D, Briggs J M, Wade R C, Davis M E, Luty B A, Ilin A, Antosiewicz J, Gilson M K, Bagheri B, Scott L R, McCammon J A. Electrostatics and diffusion of molecules in solution: Simulations with the University of Houston Brownian Dynamics program[J]. *Comput Phys Commun*, 1995, **91**:57 – 95.
- [203] Friedman H L. A hydrodynamic effect in the rates of diffusion-controlled reactions[J]. *J Phys Chem*, 1996, **70**:3931 – 3933.
- [204] Sharp K, Fine R, Honig B. Computer simulations of the diffusion of a substrate to an active site of an enzyme[J]. *Science*, 1987, **236**:1460 – 1463.
- [205] Sines J J, Allison S A, McCammon J A. Point charge distributions and electrostatic steering in enzyme/substrate encounter: Brownian dynamics of modified copper/zinc superoxide dismutases[J]. *Biochemistry*, 1990, **29**(40):9403 – 9412.
- [206] Argese E, Viglino P, Rotilio G, Scarpa M, Rigo A. Electrostatic control of the rate-determining step of the copper, zinc superoxide dismutase catalytic reaction[J]. *Biochemistry*, 1987, **26**(11):3224 – 3228.
- [207] Getzoff E D, Cabelli D E, Fisher C L, Parge H E, Viezzoli M S, Banci L, Hallewell R A. Faster superoxide dismutase mutants designed by enhancing electrostatic guidance[J]. *Nature*, 1992, **358**:347 – 351.
- [208] Quinn D M. Acetylcholinesterase: Enzyme structure, reaction dynamics, and virtual transition states[J]. *Chem Rev*, 1987, **87**:955 – 979.
- [209] Antosiewicz J, Wlodek S T, McCammon J A. Acetylcholinesterase: Role of the enzyme's charge distribution in steering charged ligands toward the active site[J]. *Biopolymers*, 1996, **39**(1):85 – 94.
- [210] Dillet V, Van Etten, Bashford D. Stabilization of charges and protonation states in the active site of the protein tyrosine phosphatases: A computational study[J]. *J Phys Chem B*, 2000, **104**:11321 – 11333.
- [211] Kolmodin K, Åqvist J. The catalytic mechanism of protein tyrosine phosphatases revisited[J]. *FEBS Lett*, 2001, **498**(2/3):208 – 213.
- [212] Schreiber G, Fersht A R. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering[J]. *Biochemistry*, 1993, **32**:5145 – 5150.
- [213] Zauhar R J. The incorporation of hydration forces determined by continuum electrostatics into molecular mechanics simulations [J]. *Comput Chem*, 1991, **12**:575 – 583.
- [214] Lu B Z, Wang C X, Chen W Z, Wan S Z, Shi Y Y. A stochastic dynamics simulation study associated with hydration force and friction memory effect[J]. *J Phys Chem*, 2000, **104**(29):6877 – 6883.
- [215] Lu B Z, Chen W Z, Wang C X. 边界元法与广义 Langevin 动力学相结合的模拟方法[J]. *科学通报*, 2000, **45**(14):1482 – 1486.
- [216] Lu B Z, Chen W Z, Wang C X, Xu X J. Protein molecular dynamics with electrostatic force entirely determined by a single Poisson-Boltzmann calculation[J]. *Proteins*, 2002, **48**(3):497 – 504.
- [217] Wan S Z, Wang C X, Xiang Z X, Shi Y Y. Stochastic dynamics simulation of alanine dipeptide: Including solvation interaction determined by boundary element method[J]. *J Phys Chem*, 1997, **18**:1440 – 1449.
- [218] David L, Luo R, Gilson M K. Comparison of generalized Born and Poisson models, energetics and dynamics of HIV protease [J]. *Comp Chem*, 2000, **21**(4):295 – 309.
- [219] Wang J, Tan C H, Emmanuel C, Luo R. Quantitative analysis of Poisson-Boltzmann implicit solvent in molecular dynamics [J]. *Phys Chem Chem Phys*, 2010, **12**:1194 – 1202.
- [220] Geng W H, Wei G W. Multiscale molecular dynamics via the matched interface and boundary (MIB) method[J]. *Journal of Computational Physics*, 2011, **230**:435 – 457.
- [221] Xin W D, Juffer A H. A boundary element formulation of protein electrostatics with explicit ions [J]. *Journal of Computational Physics*, 2007, **223**:416 – 435.
- [222] Chen J, Brooks C L, Khandogin J. Recent advances in implicit solvent-based methods for biomolecular simulations[J]. *Curr Opin Structural Biol*, 2008, **18**(2):140 – 148.
- [223] Nielson J E, Vriend G. Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based pKa calculations[J]. *Proteins*, 2001, **43**:403 – 412.
- [224] Georgescu R, Alexov E, Gunner M. Combining conformational flexibility and continuum electrostatics for calculating pKa in

- proteins[J]. *Biophysical Journal*, 2002, **83**:1731 – 1748.
- [225] Lee M S, Salsbury F R, Brooks C L. Constant-pH molecular dynamics using continuous titration coordinates[J]. *Proteins*, 2004, **56**(4):738 – 752.
- [226] Mongan J, Case D A, McCammon J A. Constant pH molecular dynamics in generalized Born implicit solvent[J]. *J Comput Chem*, 2004, **25**(16):2038 – 2048.
- [227] Haas M A, Yilmaz A, Christensen H E, Led J J. Histidine side-chain dynamics and protonation monitored by ¹³C CPMG NMR relaxation dispersion[J]. *J Biomol NMR*, 2009, **44**:225 – 233.
- [228] Dullweber F, Stubbs M T, Musil D, Sturzebecher J, Klebe G. Factorising ligand affinity: A combined thermodynamic and crystallographic study of trypsin and thrombin inhibition[J]. *J Mol Biol*, 2001, **313**:593 – 614.
- [229] Czodrowski P, Sotriffer C A, Klebe G. Protonation changes upon ligand binding to trypsin and thrombin: Structural interpretation based on pKa calculations and ITC experiments[J]. *J Mol Biol*, 2007, **367**:1347 – 1356.
- [230] Horn R. Conversation between voltage sensors and gates of ion channels[J]. *Biochemistry*, 2000, **39**:15653 – 15658.
- [231] Grabe M, Lecar H, Jan Y N, Jan L Y. A quantitative assessment of models for voltage-dependent gating of ion channels[J]. *Proc Natl Acad Sci USA*, 2004, **101**:17640 – 17645.

Electrostatic Calculation in Biomolecular Modeling

PENG Bo, LI Hanlin, LU Benzhuo

(*The State Key Laboratory of Scientific and Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China*)

Abstract: Biological systems could be investigated at molecular level by simulating interactions of molecules. Among various molecular interaction, electrostatic interaction is of central importance to understand solvation properties of biomolecules, and to explain solvation effects such as folding, binding, enzyme catalysis and so on. In this work, Poisson-Boltzmann model for biomolecular solvation calculation is reviewed, including background, development, theories, numerical calculations and applications in molecular biology. Besides, as an approximated model, generalized Born model is discussed.

Key words: Poisson-Boltzmann equation; generalized Born model; solvation; boundary element; finite element