

密级:



中国科学院大学
University of Chinese Academy of Sciences

博士学位论文

高效分子模拟方法的发展及在生物体系的应用研究

作者姓名: 王 进 安

指导教师: 朱 维 良 研究员

中国科学院上海药物研究所

学位类别: 理学 博士

学科专业: 药物设计学

培养单位: 中国科学院上海药物研究所

2013 年 5 月

**The development of highly efficient molecular dynamic
simulation methods and their application in biological
systems**

By

Jin'an Wang

A Dissertation Submitted to

University of Chinese Academy of Sciences

In partial fulfillment of the requirement

For the degree of

Doctor of Drug Design

Shanghai Institute of Materia Medica,

Chinese Academy of Science, Shanghai, 201203

May 2013

目录

目录.....	I
摘要.....	IV
Abstract.....	VII
第一章 可极化力场及增强型模拟方法的发展与应用	1
1.1 引言.....	1
1.2 分子动力学模拟的基本原理.....	2
1.3 可极化力场的发展.....	3
1.3.1 浮动电荷模型.....	4
1.3.2 经典 Drude oscillator 模型.....	6
1.3.3 诱导偶极模型.....	7
1.3.4 基于量化电荷的可极化方法.....	8
1.3.5 其他的基于量化方法的可极化方法.....	9
1.4 增强型取样方法的发展.....	10
1.4.1 基于温度的副本互换方法.....	11
1.4.2 基于 Hamilton 的副本互换方法.....	12
1.4.3 单轨迹副本互换方法.....	12
1.4.4 基于准动态 (metadynamic) 的模拟方法	13
1.5 小结.....	14
参考文献.....	16
第二章 基于显隐性水模型速度优化的副本互换动力学方法发展	27
2.1 引言.....	27
2.2 研究方法.....	31
2.2.1 理论计算方法.....	31
2.2.2 方法测试.....	33

2.3	结果与讨论	34
2.3.1	改进的副本互换方法的效率	34
2.3.2	改进的副本互换方法的准确性	36
2.4	小结	38
	参考文献	39
第三章	简振模式分析与伞取样研究蛋白质构象变化方法发展	45
3.1	引言	45
3.2	研究方法	48
3.2.1	简振模式分析	48
3.2.2	使用 NMA 产生构象变化的中间结构	49
3.2.3	伞取样模拟	50
3.3	结果与讨论	51
3.3.1	AdK 的构象变化机制	51
3.3.2	CaM 构象变化机制研究	55
3.3.3	激酶 p38 α 构象变化机制研究	57
3.4	小结	61
	参考文献	63
第四章	基于 QM/MM 研究蛋白和配体间的可极化效应	70
4.1	引言	70
4.2	研究方法	72
4.2.1	QM/MM 电荷的计算	72
4.2.2	分子动力学模拟	74
4.2.3	MM/GBSA 计算结合自由能	77
4.3	结果与讨论	78
4.3.1	雌激素受体与 W23 和 W23H 结合动力学分析	78
4.3.2	MM/GBSA 计算雌激素受体与 W23 和 W23H 的结合自由能	81

4.3.3	BTN 与亲和素及突变体 F130L 的动力学分析	83
4.3.4	MM/GBSA 计算 BTN 与亲和素及突变体 F130L 的结合自由能	87
4.4	小结	87
	参考文献	89
第五章 铵离子通道 AmtB 转运氨和甲胺分子机制的模拟研究		95
5.1	引言	95
5.2	模型和方法	98
5.2.1	利用 Gromacs3.3.1 模拟 AmtB 及突变体 W212 和 S263 的方法	98
5.2.2	利用 Gromacs4.5.3 模拟 AmtB 组氨酸突变体的方法	100
5.2.3	伞取样模拟	101
5.3	结果与讨论	102
5.3.1	野生型 AmtB 传导甲胺和氨分子	102
5.3.2	残基 W212 的重要作用	106
5.3.3	残基 Ser263 及水分子的作用	110
5.3.4	AmtB 入口动力学行为	111
5.3.5	配体分子离开通道的机制	112
5.3.6	组氨酸对体内活性的影响	113
5.3.7	动力学模拟组氨酸突变体通过配体机制	116
5.3.8	突变体 H168A 传导氨和甲胺的异同	118
5.3.9	突变体 H318A 传导氨和甲胺的异同	120
5.3.10	双突变 H318A/H168A 失活的机制研究	122
5.4	小结	125
	参考文献	127
发表论文		135
致谢		137

摘要

分子动力学模拟作为在原子水平上研究蛋白质动态行为的重要理论工具，近年来随着计算机硬件的快速发展，在蛋白质结构功能模拟、药物设计等多个领域取得了重要进展。然而分子动力学模拟的应用仍然受到各种限制，如分子力场的准确性，取样方法的有效性等等。围绕这些问题，本论文主要开展了如下 4 个方面的工作：1) 发展了一种基于显隐性水模型的副本互换动力学方法并将之用于蛋白质折叠的研究；2) 发展了将简振模式分析和伞取样模拟联合使用的方法，来研究蛋白质构象变化的路径及相应的自由能变化；3) 利用量子力学/分子力学 (QM/MM) 联用方法研究配体与蛋白质结合过程中起重要作用的极化效应；4) 利用分子动力学模拟研究铵离子通道 (AmtB) 转运氨和甲胺的分子机制。

副本互换是目前动力学模拟研究蛋白质折叠的最常用的方法，其缺点在于需要消耗大量的计算资源，从而限制了这一方法在较大蛋白质体系的应用。本论文第 2 章，针对这一问题，我们发展了一种基于显隐性水模型的速度优化的副本互换方法，可以有效减少所需副本数目，从而明显减少所需的计算资源。整个模拟体系是在真实的显性水溶液中进行，而以蛋白质在隐式水模型中的势能作为交换的判据，然后通过速度调节消除使用这 2 种水模型所带来的能量差别。此方法在小蛋白 Trp-cage 的折叠模拟中进行测试，结果显示新方法在不影响模拟准确性的情况下可以将副本数目从 40 下降到 10，从而显著地提高了模拟效率。

目前分子动力学模拟所能模拟的时间小于蛋白质大规模构象变化实际发生的时间，因此近年来高效准确模拟蛋白质构象变化的方法得到了广泛关注和发展的。在论文第 3 章中我们发展了一种将简振模式分析和伞取样模拟相结合使用的方法，来研究蛋白质构象变化的路径及相应的自由能变化。这一方法在三个

典型蛋白体系(磷酸转移酶 Adenylate Kinase (AdK)、钙调蛋白 (CaM) 和人的 MAPK 激酶 p38 α)中进行测试, 获得的计算结果与实验数据高度吻合, 显示此种方法可以普遍适用于蛋白质的构象变化。此外, 模拟结果表明体系 AdK 和 CaM 在没有配体结合时是高度柔性的, 可以在很低的能垒范围内发生较大的构象变化, 而配体的结合能够打破这个平衡, 使体系更倾向于形成结合构象, 因此这 2 个体系的构象变化遵循的是群体移动模型。而激酶 p38 α 中配体的存在则不仅稳定了结合构象, 还降低了 2 种状态转变的能垒, 显示这个体系更可能采用的是诱导契合模型。因此, 现有的结果显示不同蛋白的构象变化可能采取不同的机制, 以适应不同蛋白质的功能。

蛋白质与配体的相互作用在蛋白质结构功能研究中起重要作用。而静电极化是蛋白质-配体的作用的一个重要组成部分, 因此近年来受到越来越多的关注。在本论文第 4 章中, 不同于经典分子动力学模拟中常用的固定的点电荷, 我们利用 QM/MM 方法获得了包含有原子间极化效应的原子电荷并将其应用于分子动力学模拟。通过对 2 个体系(雌激素受体和生物素)的结合自由能的计算, 我们发现 QM/MM 电荷可以有效区分这些体系的结合自由能的差异, 而常规的 AMBER03 力场则不能, 显示了极化效应对 MM/GBSA 计算具有重要意义。

AmtB是酵母中将氨从细胞外转运进细胞内的重要通道蛋白之一。实验上, 研究AmtB转运机制的常用方法为同位素通透实验, 但由于氮的同位素半衰期短, 实际上使用的是氨的类似物甲胺。本论文第5章, 通过分子动力学模拟来研究AmtB转运甲胺和氨的分子机制, 由此考察甲胺作为氨替代物的合理性。模拟结果显示野生型的AmtB传递甲胺和氨的机制是类似的, 即由H168, H318和Y32组成的氢键网络和W212与配体形成的NH- π 作用为通道内配体的传递提供了主要的动力。最终, 位于通道出口处的S263通过与配体形成的氢键将其传递到胞内。另外, 通过分子动力学模拟与生物实验相结合的方法, 重点研究了在AmtB中高度保守的组氨酸(H168和H318)在配体传递中的作用。生物实验结果显示

组氨酸的单突变体（H168A、H318A）都不能传导甲胺，却都可以让酵母生长，即可以转移氨；而双突变体（H168A/H318A）则既不能传递甲胺，也不能转运氨。分子动力学模拟结果显示组氨酸的突变造成通道内疏水性强度增强，从而限制疏水性更强的甲胺使之不能通过，而由于氨的运动受环境疏水性的影响较小，因此仍然可以通过。对应地，平均力势的计算也显示单突变体在传递甲胺时存在一个高达6 kcal/mol的势阱，从而使其失活。另一方面，在双突变失活中则是由于F215和F107组成的开口处的动力学行为发生改变，使通道开口更易于处在张开的状态，从而降低了配体从通道内回到胞外的能垒，失去了通道传导配体的单向选择性，最终导致甲胺和氨的不能传递。这些模拟结果解释了实验研究中所观察到的AmtB野生型及突变体蛋白对甲胺及氨的不同转运能力，成功地从分子和原子水平上阐述了AmtB的功能机制。

关键词：分子动力学模拟，副本互换分子动力学，蛋白质折叠，QM/MM，平均力势，铵离子通道 AmtB，简振模式分析

Abstract

Molecular dynamic (MD) simulation is an efficient tool in the theoretical study on the dynamic mechanics of proteins at atomic level. Recently, based on the rapid development of computational power, MD has been widely used in biological field, such as protein structure and function simulations, drug design and so on. However, the application of MD simulation is still limited by the accuracy of force field and the effectiveness of sampling methods. Around these issues, this doctoral dissertation consists of 4 parts: 1) the development of a replica exchange molecular dynamic (REMD) method with hybrid explicit/implicit solvent; 2) using an approach combining normal mode analysis and umbrella sampling molecular dynamics simulation method, namely NUMD, to predict the conformational transition pathway and to calculate its free energy landscape; 3) studying the important role of polarizing effect in the protein-ligand interaction using QM/MM; 4) studying the conduction mechanism of methylamine and ammonia molecules through AmtB.

REMD is one of the most important methods to investigate the protein folding. The requirement of large computer resource has limited its application in large protein systems. Chapter 2 presents a scheme for REMD of protein simulation in explicit solvent that minimizes the number of required replicas with velocity rescaling approach. This approach relies on a hybrid method where the protein evolves at each temperature in an explicit solvent, but replica exchange moves utilize an implicit solvent term. The explicit and implicit solvent terms are coupled through the velocity rescaling. The application of the new approach on the folding of trp-cage protein shows that the required number of replica has been dramatically decreased from 40 to 10, leading to the significant improvement of simulation efficiency.

As the timescales of protein conformational changes are usually much longer than that the current simulation methods could cope with, it is of great importance to develop a highly efficient method for accurately exploring the intrinsic or substrate induced protein conformational change with acceptable computational resources. In Chapter 3, the NUMD is applied on three typical protein systems composed of up to more than 300 amino acids. The great agreement between our simulation results and the experimentally determined conformations and thermodynamic parameters for all three proteins indicates the general applicability of NUMD and its high reliability in describing protein conformational change pathway, mechanism, and free energy profile. It is observed that the substrate-free phosphotransferase enzyme adenylate kinase (AdK) and calmodulin can sample a wide ensemble of conformations along the minimum free energy transition pathway. The binding of respective substrates to the two proteins biases the sampled ensemble toward the corresponding substrate-bound conformations, suggesting that the conformational changes of AdK and calmodulin follow the population-shift mechanism. On the contrary, the simulation of p38 α mitogen-activated protein kinase shows that the binding of substrate changes the free energy barrier between the two functional states and promote the transition from the substrate-free to substrate-bound state, which is more like an induced-fit mechanism. Therefore, the present study strongly suggests that the detailed mechanism of substrate binding and the associated conformational transition is not uniform for all kinds of proteins but correlated to their respective biological functions.

The protein-ligand interaction is one of the important questions in the study of protein structure and function. Recently, the important roles of polarization in the protein-ligand binding have been received increasing attention. Chapter 4 describes a

scheme for MD simulation using QM/MM charges to reflect the polarization. This approach has been tested in two systems: estrogen receptor and biotin-avidin interaction. The binding free energy calculated by MM/GBSA shows that the QM/MM charge could distinguish the difference of these systems, but AMBER03 force field could not. The result also reveals the importance of polarization in calculating the binding free energy using MM/GBSA.

AmtB is one of the ammonium transporter proteins facilitating the ammonium transport across the cellular membranes. In experiments, the substrate used in *in vitro* studies is the radio labeled [^{14}C]methylammonium, rather than ammonium itself. In chapter 5 of the thesis, MD simulations are used to explore the conduction mechanism of methylamine and ammonia molecules through AmtB. The results demonstrate that methylamine could be automatically transported in a very similar way as ammonia in wild type AmtB. The driving force for the conduction is mainly the hydrogen bond network comprising H168, H318 and Y32, working in coordination with NH- π interaction with residue W212. Then, S263 translocates the substrates from the exit gate into the cytoplasm by hydrogen bond interaction. Besides, the role of the two conserved histidines (H168, H318) in ammonia and methylamine transport is evaluated by using a combination of *in vivo* studies, MD simulation, and potential of mean force (PMF) calculation. The *in vivo* results show that a single change of either of the conserved histidines to alanine leads to the failure of transporting methylamine but not ammonia, whereas double histidine variant completely loses its ability to transport both methylamine and ammonia. MD simulations clearly shows that a single histidine variant (H168A or H318A) of AmtB confines the rather hydrophobic methylamine more strongly than ammonia around the mutated sites, resulting in the dysfunction in conducting the former but not the

latter molecule. PMF calculations further reveal that the single histidine variants form a potential energy well of up to 6 kcal/mol for methylamine, impairing conduction of this substrate. Unlike the single histidine variants, the double histidine variant, H168A/H318A, of AmtB is found to lose its unidirectional property of transporting both ammonia and methylamine. This could be attributed to a greatly increased frequency of opening of the entrance gate formed by F215 and F107, with a resultant lowering of the energy barrier for substrate to return to the periplasm. These observations could explain the different conduction performance between ammonia and methylamine which is observed in *in vivo* experiments and provide the detailed mechanism of AmtB transporting substrates at atomic level.

Key words: Molecular dynamics simulation, Replica exchange molecular dynamic, Protein folding, QM/MM, Potential of mean force, Ammonium Transporter, Normal mode analysis

第一章 可极化力场及增强型模拟方法的发展与应用

1.1 引言

蛋白质是生命体中最重要的一类大分子，不仅是构成生物体的重要组成成分，而且参与了大部分的生命活动，如催化，调节和免疫等¹。随着对蛋白质结构和功能的研究，逐渐发现蛋白质功能的发挥不仅与其静态结构相关，而且经常需要发生相应的构象变化^{2, 3}。因此，对蛋白质功能的研究已经从传统的蛋白质结构和功能之间的研究，逐渐转为蛋白质结构，动态行为和功能的研究。对蛋白质的动态行为的研究成为连接其结构和功能的桥梁^{2, 4}。

随着X-晶体衍射⁵、核磁共振^{6, 7}和低温电镜⁸等实验技术的迅猛发展，蛋白质及其与配体结合的复合物的三维结构测定不断取得突破。截止到2013年5月7日，在蛋白质晶体数据库⁹(www.pdb.org)中的生物大分子的结构数目已经达到90206。而且在蛋白质晶体数据库中还存在着为数不少的具有不同的结构的同一蛋白质，但是晶体结构只能告诉我们这些静态的结构，却不能获得它们是如何转变的。随着新实验技术的发展，如实时X-晶体衍射方法¹⁰⁻¹²，单分子技术^{13, 14}，多维核磁共振^{6, 7}，原子力显微镜¹⁵等，已经可以获得一些重要蛋白质动态变化的信息。然而，目前的这些实验方法都不是一个直接的方法，而且很多蛋白质起重要生理功能的构象变化的时间尺度常在实验的精度范围之外，因此现有的实验技术并不能在分子和原子水平上详细的阐明蛋白质的结构变化和动力学行为。

近年来随着计算机技术和相应算法的发展，可以直接模拟物质微观特性和探究原子水平相互作用的分子模拟方法得到广泛应用，在蛋白质折叠、蛋白质与药物分子相互作用等领域都取得了巨大的进展。分子动力学模拟方法最早是由Wainwright¹⁶和Rahman¹⁷等人在20世纪30年代前后独立发展起来的。近年来，

随着分子动力学理论和计算机硬件的发展，分子动力学模拟研究体系的空间尺度和时间尺度都大大增加。如今，已经可以在较大的体系和较长的时间尺度上对生物体系的动力学行为进行模拟：模拟体系的规模从最初的 500 多个原子发展到目前的百万个原子以上，时间尺度也从最初的皮秒发展到现在的毫秒级别¹⁸⁻²¹。在生物学的众多领域，如酶催化，药物分子设计和蛋白质设计等发挥着越来越大的作用^{22, 23}。

分子动力学模拟所采用的力场和构象采样方法是分子动力学模拟中最核心的两个方面。本章将简单介绍分子动力学模拟的原理，并着重介绍近年来受到广泛关注的可极化力场和增强型取样方法在生物体系中的应用和发展。

1.2 分子动力学模拟的基本原理

分子动力学模拟是在分子力学基础上对模拟体系的运动随时间演化的一种描述方法。在处理过程中，首先需要有一个表征体系势能的函数，即分子力场，然后通过对体系势能进行求导获得各个原子的受力情况，再通过牛顿力学建立运动学方程，最终得到体系随时间的演化过程。

在分子力学中广泛使用的势能函数通常有如下形式：

$$U = U_{bonded} + U_{nobonded} \quad (\text{式 1.1})$$

如式 1.1 所示，分子势能主要包括两个方面的作用，成键作用力和非键作用力。其中成键作用包括键长，键角和二面角相互作用，而远程作用力主要包括范德华作用力和静电相互作用，如式 1.2 和 1.3 所示。

$$\begin{aligned} U_{bonded} &= U_{bond} + U_{angle} + U_{dihedral} \\ &= \sum_{bond} K_r (r - r_{eq})^2 + \sum_{angle} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedral} \frac{V_n}{2} (1 + \cos[n\phi - \gamma]) \end{aligned} \quad (\text{式 1.2})$$

$$U_{nonbonded} = U_{vdw} + U_{ele} = \sum_{i < j}^{atoms} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \sum_{i < j}^{atoms} \frac{q_i q_j}{\epsilon r_{ij}} \quad (\text{式 1.3})$$

有了分子势能后，就可以通过势能对粒子位置的求导得到每个粒子的受力情况，即

$$F_i = -\frac{\partial U}{\partial r_i} \quad (\text{式 1.4})$$

在分子动力学模拟中，每个粒子都服从经典牛顿力学，因而每个粒子的受力与运动情况都可以用牛顿第二定律来描述：

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} \quad (\text{式 1.5})$$

其中 F_i 为其余所有粒子对于粒子 r_i 的作用力， m_i 为粒子质量， U 为体系中粒子的势能函数。利用式 1.4 和式 1.5，就得到体系随时间演化的轨迹，进而可以得到体系的各种宏观性质的时间平均。

1.3 可极化力场的发展

目前在研究蛋白质结构和功能方面应用得比较广泛的力场有CHARMM²⁴，AMBER²⁵，OPLS²⁶和GROMOS²⁷等。但是这些分子力场都有2个比较大的缺陷，首先是分子力场体现的是原子的性质，不能考虑电子效应；其次就是这些力场中对电荷的描述采取固定点电荷的方式（即每个原子被视为一个中间带点电荷的球体），不能精确考虑原子间的极化效应。而最近的研究表明这种原子间的极化作用在蛋白质和配体的相互作用中起重要作用，如Hensen等²⁸利用量子力学/分子力学联用方法（QM/MM）分别计算了HIV-1蛋白酶和抑制剂 nelfinavir，mozenavir和tripnavir的结合自由能，结果显示酶环境对配体的极化作用占到总静电相互作用能的39%。此外，Ren等^{29, 30}利用可极化力场AMOEBA研究胰蛋白酶与苯甲脒类抑制剂相互作用时，也发现了静电极化作用在这类抑制剂的特异性

结合中起重要作用。

针对普通力场不能有效考虑蛋白周围环境所引起的电荷转移和极化效应所带来的影响的缺陷，近年来可极化力场得到了快速的发展和重视，成为分子力场研究的热点之一³¹⁻³⁴。而且现有的大部分力场都发展了各自的极化的版本，如 AMBER ff02，AMBER ff02 EP^{35, 36}，CHARMM³⁷⁻⁴⁰，PIPF-CHARMM⁴¹，OPLS/PFF⁴²和GROMOS⁴³等。下面根据可极化力场所采用的不同模型，分 4 个部分简要介绍一下。

1.3.1 浮动电荷模型

浮动电荷模型是基于原子电负性均衡原理而发展起来的，其基本原理是电荷可以在体系内各原子之间自由流动，最终会使体系内每个原子的电负性达到收敛。因此，在浮动点电荷模型中，各原子的电荷可以在其他原子的影响下在各原子间转移，从而可以很好地考虑环境的极化效应。

目前使用浮动电荷模型的力场有1992年Rapped等发展的通用力场⁴⁴和Friesner课题组的发展的力场⁴⁵。最近Brooks和Patel等也利用类似思想，将浮动电荷模型引入到CHARMM中，发展了基于浮动电荷的CHARMM极化力场^{39, 40}。国内辽宁师范大学的杨忠志等^{46, 47}也基于浮动电荷的思想，发展了原子-键电负性均衡方法，可以快速准确计算体系随外势变化的电荷分布，并将其应用到有机分子，多肽，蛋白和配体的相互作用的研究中，取得了较好的效果。

在浮动电荷模型中，首先将原子的能量以电荷为参数进行泰勒展开，这里只考虑前面两项：

$$E(Q) = E^0 + Q\left(\frac{\partial E}{\partial Q}\right)_0 + \frac{1}{2}Q^2\left(\frac{\partial^2 E}{\partial Q^2}\right)_0 \quad (\text{式1.6})$$

以一个中性的原子的能量作为参考点，假定当此原子具有+1和-1电荷时的能量分别为E(+1)（电离能，IP）和E(-1)（电子亲和能，EA），将其代入式1.6中，可以得到：

$$\left(\frac{\partial E}{\partial Q}\right)_0 = \frac{1}{2}(IP + EA) = \chi^0 \quad (\text{式1.7})$$

和

$$\left(\frac{\partial^2 E}{\partial Q^2}\right)_0 = IP - EA = J^0 \quad (\text{式1.8})$$

其中 χ^0 为使用Muliken定义的电负性，二阶偏导 $\left(\frac{\partial^2 E}{\partial Q^2}\right)_0$ 表示价键轨道中2个电子之间的库伦斥力常数，即原子中电子间排斥的力常数（原子的硬度），一般使用 $1/2 J^0$ ，此量表征了电子在原子中流动的能力。

因此：单个原子的能量就可以表示为：

$$E(Q) = E^0 + \chi^0 Q + \frac{1}{2} J^0 Q^2 \quad (\text{式1.9})$$

将式1.9推广到的M个具有 N_i 个原子分子体系。此时体系总能量为：

$$E(Q, r) = \sum_{i=1}^M \sum_{\alpha=1}^{N_i} \chi_{i\alpha}^0 Q_{i\alpha} + \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M \left(\sum_{\alpha=1}^{N_i} \sum_{\beta=1}^{N_j} \eta_{i\alpha j\beta} Q_{i\alpha} Q_{j\beta} + V(r_{i\alpha j\beta}) \right) \quad (\text{式1.10})$$

其中 $V(r_{i\alpha j\beta})$ 表示了体系的非静电相互作用，包括键长，键角，二面角，范德华相互作用等力场中的其他项。其中优化的 $Q_{i\alpha}$ 是通过迭代下式得到：

$$\frac{\partial E}{\partial Q_{i\alpha}} = 0, i = 1, N_i \quad (\text{式1.11})$$

即通过原子的电负性作为判断收敛的依据。

从式1.11可以看到，在浮动电荷模型中不仅可以保持分子整体的电性不变，而且电荷可以在原子间自由转移，所以可以很方便的考虑极化效应。同时由于其中每个原子的电荷都是依赖于体系中其他原子的位置和电荷，因此在模拟过程中每一步都得重新计算，最终导致这个方法的计算量是非常巨大的。

浮动电荷模型中使用的原子电负性和弹性系数 J^0 通常来自实验数据或量化

计算获得。此方法的一大缺陷就是如何选择合适的参数，尤其是弹性系数。模拟的结果发现此模型容易高估电荷的转移效应，会出现超远距离的电荷转移效应^{45, 48}。此外，电子转移是在原子对之间完成的，不能考虑它们之间的立体效应，所以不能很好反映平面或线性分子的非平面极化效应，目前主要通过加入一些非平面的位点来解决^{45, 47, 49, 50}。

1.3.2 经典 Drude oscillator 模型

经典Drude oscillator模型，也称为壳层模型，这一模型是通过将每个原子分为2个粒子的方法而引入极化效应⁴⁸。在此模型中将具有净电荷 q_i 的原子分解为具有电荷为 $q_i - q_d$ 的带电核和具有电荷为 q_d 的带电壳（Drude粒子）。二者通过一个平衡距离为 d ，弹性系数为 k_d 的弹簧连接。在模拟过程中 q_i 和 q_d 是固定不变的，而体系对外界环境的极化效应是通过调整这2个粒子之间的距离来实现的。

在此模型中主要通过弹簧的弹性系数，平衡距离和Drude粒子带电大小来反映不同的原子的可极化性。假设原子的可极化率为 α_i ，则它有如下关系：

$$\alpha_i = q_{i,D}^2 / k \quad (\text{式1.12})$$

而这个原子的电荷和弹簧系数是通过拟合分子的可极化率数据和分子之间的相互作用能量等性质获得，不过对不同类型的原子的参数化仍是比较大的问题。

在经典Drude oscillator模型中势能有如下表达式：

$$U(r, d) = U_{self}(d) + U_{bond}(r) + U_{ele}(r, d) + U_{LJ}(r) \quad (\text{式1.13})$$

其中 $U_{self}(d)$ 为原子的带电核与Drude粒子间相互作用($1/2kd^2$)， $U_{bond}(r)$ 为键能， $U_{LJ}(r)$ 为范德华能量，其中 $U_{ele}(r, d)$ 包括了所有粒子（Drude粒子和原子核）的静电相互作用，用式1.14表示：

$$U_{elec} = \sum_i \sum_{j>i} \left[\frac{(q_i - q_{D,i})(q_i - q_{D,i})}{|r_i - r_j|} + \frac{(q_i - q_{D,i})q_{D,j}}{|r_i - r_j - d_j|} + \frac{q_{D,i}(q_j - q_{D,j})}{|r_i + d_i - r_j|} + \frac{q_{D,i}q_{D,j}}{|r_i + d_i - r_j - d_j|} \right] \quad (\text{式1.14})$$

由于d远小于r,所以上式可以简化为:

$$U_{elec} \approx \sum_i \sum_{j>i} \left[\frac{q_i q_j}{r_{ij}} + (q_{D,i} d_i q_j - q_i q_{D,j} d_j) \nabla_i \left(\frac{1}{r_{i,j}} \right) - q_{D,i} d_i q_{D,j} : \nabla_i \nabla_i \left(\frac{1}{r_{i,j}} \right) \right] \quad (\text{式1.15})$$

目前此模型方法已经在CHARMM力场^{37, 38, 51}和GROMOS力场⁴³中得以实现,最近Schulten等人⁵²将Drude模型方法在NAMD中实现,显示了很好的并行效率,并在动力学模拟中得到应用⁵³。

1.3.3 诱导偶极模型

诱导偶极模型中的极化效应是通过产生诱导偶极的方式获得的,此方法也得到了广泛的关注,如Kaminski等人发展的OPLS/PFF⁵⁴, Ren等发展的AMOEBA力场⁵⁵⁻⁵⁷以及Cieplak等发展的AMBER ff02, ff02EP^{35, 36}都是利用这一模型。下面以AMOEBA力场为例,简单介绍一下这种模型。

AMOEBA 力场中处理成键相互作用,范德华相互作用和普通的力场是一致的,而关键的区别在于静电作用的处理上,它采用的是中央多极矩展开形式,如式 1.16 所示。

$$U_{ele}(\mathbf{r}_{ij}) = \mathbf{M}_i^T \mathbf{T}_{ij} \mathbf{M}_j \quad (\text{式 1.16})$$

其中 M 为多极的张量表示,即

$$\mathbf{M}_i = [q_i, \mu_{i,x}, \mu_{i,y}, \mu_{i,z}, Q_{i,xx}, Q_{i,xy}, Q_{i,xz}, \dots, Q_{i,zz}] \quad (\text{式 1.17})$$

q 为电荷,μ 为偶极,Q 为四极, T_{ij} 为原子 i 和 j 之间的多极相互作用张量,有式 1.18 计算所得。

$$T_{ij} = \begin{bmatrix} 1 & \frac{\partial}{\partial x_j} & \frac{\partial}{\partial y_j} & \frac{\partial}{\partial z_j} & L \\ \frac{\partial}{\partial x_j} & \frac{\partial^2}{\partial x_i \partial x_j} & \frac{\partial^2}{\partial x_i \partial y_j} & \frac{\partial^2}{\partial x_i \partial z_j} & L \\ \frac{\partial}{\partial y_j} & \frac{\partial^2}{\partial y_i \partial x_j} & \frac{\partial^2}{\partial y_i \partial y_j} & \frac{\partial^2}{\partial y_i \partial z_j} & L \\ \frac{\partial}{\partial z_j} & \frac{\partial^2}{\partial z_i \partial x_j} & \frac{\partial^2}{\partial z_i \partial y_j} & \frac{\partial^2}{\partial z_i \partial z_j} & L \\ M & M & M & M & O \end{bmatrix} \quad (\text{式 1.18})$$

使用这种中央多极矩展开形式（电荷，偶极，四极等），不仅能更真实地反映原子周围空间的静电势分布情况，而且原子之间的极化作用可以通过诱导多极的方式方便计算，即

$$M_i = M_i^P + M_i^{ind} \quad (\text{式 1.19})$$

$$M_{i,\alpha}^{ind} = \alpha_i \left(\sum_{\{j\}} T_{\alpha}^{ij} M_j^P + \sum_{\{j\}} T_{\alpha}^{ij'} M_j^{ind} \right) \quad (\text{式 1.20})$$

其中 α_i 是原子*i*的极化率，通过式 1.20 迭代计算可以得到原子的诱导多极矩，最终可以计算出原子受到的静电相互作用。目前Ren等利用此AMOEBA力场计算得到的金属离子溶剂化能和蛋白质-配体结合自由能，都与实验结果非常符合^{29,30,58}，显示此种方法的合理性。不过此方法需要对体系内所有的原子进行式 1.20 的迭代计算，在一些体系中其收敛会存在问题，而且可以看出对于大的体系其计算量也是巨大的。

1.3.4 基于量化电荷的可极化方法

随着计算机硬件的快速发展和量子力学方法的进步，目前其能处理的体系大小已经有了很大进展，但使用量子力学来研究整个生物体系还是非常困难。近年来很多研究组致力于发展计算量随体系尺度呈线性增长的线性标度算法，基于分而治之的思想，华东师范大学的张增辉教授等发展了共轭帽分子分割法

(MFCC) 来对生物大分子进行研究，其基本原理是将生物体系划分成很多碎片体系，只需对每个碎片分子体系分别进行量化计算，而每个量化计算的原子数目并不大，所以使用并行算法可以比较大的生物体系。他们成功的将此方法应用到蛋白质配体，DNA与配体等多个体系的结合自由能计算和作用机制的研究⁵⁹。为了与现有的动力学模拟相结合，张增辉课题组发展了基于MFCC-PB的特殊极化蛋白电荷的方法⁶⁰，并将其应用到蛋白质和配体的分子动力学模拟中^{61, 62}。他们发现对生物素（biotin）体系来说，在配体结合自由能中起重要贡献的是静电极化作用，而不是之前认为的范德华相互作用。Ryde课题组⁶³也利用类似的思想考察了特异构象的电荷对配体和蛋白的相互作用的影响，结果也显示了静电极化在MM/PBSA计算中所起的重要作用。

1.3.5 其他的基于量化方法的可极化方法

SIBFA⁶⁴方法是另外一种基于量化方法的可极化力场，在此力场中将势能分为五项， $\Delta E_{\text{int}} = E_{MTP} + E_{rep} + E_{pol} + E_{ct} + E_{disp}$

其中 E_{MTP} 为电多极作用， E_{rep} 为短程排斥能， E_{pol} 为极化化能， E_{ct} 为电荷转移能， E_{disp} 为色散能。

E_{MTP} 是通过将分子划分为分子碎片，通过量化计算得到静电势并将其拟合到各原子或者键的中心，得到静电多极子。然后使用式 1.16 计算得到体系整体的静电相互作用。 E_{rep} 为键-键，键-孤对电子对和孤对电子对之间的相互作用，是通过使用 S^2/R 的函数形式计算得到，其中 S 表征两个相互作用的离域轨道的重叠。 E_{ct} 和 E_{pol} 直接使用在DFT或者HF计算得到。而 E_{disp} 则通过 $1/R^6, 1/R^8, 1/R^{10}$ 的函数形式来计算。Gresh等⁶⁵利用这一可极化力场计算了FAK 激酶与配体的相互作用能，结果发现结合位点中的水在配体和蛋白的相互作用中起了重要的作用。

1.4 增强型取样方法的发展

在分子动力学模拟的发展过程中，其应用的体系大小及模拟的时间尺度都有着长足的进步。如在 1998 年，Duan Y 等²⁰就利用 256 个核的Cray T3D超级计算机对 36 个残基的HP36 小蛋白进行了长达 1 微秒的折叠机制的分子动力学模拟。近年来，随之分子动力学模拟专用计算机Anton及其配套的动力学软件Desmod的发展⁶⁶，分子动力学模拟的尺度得到大大提高。如在 2010 年De Shaw 等⁶⁷就在熔融温度下对WW结构域和小蛋白质FiP35 进行了长达 100 微妙以上的折叠与去折叠研究，在动力学模拟中发现FiP35 的平均折叠时间为 10 ± 3 微秒，这与实验报道的 14 微秒的折叠时间非常符合，显示在描述蛋白质折叠动力学方面的可靠性。

虽然长时间的动力学模拟在蛋白质折叠及构象变化的研究中取得了很大的发展，但是由于蛋白质构象空间巨大，构象变化之间需要跨越较大能垒，而这导致了在常规动力学模拟中要观察到相应的变化所需的模拟时间大大增加。如上面成功的例子，都是利用了超级计算机相当长的模拟时间，所花费的计算资源是极其巨大的。而且随着蛋白质原子数目的增大，所需计算量更是成指数增长，所以目前在室温下成功折叠的蛋白质的残基数都在 1 百个以下。另外，虽然目前的常规动力学模拟可以达到毫秒级别¹⁸，可是对于蛋白质折叠或者大规模的构象变化的时间尺度而言，还是太短。因此近年来理论计算家们致力于各种增强型取样方法的发展⁶⁸⁻⁷⁰，这些增强取样方法可以实现在较短的模拟时间范围内捕捉到蛋白质动态行为，因此极大地缩短模拟所需的时间⁷¹⁻⁷³。在增强型取样方法中，基于广义系综的取样算法尤其引起人们重视⁷⁴。其中副本互换分子模拟方法（replica exchange molecular dynamics (REMD)）⁷⁵就是基于广义系综发展起来的重要增强型取样方法，下面简单介绍一下副本互换的方法。

1.4.1 基于温度的副本互换方法

副本互换的基本思想是利用高温副本得到高能状态的构象，然后通过构象交换机制，将陷于局部极小值的构象交换出来，从而达到跨过能垒的目的，其已经在蛋白质折叠，构象变化及蛋白质与配体相互作用等众多研究领域得到了广泛应用⁷⁶⁻⁸¹。如Hansmann等人利用副本互换方法成功地将具有 67 个残基的Top67 蛋白折叠到与晶体结构非常近似的结构(RMSD~3Å)⁸²。

副本互换的具体做法是让分别具有能量 E_1 和 E_2 的两个副本处在两个不同温度 T_1 和 T_2 ，让它们具有一定的概率进行发生交换。其中这两个副本通过式 1.22 决定是否交换。

$$\omega(1 \leftrightarrow 2) = \min(1, \exp(\Delta\beta\Delta E)) \quad (\text{式 1.22})$$

其中 $\beta = 1/k_B T$ 为温度因子。

REMD方法所需的副本数目是与体系的自由度的平方根成正比，而在真实的模拟体系中，蛋白质一般都在水溶液中，这样体系的自由度比蛋白质本身大很多，所以在常规的REMD所需的副本数目是相当巨大的。为了缩短副本数目，提高副本互换的效率，目前出现了很多加速的方法，其中杂化REMD方法（hybrid REMD）^{83, 84}和溶质温度方法（solute tempering）⁸⁵⁻⁸⁷是比较有代表性的方法。在杂化REMD模拟中，体系的模拟是在真实水溶液中，但在进行交换时，使用的是蛋白在隐性水模型中的势能，从而避免了体系中水的自由度对副本数目的影响，这样体系所需要的副本数目就只与蛋白质原子数目相关，因此大大减少了所需的副本数目，大幅提高交换效率。Okur等⁸³将这一方法应用到ALA₁₀肽链的折叠模拟研究，结果发现只需使用 8 个副本，就可以完成常规REMD40 个副本的效率，而且此方法在体系变大的时候会更加有效。

另一种方法则是溶质温度方法⁸⁵⁻⁸⁷，它与杂化方法有点类似，不过此方法是将体系的势能分成 3 项： $E_{pot} = P_{pp} + P_{pw} + P_{ww}$ 。其中 P_{pp} 为蛋白的势能， P_{pw} 为蛋

白和水相互作用的势能，而 P_{ww} 为体系中水的势能。在此方法中判断是否交换的判据是 $P_{pp} + \frac{1}{2}P_{pw}$ ，这样不仅可以考虑蛋白的势能，还能将蛋白和水之间的相互作用能考虑进去，不仅提高效率，还能提高准确性，实际测试的结果显示此方法也大大减少了计算所需的副本，提高了计算速度。如Huang等人⁸⁶将这一方法应用到 α 螺旋， β -折叠结构和小蛋白Trp-Cage的折叠研究中，发现与常规的副本互换方法相比，都显著减少了副本数目，而获得的能量图景并没有大的差别，显示了这一方法的可靠性和有效性。

1.4.2 基于 Hamilton 的副本互换方法

在常规的副本互换方法中，不同的副本处于不同的温度，这样的方法也称为温度副本互换方法。而在Hamilton 副本互换方法⁸⁸中，不同的副本是在同样的温度下进行模拟，但是每个副本有具有不同的Hamilton形式。其中一种较常见的方式是疏水加速副本互换方法（Hydrophobic aided replica exchange）⁸⁹。

此方法中将体系势能写成下式：

$$U = U_{other} + \lambda(U_{LJ6}^{prot-water} + U_{elec}^{prot-water}) \quad (\text{式 1.23})$$

在不同的副本中 λ 值是不一样的，此值的范围为0到1，在此方法中主要是对疏水作用项 ($U_{LJ6}^{prot-water} + U_{elec}^{prot-water}$) 进行调整。当 $\lambda=1$ 时对应着水溶液的环境，即我们实际感兴趣的环境，而当 $\lambda<1$ 时，则模拟了一个有机溶液环境，在这样的环境中蛋白很容易去折叠。通过让不同的副本处于不同的环境下，通过交换，能够有力的促进能垒的跨越。如Zhou等⁸⁹将这一方法应用到GB1肽链的折叠研究，发现只要6个副本就可以在5ns的模拟时间内将这一蛋白成功折叠，而作为对照的常规的REMD方法，则需要64个副本，从而获得极大的效率提升。

1.4.3 单轨迹副本互换方法

由于常规REMD法需要多个副本同时模拟各个温度的情况，导致巨大的计算量，Ma等⁹⁰基于REMD的原理发展了一种力学积分的方法，可以在一条轨迹里

对所有温度进行模拟，然后通过统计力学的方法得到常温时的几率，从而只需要模拟一个副本，就可以获得所有温度的统计力学性质，从而使计算量大大减少。最近他们将此方法与限制动力学方法相结合⁹¹，成功应用到具有 α 螺旋的蛋白质的折叠研究中，测试了从 67 到 108 个残基的 4 个蛋白，发现它们都可以在微妙级别的时间尺度内成功折叠，且所获得的结构与晶体结构的RMSD都可以达到 2 Å 以下。

1.4.4 基于准动态（metadynamic）的模拟方法

Metadynamic⁹²是用来计算自由能变化的方法，其的基本原理是：当我们研究一个具有N个原子的体系的时候，其中体系的坐标为 $\mathbf{r}$ ， $\mathbf{r} \in R^{3N}$ ，而这个体系的势能为 $U(\mathbf{r})$ 。我们引入一组d集合变量（collective variables, CV），s是在模拟实际得到的这些集合变量的值，有 $s \in R^d$ ，因此以这些变量作为函数的helmholtz自由能函数就可以表达为：

$$F(s) = -\frac{1}{\beta} \ln \int d\mathbf{r} e^{-\beta U(\mathbf{r})} \delta(s - s(\mathbf{r})) + C \quad (\text{式 1.24})$$

其中 C 是一常数。自由能 F(s)包括了体系的热力学的所有信息，因此可以根据式 1.24 利用 CV 的概率来计算获得。此外，适当选取这些参数就以通过非平衡态的模拟得到自由能的变化曲线。

在非限制的常规动力学模拟中计算自由能的一个简单方法是统计这些 CV 出现的概率来获得，如式 1.25 所示。

$$\tilde{F}(s) = -\frac{1}{\beta} \ln N(s) \quad (\text{式 1.25})$$

但是由于能垒的存在，一些中间状态的低概率发生的现象是难以在短时间的常规的动力学模拟中获得，因此需要的模拟时间很长。另外一种更有效的方法是人为的加入一个限制力，是体系在这一构象提高在这一空间中发生的概率，

而非限制的自由能可以通过下式计算得到:

$$\tilde{F}(s) = -\frac{1}{\beta} \ln N(s) - V(s) \quad (\text{式 1.26})$$

因此通过选择合适的CV, 就可以获得这些空间中难以在常规动力学中观察到的现象。如Fabio Pietrucci等⁹³发展了一种新的集合参数, 将其用于 β 片蛋白 SH3 和GB1 的折叠研究中, 发现通过使用这种新的CV, 可以有效地搜寻到具有 β 片的结构的蛋白折叠状态。而Stefano Piana等人利用副本互换的原理, 发展了一种限制势能互换的准动态模拟方法 (Bias-Exchange Metadynamics (BE-META))⁹⁴, 将其成功用于Trp-Cage小蛋白的折叠研究中。Giovanni Bussi等人则将上述方法和副本互换方法进行结合, 发展了一种新的温度平行-准动态模拟方法 (parallel tempering plus metadynamics)^{95, 96}, 它们成功的将这一方法应用到GB1 的折叠研究中, 取得了比传统的副本互换高 3 倍以上的效率。

1.5 小结

虽然单分子技术, 多维核磁共振, 原子力显微镜等实验技术的发展和运用使得捕捉蛋白质动态行为成为可能。然而, 目前的实验方法还不能完全在分子和原子水平上阐明蛋白质的结构变化和动力学行为。随着计算机硬件和新型采样方法的发展, 分子动力学模拟作为研究蛋白质动态行为的重要手段将越来越广泛的参与到蛋白质功能的研究中, 并将为实验工作者提供更大的帮助。

与此同时, 随着新的可极化力场的发展和相应算法的发展和完善, 将来分子动力学模拟势必朝着越来越精确的方向发展。虽然目前可极化力场显示了其在处理静电极化中的效果, 且在一些体系中成功应用, 吸引了大批科学家的注意。但是我们也要看到, 目前出现的可极化力场还没有像传统的力场那样经过大量模拟的验证, 只是应用到了有限的体系上, 所以它的适用性还需大量模拟实践的验证。而且可以看到的一点是, 目前的可极化力场普遍存在计算量巨大

及参数化的问题，随着计算机硬件和模拟技术的发展，这些问题也将逐渐得到解决。

参考文献

1. Henzler-Wildman, K. and D. Kern, *Dynamic personalities of proteins*. Nature, 2007. **450**(7172): p. 964-72.
2. Henzler-Wildman, K.A., et al., *Intrinsic motions along an enzymatic reaction trajectory*. Nature, 2007. **450**(7171): p. 838-U13.
3. Boehr, D.D., R. Nussinov, and P.E. Wright, *The role of dynamic conformational ensembles in biomolecular recognition*. Nat. Chem. Biol., 2009. **5**(11): p. 789-796.
4. Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules*. Nat. Struct. Mol. Biol., 2002. **9**(9): p. 646-652.
5. Fourme, R., E. Girard, and K. Akasaka, *High-pressure macromolecular crystallography and NMR: status, achievements and prospects*. Curr. Opin. Struct. Biol., 2012. **22**(5): p. 636-642.
6. Palmer 3rd, A., *NMR characterization of the dynamics of biomacromolecules*. Chem. Rev., 2004. **104**(8): p. 3623.
7. Boehr, D.D., et al., *The dynamic energy landscape of dihydrofolate reductase catalysis*. Science, 2006. **313**(5793): p. 1638-1642.
8. Dubochet, J., *Cryo - EM—the first thirty years*. J. Microsc., 2012. **245**(3): p. 221-224.
9. Berman, H.M., et al., *The Protein Data Bank*. Nucleic Acids Res, 2000. **28**(1): p. 235-242.
10. Schotte, F., et al., *Watching a protein as it functions with 150-ps time-resolved X-ray crystallography*. Science, 2003. **300**(5627): p. 1944-1947.
11. Schotte, F., et al., *Watching a signaling protein function in real time via 100-ps time-resolved Laue crystallography*. Proc. Natl. Acad. Sci. USA, 2012.

- 109**(47): p. 19256-19261.
12. Neutze, R. and K. Moffat, *Time-resolved structural studies at synchrotrons and X-ray free electron lasers: opportunities and challenges*. Curr. Opin. Struct. Biol., 2012. **22**(5): p. 651-659
 13. Ha, T., et al., *Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism*. Proc. Natl. Acad. Sci. USA, 1999. **96**(3): p. 893-898.
 14. Michalet, X., S. Weiss, and M. Jäger, *Single-molecule fluorescence studies of protein folding and conformational dynamics*. Chem. Rev., 2006. **106**(5): p. 1785.
 15. Baclayon, M., W.H. Roos, and G.J. Wuite, *Sampling protein form and function with the atomic force microscope*. Mol. Cell. Proteomics, 2010. **9**(8): p. 1678-1688.
 16. Alder, B.J. and T.E. Wainwright, *Studies in Molecular Dynamics. I. General Method*. J. Chem. Phys, 1959. **31**(2): p. 459-466.
 17. Rahman, A., *Correlations in the Motion of Atoms in Liquid Argon*. Phys. Rev., 1964. **136**(2A): p. A405-A411.
 18. Lindorff-Larsen, K., et al., *How Fast-Folding Proteins Fold*. Science, 2011. **334**(6055): p. 517-520.
 19. McCammon, J.A., B.R. Gelin, and M. Karplus, *Dynamics of folded proteins*. Nature, 1977. **267**(5612): p. 585.
 20. Duan, Y. and P.A. Kollman, *Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution*. Science, 1998. **282**(5389): p. 740-744.
 21. Freddolino, P.L., et al., *Molecular dynamics simulations of the complete*

- satellite tobacco mosaic virus*. *Structure*, 2006. **14**(3): p. 437-449.
22. Karplus, M. and J. Kuriyan, *Molecular dynamics and protein function*. *Proc. Natl. Acad. Sci. USA*, 2005. **102**(19): p. 6679-6685.
23. Garcia-Viloca, M., et al., *How Enzymes Work: Analysis by Modern Rate Theory and Computer Simulations*. *Science*, 2004. **303**(5655): p. 186-195.
24. Zhu, X., P.E.M. Lopes, and A.D. MacKerell, *Recent developments and applications of the CHARMM force fields*. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2012. **2**(1): p. 167-185.
25. Cornell, W.D., et al., *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules*. *J. Am. Chem. Soc*, 1995. **117**(19): p. 5179-5197.
26. Jorgensen, W.L., D.S. Maxwell, and J. Tirado-Rives, *Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids*. *J. Am. Chem. Soc*, 1996. **118**(45): p. 11225-11236.
27. Oostenbrink, C., et al., *A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6*. *J. Comput. Chem.*, 2004. **25**(13): p. 1656-1676.
28. Hensen, C., et al., *A combined QM/MM approach to protein-ligand interactions: polarization effects of the HIV-1 protease on selected high affinity inhibitors*. *J. Med. Chem.*, 2004. **47**(27): p. 6673-6680.
29. Jiao, D., et al., *Calculation of protein-ligand binding free energy by using a polarizable potential*. *Proc. Natl. Acad. Sci. USA*, 2008. **105**(17): p. 6290-6295.
30. Jiao, D., et al., *Trypsin-ligand binding free energies from explicit and implicit*

- solvent simulations with polarizable potential*. J. Comput. Chem., 2009. **30**(11): p. 1701-1711.
31. Ponder, J.W. and D.A. Case, *Force fields for protein simulations*. Adv. Protein. Chem., 2003. **66**: p. 27-85.
32. Cieplak, P., et al., *Polarization effects in molecular mechanical force fields*. J. Phys. Condens. Matter., 2009. **21**(33): p. 333102.
33. Lopes, P.E.M., B. Roux, and A.D. MacKerell, *Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: theory and applications*. Theor. Chem. Acc., 2009. **124**(1): p. 11-28.
34. Warshel, A., M. Kato, and A.V. Pisliakov, *Polarizable force fields: history, test cases, and prospects*. J. Chem. Theor. Comput., 2007. **3**(6): p. 2034-2045.
35. Cieplak, P., J. Caldwell, and P. Kollman, *Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases*. J. Comput. Chem., 2001. **22**(10): p. 1048-1057.
36. Wang, Z.X., et al., *Strike a balance: Optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides*. J. Comput. Chem., 2006. **27**(6): p. 781-790.
37. Lamoureux, G., A.D. MacKerell Jr, and B. Roux, *A simple polarizable model of water based on classical drude oscillators*. J. Chem. Phys, 2003. **119**: p. 5185.
38. Lamoureux, G. and B. Roux, *Modeling induced polarization with classical drude oscillators: Theory and molecular dynamics simulation algorithm*. J.

- Chem. Phys, 2003. **119**: p. 3025.
39. Patel, S. and C.L. Brooks, *CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations*. J. Comput. Chem., 2003. **25**(1): p. 1-16.
40. Patel, S., A.D. Mackerell Jr, and C.L. Brooks III, *CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model*. J. Comput. Chem., 2004. **25**(12): p. 1504-1514.
41. Xie, W., et al., *Development of a polarizable intermolecular potential function (PIPF) for liquid amides and alkanes*. J. Chem. Theor. Comput., 2007. **3**(6): p. 1878-1889.
42. Friesner, R.A., *Modeling polarization in proteins and protein-ligand complexes: methods and preliminary results*. Adv. Protein. Chem., 2005. **72**: p. 79-104.
43. Geerke, D.P. and W.F. van Gunsteren, *On the calculation of atomic forces in classical simulation using the charge-on-spring method to explicitly treat electronic polarization*. J. Chem. Theor. Comput., 2007. **3**(6): p. 2128-2137.
44. Rappe, A., et al., *UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations*. J. Am. Chem. Soc., 1992. **114**(25): p. 10024-10035.
45. Banks, J.L., et al., *Parametrizing a polarizable force field from ab initio data. I. The fluctuating point charge model*. J. Chem. Phys., 1999. **110**: p. 741.
46. Yang, Z.Z., Y. Wu, and D.X. Zhao, *Atom-bond electronegativity equalization method fused into molecular mechanics. I. A seven-site fluctuating charge and flexible body water potential function for water clusters*. J. Chem. Phys.,

2004. **120**: p. 2541.
47. Yang, Z.Z. and Q. Zhang, *Study of peptide conformation in terms of the ABEEM/MM method*. J. Comput. Chem., 2005. **27**(1): p. 1-10.
48. Rick, S.W. and S.J. Stuart, *Potentials and algorithms for incorporating polarizability in computer simulations*. Reviews in computational chemistry, 2002. **18**: p. 89-146.
49. Chen, S.L., D.X. Zhao, and Z.Z. Yang, *An estimation method of binding free energy in terms of ABEEM/MM and continuum electrostatics fused into LIE method*. J. Comput. Chem., 2010. **32**(2): p. 338-348.
50. Chelli, R., et al., *Electrical response in chemical potential equalization schemes*. J. Chem. Phys., 1999. **111**(18): p. 8569-8575.
51. Lamoureux, G., et al., *A polarizable model of water for molecular dynamics simulations of biomolecules*. Chem. Phys. Lett, 2006. **418**(1): p. 245-249.
52. Jiang, W., et al., *High-performance scalable molecular dynamics simulations of a polarizable force field based on classical Drude oscillators in NAMD*. J. Phys. Chem. Lett., 2011. **2**(2): p. 87.
53. Baker, C.M. and A.D. MacKerell Jr, *Polarizability rescaling and atom-based Thole scaling in the CHARMM Drude polarizable force field for ethers*. J. Mol. Model., 2010. **16**(3): p. 567-576.
54. Kaminski, G.A., et al., *Development of a polarizable force field for proteins via ab initio quantum chemistry: first generation model and gas phase tests*. J. Comput. Chem., 2002. **23**(16): p. 1515-1531.
55. Ren, P. and J.W. Ponder, *Polarizable atomic multipole water model for molecular mechanics simulation*. J. Phys. Chem. B, 2003. **107**(24): p. 5933-5947.

-
56. Ponder, J.W., et al., *Current status of the AMOEBA polarizable force field*. J. Phys. Chem. B, 2010. **114**(8): p. 2549-2564.
57. Schnieders, M.J., et al., *The Structure, Thermodynamics, and Solubility of Organic Crystals from Simulation with a Polarizable Force Field*. J. Chem. Theor. Comput, 2012. **8**(5): p. 1721-1736.
58. Grossfield, A., P. Ren, and J.W. Ponder, *Ion solvation thermodynamics from simulation with a polarizable force field*. J. Am. Chem. Soc., 2003. **125**(50): p. 15671-15682.
59. Mei, Y., D.W. Zhang, and J. Zhang, *New method for direct linear-scaling calculation of electron density of proteins*. J. Phys. Chem. A, 2005. **109**(1): p. 2-5.
60. Ji, C., Y. Mei, and J.Z.H. Zhang, *Developing Polarized Protein-Specific Charges for Protein Dynamics: MD Free Energy Calculation of pKa Shifts for Asp26/Asp20 in Thioredoxin*. Biophys. J., 2008. **95**(3): p. 1080-1088.
61. Tong, Y., et al., *Electrostatic Polarization Makes a Substantial Contribution to the Free Energy of Avidin-Biotin Binding*. J. Am. Chem. Soc., 2010. **132**(14): p. 5137-5142.
62. Mei, Y., et al., *Electrostatic polarization is critical for the strong binding in streptavidin-biotin system*. J. Comput. Chem, 2012. **33**(15): p. 1374-1382.
63. Soderhjelm, P., et al., *Estimates of ligand-binding affinities supported by quantum mechanical methods*. Interdisciplinary Sciences: Computational Life Sciences, 2010. **2**(1): p. 21-37.
64. Gresh, N., et al., *Anisotropic, polarizable molecular mechanics studies of inter-and intramolecular interactions and ligand-macromolecule complexes. A bottom-up strategy*. J. Chem. Theor. Comput, 2007. **3**(6): p. 1960-1986.
-

-
65. de Courcy, B., et al., *Polarizable Water Molecules in Ligand-macromolecule Recognition. Impact on the Relative Affinities of Competing Pyrrolopyrimidine Inhibitors for FAK Kinase*. J. Am. Chem. Soc., 2010. **132**(10): p. 3312-3320.
 66. Shaw, D.E., et al. *Millisecond-scale molecular dynamics simulations on Anton*. in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*. 2009: ACM.
 67. Shaw, D.E., et al., *Atomic-Level Characterization of the Structural Dynamics of Proteins*. Science, 2010. **330**(6002): p. 341-346.
 68. Hansmann, U.H.E. and L.T. Wille, *Global Optimization by Energy Landscape Paving*. Phys. Rev. Lett., 2002. **88**(6): p. 068105.
 69. Schug, A., T. Herges, and W. Wenzel, *Reproducible Protein Folding with the Stochastic Tunneling Method*. Phys. Rev. Lett., 2003. **91**(15): p. 158102.
 70. Hansmann, U.H.E. and Y. Okamoto, *New Monte Carlo algorithms for protein folding*. Curr. Opin. Struct. Biol., 1999. **9**(2): p. 177-183.
 71. Zhou, R., *Trp-cage: Folding free energy landscape in explicit water*. Proc. Natl. Acad. Sci. USA., 2003. **100**(23): p. 13280-13285.
 72. García, A.E. and J.N. Onuchic, *Folding a protein in a computer: An atomic description of the folding/unfolding of protein A*. Proc. Natl. Acad. Sci. USA., 2003. **100**(24): p. 13898-13903.
 73. Im, W. and C.L. Brooks Iii, *De novo Folding of Membrane Proteins: An Exploration of the Structure and NMR Properties of the fd Coat Protein*. J. Mol. Bio., 2004. **337**(3): p. 513-519.
 74. Lee, J., *New Monte Carlo algorithm: Entropic sampling*. Phys. Rev. Lett., 1993. **71**(2): p. 211-214.

-
75. Hansmann, U.H.E., *Parallel tempering algorithm for conformational studies of biological molecules*. Chem. Phys. Lett., 1997. **281**(1–3): p. 140-150.
76. Hansmann, U.H.E., Y. Okamoto, and F. Eisenmenger, *Molecular dynamics, Langevin and hybrid Monte Carlo simulations in a multicanonical ensemble*. Chem. Phys. Lett. 1996. **259**(3): p. 321-330.
77. Mohanty, S., et al., *Simulation of Top7-CFr: A transient helix extension guides folding*. Proc. Natl. Acad. Sci. USA, 2008. **105**(23): p. 8004-8007.
78. Dror, R.O., et al., *Biomolecular simulation: a computational microscope for molecular biology*. Annu. Rev. Biophys., 2012. **41**: p. 429-452.
79. Okamoto, Y., *Generalized-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations*. arXiv preprint cond-mat/0308360, 2003.
80. Kokubo, H., T. Tanaka, and Y. Okamoto, *Ab Initio prediction of protein - ligand binding structures by replica - exchange umbrella sampling simulations*. J. Comput. Chem., 2011. **32**(13): p. 2810-2821.
81. Deng, N.-j., et al., *Insights into the dynamics of HIV-1 protease: a kinetic network model constructed from atomistic simulations*. J. Am. Chem. Soc., 2011. **133**(24): p. 9387-9394.
82. Meinke, J.H. and U.H.E. Hansmann, *Free - energy - driven folding and thermodynamics of the 67 - residue protein GS - α 3W—A large - scale Monte Carlo study*. J. Comput. Chem. 2009. **30**(11): p. 1642-1648.
83. Okur, A., et al., *Improved efficiency of replica exchange simulations through use of a hybrid explicit/implicit solvation model*. J. Chem. Theor. Comput., 2006. **2**(2): p. 420-433.
84. Okur, A., L. Wickstrom, and C. Simmerling, *Evaluation of salt bridge*
-

- structure and energetics in peptides using explicit, implicit, and hybrid solvation models*. J. Chem. Theor. Comput., 2008. **4**(3): p. 488-498.
85. Liu, P., et al., *Replica exchange with solute tempering: A method for sampling biological systems in explicit water*. Proc. Natl. Acad. Sci. USA, 2005. **102**(39): p. 13749-13754.
86. Huang, X., et al., *Replica exchange with solute tempering: efficiency in large scale systems*. J. Phys. Chem. B, 2007. **111**(19): p. 5405-5410.
87. Camilloni, C., et al., *Exploring the protein G helix free - energy surface by solute tempering metadynamics*. Proteins: Structure, Function, and Bioinformatics, 2007. **71**(4): p. 1647-1654.
88. Fukunishi, H., O. Watanabe, and S. Takada, *On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction*. J. Chem. Phys., 2002. **116**(20): p. 9058-9067.
89. Liu, P., et al., *Hydrophobic aided replica exchange: an efficient algorithm for protein folding in explicit solvent*. J. Phys. Chem. B, 2006. **110**(38): p. 19018-19022.
90. Zhang, C. and J. Ma, *Enhanced sampling and applications in protein folding in explicit solvent*. J. Chem. Phys., 2010. **132**(24): p. 244101.
91. Zhang, C. and J. Ma, *Folding helical proteins in explicit solvent using dihedral-biased tempering*. Proc. Natl. Acad. Sci. USA, 2012. **109**(21): p. 8139-8144.
92. Sutto, L., S. Marsili, and F.L. Gervasio, *New advances in metadynamics*. Wiley Interdisciplinary Reviews: Computational Molecular Science, 2012. **2**: p. 771-779.
93. Pietrucci, F. and A. Laio, *A collective variable for the efficient exploration of*

- protein beta-sheet structures: application to sh3 and gbl*. J. Chem. Theor. Comput., 2009. **5**(9): p. 2197-2201.
94. Piana, S. and A. Laio, *A bias-exchange approach to protein folding*. J. Phys. Chem. B, 2007. **111**(17): p. 4553-4559.
95. Bonomi, M. and M. Parrinello, *Enhanced sampling in the well-tempered ensemble*. Phys. Rev. Lett., 2010. **104**(19): p. 190601.
96. Bussi, G., et al., *Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics*. J. Am. Chem. Soc., 2006. **128**(41): p. 13435-13441.

第二章 基于显隐性水模型速度优化的副本互换动力学方法发展

2.1 引言

近年来,随着人类基因组和蛋白质组学的发展,DNA和蛋白质的序列信息有了迅猛增长,但由于蛋白质表达、纯化及结晶方面的困难及NMR研究大蛋白的困难,蛋白质三维结构信息的增长速度与序列信息相比,明显滞后很多,尤其是在药物设计方面起重要功能的膜蛋白¹。而且蛋白质分子的错误折叠与很多疾病密切相关,如阿尔茨海默病²,亨廷顿舞蹈症³,二型糖尿病⁴等疾病。因此研究多肽链如何折叠成为具有特定结构的蛋白质折叠成为当前研究蛋白质功能的热点问题之一⁵。

目前存在着几个重要的蛋白质折叠的学说,最早是由Anfinsen在20世纪60年代提出的“自组装学说”^{6,7}。实验发现经过还原变性的牛胰核糖核酸酶在体外不需要任何其他物质的帮助就可恢复其天然结构^{6,7}。根据这一实验现象,“自组装学说”认为多肽链的一级序列已经包含了其三维结构的全部信息,蛋白质的天然结构是其在特定条件下热力学最稳定的结构。

随着研究的深入,人们发现许多蛋白质在体外的变性复性过程并不是完全可逆的,而且许多蛋白质在体外还会发生错误折叠甚至不能折叠,且很多蛋白质在体外的折叠速度比活细胞内慢得多,在细胞体内蛋白质的折叠往往需要有其他辅助因子的参与。根据这些实验结果,Ellis于1987年提出了蛋白质折叠的辅助性组装学说⁸。近年来随着分子伴侣的发现和深入研究,这一假说被越来越多的研究证实。这些现象使人们意识到蛋白质折叠不仅是Anfinson学说认为的纯粹的热力学问题,折叠的动力学特性在折叠过程中也可能起重要作用。

而对于蛋白质折叠是热力学控制的还是动力学控制的问题则引起了科学家

的更大兴趣，若蛋白质折叠是热力学控制的，则蛋白质折叠达到天然状态完全由能量决定，而与失活条件以及折叠过程的起始和中间状态无关，但找到能量最低状态则需要长时间的构象搜索，如图2.1a所示。但是这个和实验上观察到的小蛋白较短的时间尺度内（微秒到毫秒）完成折叠是相互矛盾的^{9,10}。而动力学控制则认为蛋白质折叠是路径依赖性的，在一定条件下蛋白质无法达到全局能量最低构象，而只能达到局部的能量最低点。

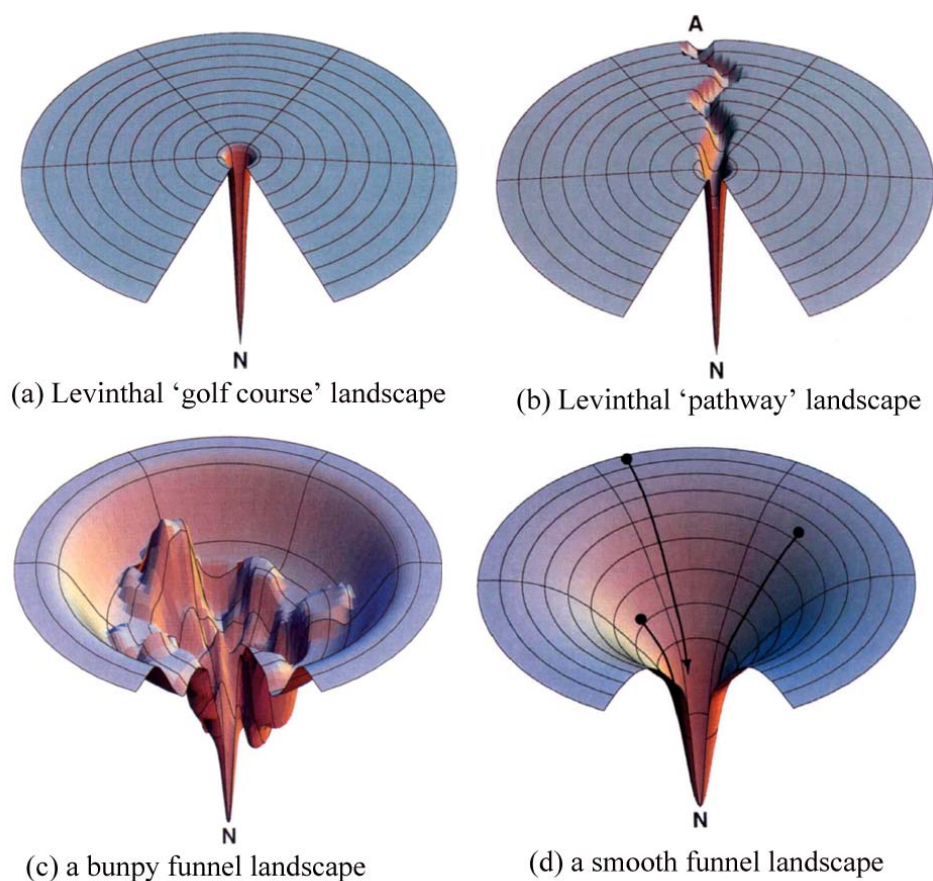


图2.1 蛋白质折叠的几种能量图景假说，图片来自文献¹¹。

为了解答蛋白质快速折叠的动力学问题，物理学家Levinthal¹² 在1968年首

先提出折叠路径的概念。他认为蛋白质无需在天文数字般的构象中进行搜索，蛋白质的折叠有它的特殊路径，如图2.1b所示，因而折叠速率大大加快。根据这一理论，Kaiand¹³和Elson¹⁴等利用同位素标记等实验手段，在实验中观察到了一些中间状态，从而支持了此学说。

到了二十世纪90年代中后期，折叠路径的概念逐渐被能量图景(energy landscape)的新观点所取代^{10, 11, 15-20}。这种新观点认为，折叠的形貌像是一个漏斗，蛋白质的构象会沿着漏斗的表面地貌汇聚到能量最低状态（图2.1c）。由于蛋白质折叠是一个寻找低能构象的过程，所以在较短时间内可以找到它的能量最低状态。而蛋白在折叠过程中可能存在许多条路径，这取决于具体蛋白的能量图景。在能量图景上存在的许多凹陷点，代表折叠的局部低能状态，这些局部低能状态的存在，使蛋白质折叠过程中存在着能垒，阻碍蛋白最稳定分子构象的获得，从而可以在实验上观察得到某些亚稳态的中间体^{13, 14}。理想情况下从直链到天然态的折叠没有能垒，就完全是一个漏斗（如图2.1d所示）。

随着计算机硬件和动力学软件的发展，分子动力学模拟在蛋白质折叠研究中起着越来越大的作用，如在1998年，Duan等²¹利用超级计算机Cray T3D进行了长达1微秒的HP36蛋白折叠的动力学模拟。如表2.1所示蛋白质折叠的分子动力学模拟的时间尺度和体系都有了很大的提高，尤其是分子动力学模拟专用计算机Anton及其配套的动力学软件Desmod的发展²²。De Shaw等²³在2011年对12个常用的模式蛋白进行了长时间的折叠研究。结果发现这12个体系都可以在融化温度附近进行多次折叠和反折叠，显示了现有力场的和长时间动力学的准确性。

表 2.1 近年来使用长时间动力学模拟研究蛋白折叠

年	模拟时间 (us)	蛋白	平台 (软件)
2007	2	Villin HP-35	GROMACS ²⁴

2008	10	WW domain	NAMD ²⁵
2010	100	Fip35	Anton ²⁶
2011	2936	NTL9	Anton ²³

虽然长时间的常规动力学模拟在蛋白质折叠中取得了长足的发展，但是由于蛋白质构象空间巨大，具有非常多的局部极小值点，而它们之间具有较大的能垒，在常规动力学模拟中跨过这些能垒，所需的计算量就会非常巨大。如表 2.1 所示的成功例子，都是利用了超级计算机相当长的时间，所花费的代价是极其巨大的。而且随着蛋白质原子数目的增大，所需计算量更是指数增长，所以目前在室温下成功折叠的蛋白质的残基数还在 100 以下²⁷。另外，虽然目前的常规动力学的时间尺度可以达到毫秒级别²³，但对于蛋白质折叠的时间尺度而言，还是太短。为了解决普通的动力学模拟不能有效跨过能垒的问题，近年来建立和发展了很多基于广义系综的统计力学的增强型取样算法²⁸⁻³⁶，并且取得了很大的进展，但是这一问题并未完全解决。

在蛋白质折叠中广泛使用的增强型方法如温度副本互换³⁷，其基本思想是在高温副本中得到高能状态的构象，通过构象交换机制，将陷于局部极小值的构象交换出来，从而达到跨过能垒的目的，其已经在蛋白质折叠模拟，蛋白质与配体相互作用等研究中得到广泛应用³⁸⁻⁴²。可是目前这种温度副本互换的方法，也存在一些问题：由于水分子在蛋白质折叠中的重要作用，在实际模拟过程中蛋白质都是处于水溶液环境中，这样整个体系的原子数目相对于蛋白质原子数目就大很多，而在副本互换中所需的副本数目和体系的自由度的平方根成正比，这样模拟所需要的副本数目就会大大增加^{43, 44}。

近年来，如何寻找更加有效的副本交换方法成为取样方法研究的热点⁴⁵⁻⁴⁹。在本章中我们发展了一种新的副本互换的方法，即在显性水环境下进行动力学模拟下，而只使用隐性水模型下的蛋白质分子的势能来决定是否交换，通过速

度调节来消除显隐性水模型中的能量之差，从而使副本交换的副本数目只取决于蛋白质的原子数目，而与体系中的水分子数目无关，从而使体系所需的副本数目大大减少。这一方法通过修改动力学软件AMBER10中的源程序来实现，并在小蛋白质Trp-cage的折叠模拟中进行了测试，模拟的结果发现可以将使用的副本数目从40下降到10，提高了效率，而且新方法所获得的蛋白质折叠的基本性质与常规的副本互换方法获得的结果并没有区别，从而验证了我们方法的准确性。

2.2 研究方法

2.2.1 理论计算方法

副本交换方法最早是由Okamoto和Hansmann等人⁵⁰提出的，并在蛋白质的折叠的蒙特卡洛模拟中得到应用，之后在蛋白质折叠中广为应用^{39, 51}。其方法如下：

处在两个不同温度 T_1 和 T_2 下的副本分别具有能量 E_1 和 E_2 ，他们具有一定的概率可以发生互换，其中这两个副本互换是否发生由式 1.22 ($\omega(1 \leftrightarrow 2) = \min(1, \exp(\Delta\beta\Delta E))$) 决定，其中 $\beta = 1/k_B T$ 为温度因子。

在动力学模拟中体系能量为动能和势能之和，即 $E(x, v) = E_{pot}(x) + E_{kin}(v)$ ，

而 $E_{kin}(v) = \frac{1}{2} \sum_i m_i v_i^2$ ，这样可以通过使用速度调节因子来对动能进行调节，如式

2.1 所示。

$$E_{kin}(rv) = r^2 E_{kin}(v) \quad (\text{式 2.1})$$

这样通过调整速度因子 $r(1,2)$ ($r(1,2) = \sqrt{T(2,1)/T(1,2)}$)，就可以使 $v_{(1,2)}^{new} = v_{(2,1)}^{old}$ ，从而得到 $\Delta E_{kin} = 0$ 。这样我们可以通过调整速度因子，从而使副本的互换完全由体系的势能决定，即：

$$w(1 \leftrightarrow 2) = \exp(\Delta\beta\Delta E_{pot}) \quad (\text{式 2.2})$$

在真实的模拟体系中，蛋白质一般都在水溶液中，因此我们可以将体系的势能分成 3 项，有 $E_{pot} = P_{pp} + P_{pw} + P_{ww}$ ，其中 P_{pp} 为蛋白的势能， P_{pw} 为蛋白和水相互作用的势能， P_{ww} 为体系中水的势能。同理，可以将体系的总动能分成蛋白的动能和水的动能这两项，即 $E_{kin} = K_p + K_w$ ，其中 K_p 为蛋白动能， K_w 为体系中水的动能。

在模拟体系中由于显性水的存在， P_{pw} 项远大于其他两项，它的存在使体系模拟所需的副本数目大大增加，因此我们使用隐形水模型中的蛋白势能作为是否发生交换的依据。即

$$\omega(1 \leftrightarrow 2) = \min(1, \exp(\Delta\beta\Delta(P_{pp} + P_{is}))) \quad (\text{式 2.3})$$

其中 P_{is} 是体系在隐形水中溶剂化能。这样体系所需的副本数目就会大大减少。不过由于显性水和隐形水之间的差别，直接使用式 2.3，不能保证交换前后的 $\Delta E = 0$ ，从而会造成模拟的系综发生变化。因此我们在模拟过程中，为了使模拟具有严格的物理意义，通过调整速度因子，消除显性水和隐性水之间的区别，从而保证 $\Delta E = 0$ ，得到正确的系综。

我们需要对速度进行如下调整：

$$v^{(2)} \leftrightarrow \hat{v}^{(1)} = v^{(2)} \sqrt{\frac{\hat{E}_{kin}^{(1)}}{E_{kin}^{(2)}}} = v^{(2)} \sqrt{\frac{E_{kin}^{(1)} - \Delta H}{E_{kin}^{(2)}}} \quad (\text{式 2.4})$$

$$v^{(1)} \leftrightarrow \hat{v}^{(2)} = v^{(1)} \sqrt{\frac{\hat{E}_{kin}^{(2)}}{E_{kin}^{(1)}}} = v^{(1)} \sqrt{\frac{E_{kin}^{(2)} + \Delta H}{E_{kin}^{(1)}}} \quad (\text{式 2.5})$$

其中 $E_{kin}^{(1)}$ 为在温度 T1 下的副本体系的动能，H 为显性水和隐形水的溶剂化

能之差差值，即

$$H = P_{pw} + P_{ww} - P_{is} \quad (\text{式 2.6})$$

这样就通过式 2.4 和 2.5 就可以保证使用式 2.3 作为交换判据的合理性，证明如下：首先定义： $Q = P_{pp} + P_{is}$ 那么由式 2.6 就可以得到

$$E_{pot} = Q + H \quad (\text{式 2.7})$$

实际上以 $\omega(1 \leftrightarrow 2) = \min(1, \exp(D))$ 作为判定交换的依据，则 D 可以由下式求得：

$$\begin{aligned} D &= -\beta_2 E_{pot}^{(1)} - \beta_2 \hat{E}_{kin}^{(2)} - \beta_1 E_{pot}^{(2)} - \beta_1 \hat{E}_{kin}^{(1)} + \beta_1 E_{pot}^{(1)} + \beta_1 E_{kin}^{(1)} + \beta_2 E_{pot}^{(2)} + \beta_2 E_{kin}^{(2)} \quad (\text{式 2.8}) \\ &= \Delta\beta\Delta Q - \beta_1 (\hat{E}_{kin}^{(1)} - E_{kin}^{(1)} + \Delta H) - \beta_2 (\hat{E}_{kin}^{(2)} - E_{kin}^{(2)} - \Delta H) \end{aligned}$$

其中 $E_{kin}^{(1)}$ 和 $\hat{E}_{kin}^{(1)}$ 为副本 1 交换前后的动能，这样我们对速度进行式 2.4 和 2.5 的调整就可以得到， $\hat{E}_{kin}^{(1)} = E_{kin}^{(1)} - \Delta H$ 和 $\hat{E}_{kin}^{(2)} = E_{kin}^{(2)} + H$ ，从而式 2.8 化简为

$$D = \Delta\beta\Delta Q \quad (\text{式 2.9})$$

从而可以使用隐性水模型下的势能来作为交换的判据，也可以保证交换前后的 $\Delta E = 0$ ，从而得到了具有严格物理意义的系综。

2.2.2 方法测试

我们使用的测试的例子是 Trp-cage，一个只有 20 个残基的小蛋白（PDB 号为 1l2y⁵²），由于其折叠速率很快，经常被用来作为增强型取样方法的测试体系^{29, 53-56}。在本研究中分别将 N 端和 C 端用 ACT 和 NME 来饱和，使用的序列为 (Ac-NLYIQWLKDGGPSSGRPPPS-NME)。

Paschek 等人⁵⁷利用常规的副本互换的方法详细研究了 this 蛋白在显性水溶

液中的折叠机理。因此我们的测试结果可以和这一研究进行比较。在模拟中我们采用AMBERff94力场来模拟蛋白质的原子相互作用，用TIP3P水模型在模拟水分子的运动。蛋白质原子总数为313个，进行25ps的300K模拟得到坍塌的结构。然后将这个坍塌的结构加上抗衡离子，再将这个结构放在充满2687个TIP3P水的盒子里，作为模拟的体系。这一起始结构如图2.2所示，它的主链重原子的RMSD值达到6 Å，而且没有任何二级结构。将这一体系在1个大气压300K下平衡100ps，得到做副本互换的起始结构，

在常规的副本互换方法中的温度与Paschek等人⁵⁷的模拟一致，一共使用 40 个副本，温度设置如下：280.00，284.1，288.2，292.4，296.7，301.1，305.6，310.2，314.9，319.7，324.6，329.6，334.7，340.0，345.4，351.0，356.6，362.5，368.4，374.6，380.9，387.3，394.0，400.8，407.8，415.1，422.5，430.1，438.0，446.0，454.3，462.8，471.6，480.6，489.8，499.3，509.0，519.0，529.2 和 539.7 K。

为了测试我们的新方法，我们使用了 10 个副本，温度设置如下：280.00，301.18，323.96，348.46，374.82，403.17，433.66，466.47，501.75 和 539.7 K。在所有的模拟中使用SHAKE⁵⁹算法来限制含氢的键长，使用的时间步长为 2.fs。在模拟中长程静电相互作用和范德华相互作用是使用IPS^{60, 61}算法计算，它们的截断半径都设置为 10 Å，交换间隔为 800 步（1.6 ps），对每个副本进行长达 100ns的动力学模拟。

2.3 结果与讨论

2.3.1 改进的副本互换方法的效率

在副本互换中，各个副本是通过构象交换从而在各个温度进行模拟，从而加强取样，因此取样的有效性取决于副本在温度空间中的游走。副本之间的交换率在某种程度上显示了此方法的效率。如图 2.2 所示，在常规的 40 个副本的

模拟中，第一个副本可以在各个温度间进行变化。计算发现这个副本在 100 ns 的模拟中经历了 6 轮从最低温度到最高温度的变化，全部 40 个副本一共观察到 220 次这样的变化。各个温度之间的交换率基本都在 18.1%，这点和Paschek等⁵⁷的结果基本类似，说明虽然使用不同的软件、控温方法和长程作用处理办法，但是由于使用的力场和方法是一样的，模拟结果并没有大的变化。接下来我们集中研究比较采用新方法和常规副本互换方法的模拟结果。

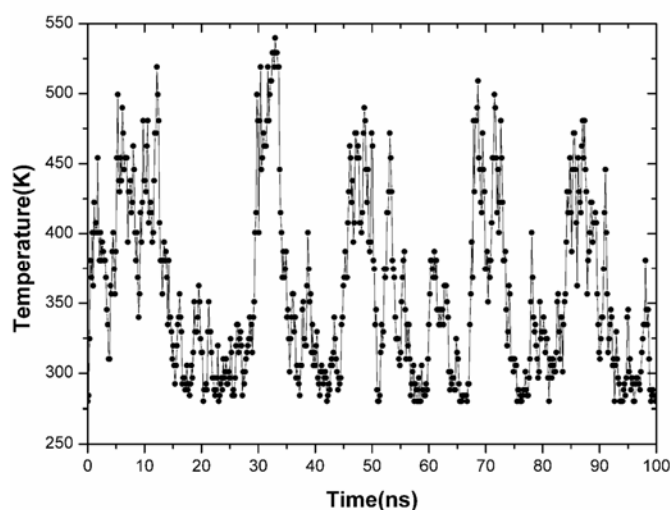


图 2.2 在常规副本互换模拟中副本 1 的模拟温度随时间的变化情况。

通过增加副本数目，减小副本之间的温度间隔可以增加副本之间的交换率，从而提高取样效率，然而与此同时，所需的计算量也会大大增加。从图 2.3 中可以看到虽然新方法中采用 10 个副本，第一个副本也可以在全部温度空间中进行变化。在这个副本中我们观察到 50 次最低温度到最高温度的变化，全部 10 个副本共观察到 404 次这样的变化。对应地，各个温度之间的交换率大概为 13.1%。相反地，使用 10 个副本的常规副本互换方法，我们发现各个副本之间没有发生任何交换，从而可以看到我们方法可以有效地提高取样效率。

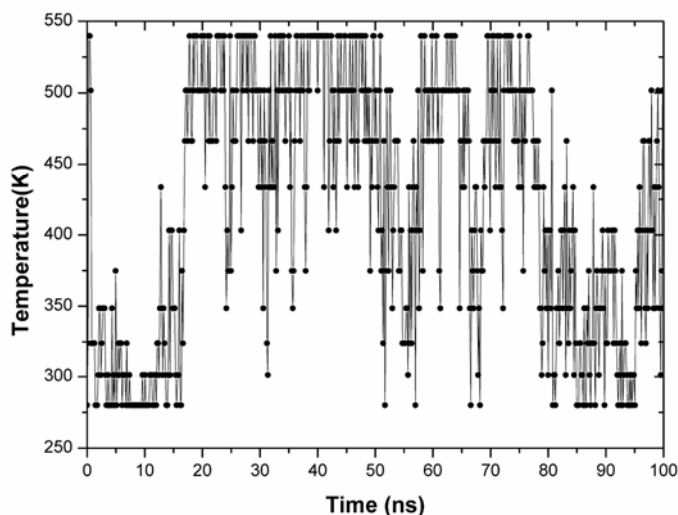


图 2.3 在修改的副本互换模拟中副本 1 的模拟温度随时间的变化情况。

2.3.2 改进的副本互换方法的准确性

为了测试我们方法的准确性，我们选取了两个表征蛋白结构的指标：近天然结构(主链的 $\text{RMSD} < 3.0 \text{ \AA}$ 的结构)百分比和天然的原子接触(native contact)，进行分析来测试我们方法的准确性。在这里所做的分析都是针对平衡后的 50ns 的轨迹。从图 2.4 中我们发现新方法（10 个副本）和传统方法（40 个副本）在近天然结构的百分比和天然的原子接触所占的比例在各个温度区间的结果都是基本一致，说明我们 10 个副本的计算结果与常规的 40 个副本的结果基本类似，从而可以验证我们方法的准确性。此外，从图 2.4 中还发现新方法在低温区间（ $T < 350 \text{ K}$ ）获得的数值与常规方法的数值非常接近，这 2 种方法的最大差别在高温区间。由此可见使用新方法不会影响体系在室温下的性质，而蛋白质折叠研究通常是在常温下进行的，可见此方法适合于蛋白质的折叠研究。

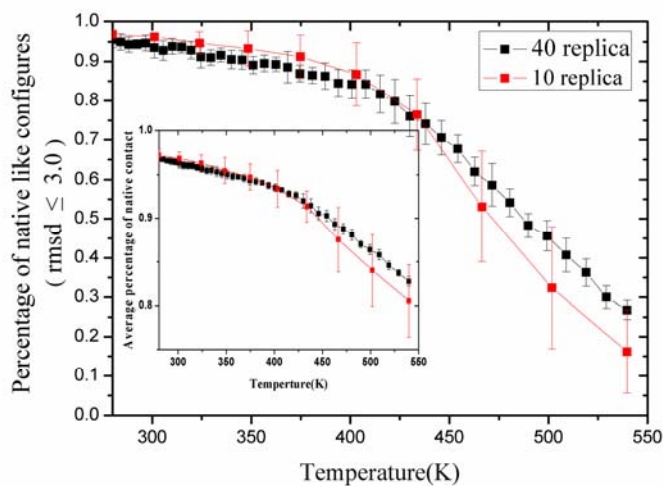


图 2.4 近天然结构所占的比重随温度的变化曲线。内图是平均天然的原子接触（native contact）所占的比例随温度的变化。其中误差线计算是通过将后 50ns 的轨迹每隔 10ns 分成 5 份，然后通过这五个平均值计算得到的。

另外我们将模拟得到的最稳定的结构和晶体结构进行叠合比较，如图 2.6 所示，发现其与天然结构的 RMSD 只有 0.71 Å，可见我们的新方法可以有效的捕捉到晶体结构。

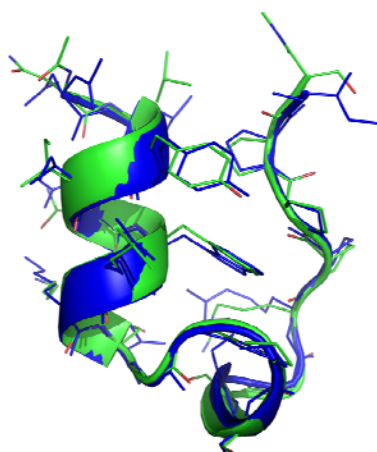


图 2.5 晶体结构和最稳定结构的重叠，它们之间的主链 RMSD 为 0.71 Å。

2.4 小结

本章中我们发现充分地运用体系动能在副本互换中的作用,可以有效地提高副本互换的效率,从而提高计算效率。本章方法建立了在显式水模拟中使用蛋白隐性模型中的能量作为交换依据的方法,此方法通过在每一步的交换过程中调整副本之间的速度,从而可以获得准确的系综,并且可以有效地提高计算方法的效率。在这个 Trp-Cage 的小体系中,我们成功地将副本数目从 40 个减小到 10 个,使计算效率得到大大的提高。而且从原理上讲,此方法在大的蛋白质体系中所能起的加速作用会更加明显,这一预测有待进一步的研究。

参考文献

1. Berman, H.M., et al., *The Protein Data Bank*. Nucleic Acids Res., 2000. **28**(1): p. 235-242.
2. Selkoe, D.J., *Alzheimer's Disease: Genes, Proteins, and Therapy*. Physiol. Rev., 2001. **81**(2): p. 741-766.
3. Hatters, D.M., *Protein misfolding inside cells: the case of huntingtin and Huntington's disease*. IUBMB life, 2008. **60**(11): p. 724-728.
4. Chiti, F. and C.M. Dobson, *Protein Misfolding, Functional Amyloid, and Human Disease*. Annu. Rev. Biochem., 2006. **75**(1): p. 333-366.
5. Dill, K.A. and J.L. MacCallum, *The Protein-Folding Problem, 50 Years On*. Science, 2012. **338**(6110): p. 1042-1046.
6. Anfinsen, C.B., et al., *The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain*. Proc. Natl. Acad. Sci. USA, 1961. **47**(9): p. 1309.
7. Anfinsen, C.B., *Principles that Govern the Folding of Protein Chains*. Science, 1973. **181**(4096): p. 223-230.
8. Ellis, J., *Proteins as molecular chaperones*. Nature, 1987. **328**(6129): p. 378.
9. Dill, K.A., *Theory for the folding and stability of globular proteins*. Biochemistry, 1985. **24**(6): p. 1501-1509.
10. Karplus, M., *The Levinthal paradox: yesterday and today*. Folding Des., 1997. **2**: p. S69-S75.
11. Dill, K.A. and H.S. Chan, *From Levinthal to pathways to funnels*. Nat. Struct. Biol., 1997. **4**(1): p. 10-19.
12. Levinthal, C., *Are there pathways for protein folding*. J. Chim. Phys, 1968. **65**(1): p. 44-45.

13. Ikai, A. and C. Tanford, *Kinetic evidence for incorrectly folded intermediate states in the refolding of denatured proteins*. Nature, 1971. **230**: p. 100-102.
14. Tsong, T., R. Baldwin, and E. Elson, *The sequential unfolding of ribonuclease A: detection of a fast initial phase in the kinetics of unfolding*. Proc. Natl. Acad. Sci. USA, 1971. **68**(11): p. 2712-2715.
15. Dobson, C.M., A. Šali, and M. Karplus, *Protein folding: a perspective from theory and experiment*. Angew. Chem. Int. Ed., 1998. **37**(7): p. 868-893.
16. Schonbrun, J. and K.A. Dill, *Fast protein folding kinetics*. Proc. Natl. Acad. Sci. USA, 2003. **100**(22): p. 12678-12682.
17. Wolynes, P., Z. Luthey-Schulten, and J. Onuchic, *Fast-folding eriments and the topography of protein folding energy landscapes*. Chem. Biol., 1996. **3**(6): p. 425-432.
18. Baldwin, R.L., *The nature of protein folding pathways: the classical versus the new view*. J. Biomol. NMR, 1995. **5**(2): p. 103-109.
19. Sali, A., E. Shakhnovich, and M. Karplus, *How does a protein fold?* Nature, 1994. **369**(6477): p. 248-251.
20. Sali, A., E. Shakhnovich, and M. Karplus, *Kinetics of protein folding. A lattice model study of the requirements for folding to the native state*. J. Mol. Biol., 1994. **235**(5): p. 1614-1636.
21. Duan, Y. and P.A. Kollman, *Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution*. Science, 1998. **282**(5389): p. 740-744.
22. Shaw, D.E., et al. *Millisecond-scale molecular dynamics simulations on Anton*. in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*. 2009: ACM.

23. Lindorff-Larsen, K., et al., *How Fast-Folding Proteins Fold*. Science, 2011. **334**(6055): p. 517-520.
24. Ensign, D.L., P.M. Kasson, and V.S. Pande, *Heterogeneity Even at the Speed Limit of Folding: Large-scale Molecular Dynamics Study of a Fast-folding Variant of the Villin Headpiece*. J. Mol. Biol., 2007. **374**(3): p. 806-816.
25. Freddolino, P.L., et al., *Ten-Microsecond Molecular Dynamics Simulation of a Fast-Folding WW Domain*. Biophysical Journal, 2008. **94**(10): p. L75-L77.
26. Shaw, D.E., et al., *Atomic-Level Characterization of the Structural Dynamics of Proteins*. Science, 2010. **330**(6002): p. 341-346.
27. Kubelka, J., J. Hofrichter, and W.A. Eaton, *The protein folding 'speed limit'*. Curr. Opin. Struct. Biol., 2004. **14**(1): p. 76-88.
28. Hansmann, U.H.E. and L.T. Wille, *Global Optimization by Energy Landscape Paving*. Phys. Rev. Lett., 2002. **88**(6): p. 068105.
29. Schug, A., T. Herges, and W. Wenzel, *Reproducible Protein Folding with the Stochastic Tunneling Method*. Phys. Rev. Lett., 2003. **91**(15): p. 158102.
30. Hansmann, U.H.E. and Y. Okamoto, *New Monte Carlo algorithms for protein folding*. Curr. Opin. Struct. Biol., 1999. **9**(2): p. 177-183.
31. Nadler, W., J.H. Meinke, and U.H.E. Hansmann, *Folding proteins by first-passage-times-optimized replica exchange*. Phys. Rev. E, 2008. **78**(6).
32. Kar, P., W. Nadler, and U.H.E. Hansmann, *Microcanonical replica exchange molecular dynamics simulation of proteins*. Phys. Rev. E, 2009. **80**(5).
33. Neuhaus, T., M. Magiera, and U. Hansmann, *Efficient parallel tempering for first-order phase transitions*. Phys. Rev. E, 2007. **76**(4): p. 045701.
34. Zhou, R., *Trp-cage: Folding free energy landscape in explicit water*. Proc. Natl. Acad. Sci. USA, 2003. **100**(23): p. 13280-13285.

35. García, A.E. and J.N. Onuchic, *Folding a protein in a computer: An atomic description of the folding/unfolding of protein A*. Proc. Natl. Acad. Sci. USA, 2003. **100**(24): p. 13898-13903.
36. Im, W. and C.L. Brooks Iii, *De novo Folding of Membrane Proteins: An Exploration of the Structure and NMR Properties of the fd Coat Protein*. J. Mol. Biol., 2004. **337**(3): p. 513-519.
37. Hansmann, U.H.E., *Parallel tempering algorithm for conformational studies of biological molecules*. Chem. Phys. Lett., 1997. **281**(1-3): p. 140-150.
38. Meinke, J.H. and U.H.E. Hansmann, *Free - energy - driven folding and thermodynamics of the 67 - residue protein GS - α 3W—A large - scale Monte Carlo study*. J. Comput. Chem., 2009. **30**(11): p. 1642-1648.
39. Mohanty, S., et al., *Simulation of Top7-CFr: A transient helix extension guides folding*. Proc. Natl. Acad. Sci. USA, 2008. **105**(23): p. 8004-8007.
40. Okamoto, Y., *Generalized-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations*. arXiv preprint cond-mat/0308360, 2003.
41. Kokubo, H., T. Tanaka, and Y. Okamoto, *Ab Initio prediction of protein - ligand binding structures by replica - exchange umbrella sampling simulations*. J. Comput. Chem., 2011. **32**(13): p. 2810-2821.
42. Zhang, C. and J.P. Ma, *Enhanced sampling in generalized ensemble with large gap of sampling parameter: Case study in temperature space random walk*. J. Chem. Phys., 2009. **130**(19), 194112.
43. Rosta, E., N.V. Buchete, and G. Hummer, *Thermostat artifacts in replica exchange molecular dynamics simulations*. J. Chem. Theor. Comput., 2009. **5**(5): p. 1393-1399.

-
44. Kim, J., T. Keyes, and J.E. Straub, *Generalized Replica Exchange Method*. J. Chem. Phys., 2010. **132**(22), 224107.
 45. Sindhikara, D.J., D.J. Emerson, and A.E. Roitberg, *Exchange often and properly in replica exchange molecular dynamics*. J. Chem. Theor. Comput., **2010.6** (9), p. 2804-2808
 46. Zuckerman, D.M. and E. Lyman, *A second look at canonical sampling of biomolecules using replica exchange simulation*. J. Chem. Theor. Comput., 2006. **2**(6): p. 1693-1693.
 47. Liu, P., et al., *Hydrophobic aided replica exchange: an efficient algorithm for protein folding in explicit solvent*. J. Phys. Chem. B, 2006. **110**(38): p. 19018-19022.
 48. Huang, X.H., et al., *Replica exchange with solute tempering: Efficiency in large scale systems*. J. Phys. Chem. B, 2007. **111**(19): p. 5405-5410.
 49. Kouza, M. and U.H.E. Hansmann, *Velocity scaling for optimizing replica exchange molecular dynamics*. J. Chem. Phys., 2011. **134**(4), 044124.
 50. Hansmann, U.H.E., Y. Okamoto, and F. Eisenmenger, *Molecular dynamics, Langevin and hybrid Monte Carlo simulations in a multicanonical ensemble*. Chem. Phys. Lett., 1996. **259**(3): p. 321-330.
 51. Dror, R.O., et al., *Biomolecular simulation: a computational microscope for molecular biology*. Annu. Rev. Biophys., 2012. **41**: p. 429-452.
 52. Neidigh, J.W., R.M. Fesinmeyer, and N.H. Andersen, *Designing a 20-residue protein*. Nat. Struct. Mol. Biol, 2002. **9**(6): p. 425-430.
 53. Neidigh, J.W., R.M. Fesinmeyer, and N.H. Andersen, *Designing a 20-residue protein*. Nat. Struct. Biol., 2002. **9**(6): p. 425-430.
 54. Okur, A., et al., *Improved efficiency of replica exchange simulations through*

- use of a hybrid explicit/implicit solvation model*. J. Chem. Theor. Comput., 2006. **2**(2): p. 420-433.
55. Snow, C.D., B. Zagrovic, and V.S. Pande, *The Trp cage: folding kinetics and unfolded state topology via molecular dynamics simulations*. J. Am. Chem. Soc., 2002. **124**(49): p. 14548-14549.
56. Paschek, D., S. Hempel, and A.E. Garcia, *Computing the stability diagram Trp-cage miniprotein of the*. Proc. Natl. Acad. Sci. USA, 2008. **105**(46): p. 17754-17759.
57. Paschek, D., S. Hempel, and A. Garcia, *Computing the stability diagram of the Trp-cage miniprotein*. Proc. Natl. Acad. Sci. USA, 2008. **105**(46): p. 17754.
58. Cornell, W.D., et al., *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules*. J. Am. Chem. Soc., 1996. **118**(9): p. 2309-2309.
59. Ryckaert, J.P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes*. J. Comput. Phys., 1977. **23**(3): p. 327-341.
60. Wu, X. and B.R. Brooks, *Isotropic periodic sum: A method for the calculation of long-range interactions*. J. Chem. Phys., 2005. **122**: p. 044107.
61. Klauda, J.B., et al., *Long-range Lennard-Jones and electrostatic interactions in interfaces: application of the isotropic periodic sum method*. J. Phys. Chem. B, 2007. **111**(17): p. 4393-4400.

第三章 简振模式分析与伞取样研究蛋白质构象变化方法发展

3.1 引言

蛋白质是生命活动中最重要的一类的大分子，参与了如免疫、催化、信号传导等大部分生命活动。当蛋白质在发挥功能时，经常伴随着比较大的构象变化¹。另外，一些重要的配体和药物分子能够特异结合靶蛋白的构象，从而发挥调节靶蛋白的功能的作用。因此对蛋白质的构象变化及其与配体间相互作用的研究，就成为研究蛋白质结构功能及药物设计的重要问题²。

目前在实验上用来研究蛋白质构象变化的手段主要有X-衍射晶体技术、单分子荧光技术及动态NMR技术等。而传统的X衍射晶体方法只能捕捉到一定条件下的稳定构象，而处于高能状态的中间构象却极其难以获得，因而也无法获得这些蛋白的构象转变路径。随着技术的发展，近年来实时X衍射晶体方法（the time-resolved X-ray structural methods）^{3,4}也被广泛应用与蛋白质构象变化研究。虽然此方法可以有效获得蛋白质的动态变换过程，但是这个技术还受很多因素的制约，如所使用的光源、所能研究的时间尺度及目标蛋白质的稳定性等。而NMR⁵和单分子荧光技术都需要表达蛋白质，而很多蛋白是难以表达的，而且单分子荧光技术还需要对蛋白进行修饰，从而会造成一些意想不到的影响。

基于生物实验方面的难度，近年来随着计算机硬件及分子模拟的发展，分子动力学作为在分子水平研究蛋白质动态行为的理论工具而得到广泛应用⁶⁻⁸。尽管近年来常规动力学模拟已经能够到达毫秒级别⁹，但是使用常规动力学研究大规模构象变化还是非常困难的。因为这种大规模构象变化经常发生在微妙到秒的时间尺度范围，远大于现有的常规分子动力学模拟所能达到的时间尺度，从而使得利用常规分子动力学模拟研究此类问题受到极大的限制。因此增强型模拟方法得到了广泛¹⁰⁻¹³。这些增强型取样方法可以分成2种类型，一种是以靶

向动力学或者伞取样方法为主，通过在体系中额外加入外力，人为的增加在高能区域的构象分布，从而可以观察到从一个蛋白构象向另一个构象的转变过程。但由于此类方法加入了额外的力或势能，因此通过这些方法计算得到的构象变化路径和能量非常依赖于所加入的力和势能¹⁴⁻¹⁶。另一种方法是使用副本互换的方法，通过改变体系温度的方法¹³，使体系容易跨过能垒。此种方法虽然没有额外加入外力的作用，但是由于需要在所有温度空间同时进行模拟，其所需的计算量是非常巨大的。

另外为了提高研究的体系大小及时间尺度，出现了很多粗粒化的方法。由于对体系进行了简化，大大提高模拟的时间和空间尺度，如在动力学模拟中使用的Martini力场等，由于使用简化的力场和原子模型，其所能模拟的体系的大小和尺度大大增加，已经在GPCR构象变化，膜自主装等方面得到应用¹⁷。另一种在构象变化中广泛使用的方法是简振模式分析（normal mode analysis, NMA）¹⁸。由于NMA计算得到的正则模式（mode）与蛋白发挥功能的构象变化密切相关，因此在实际工作中出现了利用这些正则模式产生构象变化路径的方法，通过与生物实验相结合，其已经在酶催化、离子通道转运机制等研究中得到应用和验证¹⁹⁻²¹。但是此种方法只能获得构象变化的路径，却不能获得相应的自由能变化信息，而构象变化过程中的自由能变化对于深入理解这些蛋白的功能和药物设计是至关重要的。

为了能够研究蛋白质构象变化，并且获得构象变化过程中的自由能变化，我们发展了一种首先使用NMA产生构象变化的路径，然后使用伞取样模拟方法计算得到路径的能量^{22, 23}的方法（NUMD）。计算结果显示这种方法可以有效结合NMA快速获取蛋白构象变化路径和伞取样准确计算自由能变化的优点，因此其可以在蛋白质构象变化中进行应用。我们将此方法应用到 3 个典型的构象变化的体系，分别是来自细菌的磷酸转移酶Adenylate Kinase (AdK)、人的MAPK激酶p38 α 和钙调蛋白（CaM）。为了验证我们方法不具有体系依赖性，这里所选

取的蛋白是具有不同的二级结构，而且它们在生物体中起着不同的功能：其中AdK^{24,25}是磷酸转移酶用来催化磷酸转移反应的($Mg^{2+}ATP + AMP \leftrightarrow Mg^{2+}ADP + ADP$)；而CaM^{26, 27}是细胞中多种细胞信使的调节蛋白，可以调节体内近 100 多个蛋白；激酶p38 α 激酶²⁸在细胞主要对外界压力起作用的通路上起重要作用，因此其抑制剂可以作为抗炎的药物。

AdK主要由 3 个结构域组成，即由残基 118 到 167 组成结合域LID (ATP结合域)，由残基 30 到 67 组成的NMP结构域 (AMP结合域)和连接这 2 个结构域的CORE结构域 (残基 1-29, 68-117 及 161-214)。在晶体数据库中发现AdK具有打开和关闭两种构象，即当没有与配体结合的时候，结构域LID与CORE的距离较远，而使结合口袋处于打开的状态，其PDB 号是 4AKE²⁴。当抑制剂AP₅A与AdK结合时，它的LID会与CORE靠近，使结合口袋处于关闭的构象(1AKE²⁵) (如图 3.1 所示)。第二个体系CaM的N端和C端分别具有一个helix-loop-helix motifs的结构域，中间通过可以形成螺旋的一段残基相连。如图 3.3 所示，当CaM没有结合钙离子的时候，N端和C端的helix-loop-helix motifs处于比较紧密的关闭构象(1CFD¹⁷)，而当结合了钙离子后这个helix-loop-helix motifs就暴露在溶液外面，而使CaM处于打开的构象(1CLL²⁹)，从而有利于其与靶蛋白相结合。与AdK和CaM整体的构象变化不同，激酶p38 α 主要结构变化集中在它的活性loop中 (残基 166-177)。如图 3.5 所示，激酶p38 α 的ATP和配体的结合口袋是由处在N索结构的螺旋C与活性loop和富含甘氨酸-loop(glycine-rich loop)一起形成的。在活性loop中包括了在激酶中高度保守的DFG结构 (Asp168-Phe169-Gly170)，显示它们可以形成将苯丙氨酸处于包埋状态的DFG-in构象 (1P38³⁰)和处于与ATP结合的DFG-out构象 (1W83³¹)。目前实验上通过NMR实验³²，估计产生这种构象变化的时间尺度在微妙以上，因此用常规动力学模拟中获得其构象变化是极其困难的。我们将发展的方法应用到这些体系中，结果发现我们产生的中间结构与生物实验和其他模拟结果高度吻合，因此本方法可以广泛应用到大规

模的构象变化的体系中去。

3.2 研究方法

3.2.1 简振模式分析

经典的正则模式分析方法中，首先假定给定分子处在局域极小值点附近的势能是可以二次型分布近似的。由此，分子的运动可以被分解为一系列独立的简谐振动模式，即正则模式¹⁸。这些正则模式常被视作是分子的运动模式，并以它们的线性组合来描述分子的整体运动。一个含有N个原子的分子具有 3N个自由度，其二阶力矩阵H为 3N*3N维，通过对二阶力矩阵H对角化就可以获得每个振动模式(3N维矢量)。矩阵H的矩阵元可以写作：

$$H_{ij} = \frac{\partial^2 V}{\partial x_i \partial x_j} \quad (\text{式 3.1})$$

V是体系势能， x_i 和 x_j 是原子i和j的坐标。通过将H矩阵进行对角化变化得到下式

$$H = U \Lambda U^T \quad (\text{式 3.2})$$

其中 U^T 是正交矩阵U的转置矩阵，U的每一列就是矩阵H的本征矢量，而对角矩阵 Λ 是矩阵H的本征值，或称为正则模式的力常数，这样一共可以获得 3N个振动模式，其中频率最小的 6 个是本征值为 0 的模式，对应的是平动和转动。从这些正则模式还可以计算各结构单元的动力学性质，如不同原子之间的波动的相关性和蛋白质分子构象变化的方向等。

基于全原子的正则模式分析的缺点主要是计算量非常大，主要体现在 2 个方面，首先是需要将体系优化到局域极小值点附近，其次是对H矩阵的对角化。所以近年来，基于粗粒化的简振模式分析方法被广泛应用。如只利用Ca原子坐标的弹性网络模型，有基于高斯网络模型(GNM)^{33, 34}和各向异性网络模型(ANM)³⁵⁻³⁷两种。另外一种是通过限制体系中的某些运动自由度的方法，如iMod¹⁹方法，可

以大大减少体系自由度，通过使用内坐标的形式，使计算量大大减少，其在构象变化的研究中也得到了应用。因为简振模式代表着蛋白质运动的方向，计算得到正则模式后可以利用式 3.3 计算得到蛋白质随着某一个正则模式运动后的构象

$$x_i' = x_i + s u_i^k \quad (\text{式 3.3})$$

其中 x_i 和 x_i' 为蛋白质沿着简振模式运动前后的构象， s 为运动的幅度， u_i^k 代表其中第 i 个原子的第 k 个模式。

3.2.2 使用 NMA 产生构象变化的中间结构

研究中计算 NMA 使用的软件是 iMOD¹⁹。其中 AdK 蛋白的起始结构为没有结合抑制剂的 4AKE²⁴，而目标结构为结合了抑制剂的结构 1AKE²⁵。CaM 的起始结构则为没有钙离子结合的晶体结构 1CFD¹⁷，而它的目标结构为钙离子结合的结构 1CLL²⁹。同样的 p38 α 的起始和目标结构分别为 DFG-in 的结构 1P38³⁰ 和 DFG-out 结构 1W83³¹。

为了产生从起始构象到目标构象的中间结构，我们利用 NMA 迭代产生结构，使产生的结构逐渐与目标结构接近。在第 k 代中中间结构 $R^{(k)}$ 由下式计算得到：

$$R^{(k)} = R^{(k-1)} + v^{(k)} = R^{(k-1)} + S^{(k)} \sum_i^{m^{(k)}} (d^{(k-1)} \cdot u_i^{(k)}) u_i^{(k)} \quad (\text{式 3.4})$$

其中 $R^{(k-1)}$ 是第 $k-1$ 代产生的结构， $v^{(k)}$ 是沿着 $m^{(k)}$ 个简谐模式的偏移位移。 $S^{(k)}$ 是步长，为了控制迭代数目，在这里设置的是 10。 $d^{(k-1)}$ 表示第 k 代的起始结构与目标结构的三维坐标之差， $u_i^{(k)}$ 表示第 k 代的第 i 个正则模式， $d^{(k-1)} \cdot u_i^{(k)}$ 则表征了正则模式与结构差之间的相关性。我们通过将 $d^{(k-1)} \cdot u_i^{(k)}$ 作用在正则模式中然后通过步长来调节每一步的构象变化。

$$C(m^{(k)}) = \sum_i^{m^{(k)}} \cos^2(d^{(k-1)}, u_i^{(k)}) \quad (\text{式 3.5})$$

我们使用式 3.5 来确定最终使用多少个正则模式来产生中间结构。其中 $m^{(k)}$ 是所需要的最少正则模式数数目，使得这个累计 $C(m^{(k)})$ 达到设定收敛值，在本研究中使用的是 0.8。而为了使中间结构更加合理，我们对每个产生的中间结构进行 1000 步的分子模拟优化，然后再进行下一步的 NMA 计算。

3.2.3 伞取样模拟

为了获得沿着构象变化路径的能量变化，我们使用从上面NMA计算获得的一系列中间结构作为伞取样模拟的起始结构。在每一个伞取样模拟中，我们在反应坐标 ΔD_{rmsd} 中加入了一个简谐势能 E_i ：

$$E_i = k(\Delta D_{\text{rmsd}} - \Delta D_{\text{min},i})^2 \quad (\text{式 3.6})$$

其中 k 为力常数，本研究中使用 $10 \text{ kcal/mol/\AA}^2$ ，其中反应坐标 ΔD_{rmsd} 是每个结构与起始和目标结构的RMSD之差（ $\Delta D_{\text{rmsd}} = \text{RMSD}_{(X_t, X_A)} - \text{RMSD}_{(X_t, X_B)}$ ）其中 X_t 是模拟的瞬时构象， X_A 和 X_B 是其中作为参考态的起始和目标的构象， $\Delta D_{\text{min},i}$ 是 ΔD_{rmsd} 被限制的值。研究中选取这个反应坐标是因为它可以精确表针构象相对于起始和目标的结构，而且已经在多种生物体系的构象变化中得到了广泛应用^{14, 38, 39}。

在 ΔD_{rmsd} 的空间中每隔 $0.4 \text{ \AA}$ 作为一个模拟窗口，这样在AdK中有 45 个窗口，而在CaM有 58 个窗口，而p38 α 则有 61 个窗口。蛋白和小分子使用的力场分别为AMBER03 力场⁴⁰和GAFF力场⁴¹。在每个窗口中，将选取的结构首先放在充满TIP3P的水盒子里，并加入抗衡离子使溶液呈电中性。然后对体系进行优化和升温：首先利用 $10 \text{ kcal/mol/\AA}^2$ 的力将蛋白和配体的重原子固定优化，然后使

用同样的限制在 500 ps内将体系从 0 K升温到 300 K。最后将平衡好的结构使用 10 kcal/mol/Å² 将其限制在反应坐标 $\Delta D_{\min,i}$ 中进行 2 ns模拟。其中对于AdK和CaM的体系 ΔD_{rmsd} 对应的都是蛋白的所有重原子，而在p38 α 中其RMSD计算的是活性loop区内的重原子。模拟是在 300 K下进行的，时间步长为 2 fs，温度使用Langevin动力学方法进行控温。所有含氢的原子的键长使用SHAKE算法进行限制，使用周期边界条件，PME方法计算静电相互作用，非键相互作用的截断值为 10.0 Å，所有的模拟都在修改过代码的AMBER10 中进行。最后 1.5 ns的数据用来计算自由能变化，使用的程序(WHAM)²²，其中二维自由能的计算采用的方法与Banavali等¹⁴一样。

3.3 结果与讨论

3.3.1 AdK 的构象变化机制

从X晶体的结构来看，AdK的构象从打开到关闭状态的转变可能是由于配体结合并诱导产生的⁴²。但是，最近Hanson等⁴³利用单分子FRET实验研究了空的AdK的柔性，结果发现在水溶液中，即使没有配体存在，AdK也存在着从打开到关闭的构象转变。而Katherine等⁴⁴通过使用NMR、FRET和动力学模拟等实验结果也表明AdK是一个非常柔性的蛋白。而且在实验中还发现AdK从完全打开的构象到部分关闭的构象是很容易发生的，发生的时间大概在纳秒级别⁴⁴，而蛋白要从完全打开的状态到完全关闭的状态则需要微妙到毫秒的时间，因此他们的结果显示AdK可能采取群体移动模型。

我们使用发展的NUMD方法计算得到了空的及有配体的AdK随反应坐标 ΔD_{rmsd} 的自由能图景。如图3.1所示：当没有配体时，处于关闭（ $\Delta D_{\text{rmsd}} > 6.0$ Å）的构象具有较高的能量，是不稳定的；而处于打开的构象（ $\Delta D_{\text{rmsd}} < -6.0$ Å）则相对比较稳定。这个结果与无配体结合的AdK的晶体结构只存在打开构象的实验事实是一致的。另外从图3.1上还可以看出AdK的构象可以在一个很宽的范围

进行转变, 因为 ΔD_{rmsd} 从 $-6.5 \text{ \AA}$ 到 $\sim 1.0 \text{ \AA}$ 的转变过程中并不存在明显的能垒。这一结果显示空的AdK蛋白的柔性非常大, 可以很快从打开的构象转变为部分关闭的构象, 这点和Hanson等⁴³的实验结果是高度吻合的。为了分析各个结构域的稳定性, 我们通过对符合条件 $-1.0 \text{ \AA} \leq \Delta D_{\text{rmsd}} \leq 0.0 \text{ \AA}$ 的结构进行聚类分析得到典型结构。如图3.2c所示, AdK的 ΔD_{rmsd} 从 $-6.5 \text{ \AA}$ 到 $\sim 1.0 \text{ \AA}$ 的过程中最大的变化集中在结构域LID中, 而另一个结构域NMP则基本没有变化, 这一结果显示结构域LID的柔性大于NMP, 这点和Katherine等的实验结果也是高度吻合的⁴⁴。

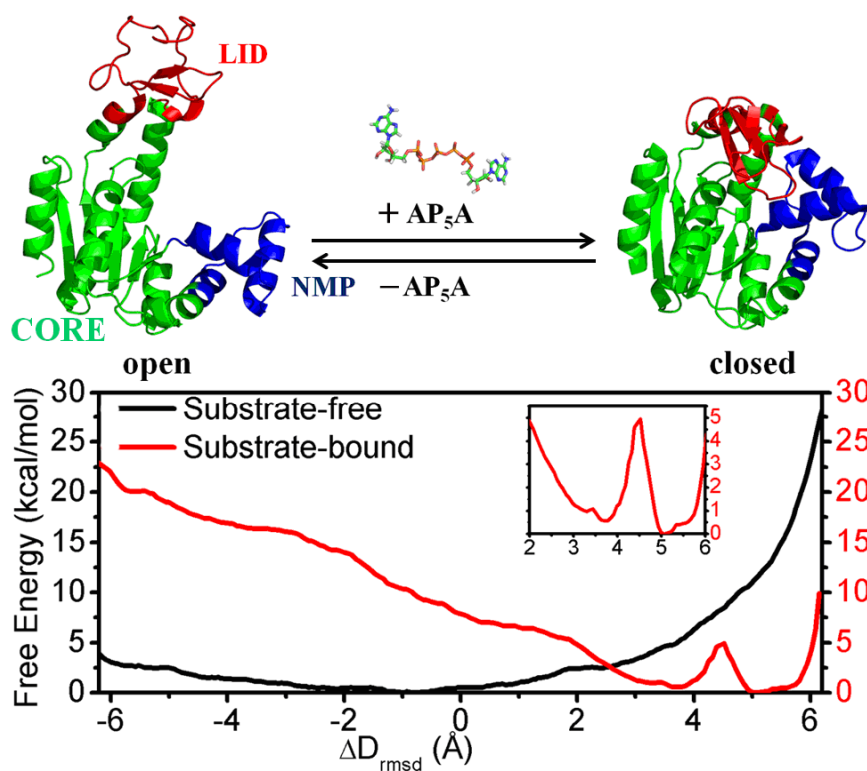


图 3.1 上图: 开和关的AdK的晶体结构 (PDB号分别为 4AKE和 1AKE), 中间为抑制剂AP₅A。下图沿着以反应坐标 ΔD_{rmsd} 的构象变化的自由能变化图 (黑线为空, 红线为有配体结合)。

另外值得注意的是当没有配体存在时, AdK的最低能量的构象并不位于晶体结构附近($\Delta D_{\text{rmsd}} = -6.5 \text{ \AA}$), 而是更接近关闭的结构 ($\Delta D_{\text{rmsd}} = -1 \text{ \AA}$)。如图 3.2c 所示, 其最主要的构象变化是LID的构象更趋向于关的结构($-1.0 \text{ \AA} \leq \Delta D_{\text{rmsd}} \leq 0.0 \text{ \AA}$), 而前者的能量比后者高了将近 4.0 kcal/mol , 这一结果与Hanson等⁴³通过FRET实验观察到的实验结果也是比较符合的。他们通过比较结构域LID和CORE之间的距离, 发现空的AdK结构域LID的构象更倾向于关闭的构象。这些结果显示AdK在溶液中是非常柔性, 即使没有配体的结合, 它的结构域LID也可以部分关闭, 而结构域NMP则相对比较稳定。

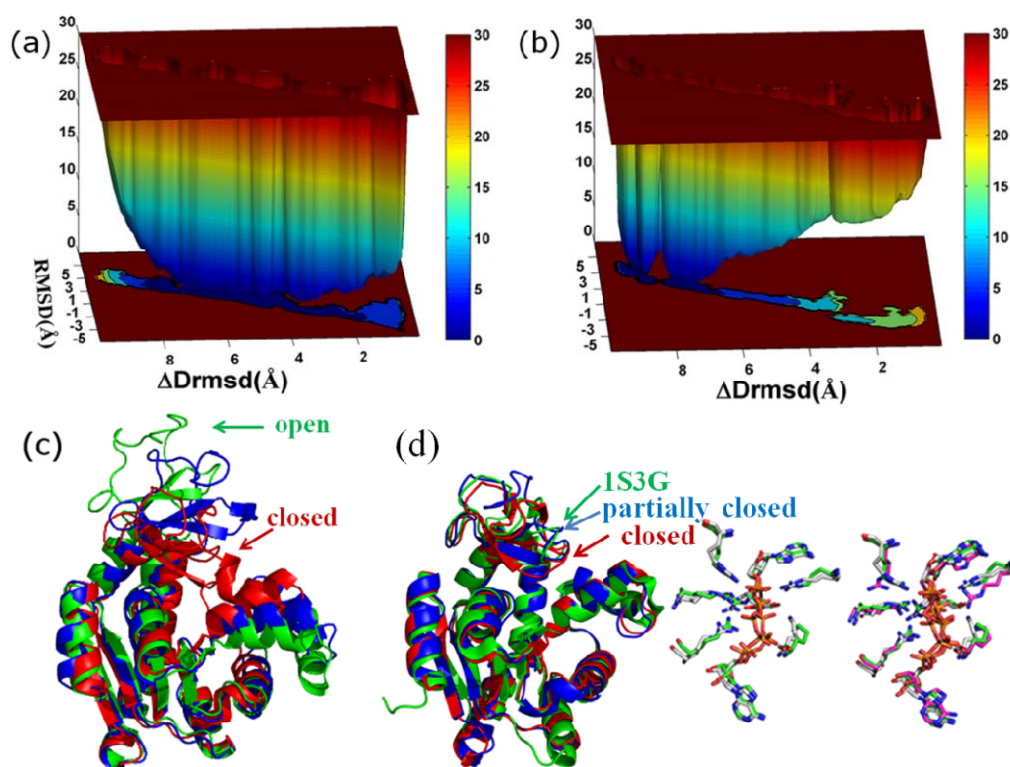


图 3.2 有无配体的AdK的三维自由能变化图 (a为空的AdK, b为结合AP₅A的AdK) 其中反应坐标分别为 ΔD_{rmsd} 和 $\text{RMSD}_{\text{open}}$ (c)空的AdK的典型构象 (蓝色) 与控的晶体结构 (绿色) 和关的结构 (红色) 的比较。(d)部分关闭的典型结构

（蓝色）与结合配体的关闭构象的比较，其中AdK的闭合构象绿色显示，同源蛋白 1S3G用红色显示。

而当AdK与抑制剂AP₅A相结合时，其自由能随着反应坐标 ΔD_{rmsd} 的情况则明显不同。如图 3.1 所示，当AdK结合抑制剂AP₅A后，其处于打开状态的构象变得非常不稳定，而关闭构象则变得相当稳定。可见抑制剂AP₅A的加入，使得AdK的构象明显的偏向于关闭构象。另外从自由能变化的曲线可以看出AdK与配体结合后呈现很明显的下山式的转变，只在 $\Delta D_{\text{rmsd}} \approx 4.5 \text{ \AA}$ 的位置存在一个较小的能垒。通过对没有配体结合的自由能曲线进行叠合，发现在 $\Delta D_{\text{rmsd}} = 2.5 \text{ \AA}$ 处这二者具有重合，因此配体与AdK的结合可能发生在这一位置。这样我们可以利用Marcus⁴⁵的电子转移理论来计算这个蛋白的活化能。从以上结果来看，由于AdK本身柔性很大，在溶液中具有很多的构象，各构象间处于动态平衡过程，当配体加入时，它可以和AdK的结合构象进行结合，从而影响AdK原来的构象变化平衡，最终导致AdK处于结合的构象，因此AdK的构象变化应该遵循群体移动模型（population-shift models）。

此外如图 3.2 所示,我们计算了以 ΔD_{rmsd} 和 $\text{RMSD}_{\text{open}}$ （相对于打开构象的RMSD）为反应坐标的三维自由能图景。为了定性比较它们的构象，我们对在动力学模拟过程中符合条件（ $3.0 \text{ \AA} < \Delta D_{\text{rmsd}} < 4.0 \text{ \AA}$ 且 $7.0 \text{ \AA} < \text{RMSD}_{\text{open}} < 8.0 \text{ \AA}$ ）的构象进行聚类分析结果如图 3.2d所示。结果显示这个构象与完全关闭的构象的主要区别就是结构域LID比晶体结构更倾向于打开的状态。而且在晶体数据库中我们发现这种模拟得到的构象也存在于AdK的同源蛋白中，如来源于圆孢芽孢杆菌的同源蛋白（PDB号为 1S3G⁴⁶）。通过将这些结构（实验上的关闭结构 1AKE，同源蛋白 1S3G和模拟得到的典型结构（ $3.0 \text{ \AA} < \Delta D_{\text{rmsd}} < 4.0 \text{ \AA}$ ））进行叠合，发现这三个结构的抑制剂AP₅A和周围主要残基的相互作用都是一样的。正是由于这些相同的结合作用，导致这个构象和晶体结构的构象一样是很稳定的。

而如图 3.1 所示, 当 $\Delta D_{\text{rmsd}} \approx 4.5 \text{ \AA}$ 左右的时候存在着一个较小的能垒, 正好对应着结构域LID打开与关闭。

Wolf-Watz等⁴⁷通过动态NMR技术研究了AdK蛋白的结构域LID的柔性与催化效应的关系, 结果显示在嗜温性大肠杆菌中的结构域LID开、关的速率分别为 $k_{\text{open}} = 286 \pm 85 \text{ s}^{-1}$, $k_{\text{close}} = 1374 \pm 110 \text{ s}^{-1}$, 而催化速率为 $k_{\text{cat}} = 263 \pm 30 \text{ s}^{-1}$; 同样的在细菌 *Aquifex aeolicus* 中 AdK 的结构域LID 开、关的转换速率分别是 $k_{\text{open}} = 44 \pm 20 \text{ s}^{-1}$ 和 $k_{\text{close}} = 1571 \pm 100 \text{ s}^{-1}$, 催化速率则为 $k_{\text{cat}} = 30 \pm 10 \text{ s}^{-1}$ 。上述结果显示AdK的催化速率与结构域LID的打开运动是密切相关的, 结构域LID的打开运动是这类酶催化的限速步骤。我们模拟的结果显示AdK的结构域LID打开与到关闭过程(ΔD_{rmsd} 从 $2.0 \text{ \AA}$ 到 $6.0 \text{ \AA}$) 的能垒约为 0.62 kcal/mol 。根据这个能垒, 可以计算出 LID 关闭的速度比打开的速度大 2.85 倍 ($k_{\text{close}} / k_{\text{open}} \propto \exp((\Delta G_{\text{open}}^* - \Delta G_{\text{close}}^*) / k_B T)$), 而这个和实验上观察到的数值 ($k_{\text{close}} / k_{\text{open}} = 4.80$) 是非常接近的, 可见我们模拟的结果与实验数据非常吻合⁴⁷。

3.3.2 CaM 构象变化机制研究

钙调蛋白在晶体结构中不仅存在打开和关闭两种状态^{17, 29}, 而且当这个蛋白与靶蛋白或者多肽相结合时, 中间的 α 螺旋会进一步弯曲, 形成一系列从完全伸展到紧密的结构⁴⁸⁻⁵⁰, 因此CaM在溶液的柔性是非常大的。我们同样计算得到了这个蛋白随着构象变化的自由能图景, 如图 3.3 所示。结果显示不管有无结合钙离子, CaM的构象之间进行转变并不存在较大的能量势阱, 表明在溶液CaM的柔性很大。结果显示在空的CaM可以很容易使其构象从 $\Delta D_{\text{rmsd}} = -10.0 \text{ \AA}$ 变化到 $6.0 \text{ \AA}$; 同样在钙离子存在时, 构象也很容易从 $\Delta D_{\text{rmsd}} = -5.0 \text{ \AA}$ 运动到 $10.0 \text{ \AA}$ 。这个结果显示溶液中不管是否存在钙离子, CaM蛋白都可以存在大量差别很大的构象。另外从图 3.3 也可以看出, 当钙离子结合的时候是倾向于打开的构象,

而没有钙离子存在的时候则倾向于关闭的构象，这点和实验中观察到的晶体结构是一致的。

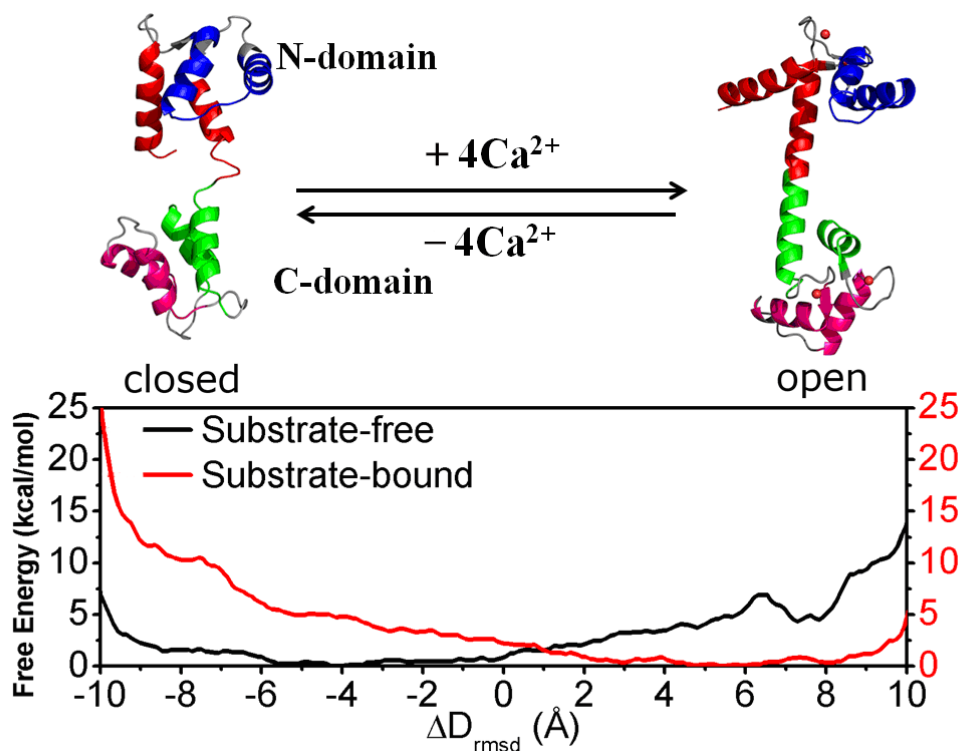


图 3.3 上图CaM的开和关的晶体结构（1CLL和 1CFD）。下图沿着以反应坐标 ΔD_{rmsd} 的构象变化的自由能变化图（黑线为无钙，红线为有钙结合）。

我们通过对PDB数据库中存在的CaM的结构进行搜索，结果发现存在着大量结构各异的构象。尤其值得关注的是我们发现没有结合靶蛋白的构象（包括晶体结构和NMR结构）大都存在于我们模拟得到的转变路径中（见图 3.4 a和 4b）。其中，CaM的NMR结构 1CFD（没有钙离子）和 1DMO（与钙离子结合）包含着构象各异的结构。在转变路径中，前者的结构集中于关闭的构象而后者包含的结构则分布更为广泛，这个与X晶体结构上捕捉到这两个结构的结果是一

致的。这些结果进一步显示我们使用的这个方法可以有效地捕捉到与实验符合的中间结构。另外从自由能图景上可以看出CaM不管有无钙离子存在都显示出极大的柔性，而钙离子的存在只是影响了它的动态平衡，显示这个蛋白的构象变化也符合群体移动模型⁴⁸。

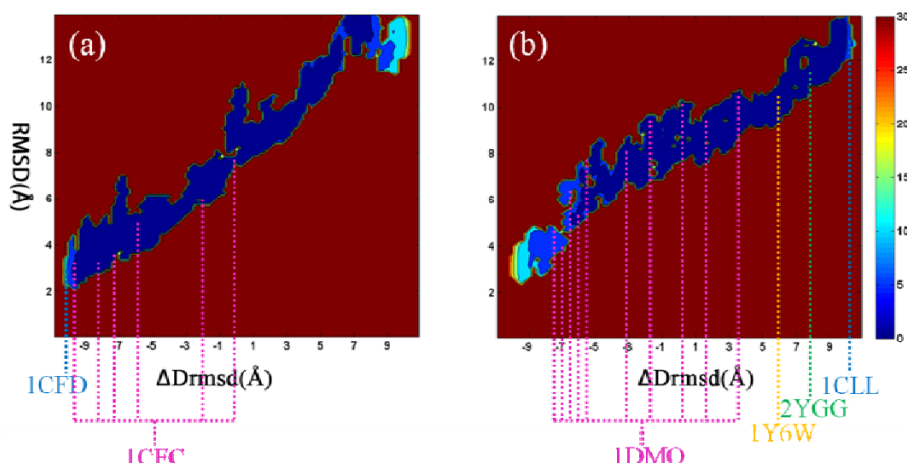


图 3.4 CaM的三维自由能变化图(a为空，b为结合钙离子)其中反应坐标分别为 ΔD_{rmsd} 和 $\text{RMSD}_{\text{open}}$

3.3.3 激酶 p38 α 构象变化机制研究

为了考察抑制剂结合导致p38 α 的构象变化的机制，我们把没有配体的DFG-in的晶体结构（1P38³⁰）和有抑制剂L11结合的处于DFG-out构象³¹作为模拟的起始和目标结构。在这个体系中由于蛋白整体并没有大的构象变化，而只有活性残基loop具有较大变化，所以为了更加清晰地分析这个loop的构象变化的能量情况，我们集中于分析活性loop区域的 ΔD_{rmsd} ，并计算获得构象变化的自由能图景。如图 3.5 所示，发现不管有无配体结合，都存在 2 个明显的局域极小值，对应着DFG-in 和DFG-out构象，它们中间存在着一个很大的能垒。这一结果与前面的 2 个体系相比，具有明显不同的图形，揭示p38 α 蛋白构象变化的机制可

能与AdK和CaM不同。

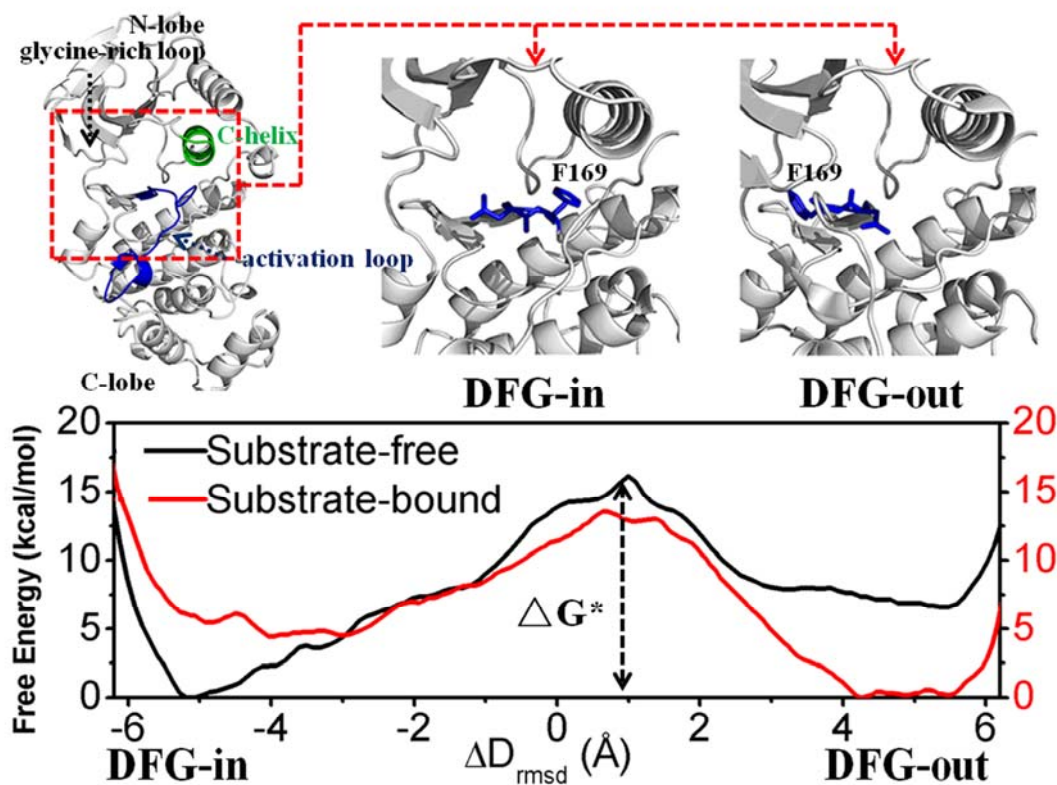


图 3.5 上图p38 α 激酶的晶体结构和DFG附件的放大图，其中DFG-in的结构是来自(1P38)，而DFG-out来自 1W83。下图沿着以反应坐标 ΔD_{rmsd} 的构象变化的自由能变化图（黑线为无配体，红线为有配体结合）

如图 3.5 所示，在没有抑制剂的时候，DFG-in状态比DFG-out状态的自由能约下降 7.0 kcal/mol，而且这两种状态转变的能垒高达 16.4 kcal/mol。由此可见没有抑制剂存在时，p38 α 的DFG-in构象不管是在动力学，还是热力学上都很稳定，这点和使用NMR很容易观察到DFG-in构象，而几乎观察不到DFG-out的构象的实验结果一致^{52, 53}。另外在PDB数据库中，p38 α 存在大量的DFG-in的构象，却只存在为数不多的DFG-out的构象的实验结果也是一致的⁵⁴。通过文献调研发

现, 这里的计算结果与其他使用动力学模拟得到的结果也非常相似, 如Lovera等⁵⁵利用MD模拟获得的c-Src和c-Abl 结合imatinib的能量差分别为 4.0 ± 0.5 和 6.0 ± 0.5 kcal/mol。此外这里计算获得的活化能也与计算获得的同源蛋白CDK5¹⁵和c-Src⁵⁶的活化能 (~ 16 - 20 kcal/mol和 ~ 20 kcal/mol) 也是高度一致的。

Vogther等³²利用在NMR中监测F169 的氨基的信号的方法, 推测空的p38 α 是处于DFG-in到DFG-out构象的动态平衡过程, 但是这个过程的发生是比较慢的过程, 时间尺度估计在微妙到毫秒范围。为了与实验结果进行比较, 我们利用Kramers理论计算了空的p38 α 从DFG-in 到DFG-out转变发生的时间, 使用的公式如下所示:

$$t = \frac{2\pi k_B T}{\omega_{\min} \omega_{\max} D_{\max}} \exp\left(\frac{\Delta G_{in \rightarrow out}^*}{k_B T}\right) \approx \frac{2\pi k_B T}{\omega_{\min}^2 D_{\min}} \exp\left(\frac{\Delta G_{in \rightarrow out}^*}{k_B T}\right) \quad (\text{式 3.7})$$

其中 $\omega_{\min}$ 和 $\omega_{\max}$ 分别代表频率, 即在自由能图中DFG-in到中间状态的曲率, $D_{\min}$ 和 $D_{\max}$ 分别为处于 2 种状态的扩散系数, $\Delta G_{in \rightarrow out}^*$ 为能垒高低, k_B 是波尔兹曼常数, T是温度。由于 $\omega_{\min}$ 可以近似为 $\omega_{\max}$, 而 $D_{\min}$ 也可近似为 $D_{\max}$, 如图 3.5 所示, $\Delta G_{in \rightarrow out}^*$ 为 16.4 kcal/mol, $\omega_{\min}$ 为 1.32 s^{-1} 。而 $D_{\min}$ 通过将空的p38 α 在显性水中进行 10ns的分子模拟计算得到, 为 $3.62 \times 10^{-9} \text{ m}^2/\text{s}$ 。利用式 3.7 计算得到构象转换时间为 5.9 ms, 和实验上观察到的时间尺度是比较符合的。另外 D.E. Shaw⁵⁷利用长时间动力学模拟和荧光淬灭实验发现同源蛋白c-Abl的DFG-in到DFG-out的构象转变发生的时间尺度在 10 毫秒量级, 与我们的计算结果也是吻合的。

而当抑制剂L11 结合到p38 α 的时候, DFG-in状态的能量比DFG-out的能量还高 4.8 kcal/mol, 而且能垒降低到 9.0 kcal/mol, 与没有配体时相比下降了近 7.2 kcal/mol (见图 3.5)。与之相反, 从DFG-out转变成DFG-in的能垒则达到了 13.0 kcal/mol, 比没有配体时高了 4.3 kcal/mol。这些结果显示, 配体L11 的结合不仅

将DFG-in转变成DFG-out的能垒下降,而且还稳定了DFG-out的构象,使蛋白处于DFG-out的构象而不能重新转变成DFG-in的构象。为了考察抑制剂对中间结构的影响,我们计算了以 ΔD_{rmsd} 和 $\text{RMSD}_{\text{DFG-in}}$ (相对于DFG-in结构的RMSD)为反应坐标的三维自由能图景(见图 3.6a和 3.6b)。从中可以看出中间状态主要位于 $0.0 \text{ \AA} < \Delta D_{\text{rmsd}} < 2.0 \text{ \AA}$ 且 $5.0 \text{ \AA} < \text{RMSD}_{\text{DFG-in}} < 8.0 \text{ \AA}$ 的范围,因此,我们对动力学轨迹中符合这个条件的构象进行聚类,得到了典型结构如图 3.6c和 3.6d)所示。

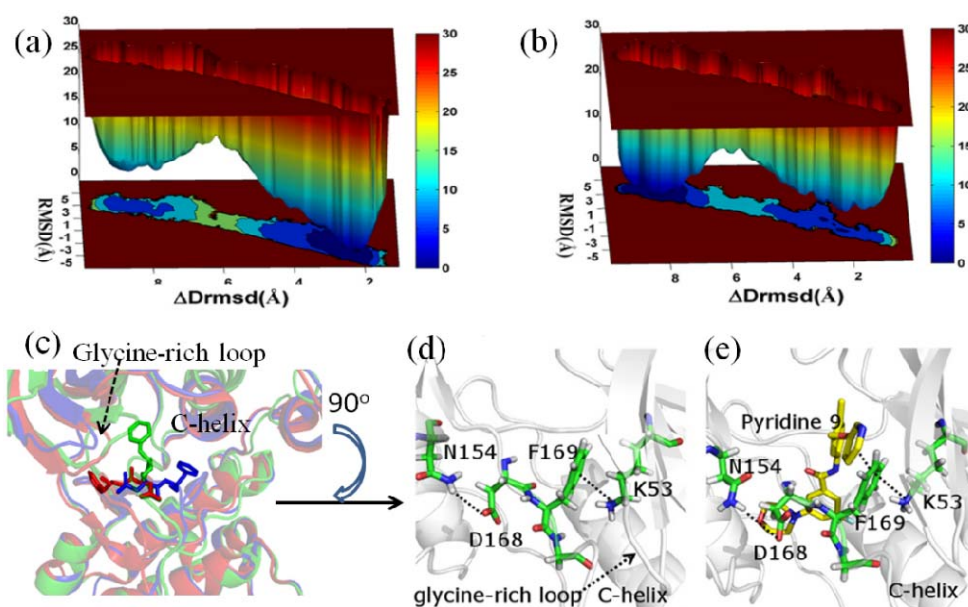


图 3.6 有无配体结合的p38 α 的三维自由能变化图(a为空,b为结合L11)其中反应坐标分别为 ΔD_{rmsd} 和 $\text{RMSD}_{\text{open}}$ 。c为将DFG-in、DFG-out和中间状态结构叠合的示意图。d和e分别为空的和结合配体的p38 α 的中间状态结构图。

从图中可以看到在没有结合配体的p38 α 中间结构中, F169 能够与K53 形成阳离子- π 相互作用, 而同时D168 的羰基可以和N154 的侧链形成氢键, 同样的作用在结合配体L11 时也存在。另外配体L11 还可以与F169 形成 π - π 相互作用,

进一步稳定中间状态的构象。Filomia等⁵⁸通过加速MD方法也同样发现中间状态的结构中存在F169 与K53 形成的阳离子- π 作用, 显示这个作用在DFG-in 到DFG-out的转变过程中起着重要作用。我们的结果显示, L11 不仅可以结合DFG-out构象, 也可以与DFG-in构象相结合, 配体的存在使得这两种构象的自由能发生改变, 同时降低了转变的能垒, 从而促使DFG-in的构象往DFG-out的构象改变, 显示p38 α 蛋白的构象变化更可能采取的是诱导契合的模型。

3.4 小结

最近几十年来, 与生物体功能密切相关的蛋白质构象变化越来越受到人们的重视, 很多实验和理论的方法被提出用来解释蛋白质构象变化的分子机制。然而由于捕捉中间状态的困难, 导致对其机制的了解并不清晰。其中最主要的问题就是: 蛋白质构象变化与配体结合的关系到底是什么? 其中最主要的理论有诱导契合模型和群体移动模型。诱导契合模型则是在 1958 年提出, 基于的理论是: 酶是具有一定柔性结构的蛋白质分子, 它的活性位点在结合底物的过程中会通过和底物分子相互作用, 不断发生一定变化以适应配体的结合⁵⁹。而群体移动模式^{42, 60, 61}则是近年来提出的一种新的酶与底物的结合模式, 其基于的假设是, 酶在溶液中同时存在不同构象, 一种构象(构象A)为适合底物结合的构象, 而另一种(构象B)则不适合, 这两种构象之间保持着动态平衡。在没有底物存在的情况下, 构象B占主导地位; 当加入底物后, 随着底物不断与构象A结合, 溶液中构象A含量下降, 两种构象之间的平衡被打破, 导致构象B不断地转化为构象A。

在本章中我们使用发展的简振模式分析与伞取样动力学模拟相结合的方法对蛋白质及其与配体结合的构象变化进行研究, 通过对构象变化的路径及自由能图景的计算来考察蛋白质构象变化是由哪些因素决定的。通过对有无配体的AdK 及 CaM 体系的研究发现这 2 个体系在溶液中都是高度柔性的, 配体的加入

只是影响了溶液中的构象平衡，使得空的状态自由能升高，变得不稳定，而对结合的构象达到了稳定的作用，从而使溶液中的构象变化平衡发生改变，最终稳定在结合构象，显示在这 2 个体系中蛋白发生的构象变化是以群体移动模型为主的。但是我们对激酶 p38 α 的研究发现它的行为与前面 2 个体系完全不同：无论有无配体存在，这个蛋白都存在着 2 个极小值，而且中间存在着一定的能垒。配体的作用不仅仅是使蛋白的结合构象更加稳定，另一个更重要的作用在于配体可以稳定中间结构，从而使得构象转变的能垒大幅下降，使得原本不易从 DFG-in 构象到 DFG-out 构象的转变变得可能，因此这个体系更符合诱导契合模型。从这些模拟体系的结果来看，群体移动模型和诱导契合模型都存在于蛋白质构象变化的体系中，而不同的蛋白会根据其功能的不同而选择不同的方式。另外我们通过与实验及其他模拟结果的比较发现本章使用的这种方法可以有效获得准确的自由能图景，其在蛋白质大规模构象变化中具有较好的应用前景。

参考文献

1. Bakan, A. and I. Bahar, *The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding*. Proc. Natl. Acad. Sci. USA, 2009. **106**(34): p. 14349-14354.
2. Grant, B.J., A.A. Gorfe, and J.A. McCammon, *Large conformational changes in proteins: signaling and other functions*. Curr. Opin. Struct. Biol., 2010. **20**(2): p. 142-147.
3. Kokkinidis, M., N.M. Glykos, and V.E. Fadouloglou, *Protein Flexibility and Enzymatic Catalysis*. Structural and Mechanistic Enzymology: Bringing Together Experiments and Computing, 2012. **87**: p. 181-218.
4. Westenhoff, S., et al., *Time-resolved structural studies of protein reaction dynamics: a smorgasbord of X-ray approaches*. Acta Crystallogr. A, 2010. **66**: p. 207-219.
5. Skjaerven, L., N. Reuter, and A. Martinez, *Dynamics, flexibility and ligand-induced conformational changes in biological macromolecules: a computational approach*. Future Med. Chem., 2011. **3**(16): p. 2079-2100.
6. Mei, Y., et al., *Electrostatic polarization is critical for the strong binding in streptavidin - biotin system*. J. Comput. Chem., 2012. **33**(15): p. 1374-1382.
7. Li, Y.L., et al., *Structure and Dynamics of a Dizinc Metalloprotein: Effect of Charge Transfer and Polarization*. J. Phys. Chem. B, 2011. **115**(33): p. 10154-10162.
8. Tong, Y., et al., *Electrostatic Polarization Makes a Substantial Contribution to the Free Energy of Avidin- Biotin Binding*. J. Am. Chem. Soc., 2010. **132**(14): p. 5137-5142.

9. Dror, R.O., et al., *Biomolecular Simulation: A Computational Microscope for Molecular Biology*. Annu. Rev. Biophys., Vol 41, 2012. **41**: p. 429-452.
10. Schlitter, J., M. Engels, and P. Kruger, *Targeted Molecular-Dynamics - a New Approach for Searching Pathways of Conformational Transitions*. J. Mol. Graph., 1994. **12**(2): p. 84-89.
11. Hamelberg, D., J. Mongan, and J.A. McCammon, *Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules*. J. Chem. Phys, 2004. **120**(24): p. 11919-11929.
12. Bussi, G., A. Laio, and M. Parrinello, *Equilibrium free energies from nonequilibrium metadynamics*. Phys. Rev. Lett., 2006. **96**(9).
13. Hansmann, U.H.E., *Parallel tempering algorithm for conformational studies of biological molecules*. Chem. Phys. Lett., 1997. **281**(1-3): p. 140-150.
14. Banavali, N.K. and B. Roux, *Free energy landscape of A-DNA to B-DNA conversion in aqueous solution*. J. Am. Chem. Soc., 2005. **127**(18): p. 6866-6876.
15. Berteotti, A., et al., *Protein Conformational Transitions: The Closure Mechanism of a Kinase Explored by Atomistic Simulations*. J. Am. Chem. Soc., 2009. **131**(1): p. 244-250.
16. Fornili, A., et al., *Energy Landscapes Associated with Macromolecular Conformational Changes from Endpoint Structures*. J. Am. Chem. Soc., 2010. **132**(49): p. 17570-17577.
17. Marrink, S.J. and D.P. Tieleman, *Perspective on the Martini model*. Chem. Soc. Rev., 2013.
18. Ma, J.P., *Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes*. Structure, 2005. **13**(3): p. 373-380.

19. López-Blanco, J.R., J.I. Garzón, and P. Chacón, *iMod: multipurpose normal mode analysis in internal coordinates*. Bioinformatics, 2011. **27**(20): p. 2843-2850.
20. Möcklinghoff, S., et al., *Design and evaluation of fragment-like estrogen receptor tetrahydroisoquinoline ligands from a scaffold-detection approach*. J. Med. Chem., 2011. **54**(7): p. 2005-2011.
21. Baugh, L., et al., *A Distal Point Mutation in the Streptavidin– Biotin Complex Preserves Structure but Diminishes Binding Affinity: Experimental Evidence of Electronic Polarization Effects?* Biochemistry, 2010. **49**(22): p. 4568-4570.
22. Hyre, D.E., et al., *Cooperative hydrogen bond interactions in the streptavidin–biotin system*. Protein Sci., 2006. **15**(3): p. 459-467.
23. Duan, Y., et al., *A point - charge force field for molecular mechanics simulations of proteins based on condensed -phase quantum mechanical calculations*. J. Comput. Chem., 2003. **24**(16): p. 1999-2012.
24. Gordon, J.C., et al., *H⁺⁺: a server for estimating pK_as and adding missing hydrogens to macromolecules*. Nucleic Acids Res., 2005. **33**(suppl 2): p. W368-W371.
25. Wang, J., et al., *Development and testing of a general amber force field*. J. Comput. Chem., 2004. **25**(9): p. 1157-1174.
26. Jiao, D., et al., *Calculation of protein–ligand binding free energy by using a polarizable potential*. Proc. Natl. Acad. Sci. USA, 2008. **105**(17): p. 6290-6295.
27. Steinbrecher, T. and A. Labahn, *Towards accurate free energy calculations in ligand protein-binding studies*. Curr. Med. Chem., 2010. **17**(8): p. 767-785.

28. Kumar, S., J. Boehm, and J.C. Lee, *p38 MAP kinases: key signalling molecules as therapeutic targets for inflammatory diseases*. *Nat Rev. Drug Discov.*, 2003. **2**(9): p. 717-726.
29. Chattopadhyaya, R., et al., *Calmodulin structure refined at 1.7 Å resolution*. *J. Mol. Biol.*, 1992. **228**(4): p. 1177-1192.
30. Wang, Z., et al., *The structure of mitogen-activated protein kinase p38 at 2.1-Å resolution*. *Proc. Natl. Acad. Sci. USA*, 1997. **94**(6): p. 2327-2332.
31. Gill, A.L., et al., *Identification of Novel p38α MAP Kinase Inhibitors Using Fragment-Based Lead Generation*. *J. Med. Chem.*, 2004. **48**(2): p. 414-426.
32. Vogtherr, M., et al., *NMR characterization of kinase p38 dynamics in free and ligand-bound forms*. *Angew. Chem.*, 2006. **45**(6): p. 993-997.
33. Yang, L.-W., et al., *iGNM: a database of protein functional motions based on Gaussian Network Model*. *Bioinformatics*, 2005. **21**(13): p. 2978-2987.
34. Yang, L.-W., et al., *oGNM: online computation of structural dynamics using the Gaussian Network Model*. *Nucleic Acids Res.*, 2006. **34**(suppl 2): p. W24-W31.
35. Eyal, E., L.-W. Yang, and I. Bahar, *Anisotropic network model: systematic evaluation and a new web interface*. *Bioinformatics*, 2006. **22**(21): p. 2619-2627.
36. Nicolay, S. and Y.-H. Sanejouand, *Functional modes of proteins are among the most robust*. *Phys. Rev. Lett.*, 2006. **96**(7): p. 078104.
37. Atilgan, A., et al., *Anisotropy of fluctuation dynamics of proteins with an elastic network model*. *Biophys. J.*, 2001. **80**(1): p. 505-515.
38. Arora, K. and C.L. Brooks, *Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism*. *Proc.*

- Natl. Acad. Sci. USA, 2007. **104**(47): p. 18496-18501.
39. Yu, H.B., et al., *Mechanochemical coupling in the myosin motor domain. I. Insights from equilibrium active-site simulations*. PLoS Comput. Biol., 2007. **3**(2): p. 199-213.
40. Duan, Y., et al., *A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations*. J. Comput. Chem., 2003. **24**(16): p. 1999-2012.
41. Wang, J.M., et al., *Development and testing of a general amber force field*. J. Comput. Chem., 2004. **25**(9): p. 1157-1174.
42. Boehr, D.D., R. Nussinov, and P.E. Wright, *The role of dynamic conformational ensembles in biomolecular recognition*. Nat. Chem. Biol., 2009. **5**(11): p. 789-796.
43. Hanson, J.A., et al., *Illuminating the mechanistic roles of enzyme conformational dynamics*. Proc. Natl. Acad. Sci. USA, 2007. **104**(46): p. 18055-18060.
44. Henzler-Wildman, K.A., et al., *Intrinsic motions along an enzymatic reaction trajectory*. Nature, 2007. **450**(7171): p. 838-U13.
45. Marcus, R.A. and N. Sutin, *Electron Transfers in Chemistry and Biology*. Biochim. Biophys. Acta., 1985. **811**(3): p. 265-322.
46. Rastelli, G., et al., *Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA*. J. Comput. Chem., 2010. **31**(4): p. 797-810.
47. Wolf-Watz, M., et al., *Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair*. Nat. Struct. Mol. Biol., 2004. **11**(10): p. 945-949.
48. Gsponer, J., et al., *A coupled equilibrium shift mechanism in*

- calmodulin-mediated signal transduction*. Structure, 2008. **16**(5): p. 736-746.
49. Zhang, M., et al., *Structural Basis for Calmodulin as a Dynamic Calcium Sensor*. Structure, 2012. **20**(5): p. 911-923.
50. Wu, G.R., et al., *Calcium-induced changes in calmodulin structural dynamics and thermodynamics*. Int. J. Biol. Macromol., 2012. **50**(4): p. 1011-1017.
51. Chen, B.W., et al., *Helix A stabilization precedes amino-terminal lobe activation upon calcium binding to calmodulin*. Biochemistry, 2008. **47**(35): p. 9220-9226.
52. Pargellis, C., et al., *Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site*. Nat. Struct. Biol., 2002. **9**(4): p. 268-272.
53. Regan, J., et al., *Pyrazole urea-based inhibitors of p38 MAP kinase: From lead compound to clinical candidate*. J. Med. Chem., 2002. **45**(14): p. 2994-3008.
54. Huang, Y.M.M., et al., *Insights from Free-Energy Calculations: Protein Conformational Equilibrium, Driving Forces, and Ligand-Binding Modes*. Biophys. J., 2012. **103**(2): p. 342-351.
55. Lovera, S., et al., *The Different Flexibility of c-Src and c-Abl Kinases Regulates the Accessibility of a Druggable Inactive Conformation*. J. Am. Chem. Soc., 2012. **134**(5): p. 2496-2499.
56. Gan, W.X., S.C. Yang, and B. Roux, *Atomistic View of the Conformational Activation of Src Kinase Using the String Method with Swarms-of-Trajectories*. Biophys. J., 2009. **97**(4): p. L8-L10.
57. Ji, C., Y. Mei, and J.Z. Zhang, *Developing Polarized Protein-Specific Charges for Protein Dynamics: MD Free Energy Calculation of pKa Shifts for Asp²⁶/Asp²⁰ in Thioredoxin*. Biophys. J., 2008. **95**(3): p. 1080-1088.

-
58. Filomia, F., F. De Rienzo, and M.C. Menziani, *Insights into MAPK p38 alpha DFG flip mechanism by accelerated molecular dynamics*. Bioorg. Med. Chem., 2010. **18**(18): p. 6805-6812.
 59. Koshland, D.E., *Application of a Theory of Enzyme Specificity to Protein Synthesis*. Proc. Natl. Acad. Sci. USA, 1958. **44**(2): p. 98-104.
 60. Tsai, C.J., B.Y. Ma, and R. Nussinov, *Folding and binding cascades: Shifts in energy landscapes*. Proc. Natl. Acad. Sci. USA, 1999. **96**(18): p. 9970-9972.
 61. Volkman, B.F., et al., *Two-state allosteric behavior in a single-domain signaling protein*. Science, 2001. **291**(5512): p. 2429-2433.

第四章 基于QM/MM研究蛋白和配体间的可极化效应

4.1 引言

在生物体内，蛋白分子和配体相互作用是蛋白质分子发挥生物功能的重要途径，如抗原抗体的结合、酶催化反应等¹。因此蛋白质与配体相互结合过程的研究就成为研究蛋白质功能和药物分子设计的一个核心问题²。由于蛋白质和配体在相互作用的过程中通常会引起体系性质的改变，如电学、光学性能或构象的变化，因此其在材料科学和分析科学中也有重要应用³。最早关于蛋白质配体相互作用的理论是E. Fisher于1894年提出的“锁钥学说”⁴。随后，1958年Koshland提出了诱导契合的概念，即受体和配体在相互作用的过程中会伴随着相应的构象变化⁴。目前诱导契合的理论已成为当今生物分子之间相互识别、相互作用和有机分子自组装等领域的基础⁴。

蛋白质和配体的结合自由能是表征其相互作用的一个重要的物理量，所以其准确获得在药物设计领域是非常重要的²。目前实验上的测定方法主要有等温滴定量热法（Isothermal Titration Calorimetry）⁵和基于表面等离子体技术的biacore技术⁶。虽然这些实验方法可以获得准确的结合热力学性质，但由于蛋白质表达纯化及实验条件的限制，这些方法是非常费时费力的，并不适合大规模的药物筛选。

近年来，随着计算化学和分子模拟方法的发展，出现了很多结合自由能的计算方法⁷⁻¹⁴。最简单的是基于经验方程的计算方法，如定量构效关系¹⁴和在分子对接程序中广泛应用的打分函数¹³。然而定量构效关系很多情况下并不能明确给出模型的物理意义，而且所得模型的预测能力很大程度上受到实验数据的限制。分子对接所采用的打分函数则很依赖于其训练集的选取，不能很好地考虑体系的柔性以及溶剂效应，而这些在结合自由能的准确计算中起重要作用¹²。

分子动力学模拟不仅可以获得蛋白质和配体的结合自由能，还可以考察蛋白质与配体相互作用的动态行为，已经成为药物设计的重要理论工具¹¹。目前通过分子动力学方法获得结合自由能的方法有MM/PB(GB)SA (Molecular Mechanics/Poisson-Boltzmann or Generalized Born Surface Area)^{9, 10}，自由能微扰和热力学积分方法^{7, 8}等。虽然基于统计力学的自由能微扰和热力学积分可以获得更准确的结果，但是此方法计算量相对较大，而通过MM/PB(GB)SA的方法计算得到的结合能和生物测试值不一定完全符合，但是由于其获得的结果与很多生物体系的结合自由能具有很好的相关性，因而在实际研究中也得到了广泛应用¹⁵。

在蛋白质和配体结合自由能计算中起关键作用的是分子力场¹⁶，最近通过比较使用可极化力场和非极化力场获得的结合自由能，发现可极化力场在结合自由能的准确计算中具有较大的优势¹⁷⁻²²，这些研究也表明原子间的极化作用在蛋白质和配体的相互作用中起着重要作用。如Ren等利用AMOEBA可极化力场研究苯甲脒类抑制剂与胰蛋白酶的相互作用时发现静电极化起着不可忽略的重要作用^{21, 22}。但是目前常用的力场，如CHARMM²³，AMBER²⁴，OPLS²⁵和GROMOS²⁶等都采用固定点电荷的描述方式，不能有效考虑原子间的极化效应，所以在蛋白质和配体相互作用的计算上具有一定的缺陷，如Tong等²⁷在研究生物素（biotin）和亲和素（avidin）的结合自由能的时候发现非极化的力场，如AMBER03和AMBER94力场都不能分辨出生物素BTN1和其类似物BTN2的结合自由能的差别，而只有通过利用量化计算获得的极化电荷后，才能正确分辨出这2个配体的结合自由能的区别。

量子力学（QM）方法可以精确考察原子间的电子效应和量化效应，其在酶催化、结合自由能的计算中已经起到了重要作用²⁸。但对于蛋白质的尺寸来说，目前量子力学所能处理的体系还是太小，故其在生命体系中的应用受到了极大限制。为了扩大量子力学应用体系出现了将量子力学与分子力学方法结合起来

的QM/MM方法^{29, 30}, 即将体系中的重要部分使用QM来精确描述, 而对体系中起环境作用的部分使用MM来描述。这种联用方法的好处在于不仅具有QM计算的准确性, 还克服了QM处理体系较小和一些小分子力场参数难以准确获得的缺陷, 扩大了量子力学和分子动力学的研究范围, 已经在实际中得到广泛应用^{31, 32}。

本章为了准确计算蛋白质和配体间的结合自由能, 首先利用 QM/MM 方法得到每个残基的精确的静电势, 通过电荷拟合的方法得到每个原子的精确电荷, 然后利用这种考虑了极化效应的电荷进行动力学模拟, 通过比较 QM/MM 电荷和 AMBER03 电荷的计算结果来考察此种方法的适用性和极化作用的影响。为了验证此方法, 我们挑选了 2 个研究体系: 一是雌激素受体(ER)结合相似的小分子 (W23 和 W23H), 另一个是野生型和突变体 F130L 的亲激素结合同一个生物素 (BTN)。计算的结果发现, 使用了 QM/MM 电荷后的得到的结合自由能与实验结果比较相符, 而使用常规的 AMBER03 力场则不能很好地区分这 4 个体系, 从而也验证了此种 QM/MM 电荷的合理性和在计算小分子和蛋白的结合自由能计算中考虑极化效应的重要性。

4.2 研究方法

4.2.1 QM/MM 电荷的计算

华东师范大学的梅晔等人基于分而治之的思想, 发展了MFCC的方法, 这种方法在多种蛋白-配体和蛋白-蛋白相互作用中得到验证^{27, 33-35}。我们利用类似的分而治之的思想, 采用分块的方法, 将具有N个残基的蛋白质以残基为基础划分成N个QM/MM计算的体系, 然后对每个体系进行QM/MM计算, 获得各自的静电势, 进而拟合得到每个原子的点电荷, 最后通过迭代的方法得到一个收敛的蛋白质全原子的电荷, 具体方法如下:

首先将具有N个残基的蛋白质以每个残基为基准划分成N个体系, 如图 4.1 所示。其中每个体系包括核心量化区, 过渡量化区, 分子力学区和溶剂化电荷

区四个部分。其中核心量化区就是分块得到的第 i 个残基的原子，过渡量化区是指与核心量化区原子直接相连的残基，将这些残基包括在量化区可以减小量化计算的边界效应，让计算得到的残基的静电势更加准确。然后将体系中的其他原子作为分子力学区原子。因为量化计算通常是在真空中进行的，而目前的研究表明溶剂化效应对体系的影响非常大。因此为了考虑溶剂化效应，我们使用软件DELPHI获得蛋白质的溶剂化电荷³⁶，然后将这些电荷以背景电荷的方式放在QM/MM计算中。所有的量化计算均使用高斯 09 中的ONIOM方法³⁷，其中量化方法采用的方法是B3LYP，基组使用 6-31G*，而分子力学使用的是AMBER力场³⁸。接下来，将量化计算得到的每个残基的静电势用RESP方法^{39,40}重新拟合得到各个原子的点电荷。最后判断此时得到的电荷和原来的电荷是否收敛，如果不收敛，利用更新的电荷重新进行QM/MM计算直到收敛为止。

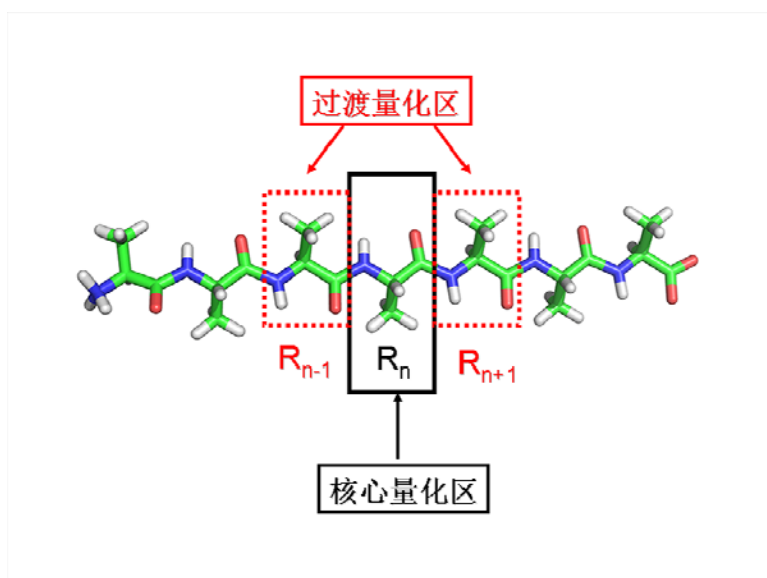


图 4.1 蛋白质分解成 QM/MM 区的示意图。第 N 个残基的原子划分成核心量化区的原子，与其相连的 2 个残基的原子划分成过渡量化区，而体系中的其他原子则划分为分子力学区。

4.2.2 分子动力学模拟

由于氟代化合物在活性和代谢方面具有的优良性质，近年来其在药物分子设计中得到广泛关注⁴¹。通过文献调研发现将雌激素受体(ER)的配体W23 的 3 个氟原子变成氢原子的话 (W23H)，这个化合物的活性比W23 下降了近 33 倍⁴²，在本章中我们想通过分子动力学模拟计算来研究这个活性丧失的机制。W23 和雌激素受体形成的复合物来自X-ray晶体结构 3OMQ⁴²，如图 4.2 所示。这个晶体结构的 1 到 263 号残基是缺失的，由于小分子结合口袋在另外一边，在模拟中没有补齐这部分残基。此外，在小分子结合域中缺失了 417-424 这 8 个残基，我们利用Discovery Studio 2.5⁴³将其补齐。而W23H-雌激素受体的复合物结构则是通过将W23 的三个氟取代成氢获得。

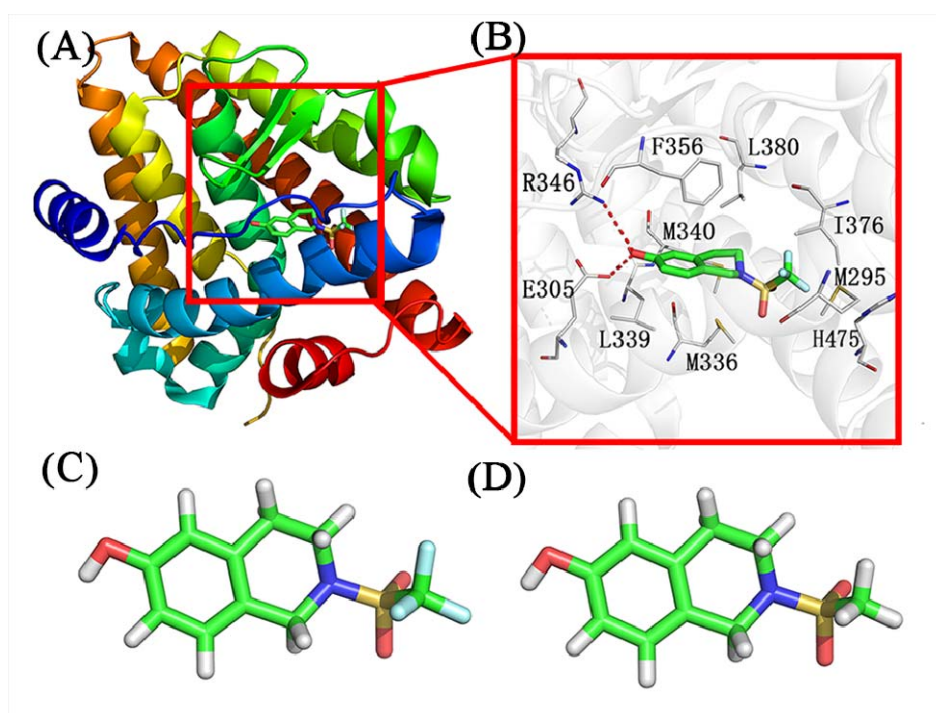


图 4.2 (A) 雌激素受体和其配体 W23 的晶体结构；(B) 为其结合口袋周围残

基的放大图; (C) 和 (D) 分别为小分子 W23 和其氢化物 W23H 的化学结构示意图。

另一个研究体系是生物素(BTN)结合野生型和F130L突变体的亲和素, 通过文献发现⁴⁴, 130 号残基的这个突变(F130L)造成了BTN活性近 975 倍的下降, 而比较晶体结构上并没有发现野生型和突变体有什么变化(如图 4.3 所示), 使用的晶体复合物分别来自晶体结构 1MK5⁴⁵和 3MG5⁴⁴。

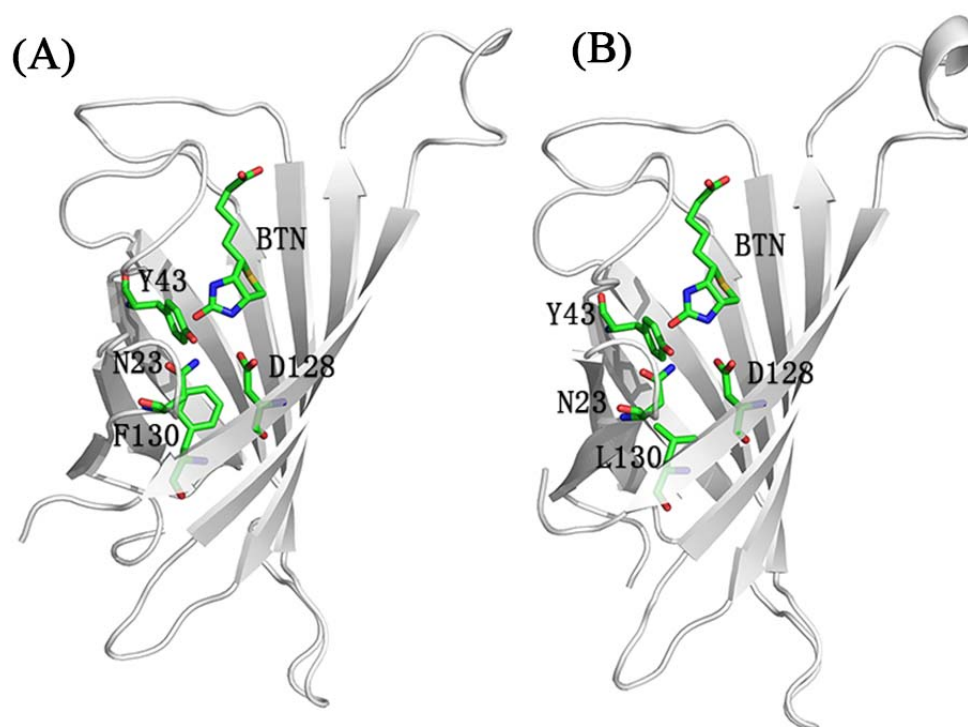


图 4.3 (A) 和 (B) 分别为野生型和突变型 F130L 的生物素亲和素的晶体结构, PDB 号分别为 1MK5 和 3MG5, 其中 Y43, N23 和 D128 是位于突变残基 F130 附近的能与 BTN 形成氢键的残基。

在所有的模拟中，蛋白质分子使用的力场都是AMBER03³⁸，晶体结构中缺失的氢原子使用tleap补齐。使用高斯 09 软件在B3LYP/6-31G*级别上分别计算了小分子BTN、W23 和W23H的静电势，通过RESP^{39, 40}方法拟合得到了这些小分子的原子电荷。其他的力场参数，如Lennard-Jones项，键扭转和角弯曲等项使用的GAFF 力场参数⁴⁶。另外为了考察我们的极化方法，我们将蛋白质原子的电荷分别使用QM/MM电荷替换，这样一共具有 8 个体系，详细情况见表 4.1。

表4.1动力学模拟体系信息

轨迹	体系	电荷	模拟时间(ns)
A1	ER-W23	AMBER03	20
A2	ER-W23	QM/MM电荷	20
A3	ER-W2H	AMBER03	20
A4	ER-W2H	QM/MM电荷	20
B1	WT-BTN	AMBER03	20
B2	WT-BTN	QM/MM电荷	20
B3	F130L-BTN	AMBER03	20
B4	F130L-BTN	QM/MM电荷	20

我们将得到的复合物和结晶水分子放在立方体的TIP3P水盒子里，其中盒子边缘距离溶质的最短距离为 10 Å，然后向体系加入适当的抗衡粒子中和体系的电荷。为了模拟pH=7 的状态，我们在模拟中使用经过H++⁴⁷滴定得到残基的质子化状态。使用PME方法⁴⁸处理长程静电相互作用，截断半径为 10 Å。含氢原子的键长的使用SHAKE算法⁴⁹来限制，设定的时间步长为 2 fs。

对每个体系分两步进行优化：第一步，以 500 kcal/mol/ Å² 的力约束溶质对体系进行 2500 步最陡下降法和 2500 步共轭梯度法优化；第二步优化时去除约束力，对系统分别进行 5000 步的最陡下降法和 5000 步的共轭梯度法优化。然后每个体系在 2 ns 内将温度从 0K 加热到 300K，随后进行 2ns 的恒温恒压动力学平衡（NPT 系综）。在升温和平稳过程中，都对蛋白和配体加上弱限制（力常数为 1 kcal/mol/ Å²）以减小体系波动。最后在 300K NPT 系综下进行 20ns 的常规分子动力学模拟。

4.2.3 MM/GBSA 计算结合自由能

此方法是将结合自由能分解成 3 部分能量，即包括分子力学势能（ E_{MM} ），溶剂化能（ G_{sol} ）和熵（ S_{MM} ），如式 4.1 表示：

$$G = U - TS = \langle E_{MM} \rangle + G_{sol} - T \langle S_{MM} \rangle \quad (\text{式 4.1})$$

获得体系的结合自由能则需要通过式 4.2 计算得到：

$$\Delta G = \Delta E_{MM} + \Delta G_{sol} - T \Delta S \quad (\text{式 4.2})$$

其中 ΔE_{MM} 和 $T \Delta S$ 分别指在真空中的使用分子力学计算的结合内能和构象熵变，而 ΔG_{sol} 则是体系的溶剂化自由能之差。在实际计算中 ΔE_{MM} 分解成范德华和静电相互作用两项，而 ΔG_{sol} 则分解成极性溶剂化能和非极性溶剂化能两项，分别使用式 4.3 和 4.4 计算得到：

$$\Delta E_{MM} = \Delta E_{ele} + \Delta E_{vdw} \quad (\text{式 4.3})$$

$$\Delta G_{sol} = \Delta G_{pol} + \Delta G_{nonpol} \quad (\text{式 4.4})$$

其中 ΔG_{pol} 是利用 Onufriev⁵⁰ 等人发展的 GB 模型计算所得，而溶剂和溶质的介电常数分别被设为 80 和 1。非极性部分 ΔG_{nonpol} 是基于溶剂可极化面积（SASA）

来计算：

$$\Delta G_{nonpol} = \gamma SASA + \beta \quad (\text{式4.5})$$

其中 $SASA$ 为体系溶剂可极表面积, 利用LCPO 算法计算得到⁵¹, 而 γ 和 β 分别设为 $0.0072 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$ 和 $0.0 \text{ kcal} \cdot \text{mol}^{-1}$ 。

$$\Delta S = \Delta S_{translational} + \Delta S_{rotational} + \Delta S_{vibrational} \quad (\text{式4.6})$$

结合过程中的熵变包括了转动、平动和振动三个部分, 构象熵变则通过式4.6计算得到, 我们利用Amber 10中的NMODE模块计算得到。

4.3 结果与讨论

4.3.1 雌激素受体与 W23 和 W23H 结合动力学分析

由于不同的配体所具有的极化效应是不一样的, 当结合不同的配体时蛋白质的静电势有着一定区别的, 如图4.4A和B所示, 当蛋白质结合了小分子W23和W23H后, 获得的QM/MM电荷与AMBER03的电荷相比, 都具有较大的区别的。而且如图4.4C所示, 即使配体的结构只具有微小的改变, 其也可造成整体蛋白的静电势的重新分配, 尤其是在离配体很近的位置。从我们获得的QM/MM电荷的区别主要集中在配体近的地方(原子序号分别在500, 2400和3500前后, 见图4.4C), 而在离配体较远的地方的其静电势受其影响则比较小。

为了评估体系是否达到动力学平衡, 我们计算了体系主链重原子的 RMSD 随模拟时间的变化曲线。从图 4.4 可以看出, 主链重原子的 RMSD 值在 10ns 后趋于平稳, 因此后面的 MM/GBSA 使用的是 10ns 后的结构。

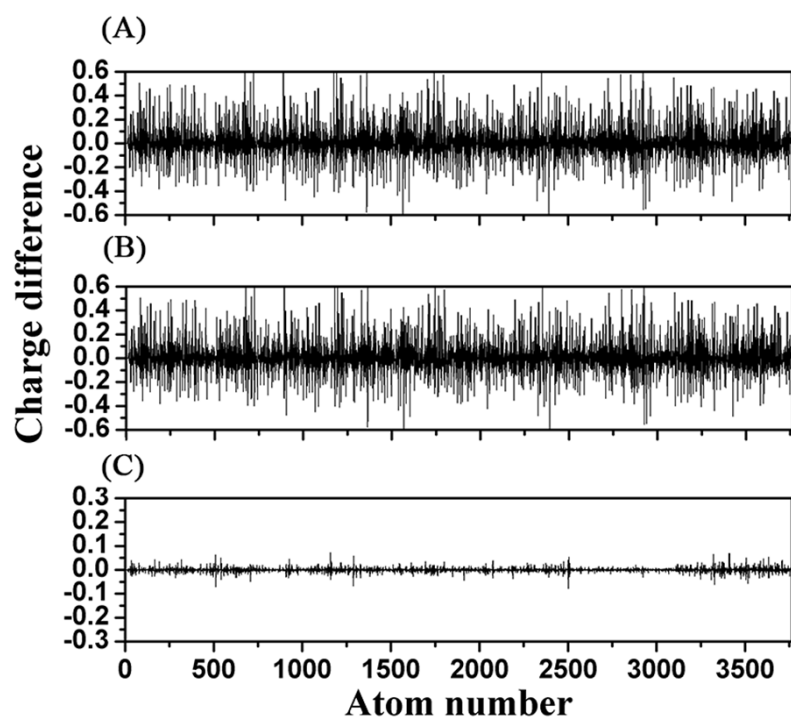


图 4.4 (A) 和 (B) 分别是量化计算得到的蛋白质各原子电荷与 AMBER03 力场中电荷之差(A 来自 W23 复合物, B 来自 W23H 复合物); (C) 是 W23 复合物和 W23H 复合物之间各原子电荷之差。

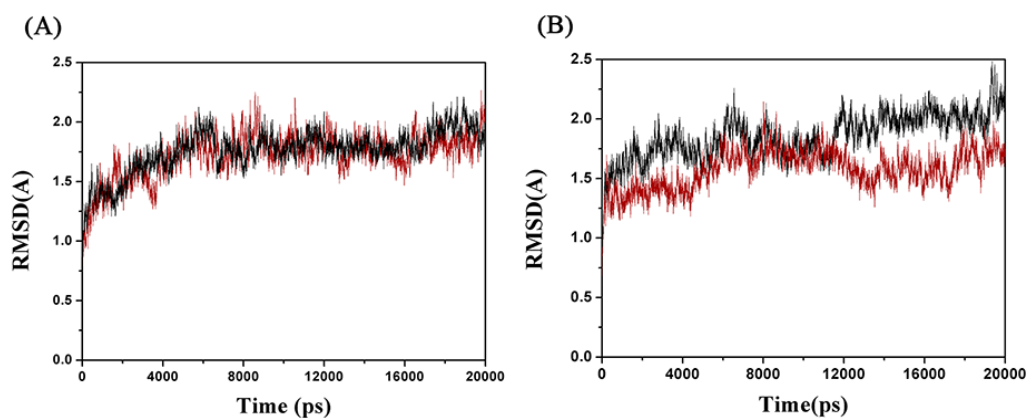


图 4.4 体系主链重原子的 RMSD 随时间变化的情况。(A) 中黑线和红线分别来自轨迹 A1 和 A2; (B) 中的黑线和红线则分别来自轨迹 A3 和 A4。

在晶体结构中发现（见图 4.2B），结合口袋中的R346 和E305 可以与配体形成的氢键。为了考察这个氢键的动力学行为，我们计算出这个氢键随时间变化的曲线，如图 4.5 和 4.6 所示。结果显示这 2 个残基与配体W23 形成的氢键在动力学模拟中都比较稳定，显示出这些氢键在配体的结合起着重要作用。而且 Sabine等⁴²发现将W23 的羟基还原成氢，使配体失去形成氢键的能力，则这个化合物就彻底失去活性。另外通过比较可以发现使用QM/MM电荷后，体系氢键的波动性与使用AMBER03 电荷相比会小一点，说明极化作用也有利于稳定氢键相互作用，这点与Tong等人²⁷的模拟结果是极其一致的。

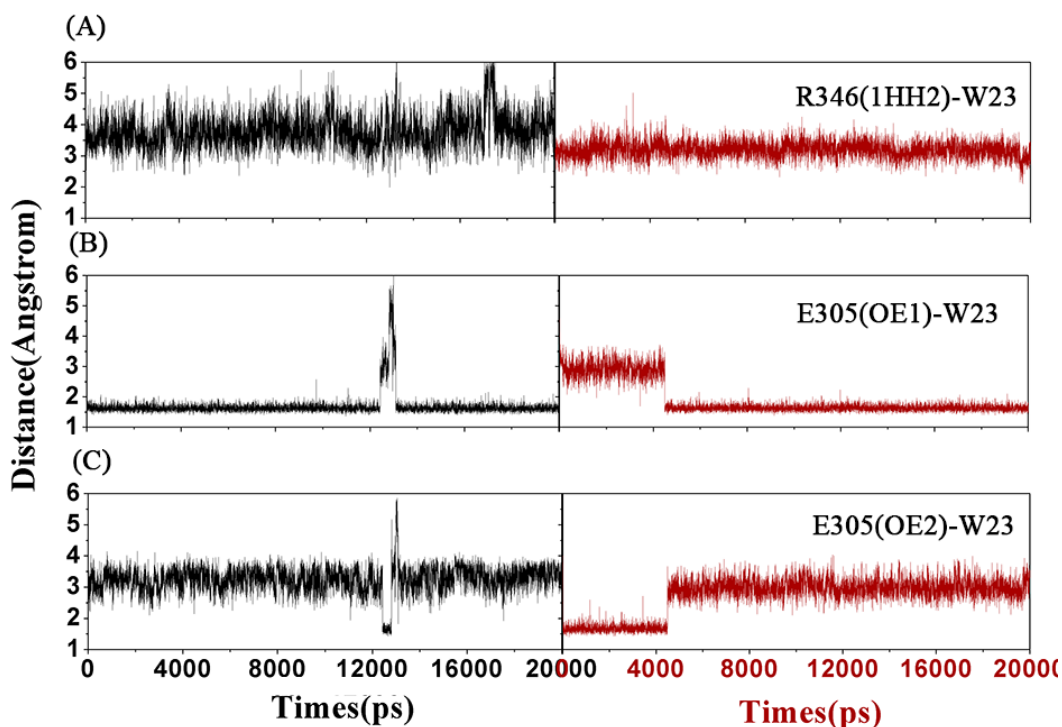


图 4.5 动力学模拟中 R346 和 E305 与配体 W23 形成的氢键随时间变化曲线（左边来自轨迹 A1,右边红色来自轨迹 A2）。（A）R346 与 W23 形成的氢键；（B）E305 中的羧基中的氧原子（OE1）与 W23 形成的氢键;（C）E305 中的羧基中的氧原子（OE2）与 W23 形成的氢键。

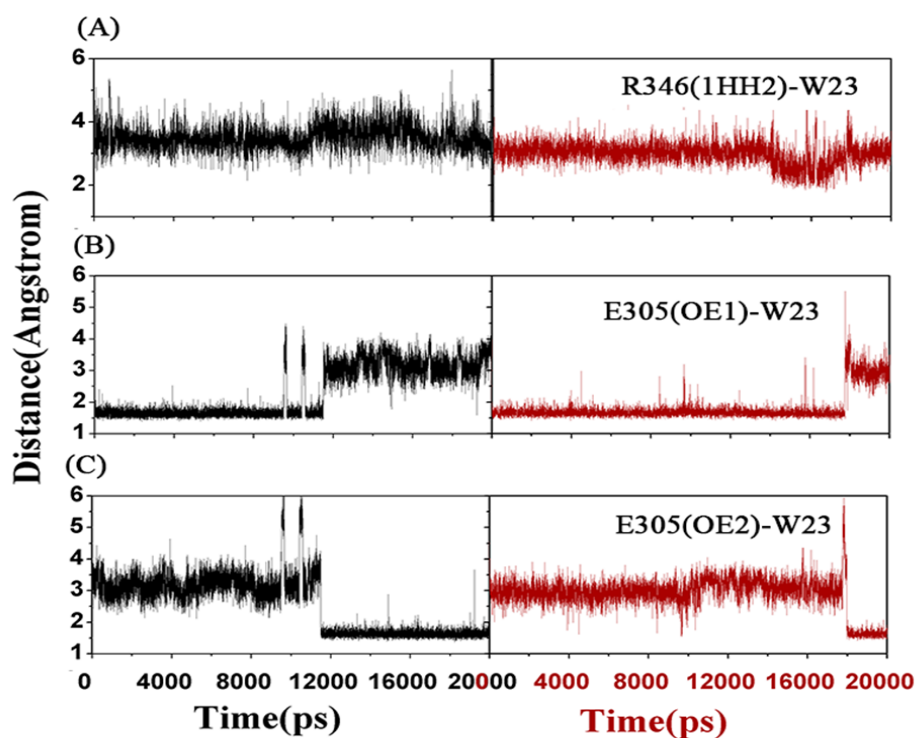


图4.6 动力学模拟中R346和E305与配体W23H形成的氢键随时间变化曲线(左边来自轨迹A3, 右边红色来自轨迹A4); (A) R346与W23形成的氢键; (B) E305中的羧基中的氧原子(OE1)与W23形成的氢键; (C) E305中的羧基中的氧原子(OE2)与W23形成的氢键。

4.3.2 MM/GBSA 计算雌激素受体与W23和W23H的结合自由能

通过MM/GBSA计算我们发现当使用AMBER03力场时, W23和W23H这两个配体的结合自由能非常接近, 只有-0.13kcal/mol之差, 而使用了QM/MM电荷后, 其结合自由能之差有-3.49kcal/mol(见表4.2)。后者与其生物活性数据($\Delta\Delta G \sim -2.09$ kcal/mol)比较吻合⁴²。

表 4.2 MM/GBSA 计算得到的 ER 与配体 (W23 和 W23H) 的结合自由能
(单位为 kcal/mol)

轨迹	ΔE_{ele}	ΔE_{vdw}	ΔE_{gas}	ΔG_{gb}	$T\Delta S$	ΔG	$\Delta\Delta G$
A1(W23)	-25.10	-32.03	-57.13	-26.38	-16.67	-9.70	-0.13
A3(W23H)	-30.53	-29.47	-60.00	37.01	-17.51	-9.57	
A2(W23)	-24.84	-32.65	-57.49	-31.5	-18.11	-13.39	-3.49
A4(W23H)	-26.85	-30.37	-57.23	-26.78	-16.88	-9.90	

进一步分析结合自由能的各项组成部分, 我们发现 AMBER03 不能有效区分这 2 个配体的相互作用, 尤其在于 ΔE_{ele} 和 ΔG_{gb} 这两项, 由于这 2 项能量都直接与电荷相关的, 由此可见静电极化作用在其中起着重要作用, 因此考虑了极化效应的 QM/MM 电荷的模拟计算具有更好的结果。

另外通过比较 ΔE_{ele} 和 ΔE_{vdw} 这 2 项也可以发现在这个体系中, 静电作用没有范德华作用起的作用大, 如对化合物 W23 而言, 静电作用为 -24.84 kcal/mol, 而范德华力作用则为 -32.65 kcal/mol, 这点和晶体中观察到配体周围都是疏水性残基, 主要通过疏水作用是一致的。如图 4.2B 所示, ER 蛋白的结合口袋是个非常疏水的口袋, W23 周围除了 R346 和 E305 外, 其余全是疏水性残基。为了定性的考察此蛋白和配体的疏水作用, 我们利用软件 ligplot4.4⁵² 对体系进行疏水分析。如图 4.7 所示, 其活性口袋附近的非极性残基如 I298, I301, L339, M336, M340 和 F356 等形成的疏水作用在动力学模拟过程中都相当稳定。尤其是其中的 F356, 可以与配体形成典型的 CH- π 相互作用, 因此在蛋白质-配体相互作用中起着重要作用。

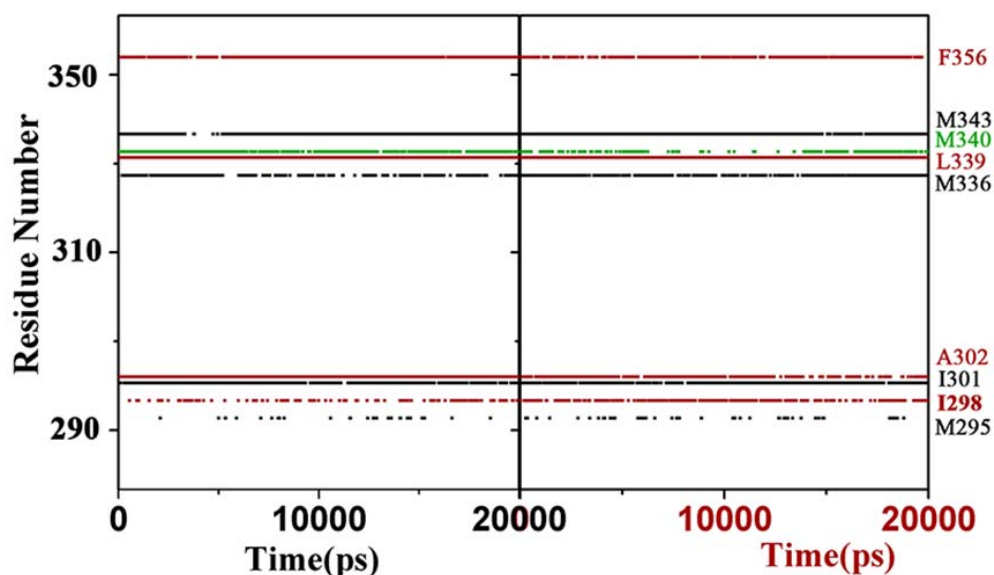


图 4.7 动力学模拟中蛋白质疏水残基与配体 W23 和 W23H 形成的疏水作用随时间变化曲线（左半边来自轨迹 A2，右半边来自轨迹 A4）。

4.3.3 BTN 与亲和素及突变体 F130L 的动力学分析

BTN与野生型和F130L的晶体结构显示BTN可以与其口袋中的多个残基形成氢键相互作用（见图4.3A和B），在其突变的位置附近可以与配体形成氢键的残基有Y43，N23和D128。

如同ER体系一样，我们首先分析了蛋白质各原子的QM/MM电荷和AMBER03电荷的区别。如图4.8所示，计算所得的电荷和AMBER03电荷也有着很大的区别。通过比较突变体和野生型中的相同原子的电荷，我们发现其电荷发生主要变化的地方主要是在发生突变的残基周围（如原子序号500，1200及1700附近）。这一发现进一步说明体系内即使只有很小的地方发生变化，也可造成体系的静电势有较大的变化，而通过QM/MM方法可以比较准确地反映出这种静电势的变化。

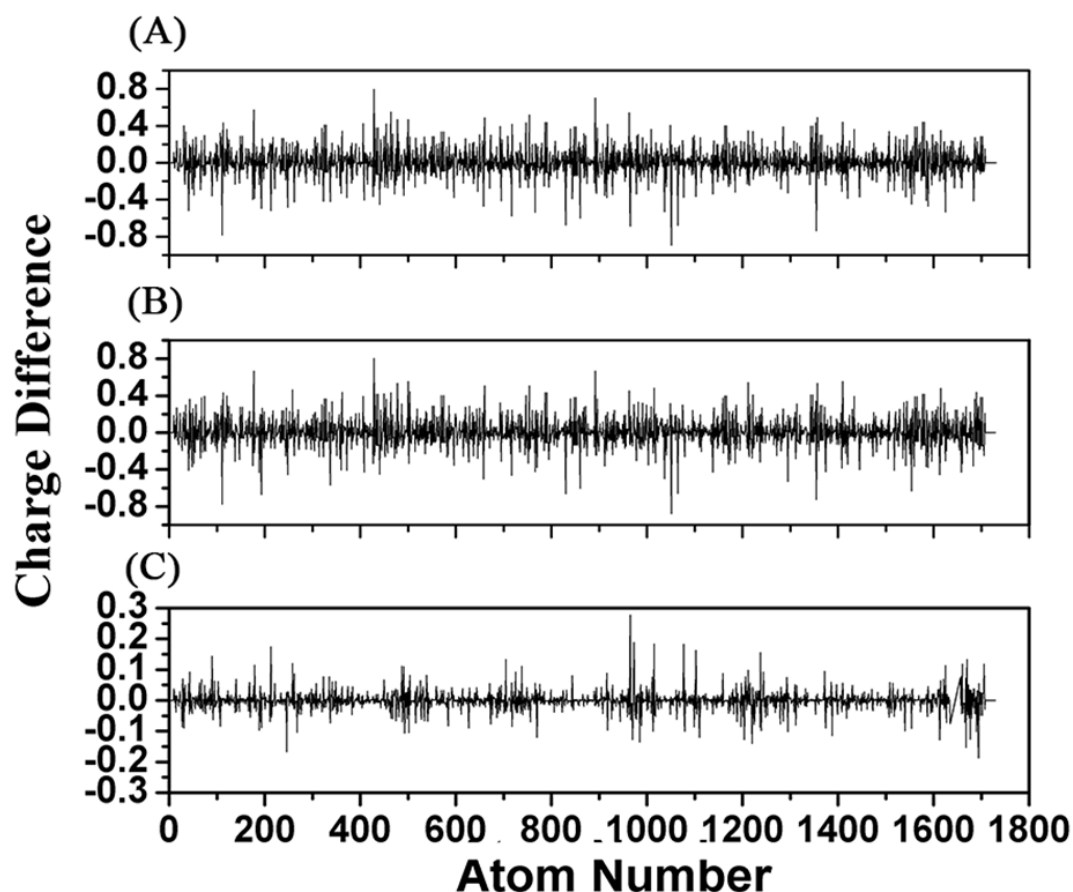


图 4.8 (A) 和 (B) 分别是量化计算得到的具有野生型和 F130L 的亲素素的蛋白质各原子电荷与 AMBER03 力场中电荷之差; (C) 是野生型和 F130L 之间各原子电荷之差 (其中去除了突变体残基的原子电荷)。

为了评估这个体系的动力学稳定性, 同样计算了其主链重原子的 RMSD 随模拟时间的变化曲线。如图 4.9 所示, 主链的 RMSD 都在 $1.5 \text{ \AA}$ 以内, 而且在 10ns 后体系基本处于平衡, 其后 10ns 的结构 RMSD 变化在 $0.5 \text{ \AA}$ 以内, 因此我们仍然取后 10ns 的结构用于 MM/GBSA 的计算。

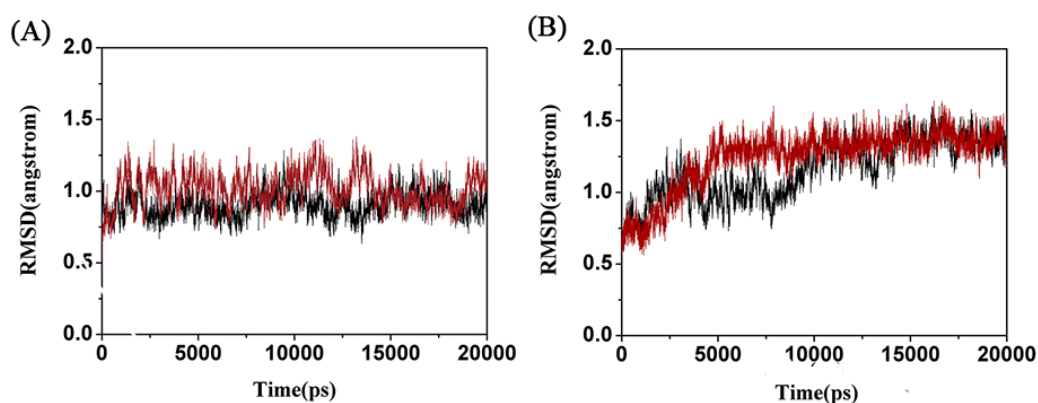


图 4.9 体系主链重原子的 RMSD 随时间变化的情况。(A) 中黑线和红线分别来自轨迹 B1 和 B2; (B) 中黑线和红线则分别来自体系 B3 和 B4。

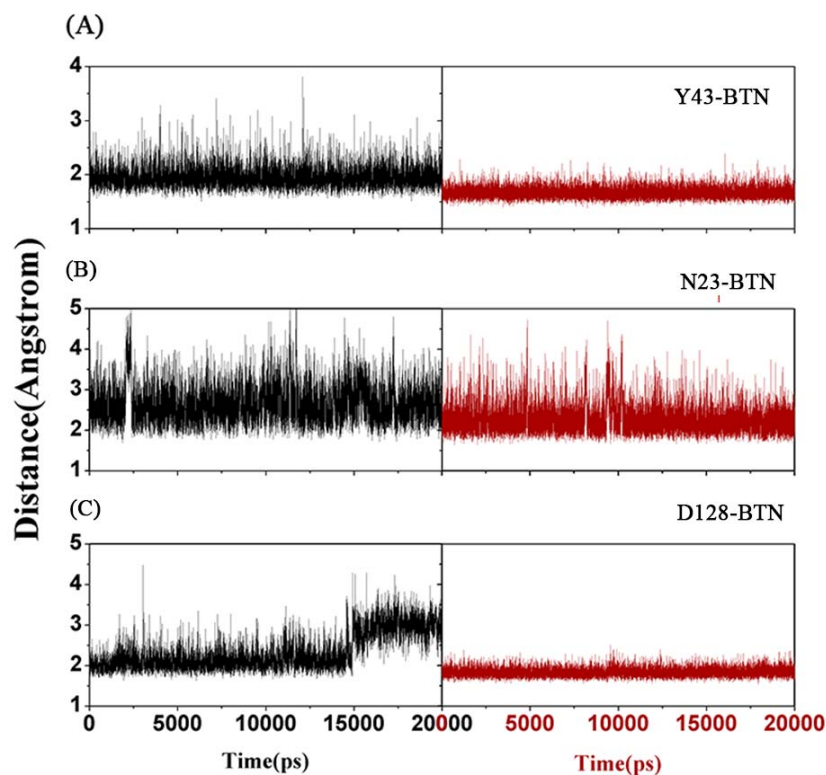


图 4.10 动力学模拟中野生型的亲和素残基 Y43, N23 和 D128 与 BTN 形成的氢键随时间变化曲线 (左半边黑线来自轨迹 B2, 右半边红线来自轨迹 B1)。

如同 ER 体系一样，我们分别比较了使用和未使用 QM/MM 电荷的氢键随时间变化曲线（见图 4.10 和 4.11），结果同样发现，使用 QM/MM 电荷后，体系氢键的稳定性加强。而 BTN 主要是通过这此氢键相互作用来维持其相互作用的。通过比较 BTN 与野生型和 F130L 的体系（图 4.10 和 4.11），我们发现 Y42 与 BTN 形成的氢键在野生型中非常稳定，而在突变体 F130L 中则不稳定，这一氢键的丢失可能导致了 F130L 活性的下降。

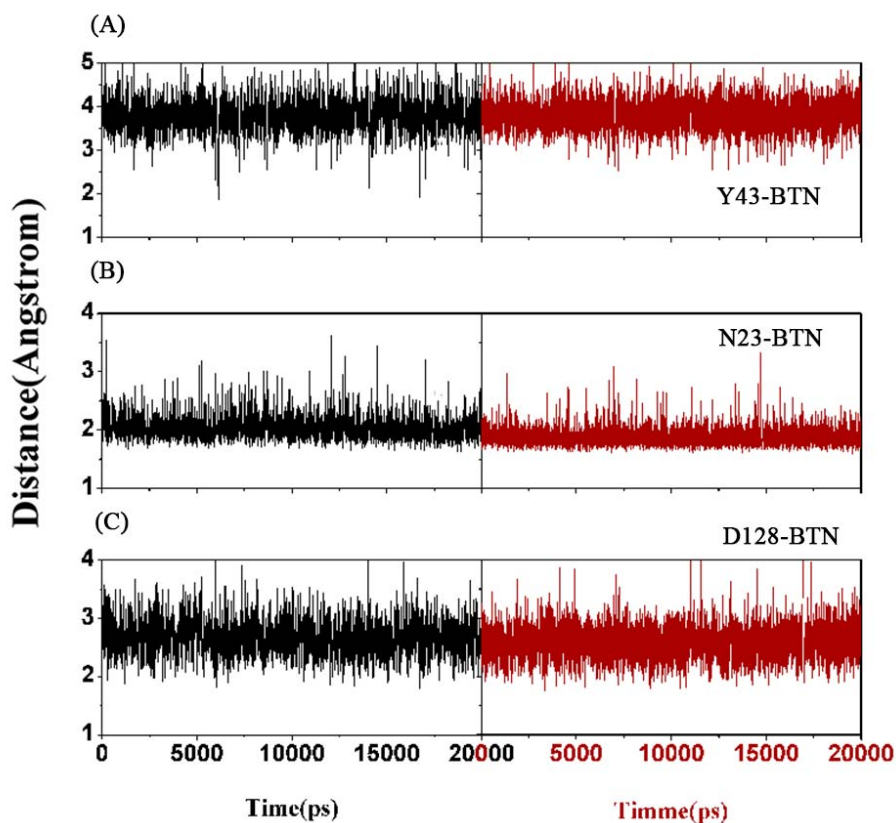


图 4.11 动力学模拟中 F130L 亲和素的残基 Y43，N23 和 D128 与 BTN 形成的氢键随时间变化曲线（左半边黑线来自轨迹 B4，右半边红线来自轨迹 B3）。

4.3.4 MM/GBSA 计算 BTN 与亲和素及突变体 F130L 的结合自由能

我们发现当使用 AMBER03 力场的时候, 这 2 个体系的结合自由能非常接近, 只有 -0.34 kcal/mol 之差, 而使用了 QM/MM 电荷后, 其结合自由能之差为 -5.34 kcal/mol, 与其生物活性数据 ($\Delta\Delta G = -4.2$ kcal/mol) 是非常符合的⁴⁴。

进一步分析结合自由能的各项 (如表 4.3 所示), 我们发现 AMBER03 不能有效区分这 2 个体系的相互作用主要在于 ΔE_{ele} 。在使用了 QM/MM 电荷后, BTN 和野生型蛋白质的静电作用 (-122.04 kcal/mol) 明显大于其与 F130L 突变体 (-99.07 kcal/mol), 但是使用 AMBER03 力场其数值则分别为 -75.62 和 -79.74 kcal/mol, 野生型的静电作用甚至弱于突变体。可见静电作用在这个配体中起着至关重要的作用, 这个和 BTN 与体系内多个残基可以形成较强的氢键相一致, 因为氢键主要是静电作用性质的, 所以在这个体系中的静电作用起着至关重要的作用。

表 4.3 MM/GBSA 计算得到的 BTN 与亲和素及突变体 F130L 的结合自由能
(单位为 kcal/mol)

轨迹	ΔE_{ele}	ΔE_{vdw}	ΔE_{gas}	ΔG_{gb}	$T\Delta S$	ΔG	$\Delta\Delta G$
B1(WT)	-75.62	-32.06	-107.67	-28.78	-21.14	-7.64	-0.34
B3(F130L)	-79.74	-32.07	-111.81	-27.8	-20.5	-7.30	
B2(WT)	-122.04	-25.63	-147.67	-45.89	-20.82	-25.07	-5.34
B4(F130L)	-99.07	-27.92	-126.99	-37.48	-17.75	-19.73	

4.4 小结

本章中所采用的基于分而治之的 QM/MM 电荷的方法可以得到考虑蛋白质和配体极化作用的电荷, 在对选取的 2 个体系的研究中发现, 使用这种电荷可

以比较精确地区分出相似配体之间或不同突变体之间的活性差异。另外研究也发现了在动力学模拟中考虑这些极化效应的必要性。研究表明 MM/GBSA 方法虽然是一种基于参数化的结合自由能的计算方法，然而将其用于合理的体系中（如用来比较同系物或不同蛋白突变体中的作用）是具有潜在使用价值的。

本章使用的 QM/MM 电荷是对单个结构进行的量化计算，而实际上蛋白质的静电势还会随着构象的变化而发生改变，所以为了更加准确地描述蛋白质的极化效应，一个改进的方法是每隔一段时间重新拟合电荷。另外目前对单个残基使用 RESP 方法来拟合电荷，因此还不能考虑残基之间的电荷转移效应，而这个在很多具有较强的极化作用的体系中经常发生。因此，本研究工作的下一个发展方向在于如何考虑残基之间的电荷转移效应。

参考文献

1. Bakan, A. and I. Bahar, *The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding*. Proc. Natl. Acad. Sci. USA, 2009. **106**(34): p. 14349-14354.
2. Gilson, M.K. and H.X. Zhou, *Calculation of protein-ligand binding affinities*. Annu. Rev. Biophys. Biomol. Struct., 2007. **36**: p. 21-42.
3. Cram, D.J., *The design of molecular hosts, guests, and their complexes*. J. Incl. Phen., 1988. **6**(4): p. 397-413.
4. Koshland, D.E., *The key-lock theory and the induced fit theory*. Angew. Chem. Int. Ed., 1995. **33**(23 - 24): p. 2375-2378.
5. Leavitt, S. and E. Freire, *Direct measurement of protein binding energetics by isothermal titration calorimetry*. Curr. Opin. Struct. Biol., 2001. **11**(5): p. 560-566.
6. Jason - Moller, L., M. Murphy, and J. Bruno, *Overview of Biacore systems and their applications*. Current Protocols in Protein Science, 2006: Chapter 19, Unit 19 13.
7. Jorgensen, W.L., *Free-Energy Calculations - a Breakthrough for Modeling Organic-Chemistry in Solution*. Accounts Chem. Res., 1989. **22**(5): p. 184-189.
8. Wong, C.F. and J.A. Mccammon, *Dynamics and Design of Enzymes and Inhibitors*. J. Am. Chem. Soc., 1986. **108**(13): p. 3830-3832.
9. Kollman, P.A., et al., *Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models*. Accounts Chem. Res., 2000. **33**(12): p. 889-897.
10. Srinivasan, J., et al., *Continuum solvent studies of the stability of DNA, RNA,*

- and phosphoramidate - DNA helices*. J. Am. Chem. Soc., 1998. **120**(37): p. 9401-9409.
11. Gago, F., *Molecular simulations of drug-receptor complexes in anticancer research*. Future Med. Chem., 2012. **4**(15): p. 1961-1970.
 12. Lin, J.-H., *Accommodating protein flexibility for structure-based drug design*. Curr. Top. Med. Chem., 2011. **11**(2): p. 171-178.
 13. Kitchen, D.B., et al., *Docking and scoring in virtual screening for drug discovery: methods and applications*. Nat. Rev. Drug Discov., 2004. **3**(11): p. 935-949.
 14. Mavromoustakos, T., et al., *Strategies in the rational drug design*. Curr. Med. Chem., 2011. **18**(17): p. 2517-2530.
 15. Rastelli, G., et al., *Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA*. J. Comput. Chem., 2010. **31**(4): p. 797-810.
 16. Steinbrecher, T. and A. Labahn, *Towards accurate free energy calculations in ligand protein-binding studies*. Curr. Med. Chem., 2010. **17**(8): p. 767-785.
 17. Chen, S.L., D.X. Zhao, and Z.Z. Yang, *An estimation method of binding free energy in terms of ABEEM/MM and continuum electrostatics fused into LIE method*. J. Comput. Chem., 2010. **32**(2): p. 338-348.
 18. Tong, Y., et al., *Electrostatic Polarization Makes a Substantial Contribution to the Free Energy of Avidin-Biotin Binding*. J. Am. Chem. Soc., 2010. **132**(14): p. 5137-5142.
 19. Mei, Y., et al., *Electrostatic polarization is critical for the strong binding in streptavidin-biotin system*. J. Comput. Chem., 2012. **33**(15): p. 1374-1382.
 20. Khoruzhii, O., et al., *Application of a polarizable force field to calculations of relative protein-ligand binding affinities*. Proc. Natl. Acad. Sci. USA, 2008.

- 105**(30): p. 10378-10383.
21. Jiao, D., et al., *Trypsin-ligand binding free energies from explicit and implicit solvent simulations with polarizable potential*. J. Comput. Chem., 2009. **30**(11): p. 1701-11.
 22. Jiao, D., et al., *Calculation of protein–ligand binding free energy by using a polarizable potential*. Proc. Natl. Acad. Sci. USA, 2008. **105**(17): p. 6290-6295.
 23. Zhu, X., P.E.M. Lopes, and A.D. MacKerell, *Recent developments and applications of the CHARMM force fields*. Wiley Interdisciplinary Reviews: Computational Molecular Science, 2012. **2**(1): p. 167-185.
 24. Cornell, W.D., et al., *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules*. J. Am. Chem. Soc., 1995. **117**(19): p. 5179-5197.
 25. Jorgensen, W.L., D.S. Maxwell, and J. Tirado-Rives, *Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids*. J. Am. Chem. Soc., 1996. **118**(45): p. 11225-11236.
 26. Oostenbrink, C., et al., *A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6*. J. Comput. Chem., 2004. **25**(13): p. 1656-1676.
 27. Tong, Y., et al., *Electrostatic Polarization Makes a Substantial Contribution to the Free Energy of Avidin– Biotin Binding*. J. Am. Chem. Soc., 2010. **132**(14): p. 5137-5142.
 28. Raha, K., et al., *The role of quantum mechanics in structure-based drug design*. Drug Discov Today, 2007. **12**(17): p. 725-731.

-
29. Kamerlin, S.C.L., M. Haranczyk, and A. Warshel, *Progress in Ab Initio QM/MM Free-Energy Simulations of Electrostatic Energies in Proteins: Accelerated QM/MM Studies of pK(a), Redox Reactions and Solvation Free Energies*. J. Phys. Chem. B, 2009. **113**(5): p. 1253-1272.
30. Senn, H.M. and W. Thiel, *QM/MM Methods for Biomolecular Systems*. Angew. Chem. Int. Ed., 2009. **48**(7): p. 1198-1229.
31. Gleeson, M.P. and D. Gleeson, *QM/MM calculations in drug discovery: a useful method for studying binding phenomena?* J. Chem. Inform. Model, 2009. **49**(3): p. 670-677.
32. Lodola, A. and M. De Vivo, *The Increasing Role of QM/MM in Drug Discovery*. Adv. Protein Chem. Struct. Biol., 2012. **87**: p. 337-362.
33. Mei, Y., et al., *Electrostatic polarization is critical for the strong binding in streptavidin - biotin system*. J. Comput. Chem., 2012. **33**(15): p. 1374-1382.
34. Li, Y.L., et al., *Structure and Dynamics of a Dizinc Metalloprotein: Effect of Charge Transfer and Polarization*. J. Phys. Chem. B, 2011. **115**(33): p. 10154-10162.
35. Ji, C., Y. Mei, and J.Z. Zhang, *Developing Polarized Protein-Specific Charges for Protein Dynamics: MD Free Energy Calculation of pK_a Shifts for Asp²⁶/Asp²⁰ in Thioredoxin*. Biophys. J., 2008. **95**(3): p. 1080-1088.
36. Rocchia, W., et al., *Rapid grid - based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects*. J. Comput. Chem., 2002. **23**(1): p. 128-137.
37. Dapprich, S., et al., *A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric*
-

- field derivatives*. J. Mol. Struct. (THEOCHEM), 1999. **461**: p. 1-21.
38. Duan, Y., et al., *A point - charge force field for molecular mechanics simulations of proteins based on condensed -phase quantum mechanical calculations*. J. Comput. Chem., 2003. **24**(16): p. 1999-2012.
39. Bayly, C.I., et al., *A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model*. J. Phys. Chem., 1993. **97**(40): p. 10269-10280.
40. Cornell, W.D., et al., *Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation*. J. Am. Chem. Soc., 1993. **115**(21): p. 9620-9631.
41. Hagmann, W.K., *The many roles for fluorine in medicinal chemistry*. J. Med. Chem., 2008. **51**(15): p. 4359.
42. Möcklinghoff, S., et al., *Design and evaluation of fragment-like estrogen receptor tetrahydroisoquinoline ligands from a scaffold-detection approach*. J. Med. Chem., 2011. **54**(7): p. 2005-2011.
43. *Accelrys Discovery Studio*, Accelrys Software Inc.: San Diego, CA 92121.
44. Baugh, L., et al., *A Distal Point Mutation in the Streptavidin– Biotin Complex Preserves Structure but Diminishes Binding Affinity: Experimental Evidence of Electronic Polarization Effects?* Biochemistry, 2010. **49**(22): p. 4568-4570.
45. Hyre, D.E., et al., *Cooperative hydrogen bond interactions in the streptavidin–biotin system*. Protein Sci., 2006. **15**(3): p. 459-467.
46. Wang, J., et al., *Development and testing of a general amber force field*. J. Comput. Chem., 2004. **25**(9): p. 1157-1174.
47. Gordon, J.C., et al., *H⁺⁺: a server for estimating pK_as and adding missing*

- hydrogens to macromolecules*. Nucleic Acids Res., 2005. **33**(suppl 2): p. W368-W371.
48. Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems*. J. Chem. Phys., 1993. **98**: p. 10089.
49. Ryckaert, J.-P., G. Ciccotti, and H.J. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n -alkanes*. J. Chem. Phys., 1977. **23**(3): p. 327-341.
50. Onufriev, A., D. Bashford, and D.A. Case, *Exploring protein native states and large - scale conformational changes with a modified generalized born model*. Proteins., 2004. **55**(2): p. 383-394.
51. Weiser, J., P.S. Shenkin, and W.C. Still, *Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO)*. J. Comput. Chem., 1999. **20**(2): p. 217-230.
52. Wallace, A.C., R.A. Laskowski, and J.M. Thornton, *Ligplot - a Program to Generate Schematic Diagrams of Protein Ligand Interactions*. Protein Eng., 1995. **8**(2): p. 127-134.

第五章 铵离子通道AmtB转运氨和甲胺分子机制的模拟研究

5.1 引言

细菌和植物的生长离不开铵盐，在体内它们会利用氨合成必需的氨基酸。当胞外氨的浓度比较低，主动扩散不足以满足细菌和植物生长时，就需要通过专门的铵离子通道来传递氨¹。而在人和动物体中，当铵盐浓度过高时，会引起细胞毒性，也需要通过相应的通道蛋白将体内多余的氨排出体外，保持细胞内的酸碱平衡，因此，氨的跨膜转运在生命活动中起着关键作用^{2,3}。

植物和真菌中负责氨转运的蛋白是铵离子通道 (Ammonium transporters, AmtS)，它在细菌和人的同源蛋白分别是甲胺通道(Methylamine permeases, MEPs)和Rh蛋白⁴⁻⁷。自从铵离子通道被发现以来，通过各种生化实验的研究，对这类通道有了比较深的理解，最近有很多关于铵离子通道的功能及转导机制的综述报道⁸⁻¹²，其中AmtB是氨通道蛋白家族中研究得较多的成员。

研究发现AmtB传递氨是浓度依赖的，当胞外铵离子浓度较高时，由于氨分子扩散就可以满足细胞的正常生长，此时细胞内的P_{II}蛋白家族的成员GlnK就会和AmtB相结合，形成复合物，从而抑制其转运氨的功能；而当胞外铵离子浓度较低时，GlnK 就不与AmtB 形成复合物，AmtB 就发挥转运铵的功能^{3,13,14}。

随着2004年AmtB的晶体结构的解析^{15,16}，近年来其家族的各种同源蛋白也相继被解出，如来自Archaeoglobus fulgidus的同源蛋白Amt1¹⁷和来自人的同源蛋白RhCG⁷等。但是有关铵离子通道转运氨的机制还是具有较大争议。目前关于AmtB传导氨的机制至少有四种，即电中性的氨^{15,16,18-20}，NH₃/H⁺协同转运^{10,20-23}，NH₄⁺/H⁺逆向转运^{24,25}和 NH₄⁺ 转运^{26,27}。而晶体结构显示此类蛋白是个狭窄的疏水性的通道，因此适合中性分子通过，最近通过理论计算的结果也支持铵离子首先被通道聚集，然后以中性分子通过的机制^{22,28-39}。

在实验上被广泛用来研究AmtB功能的有pH染料电极和同位素标记实验^{20, 26}。由于pH染料电极的分辨率及实验的难度，利用同位素标记的办法来研究AmtB机制则是更为广泛使用的方法。不过由于氮同位素的半衰期比较短，在实验上使用更多的是氨的类似物甲胺分子^{20, 40-42}。随之而来的问题就是在研究AmtB及其突变体的功能时，甲胺分子是不是个有效的氨分子的模拟物。通道和配体在原子水平上的作用又是怎么样的呢？这些问题在生物实验上是没法解释的。而且目前也缺乏系统的理论研究来比较AmtB在传导氨和甲胺分子的差异，如Hu等³⁸利用QM/MM方法研究了不同配体在通道的胞外结合位点Am1和AmtB的结合强度，结果显示氨和甲胺的结合强度在同一数量级，显示铵和甲胺都可以被AmtB聚集。然而AmtB通道传递氨和甲胺分子的机制并不清楚。为了这一问题，我们首先使用分子动力学方法（MD）模拟了14个体系，结果显示甲胺和氨都能以相似的机制通过AmtB的野生型和突变体W212F, S263C和S263D，而突变体W212A 和S263A却不能传递氨和甲胺分子。这些结果显示了配体和残基S263形成的氢键及残基W212形成的阳离子- π 在传导中起的重要作用。

另外，在AmtB的疏水通道中间并行排列着两个组氨酸（H168和H318），这两个组氨酸在铵离子通道中非常保守，英国的Mike课题组构建了14个组氨酸的突变体，利用甲胺作为探针，通过突变实验和同位素实验，发现除了突变体H168E具有传导甲胺的活性外，其他突变体都丧失了传递甲胺的能力，由此他们推测这两个残基在传递氨时也起着一样的功能⁴²。它们还解析了6个组氨酸的突变体（H168A, H168E, H168F, H318A, H318F and H168A/H318A）的结构⁴³。如图5.1所示，通过结构叠合，我们发现这些突变体和野生型的结构（1U7G¹⁵）的Ca的RMSD值都小于0.4 Å，而且整体结构和突变残基周围的残基并没有大的变化。这些现象引出了如下疑问：这两个组氨酸在原子水平上到底起什么作用？为什么AmtB的组氨酸突变体不能传导甲胺，那么是不是这些突变体也不能传递氨呢？为了回答这些问题，我们和Mike课题组合作，首先分析了在以铵盐为唯

一氮源的导入AmtB基因的酵母生长情况。实验结果表明，与甲胺的传导不一样，这些单组氨酸的突变体（H168A和H318A）都保持着野生型的活性，而双突变体(H168A/H318A)则失去了活性。为了解释这些实验现象，我们通过分子动力学模拟和分子平均力势（potential of mean force, PMF）计算，从而可以在原子水平解释这些组氨酸的作用。

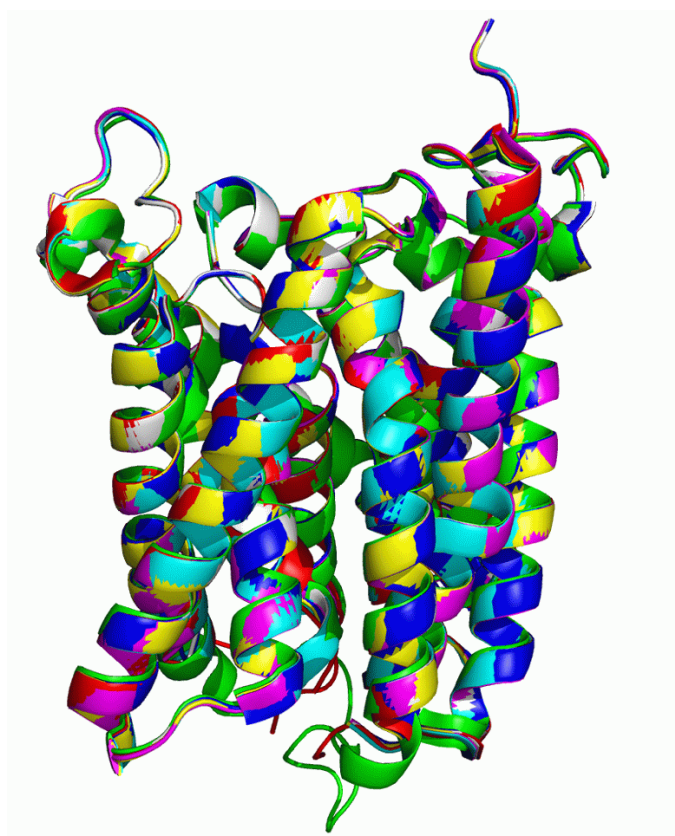


图5.1 组氨酸突变体与野生型AmtB晶体结构的叠合：H168A（2NPC）、H168E（2NOW）、H168F（2NPD）、H318A（2NPE）、H318F（2NPG）、H168A/H318A（2NPK）和野生型AmtB（1U7G）分别用黄色、紫红色、青色、蓝色、白色、红色和绿色显示。而它们的C α 的RMSD分别是0.364、0.372、0.323、0.367、0.306和0.328 Å。

5.2 模型和方法

5.2.1 利用 Gromacs3.3.1 模拟 AmtB 及突变体 W212 和 S263 的方法

Khademi组测定的分辨率为 1.3 Å 的X衍射晶体结构 (1U7G)¹⁵作为模拟的起始结构, 晶体结构中的三个突变(F68S, S126P和K255L)突变成野生型。整个通道的结构如图 5.2 所示, 其包含 11 个跨膜螺旋, 从Tieleman研究小组的网站 (<http://people.ucalgary.ca/~tieleman/download.html>) 下载得到 128 个DPPC磷脂分子的初始坐标。将其拼接得到包含 512 个磷脂分子的磷脂双层, 然后将AmtB与磷脂分子叠合, 删除与AmtB通道重合的磷脂分子, 最后得到 288 个围绕在AmtB通道周围的DPPC磷脂分子。最后定义一个合适的盒子包含全部DPPC和通道蛋白, 并在磷脂层两侧填充 14375 个SPC水分子, 加入Na⁺和Cl⁻中和体系多余电荷, 并进行优化。所有的动力学模拟采用的程序是GROMAS3.2.1^{44, 45}, 磷脂分子的力场参数来源于之前的研究⁴⁶⁻⁴⁸, 蛋白和小分子使用的是GROMOS87 力场, 而小分子氨和甲胺的电荷是利用高斯 03 计算得到的ChelpG电荷⁴⁹, 最终得到的AmtB离子通道的模拟体系模型如图 5.2 A所示。

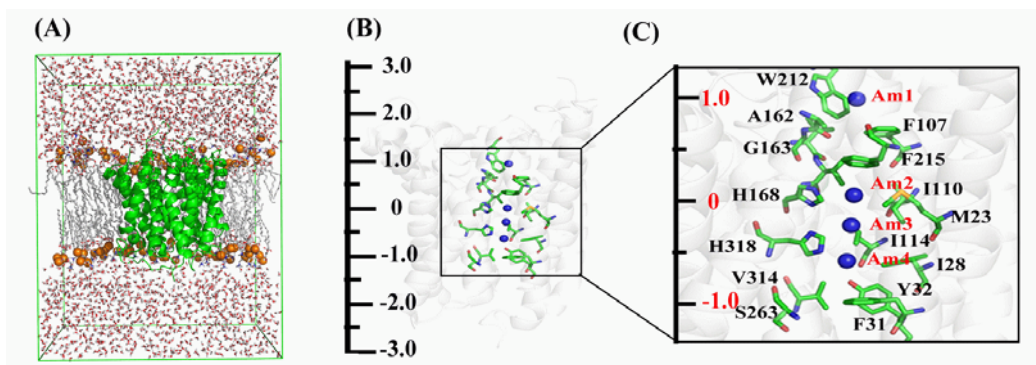


图 5.2 (A)模拟体系的侧视图, 晶体结构为 1U7G。蛋白质以绿色卡通结构显示, 磷原子以橙色球显示, 磷脂中其他原子以白色线条显示, 水分子以红色和白色棍显示。为了显示得更清晰, 体系的前半部分未显示; (B) 模拟所使用的反应坐标; (C) 晶体结构中的氨的 4 个位点及其周围重要残基。

对整个体系做完优化后, 一共运行 1ns的预处理, 逐步放开对磷脂分子,

AmtB的氢分子、侧链，主链的约束力。接着进行最长为 20ns的常规动力学模拟。模拟采用的是等温等压（NPT）系综和周期边界条件，使用Berendsen热浴⁵⁰方法将体系温度保持在 323 K，耦合常数设置为 0.1 ps；模拟体系的压力是 1 个大气压，耦合常数设置为 1 ps。在升温及动力学平衡过程中，静电相互作用计算采用PME算法⁵¹，非键相互作用的截断半径为 9 Å，所有模拟的时间积分步长都是 2 fs，LINCS算法⁵²被用来限制体系中含氢的键长的振动。氢键和疏水作用通过软件LIGPLOT 4.4.2 来计算获得⁵³。

为了确定通道中几个重要残基的作用，如表 5.1 所示，我们一共设计了 14 个模拟体系。其中体系 A1, A2 和 A3 用来研究比较野生型 AmtB 传导氨和甲胺分子的异同点。而体系 B1-B4 则用来研究 212 残基上的芳香环在传导配体中的重要作用，分别模拟了突变体 W212F 和 W212A。而体系 C1 到 C6 用来研究 263 号残基中氢键在传导过程中的作用。

表5.1 AmtB动力学模拟体系的细节

轨迹	蛋白	Am2	Am3	Am4	时间(ns)
A1	wild type	CH ₃ NH ₂	CH ₃ NH ₂	CH ₃ NH ₂	4.5
A2	wild type	CH ₃ NH ₂	-	-	2.5
A3	wild type	NH ₃	-	-	9
A'	wild type	NH ₃	NH ₃	NH ₃	8
B1	W212F	CH ₃ NH ₂	-	-	4
B2	W212F	NH ₃	-	-	3
B3	W212A	CH ₃ NH ₂	-	-	20
B4	W212A	NH ₃	-	-	20
C1	S263A	CH ₃ NH ₂	-	-	20
C2	S263A	NH ₃	NH ₃	NH ₃	20
C3	S263D	NH ₃	-	-	14

C4	S263D	CH ₃ NH ₂	-	-	15
C5	S263C	NH ₃	-	-	14
C6	S263C	CH ₃ NH ₂	-	-	16

轨迹A'和D2取自文献模拟结果³²

5.2.2 利用 Gromacs4.5.3 模拟 AmtB 组氨酸突变体的方法

模拟采用的 AmtB 结构的处理与上面一样，这里使用的力场是 CHARMM27⁵⁴，从 charmm-gui 网站(<http://www.charmm-gui.org>)下载得到 288 个 DPPC 磷脂双层的初始坐标。将 AmtB 与下载得到的 288 个 DPPC 磷脂叠合，删除与 AmtB 通道重合的磷脂分子，最后得到 181 个 DPPC 磷脂分子围绕在 AmtB 通道周围。然后，定义一个合适的盒子包含 DPPC 和蛋白原子，并在磷脂层两侧填充 13137 个 TIP3P 水分子，然后加入 Na⁺ 和 Cl⁻ 中和体系多余电荷，并进行优化。所有的动力学模拟都采用 GROMACS4.5.3 程序⁵⁵，磷脂分子蛋白和小分子使用的是 CHARMM27 力场⁵⁴，其中氨分子和甲胺分子的力场参数通过网站 SwissParam⁵⁶ 获得。

对整个体系做完优化后，首先固定蛋白，磷脂分子和配体进行 250ps 的动力学模拟，然后再将蛋白侧链，磷原子和配体固定进行 250ps 动力学模拟，最后只固定蛋白的 C α 原子和配体进行 10ns 的限制动力学模拟。最后进行最长 100ns 的常规动力学模拟。模拟采用的是 NPT 系综和周期边界条件。模拟使用 v-rescale 热浴方法⁵⁷，将体系控温在 323K，温度耦合常数为 0.1 ps；压力为 1 个大气压，压力耦合常数 1 ps。在升温及动力学平衡过程中，静电相互作用计算采用 PME 算法⁵¹，非键相互作用的截止距离设置为 12 Å；时间积分步长为 2fs，同样利用 LINCS 算法⁵² 限制体系中含有氢的键长的振动。

为了确定通道中组氨酸的作用，我们一共设计了 12 个模拟体系，见表 5.2 所示，其中体系 D1 和 D2 用来研究在 CHARMM27 力场下野生型 AmtB 传导氨和甲胺分子的异同点。为了研究 H168 和 H318 在配体传导中的作用，特设计体

系 E1, E2, F1, F2 及 G1-G6, 来分别模拟组氨酸的突变体 (H168A, H318A 和 H168A/H318A)。

表 5.2 动力学模拟 AmtB 及组氨酸突变体信息

	蛋白	配体	$t_{\text{simulation}}$ (ns)	t_{exiting} (ns)	模拟结果	实验结果
D1	wild-type	CH ₃ NH ₂	20	13.950	通过	转运
D2	wild-type	NH ₃	20	1.792	通过	转运
E1	H168A	CH ₃ NH ₂	100	-	不能通过	不能转运
E2	H168A	NH ₃	20	4.415	通过	转运
F1	H318A	CH ₃ NH ₂	100	-	不能通过	不能转运
F2	H318A	NH ₃	20	8.860	通过	转运
G1	H168A/H318A	CH ₃ NH ₂	20	5.190	进入胞外	不能转运
G2	H168A/H318A	NH ₃	40	37.460	进入胞内	不能转运
G3	H168A/H318A	CH ₃ NH ₂	100	3.700	进入胞外	不能转运
G4	H168A/H318A	NH ₃	100	78.420	进入胞内	不能转运
G5	H168A/H318A	-	20			
G6	H168A/H318A	-	20			

5.2.3 伞取样模拟

进行伞取样模拟的起始构象来源于10ns的平衡动力学。在伞取样模拟中, 为了防止体系Z轴改变造成的统计误差, 在伞取样模拟中保持盒子Z轴长度不变。

AmtB通道每隔0.15Å分成一个坐标点，然后这些点分别作为配体的中心，作为模拟的一个窗口。使用Nose'-Hoover热浴⁵⁸将体系温度控制在300K，压力控制方法使用Parrinello-Rahman方法⁵⁹。静电相互作用计算采用PME算法⁵¹，非键相互作用的截止值为10 Å，时间积分步长设置为2 fs；LINCS算法⁵²同样用于限制体系中含有氢的键长的振动。在每一个窗口中使用一个800 kJ/mol/nm的简谐力对配体的Z轴方向运动进行限制。能量优化后，固定蛋白和配体进行100 ps的动力学升温平衡。最后进行500ps的采样。为了消除前面的波动的影响，只取后面350 ps的数据来计算PMF，其中PMF的计算使用的是Kumar等发展的g_wham程序⁶⁰。

5.3 结果与讨论

5.3.1 野生型 AmtB 传导甲胺和氨分子

如表 5.1 所示我们设计了体系 A1 和 A2 来研究甲胺分子的传导机制，这两个体系的唯一区别在于在体系 A1 中有三个分别位于 Am2，Am3 和 Am4（M1，M2 和 M3）的甲胺分子，而在体系 A2 中只有一个甲胺分子（M1'）位于 Am2 上。结果显示这两条轨迹中的甲胺分子在 4.5ns 的模拟时间内都可以从 AmtB 离开。如图 5.3 所示，在体系 A1 中 M3 在 102ps 左右就离开通道，而 M1 在 1882ps 左右离开，最后一个甲胺分子 M2 在大约 2678ps 离开。而在体系 A2 中氨分子则在 1792 ps 左右离开通道。这一结果显示通道中配体分子的数目多少并不影响 AmtB 的传导性。从轨迹 A3 和 A'的模拟结果也进一步验证了这一点。另外我们利用 CHARMM27 力场也模拟了野生型 AmtB 传递甲胺和氨的情况（轨迹 D1 和 D2），也发现甲胺和氨分子都可以在 14 ns 左右的模拟时间内从 AmtB 通过(见图 5.16 A1 和 A2)，这些模拟结果显示野生型的 AmtB 是可以传导甲胺和氨分子的，这和实验结果是一致的。

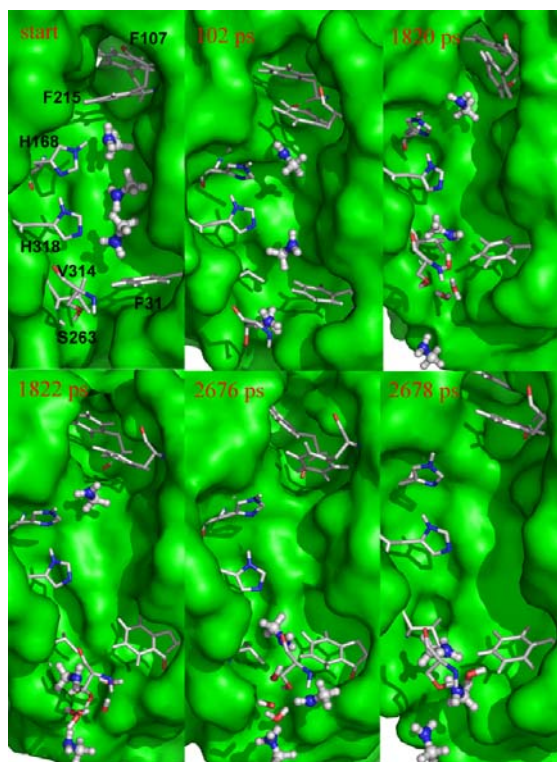


图5.3 甲胺分子在轨迹A1 中离开通道的过程。起始模型以及五个快照结构(102, 1820, 1822, 2676 and 2678 ps).的侧视图。在此侧视图中, 甲胺分子以白色和蓝色棍状结构显示, 残基F31、F107、H168、F215、S263 , V314和H318 以彩色棍状结构显示。在1820ps和2676ps时Ser263与甲胺分子之间形成水桥作用。

为了进一步分析 AmtB 传导甲胺和氨的动力学机制, 如图 5.4 所示, 我们分析了轨迹 A1 中的分子 M1 在通过 AmtB 时所形成的氢键和疏水作用随时间变化曲线。可以看到在模拟过程与配体 M1 形成氢键的主要有 5 个残基, H168, H318, Y32, S263 和 I110 (图 5.4A), 且有 11 个疏水性残基能和 M1 形成疏水作用 (图 5.4B)。甲胺分子 M2 在通道中也是通过与这些残基相互作用, 最终通过通道, 因此这些氢键和疏水作用提供了配体分子通过 AmtB 的主要动力。

轨迹A3和D2也显示了相同的残基H168, H318, Y32和S263参与形成氢键网络, 从而提供了传导氨分子的主要动力。Javelle等⁴²通过将残基 168 和 318 突变

成其他氨基酸，验证了这两个氨基酸在传到甲胺分子的重要作用。另外从轨迹 C1-C6 中可以发现S263 形成氢键的重要性，因为当我们将此残基突变成不能形成氢键的丙氨酸时，配体就不能通过；而当突变成具有形成氢键能力的半胱氨酸和天冬氨酸，在模拟时间内都可以观察到配体通过通道，从而揭示这个氢键在配体通过时的重要作用。

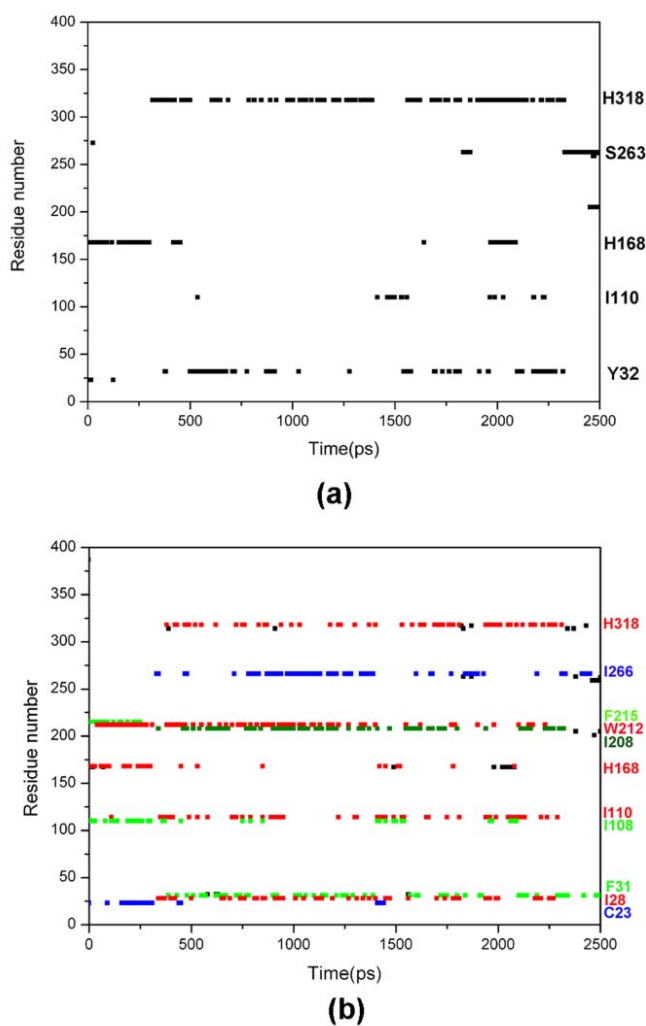


图 5.4 在轨迹 A1 中与甲胺分子(M1)形成氢键和疏水作用的残基随时间的变化。(a)形成氢键的残基随时间变化。(b)形成疏水作用的残基随时间变化。为了显示清晰，使用不同的颜色显示不同残基。

为了定量的衡量配体通过通道时的能量变化情况，我们通过伞取样获得了这两个配体通过野生型 AmtB 的 PMF（如图 5.5 所示）。从 PMF 的结果来看，甲胺和氨传导过程中最大的能垒都分布在 $z \approx 0.6$ nm 左右，对应着两个苯丙氨酸（F107 和 F215）所在的位置（见图 5.2 C）。从位点 Am1 ($z \approx 1.0$ nm) 到 Am2 ($z \approx 0$ nm) 时，甲胺和氨分子的能垒分别是 8.16 和 6.20 kcal/mol（见图 5.5）。这个结果显示氨分子更容易通过通道，这个与实验上发现的氨分子的传导速率大于甲胺的现象是相符的。而配体一旦跨过这一能垒进入通道内部，它们只需要克服较低的能垒，就可从 $z \approx 0$ 运动到 -1 nm 处。对于氨分子来说，从 $z \approx 0$ 到 -1 nm，它的 PMF 值都是正的，暗示着它可以自发的从 Am2 位置运动到出口处，并离开，这个也和我们在很短的时间（1.8 ns）就观察到氨分子离开通道的模拟结果相符。

而从 AmtB 转运甲胺的 PMF 来看，在中间有个 -1.53 kcal/mol 的局部极小值，这点正对应着 W212 所处的位置，显示这个残基的存在使甲胺易于从开口处相下运动，而且主要是由于残基的芳香环与配体形成的阳离子- π 作用，使得此处存在一个局域极小值。此外，生物实验上发现将这个残基突变成具有芳香环的苯丙氨酸，这个通道还具有一定的活性；当突变成没有芳香环的丙氨酸时，就彻底失去活性，显示了芳香环对配体传递所起的重要作用。另外，我们的轨迹 B1-B4 得到的模拟结果也与实验结果非常相符，进一步验证了这个芳香环对于配体通过起着重要作用。而从 Am4 到胞外，甲胺和氨的能垒分别是 1.0 和 3.44 kcal/mol。从这些结果来看，我们推测甲胺和氨的疏水性和体积的差异是造成甲胺低传导性的主要原因，而它们的最大能垒都集中在两个苯丙氨酸所处的位置。另外，发现这两个配体在通道传导过程中形成的氢键相互作用的残基基本一样，都是 H168, H318, Y32 和 S263。虽然甲胺和氨在 PMF 具体数值上有所差别，但是他们的整体表现并无本质差别，因此实验上使用甲胺作为氨的替代物来研究野生型 AmtB 的功能是具有其合理性的。

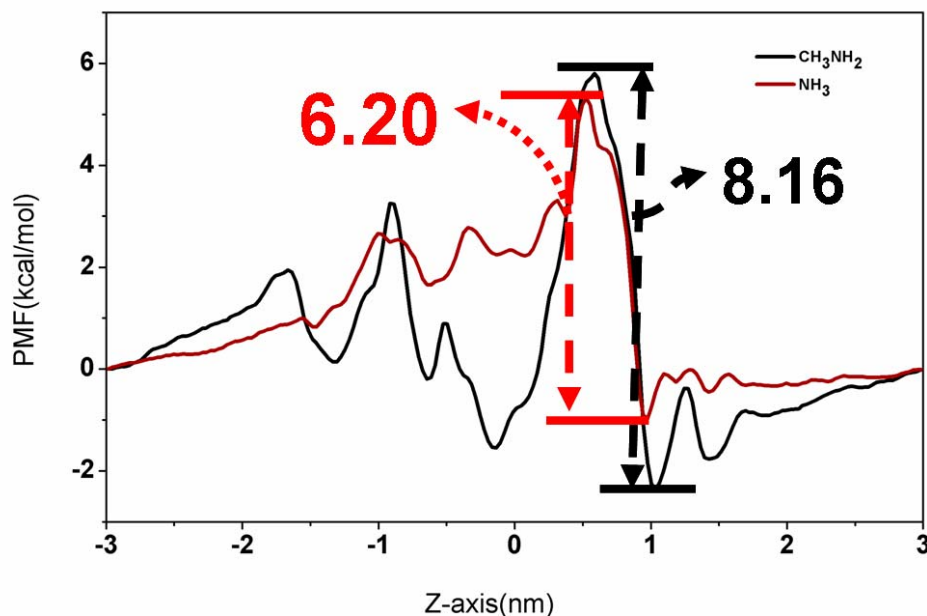


图 5.5 野生型 AmtB 传导甲胺和氨分子的 PMF 随通道位置的变化, 其中红线为氨, 黑线为甲胺。

5.3.2 残基 W212 的重要作用

通过分析轨迹A1发现, 随着模拟的开始, M1首先位于Am2时会和H168形成氢键, 不过由于W212上有芳香环, 存在阳离子- π 相互作用, 它会自发的振动到W212的芳香环平面上, 在模拟中经常可以观察到氮原子和芳香环的距离在3.2 Å左右, 显示可以形成比较强的阳离子- π 相互作用, 且由于六元环的阳离子- π 相互作用强于五元环的作用⁶¹, 因而甲胺能够自发的从色氨酸的五元环滑向六元环, 如图5.6所示。这点和PMF图上在W212附近具有一个极小值点是相符的。因此它不需要跨过很大的能垒就可以到达通道底部。随着模拟的进行, M1可以继续向下移动, 从而可以和H318及Y32形成氢键, 将M1拉向通道底部。在这一过程中M23, I28, I110, I114, I208, W212 和 F215都能与M1形成疏水作用, 从而

有利于M1从位点Am2移动到Am4。同样的M2也是通过相同的作用方式从而达到通道底部的。

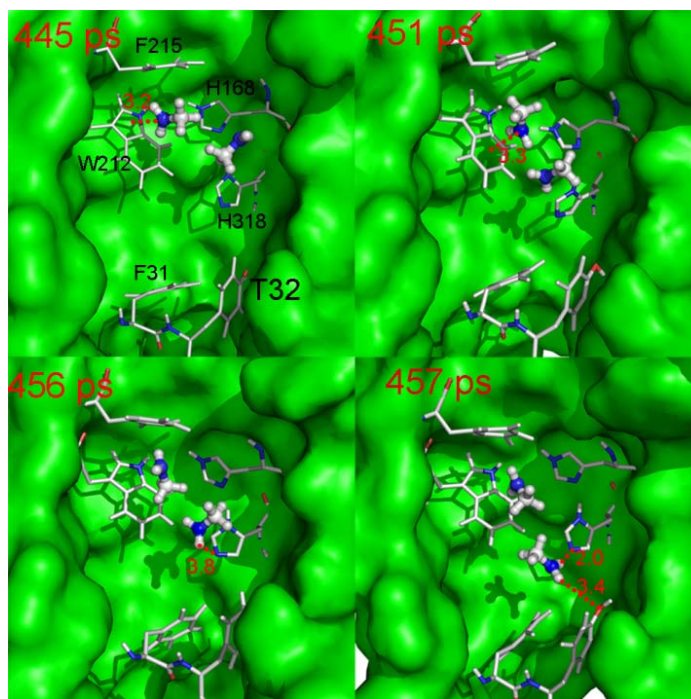


图5.6 通过阳离子- π 作用，M1从位点Am2运动到Am4。来自轨迹A1的五个快照结构(445, 451, 456 和 457 ps)的俯视图。氢键和阳离子- π 作用使用红色的虚线显示。

为了验证W212上芳香环对于传导氨的重要性，我们建立了W212F和W212A这两个突变体，且分别对甲胺和氨分子通过这两个突变体的情况进行模拟(轨迹B1-B4)，结果显示甲胺和氨分子都可以通过突变体W212F(轨迹B1和B2)，但都不能在20ns的模拟时间内通过突变体W212A（轨迹B3和B4）。进一步的分析显示，甲胺分子在924ps左右离开AmtB(轨迹A2)，而在3215ps左右离开W212F(轨迹B1)，而氨分子则在7660ps(轨迹A3)和2560ps左右离开通道（轨迹B2）。

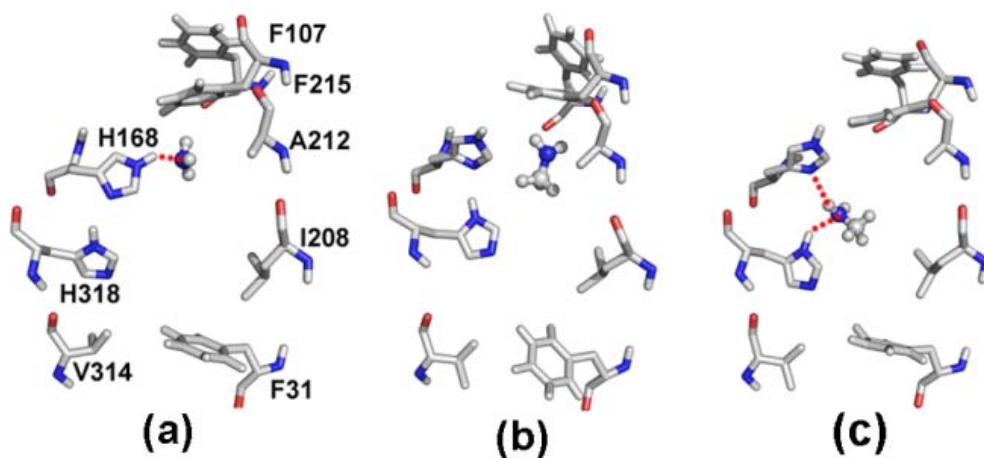


图5.7 三个来自突变体W212A的快照(A)从轨迹C2中的20ns快照结构；(B)轨迹C1中的20ns快照结构；(C)从轨迹C1中的6.5ns的快照结构。

通过分析轨迹B3可以发现：甲胺分子在Am2位置附近摆动，且有时能够插入H168和H318之间，形成两个氢键（如图5.7 B和C所示），由于这2个氢键的同时形成，限制了甲胺分子继续运动到AmtB通道底部。而在轨迹B4中发现氨分子基本上在Am2位置振动，极少向下移动（见图5.7A）。这些结果显示，没有了芳香环的存在，H168中氮原子会和氨分子形成足够强的氢键阻止其向下运动。而在野生型AmtB及突变体W212F中由于存在的阳离子- π 相互作用，使配体可以很容易通过。

同样的为了解释W212F保持了野生型的大部分活性，我们计算了在传导过程配体的氮原子与AmtB中芳香环的距离随时间的变化情况。如图5.8所示，配体从Am2传导到Am4的过程中存在着NH- π 相互作用，如在突变体W212F(轨迹B1和B2)中起作用的平均距离为 ~ 5.3 Å 和 ~ 4.0 Å，因此也应该具有和野生型类似的功能。另外，我们还发现芳香环和甲胺的距离大于与氨的距离，显示甲胺形成的NH- π 相互作用弱于氨。同时我们发现轨迹B1的平均距离比轨迹A2中的数值大 0.5 Å，显示阳离子- π 相互作用在野生型中强于W212F突变体，这个或许就是W212F活性下降的原因。而通过比较轨迹A'，A3 和B2，芳香环到氨的距离没

有什么变化，揭示了在W212F中可能传导氨的活性并未下降。

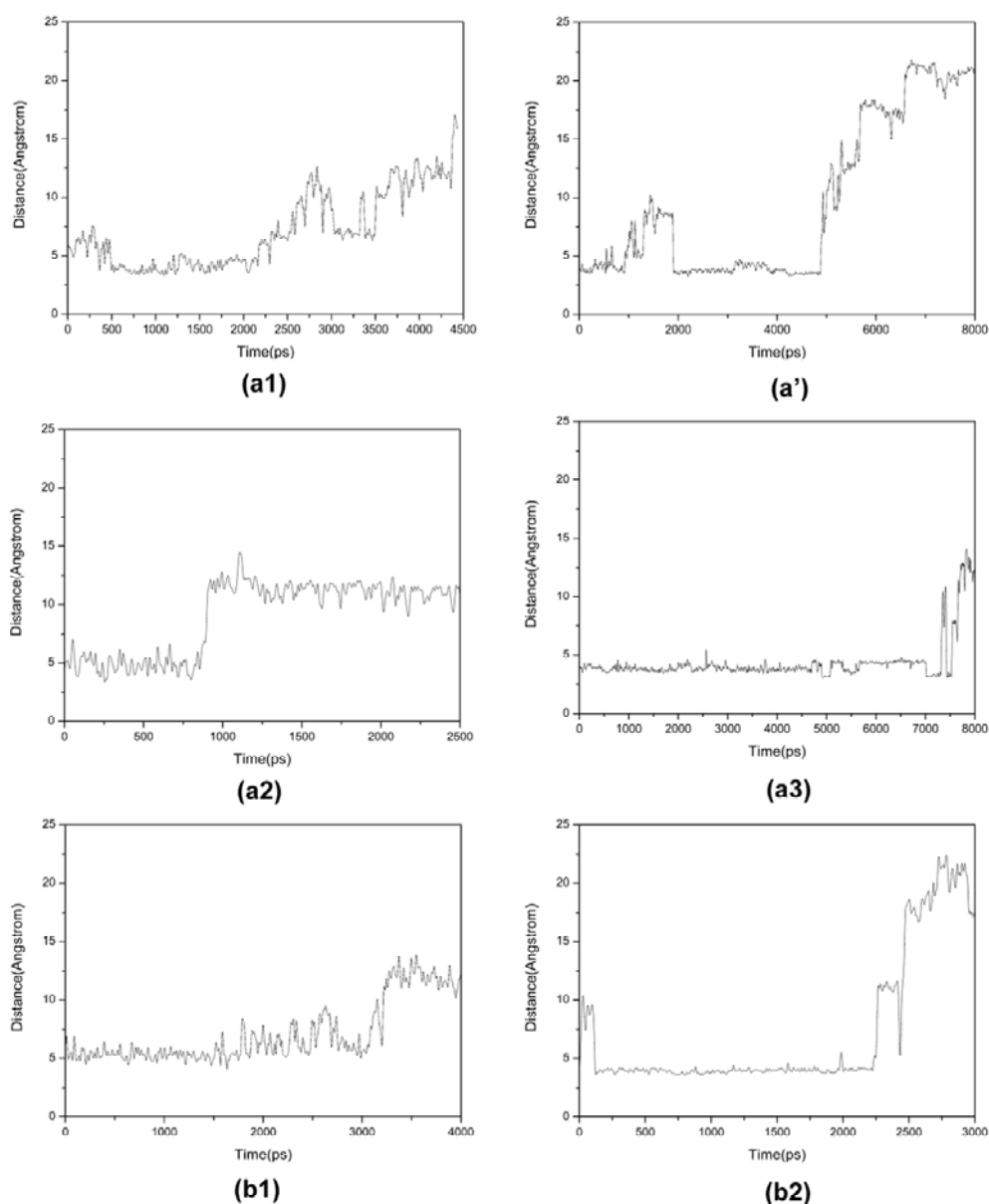


图5.8 配体（甲胺和氨）中氮原子与212号残基的芳香环的距离随时间变化。(a1) 分子M2来自轨迹A1；(a') 分子nha1来自轨迹A1 (文献32)；(a2)甲胺分子来自轨迹A2；(a3)分子氨来自轨迹A3；(b1)配体为甲胺，来自轨迹B1；(b2)配体是氨分

子，来之轨迹B2。

5.3.3 残基 Ser263 及水分子的作用

在模拟中我们发现通道的出口处常有水分子出现，而且可以和配体形成氢键，如在1295ps的时候，它可以和M1形成氢键，从而将其拉向出口，与S263形成氢键，最终将配体从由F31和V314组成的出口中拉向胞内（见图5.9）。在模拟中发现水分子可以进入通道内，不过一般只停留在Am4这个位置，而不能继续向上运动，且这些水分子也会自动离开通道，进入胞内，这点和Zheng等的晶体结构¹⁶中同一位置发现水的事实相符。

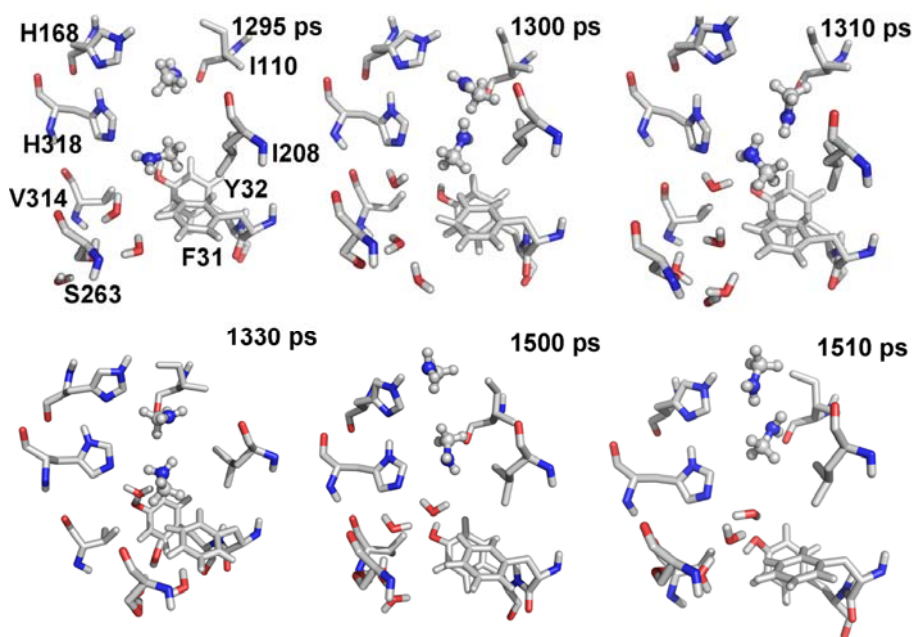


图5.9 轨迹A1水分子进入AmtB,显示了6个快照结构(1295, 1300, 1310, 1330, 1500 和1510 ps)。

为了更进一步验证S263的作用，我们设计了C1-C6六条轨迹，分别对能形成氢键的S263C，S263D和不能形成氢键的S263A突变体进行模拟，结果如图5.10

所示, 这些能形成氢键的突变体(S263C和S263D)都能传导配体, 而S263A突变体则在模拟时间20ns内没有观察到配体的离开。

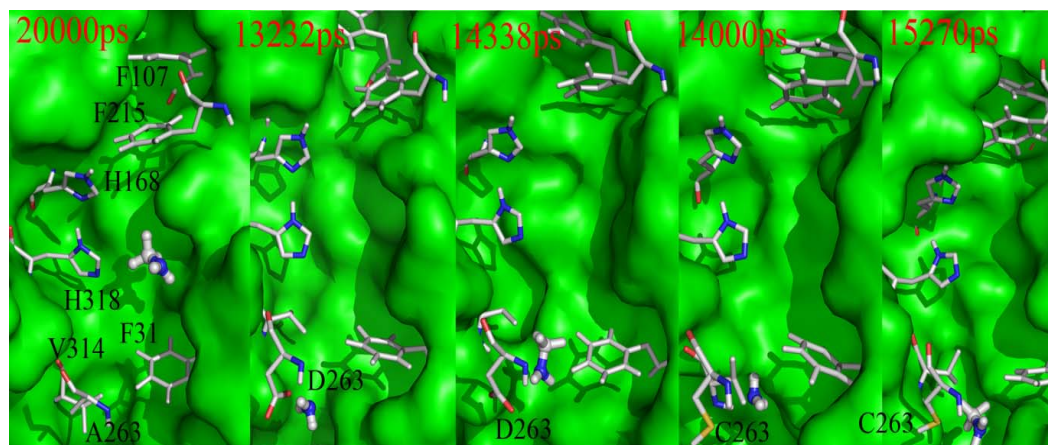


图5.10 配体(甲胺和氨)通过263号突变体的过程。显示了五个快照结构(轨迹C1- S263A的20ns、轨迹C3-S263D的13232ps、轨迹C4-S263D的14338ps和轨迹C5-S263C的14000ps, 轨迹C6-S263C的15270ps的结构)。

5.3.4 AmtB 入口动力学行为

残基A162位于入口处F107和F215的正对面, Zheng¹⁶等推测这个残基的羰基氧有向上和向下两个朝向, 从而可以与配体形成氢键, 将配体从入口Am1处转运到通道Am2处。在我们的动力学模拟中确实观察到了A162的这一构象变化。如图5.11所示, 在轨迹A1中的G163的 ϕ 角度可以在 $\sim 140^\circ$ 到 $\sim 80^\circ$ 之间变化, 从而导致A162的羰基从向上构象转为向下构象。同时我们发现这一运动模式并不仅仅只在这条轨迹中发现, 而是在所有的轨迹中都出现, 显示这一运动模式是这一蛋白的内在运动。如图5.11所示, 当A162取向上构象时, 可以看到其可以和处在位点Am1中的水分子形成氢键, 而当它取向下构象时, 在模拟中可以观察到其可以与在Am2中的配体形成氢键。因此我们推测这个残基从向上构象往向下构象转变时, 可以通过与配体形成氢键将其从Am1拉向通道内部, 这点和

Nagaard 等³⁷的动力学模拟结果也是一致的。

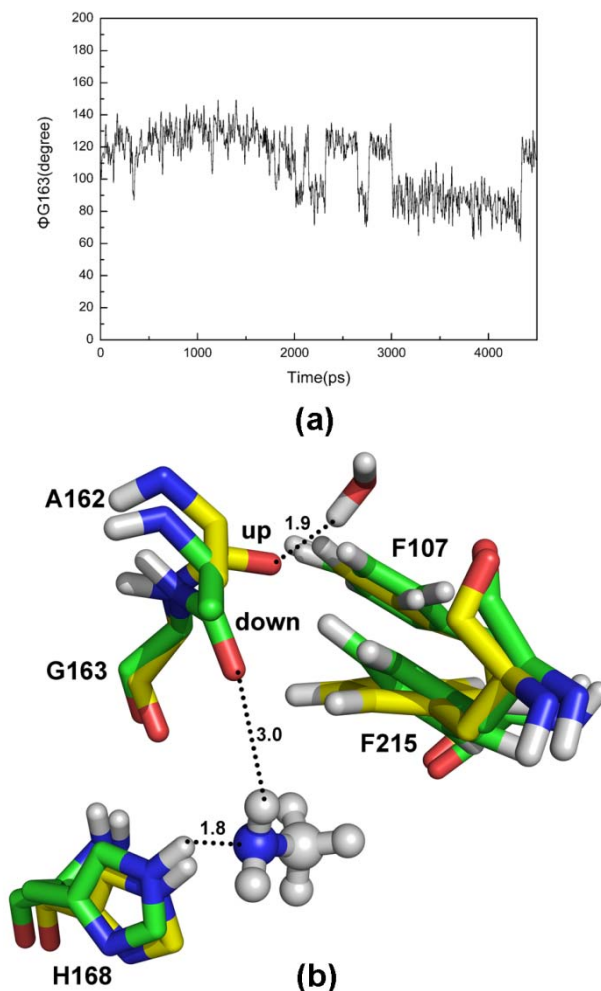


图5.11 (a) 轨迹A1中Gly163的 ϕ 角随时间的变化。(b) Ala162向上(1425ps)和向下(1758ps)两个构象的叠和。

5.3.5 配体分子离开通道的机制

当AmtB处于不同的结晶环境下时, 它的胞内出口处存在两种状态, F31 与 V314 的距离在此2中状态中变化挺大的, 即在晶体1XQF中其距离大约为8.0 Å 而在1XQE中则大约为10.5 Å。基于这个实验结果, 有研究提出了一个可能的氨

分子离开通道的门控机理：AmtB在转运氨分子的过程时，开口处会发生大的构象变化，F31 与V314 的距离需要在8.0Å 和10.5 Å 之间变化¹⁶。

之前的模拟结果显示了氨分子离开通道的过程中，F31 和V314 间的距离并不需要发生大的变化³²。在此我们也检测了F31 与V314 的距离在轨迹A1 中随时间的变化，结果显示，虽然在模拟中观察到了F31 与V314 的距离可以达到10 Å左右，但是在甲氨分子通过通道时，它们的距离都很小(0-100 ps, 1800-1900 ps和2600-2700 ps, 见图5.12)，在轨迹A2中也有相似的发现。因此我们认为在甲胺分子通过通道时，是不需要F31 与V314大的变化的。

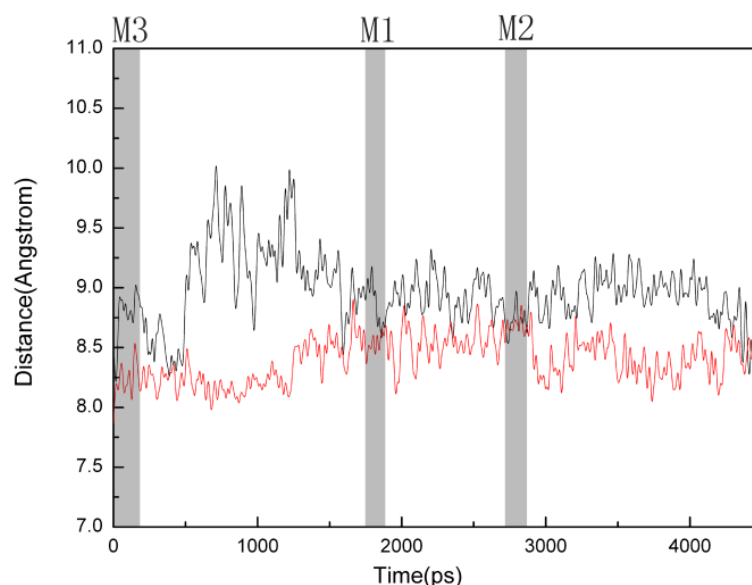


图5.12 Phe31 与Val314 的距离在轨迹A1(黑色)和 A' (红色)中随时间的变化。

5.3.6 组氨酸对体内活性的影响

我们的合作者 Mike 课题组将 AmtB 和相应的突变体基因导入到缺失所有 3 个 MEP 等位基因的 *S. cerevisiae* 的菌株中，进行细菌的生长实验。所使用的突变体有 F215A, S219A, H168A, H318A 和 H168A/H318A。其中作为对照试验的

是完全失活的突变体 F215A 和活性提高的突变体 S219A。结果如图 5.13 A 所示，在铵盐生长实验中只有双突变体 H168A/H318A 才完全失活，而其他的单组氨酸突变体（H168A 和 H318A）都可以存活。

为了排除 AmtB 是 *S.cerevisiae* 外源基因而可能导致的缺陷，他们还将内源的等位基因 MEP2 及相应的突变体导入到同样的菌株中去，结果如图 5.13 B 所示，这一结果和导入基因 AmtB 是完全一致的，即只有双突变体才失活，而单突变体都具有活性，所以此实验排除了外源基因导入的可能原因。

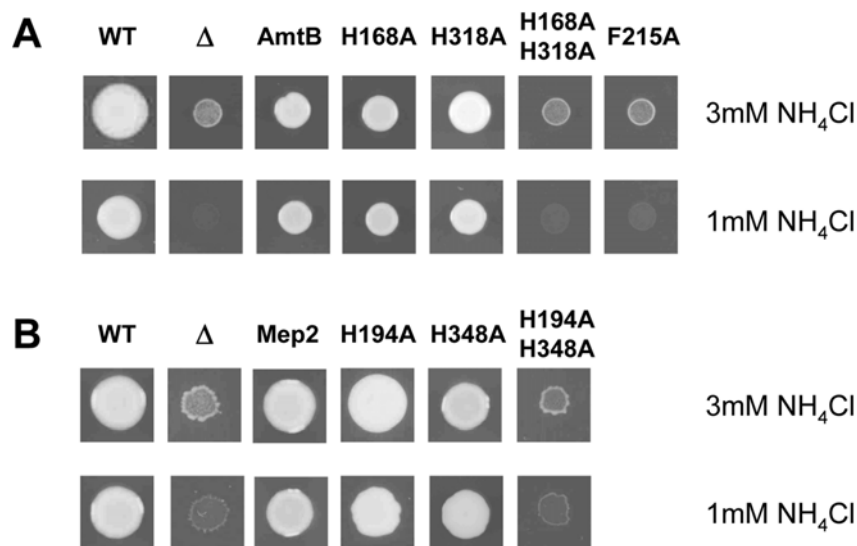


图 5.13 表达AmtB和Mep2 组氨酸突变体的*S. cerevisiae*依赖铵盐生长实验。(A) 在 30 °C，pH=6.0 包含 1mM(上列)或者 3mM（下列）NH₄Cl 的 YNB培养液中培养*S. cerevisiae* 突变体 5 天后的生长情况。WT-*S. cerevisiae* 菌株 23344c, Δ-*S. cerevisiae* 菌株 31019b缺乏所有的 3 个mep基因，AmtB-菌株 31019b表达了野生型的E.coli AmtB基因或者其突变体；(B) 与A一样，表达的是Mep2 基因或者突变体的mep2 基因。

而利用甲胺同位素实验结果如图 5.14 所示，各种组氨酸突变体都不能通过甲胺。因为氨的存在形式依赖于外界 pH 值，同时做了在 pH=4.0 时的 *S.*

cerevisiae 的生长实验，结果和 pH=6.0 的时候完全一致（如图 5.15 所示）。这些结果显示甲胺和氨的传导在组氨酸突变体中是具有很大区别的，单组氨酸突变体（H168A 和 H318A）都可以传导氨而不能传导甲胺，而双突变体既不能通过甲胺也不能通过氨。

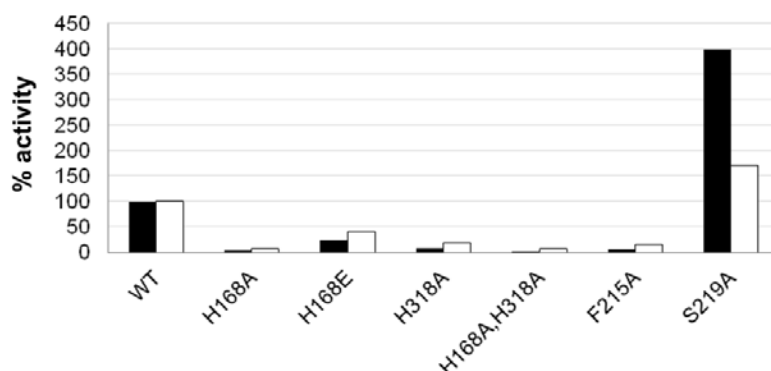


图 5.14 不同AmtB突变体的传递甲胺的活性。 $^{14}\text{CH}_3\text{NH}_2$ 在 *S. cerevisiae* 菌株 31019b（空心）中和大肠杆菌中（实心）表达AmtB后得同位素实验。

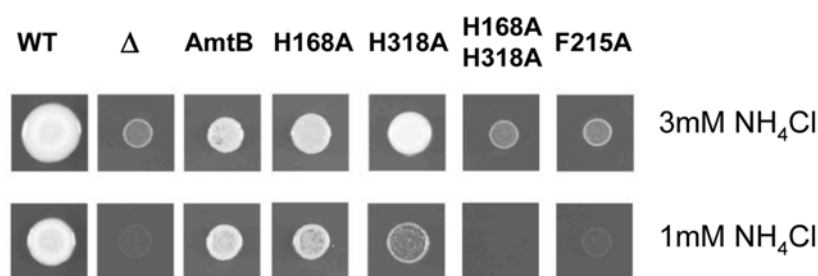


图 5.15 在 pH=4.0 时，表达AmtB及组氨酸突变体的 *S. cerevisiae* 依赖铵盐生长实验。在 30 °C，包含 3mM(上列)或者 1mM(下列) NH_4Cl 的 YNB 培养液中培养 *S. cerevisiae* 突变体 5 天后的生长情况。WT-*S. cerevisiae* 菌株 23344c, Δ -*S. cerevisiae* 菌株 31019b 缺乏所有的 3 个 mep 基因, AmtB-菌株 31019b 表达了野生型的 *E. coli* AmtB 基因及其突变体。

5.3.7 动力学模拟组氨酸突变体通过配体机制

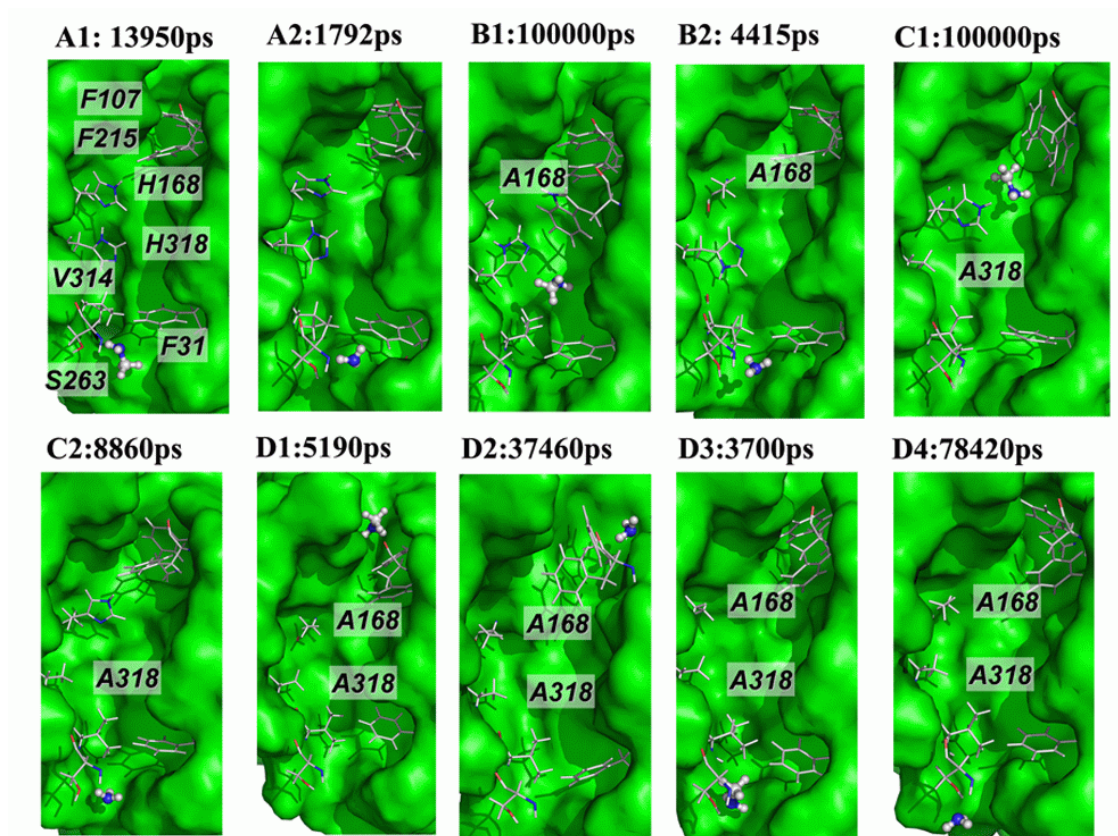


图5.16 甲胺和氨分子在各个模拟轨迹中离开通道的过程。A1-13950ps来自轨迹D1（甲胺，野生型）（表4.2）;A2-1792ps来自轨迹D2（氨，野生型）;B1-100ns来自轨迹E1(甲胺，突变体H168A); B2-4415ps来自轨迹E2(氨，突变体H168A); C1-100ns来自轨迹F1(甲胺，突变体H318A); C2-8860ps来自轨迹F2(氨，突变体H318A); D1-5190ps来自轨迹G1(甲胺，突变体H168A/H318A); D2-37460ps来自轨迹G2(氨，突变体H168A/H318A); D3-3700ps来自轨迹G3(甲胺，突变体H168A/H318A); D4-78420ps来自轨迹G4(氨，突变体H168A/H318A);在此侧视图中，甲胺分子以白色和蓝色球棍状结构显示，残基F31、F107、H168、F215、S263，V314和H318以彩色棍结构显示。

所有的组氨酸突变体的动力学模拟细节及结果如表5.2所示，甲胺和氨分子分别放在位点Am2中。结果如图5.16所示，在100ns的模拟中，甲胺不能通过H168A和H318A突变体（轨迹E1和F1），而氨分子都可以在10 ns以内通过通道。其中H168A在9 ns离开通道(轨迹E2)，在H318A中则在5 ns就离开通道(轨迹F2)。在双突变中我们发现（轨迹G1-G4），不管是甲胺还是氨都可以从通道的开口和出口离开通道，这样的结果使我们推测这个双突变体失去活性的原因是失去了通道的方向性，而在生物体内AmtB是单一方向传递配体的，从而可以将氨分子从胞外运输到胞内的。而我们的模拟结果显示这个突变体失去了方向性，因此在低浓度下这个通道是极有可能因此失活，我们的常规动力学模拟结果和Mike的实验结果非常吻合。

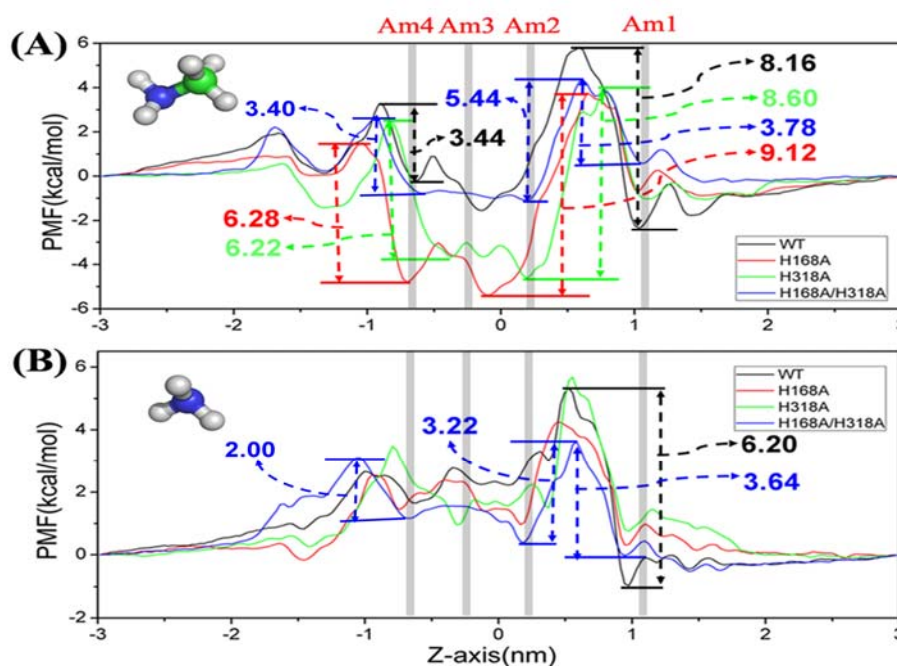


图 5.17 甲胺和氨通过野生型 AmtB 和各种突变体 (H168A, H318A, H168A/H318A) 的 PMF 随通道位置的变化曲线。其中黑色为野生型, 红色为突变体 H168A, 绿色为突变体 H318A, 青色为突变体 H168A/H318A。

5.3.8 突变体 H168A 传导氨和甲胺的异同

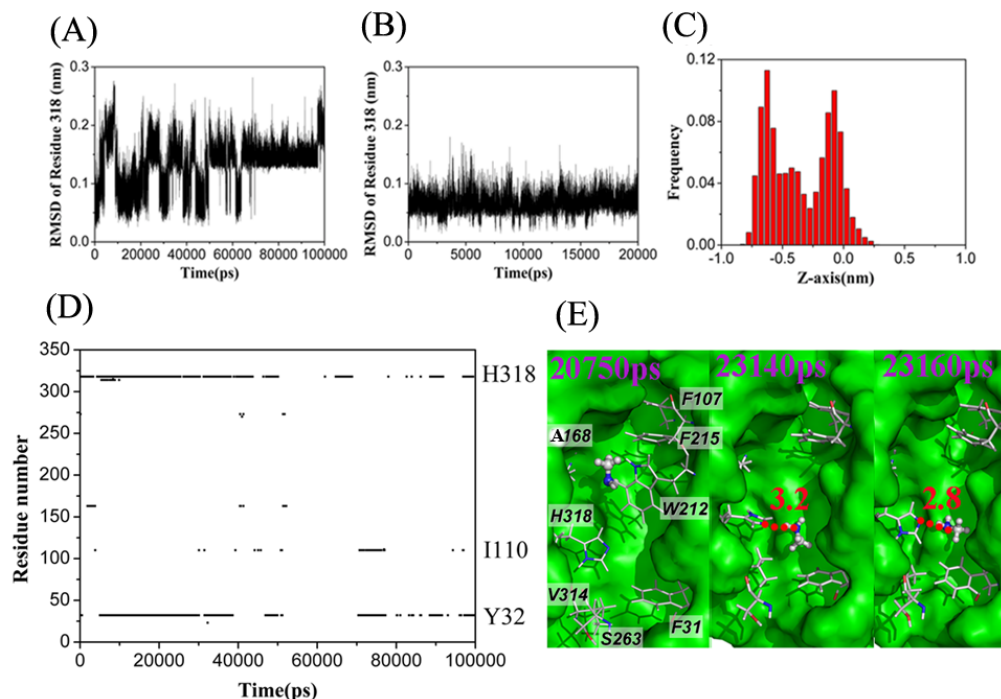


图5.18 甲胺在H168A AmtB中的影响。(A) 残基318在轨迹E1中随时间的变化情况。(B) 残基318在轨迹E2中随时间的变化情况。(C) 甲胺在通道各处所处时间的概率，数据来自轨迹E1(100ns)。(D) 在轨迹E1中与甲胺分子形成氢键的残基随时间的变化情况。(E) 20750, 23140和23160 ps的构象侧视图。显示甲胺在 $z \approx 0$ 出由于和A318附近残基的疏水作用和 $z \approx -0.6$ 处由于和H318氢键相互作用而被限制在这些位置。

在铵盐生长实验中我们发现H168A可以发挥正常的传导氨功能，却不能传导甲胺（甲胺通透实验，图5.14）。因此我们设计了体系E1和E2来考察H168A传导甲胺和氨的异同点。如图5.16所示，氨可以在5 ns以内通过H168A，而甲胺在100 ns的模拟时间内也不能通过。通过分析轨迹E1和E2，我们发现H318这个残基在甲胺中的波动明显大于氨的情况（图5.18A和B）。进一步轨迹分析发现

造成H318残基波动的原因是这个残基可以和甲胺形成比较强的氢键相互作用（图5.18E），而这个氢键可以将甲胺分子限制在 $Z=0\text{nm}$ 和 $z=-0.6\text{ns}$ 附近，这点和我们通过统计比较甲胺在通道中各点出现概率是一致的（图5.18C），在这两点附近都可以看到甲胺出现的概率最大。而在传导氨分子的时候，如图5.19所示，由于W212的芳香环能和氨分子形成较强的NH- π 相互作用，而这一作用可以将氨分子从位点Am2传到Am4附近，而到了位点Am4后，氨分子就可以通过这个通道的出口离开通道了。

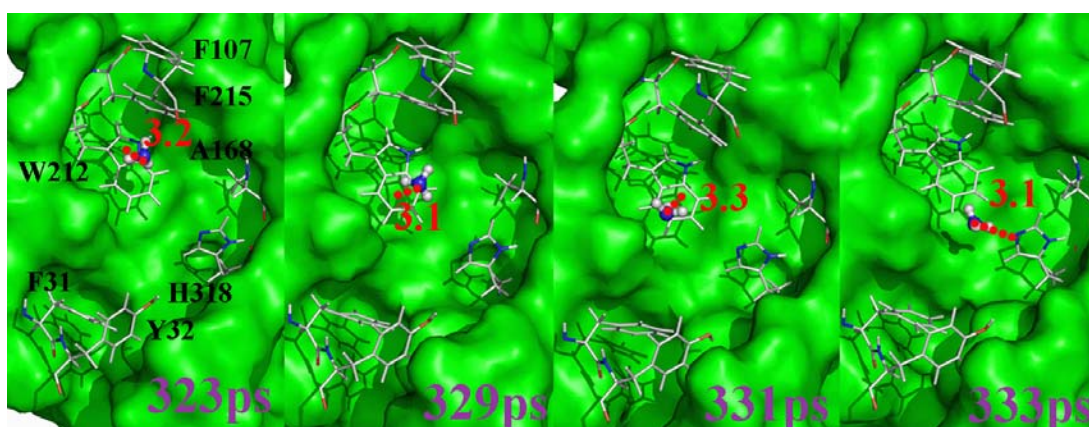


图5.19 AmtB突变体H168A通过阳离子- π 作用将氨分子从Am2转运到Am4。来自轨迹E2五个快照结构(323, 329, 331 和 333ps)的俯视图。氢键和阳离子- π 作用使用红色的虚线显示。

通过对图5.17中的H168A和野生型的PMF进行比较，发现这2个体系的PMF具有很大的区别：在H168A传递甲胺分子的时候，从 $z \approx -0.8\text{ nm}$ 到 $z \approx 0.7\text{ nm}$ 的过程中存在一个非常大的势阱，分别有 9.12 kcal/mol 和 6.29 kcal/mol 的能垒，而且在 $z \approx -0.70\text{ nm}$ 和 -0.10 nm 处的局域极小值的PMF值分别为 -4.79 和 -5.37 kcal/mol ，而这两个正好对应着常规动力学模拟中甲胺所发现的出现概率最大的两个点（图5.18 C）。可见这个巨大的势阱阻碍了甲胺的通过。而对于氨分子来说，它和野生型的没有多大区别了。我们通过自由能微扰方法分别计算了甲胺

和氨分子在水溶液和在H168A中的结合位点Am2中的自由能变化，发现在水溶液中的为 $-0.51 \pm 0.42 \text{ kcal/mol}$ ，显示在水溶液中，这2中分子的结合自由能没有什么变化。但是在结合位点中则为 $2.4 \pm 0.50 \text{ kcal/mol}$ ，可见甲胺在疏水环境中的结合能力比氨分子要强的多，因此单突变造成的通道内更疏水的环境是导致其限制甲胺通过，而不影响氨通过的主要原因。

5.3.9 突变体 H318A 传导氨和甲胺的异同

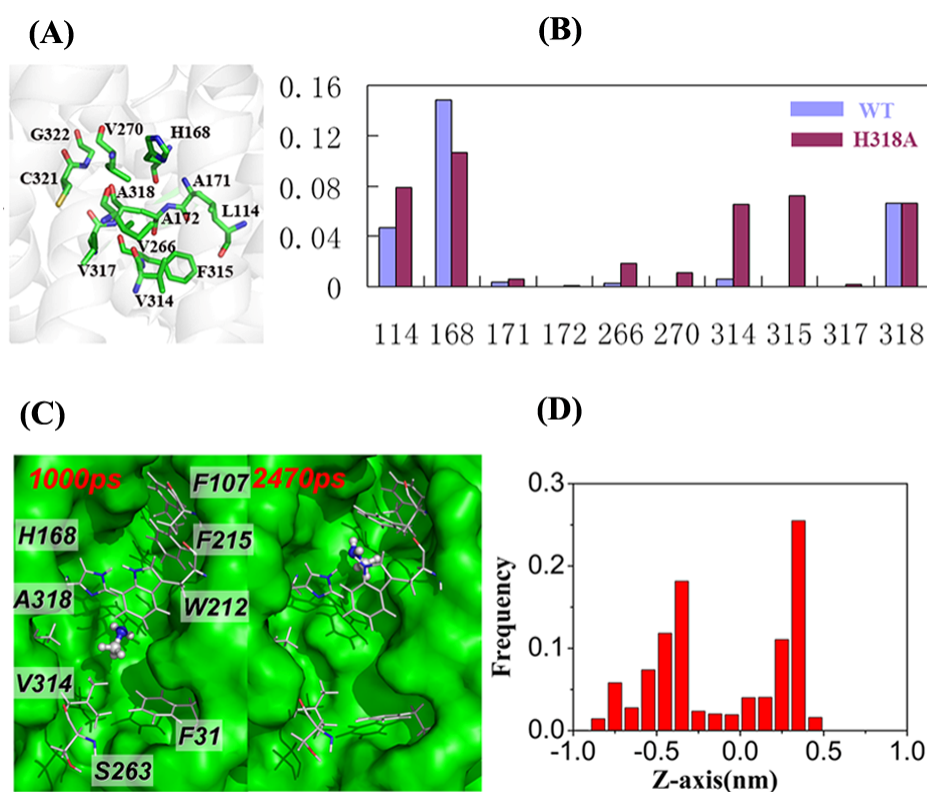


图 5.20 配体离开 AmtB 突变体 H318A 的动力学机制。(A) 在晶体结构 2NPE 中与甲胺作用的处于残基 318 附近的残基。(B) 在野生型与突变体 H318A 中甲胺与处于残基 318 附近的残基疏水作用比较 (C) 甲胺出现在通道中概率最多的地方，显示了 1000ps 和 2470ps 的结构。(D) 甲胺在通道各处所处时间的概率，数据来自轨迹 F1(100ns)。

为了研究H318A的动力学行为，我们分别进行了F1和F2的动力学模拟。如图5.16所示，甲胺在100 ns的模拟时间内不能通过通道，而氨则在10 ns内可以通过通道。通过分析轨迹E2，我们发现氨分子可以通过与残基P311的主链形成氢键，而通过这个氢键，则可以将氨分子带出出口到胞内（见图5.21）。而在传递甲胺分子的时候，我们发现甲胺分子被限于两个位点附近（ $z \approx -0.48$ nm和0.30 nm，图 5.20 C 和D）。进一步检查发现在残基318附近的残基都是疏水性的，如I114, I266, V314和F315等（见图5.20A）。我们计算了在318号残基附近的疏水残基和甲胺作用的频率，通过与野生型相比（见图5.20B），我们发现如I114, I266, V314和F31这些位在318附近的残基，与甲胺的作用频率在H318A中要高于野生型。因此，这些疏水作用使得甲胺被限制在这个位点中。

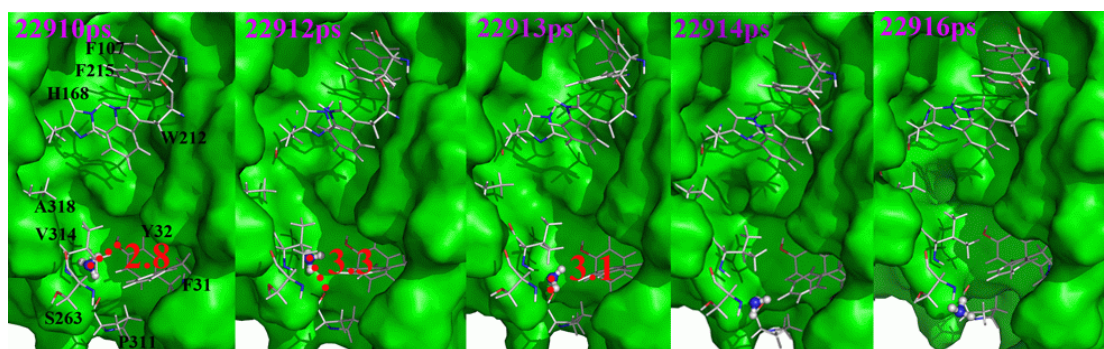


图5.21 H318A突变体中氨分子通过与Pro311形成氢键从Am4运动到胞外。来自轨迹F2的五个快照结构(22910, 22912, 22913, 22914和22915ps)的俯视图。氢键使用红色的虚线显示。

我们同样计算了甲胺和氨在H318A中的PMF。如图5.17所示，氨从 $z \approx -0.32$ nm移动到胞内 $z \approx -1.50$ nm的能垒只有2.51 kcal/mol，因此很容易通过氨。然而对于甲胺，和H168A突变体一样，甲胺也存在一个很深的势阱，其中分别存在8.60 kcal/mol 和 6.22 kcal/mol的能垒，因此显示甲胺很容易被限于 $z \approx 1.0$ nm到-1.0 nm的区域。在这一区域，同样出现有两个局域极小值（在 $z \approx -0.38$ nm的

-3.67 kcal/mol 和在 $z \approx 0.25$ nm的-4.60 kcal/mol），这也和在常规动力学中发现的两个出现概率最大的点相符。根据存在的这2个巨大的能垒，我们可以推测在这个H318A突变体中，同样传导甲胺的能力是大大下降的，而这个原因也是由于突变造成的通道疏水性的改变造成的。

5.3.10 双突变 H318A/H168A 失活的机制研究

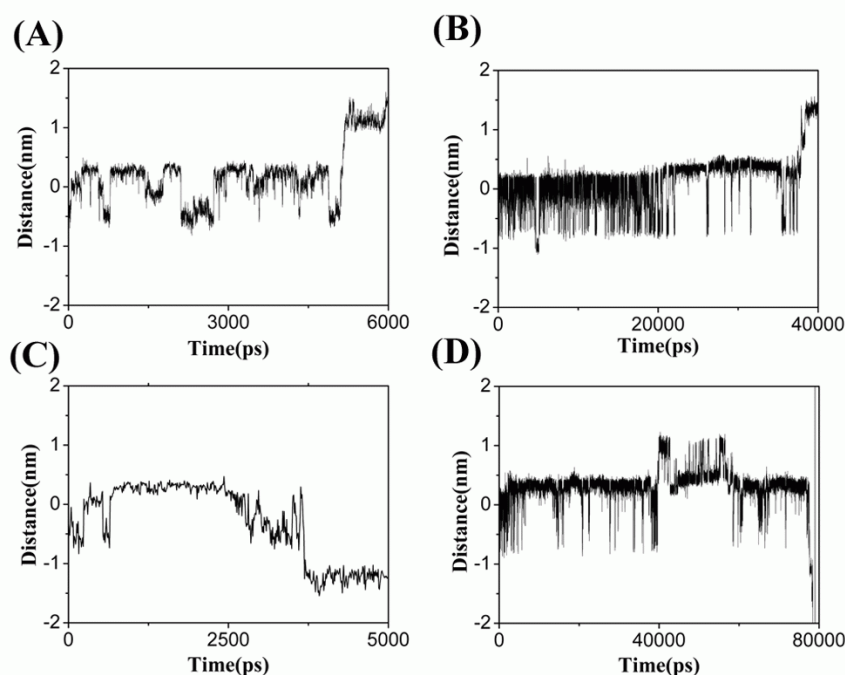


图 5.22 配体甲胺和氨在双突变在 H168A/H318A AmtB 中随 Z 轴方向随时间变化函数。A、B、C 和 D 分别来自轨迹 G1-G4。

和野生型相比，我们发现双突变体（H318A/H168A）既不能传递甲胺也不能传递氨分子。为了研究双突变体失活的机制，我们分别进行了4条动力学模拟（轨迹G1-G4）。我们发现不管是甲胺还是氨都可以从通道的上部或下部离开通道，显示出这个通道没有方向性（图5.16 D1-D4 和图5.22），因此我们推断这个突变体失活是由于其失去了单向传递配体的能力。

在AmtB的开口处有两个基本平行的苯丙氨酸(F107 和 F215)，这两个残基在开口处起着门控的作用。模拟结果显示开口处的F107 在动力学中是不稳定的，这点和之前的模拟结果也是吻合的(见图 5.23)，而且生物实验发现将这一残基突变成丙氨酸，并不影响通道的功能，这些结果显示这个残基并不是真正决定门控的残基。而生物实验也显示这个F215A突变体不仅不能传导甲胺也不能传导氨分子(图 5.13 和 5.14)，显示F215 这个苯丙氨酸才是真正起门控作用的残基。而且最近Akgun⁶²等发现F215 的二面角 $C_{\alpha}-C_{\beta}-C_{\gamma}-C_{\delta 1}$ 在配体转运过程中起着重要作用，这个角度可以控制由F107 和F215 组成的开口的开和关。他们发现这个角度在 $20^{\circ} (\pm 17^{\circ})$ 或者 $-166^{\circ} (\pm 14^{\circ})$ 的时候开口处于开的状态，而在 $80^{\circ} (\pm 16^{\circ})$ 和 $-100^{\circ} (\pm 15^{\circ})$ 范围内的时候则处于关闭的状态。

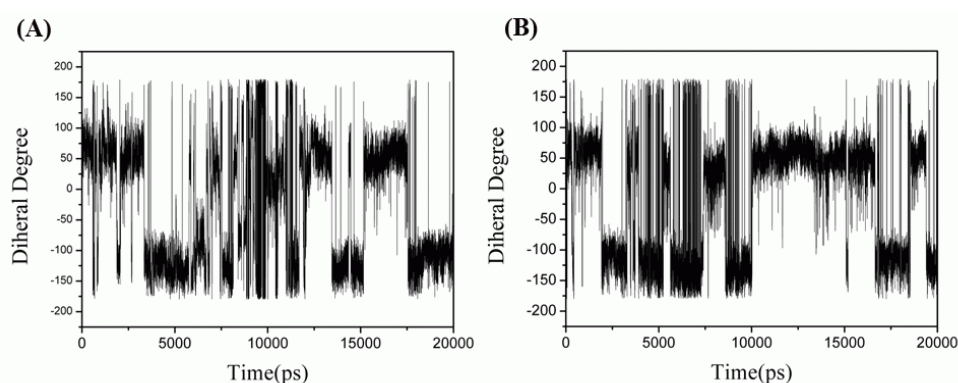


图5.23 二面角 $C_{\alpha}-C_{\beta}-C_{\gamma}-C_{\delta 1}$ 随时间变化情况。数据来源于轨迹D1和D2。

图 5.24 显示了野生型和双突变体AmtB中F215 的二面角 $C_{\alpha}-C_{\beta}-C_{\gamma}-C_{\delta 1}$ 随时间的变化情况。从图中可以看到在双突变体中这个角度与野生型相比非常不稳定，而为了考察这个变化是不是蛋白本身的运动，而非由配体诱导造成的，我们进一步模拟了没有配体的情况(轨迹G5 和G6)。结果也显示了空的突变体中这个角度也是不稳定的，这一结果显示这一运动是蛋白的本征运动，而不是配体有

道造成的。这些结果表明在双突变中由于突变造成了F215的运动变得不稳定，而使由F215所组成的开关经常处于打开的状态，从而使其失活。

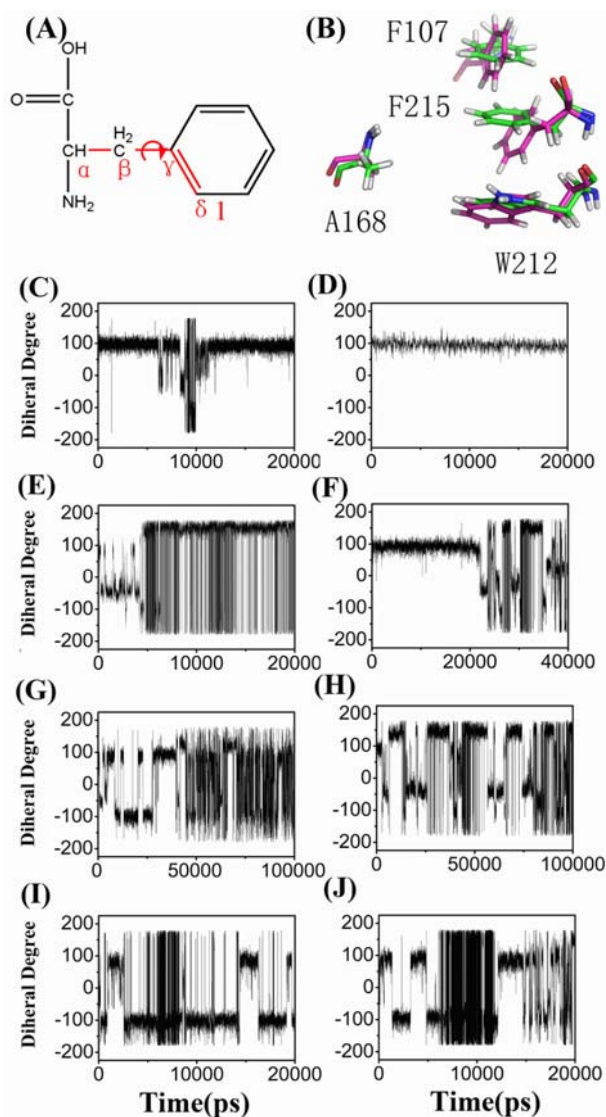


图 5.24 分析苯丙氨酸门控动力学。(A) 用于分析F215 门控的二面角 ($C_{\alpha}-C_{\beta}-C_{\gamma}-C_{\delta1}$)。(B) 两个部分平行的苯丙氨酸处于开(红色)和关(绿色)的情况(C-J)二面角 ($C_{\alpha}-C_{\beta}-C_{\gamma}-C_{\delta1}$) 随时间的变化函数。C和D分别是甲胺和氨通过野生型AmtB的情况，来自轨迹C1 和C2。E-J全部来自双突变体H168A/H318A AmtB (轨迹G1-G6)，其中E,G为甲胺，F和H为氨，I和J则没有配

体。

通过计算双突变的 PMF 发现：对于氨而言，从中间 $Z=0$ nm 处，往开口处移动和外出口处移动所需要的能垒基本一致，分别为 3.64 kcal/mol 和 3.22 kcal/mol，显示了其传导氨分子时没有方向性，这和我们在常规动力学中发现其既可以从上面开口，也可以从下面出口处离开通道的结果是一致的。而且在开口 ($z \approx -1.0$ nm) 和出口处 ($z \approx 1.0$ nm) 的 PMF 值分别为 3.10 和 3.62 kcal/mol，他们的差别只有 0.52 kcal/mol。显示氨分子在通道内部会失去单向选择性。

和氨通过的情况一样，在双突变体中甲胺开口和出口处的能垒之差是最小的，显示对于传递甲胺也是一样具有双向性的。具体说来就是从 $z \approx 1$ nm 到 0.5 nm，其能垒为 3.78 kcal/mol，只比 $z \approx 0.0$ nm 到 $z \approx 0.5$ nm 高 1.66 kcal/mol。另外在出口处 ($z \approx -1.0$ nm) 和开口处 ($z \approx -1.0$ nm) 的 PMF 值也只差 1.66 kcal/mol。而在野生型中这一差值为 2.55 kcal/mol。所有这些数据都显示这一突变体最容易失去方向选择性。这个和我们在动力学模拟中观察到的配体可以从上部和下部离开通道也是一致的。

5.4 小结

通过对我们建立的26个体系的动力学模拟，PMF计算及与实验的合作，我们系统研究了AmtB蛋白转运氨和甲胺的动力学机制和重要残基组氨酸在转运配体中的作用。我们的模拟结果显示野生型AmtB传导氨和甲胺分子的机制是相当类似的：即由H168, H318, Y32 和 S263组成的氢键网络将配体从Am2位置传导到胞内。其中W212在将配体从Am2传向Am4过程中起着重要作用，它像个中继站一样，通过NH- π 作用，将配体从Am2拉向自己的平面，从而转向Am4，然后配体可以与H318和Y32会和配体形成氢键，将其继续拉向AmtB通道的底部，而最后通过位于通道口的S263和出口处水分子的帮助，最终配体从通道内部转运到胞内。

通过生物实验和分子动力学实验探讨了组氨酸（H168和H318）在铵离子通道中的重要作用。结果显示了单突变（H168A和H318A）都不能传到甲胺，但是他们在生物体内都是具有活性的，都可以支持酵母的生长，而双突变体（H168A/H318A）则既不能传递甲胺，也不能支持酵母生长。虽然这些组氨酸突变体的结构与野生型的结构相比，非常接近，但是由于突变造成的通道内疏水性的改变，这个导致了甲胺在传递过程中出现巨大的能垒，而对于氨的传导则影响不大（见图5.17），而对于双突变，则是由于开口处苯丙氨酸的门控动力学行为的改变，导致了通道经常处于开的状态，从而降低了配体从通道内到胞外的能垒，造成了通道失去了单向选择性，从而导致了失活。从我们的实验结果也发现了使用甲胺作为氨的探针研究AmtB功能时具有一定的缺陷，结果显示虽然野生型AmtB传递甲胺和氨的机制类似，但是在一些突变体（比如H168A和H318A）中甲胺的通透性机制与氨并不相同，显示了甲胺作为氨类似物研究AmtB功能的不足，这点和Hall等⁶³利用突变实验的结果相似。

参考文献

1. Amon, J., et al., *Nitrogen control in Mycobacterium smegmatis: nitrogen-dependent expression of ammonium transport and assimilation proteins depends on the OmpR-type regulator GlnR*. J Bacteriol, 2008. **190**(21): p. 7108-16.
2. Broach, J., C. Neumann, and S. Kustu, *Mutant strains (nit) of Salmonella typhimurium with a pleiotropic defect in nitrogen metabolism*. J Bacteriol, 1976. **128**(1): p. 86.
3. Javelle, A., et al., *Ammonium sensing in Escherichia coli. Role of the ammonium transporter AmtB and AmtB-GlnK complex formation*. J. Biol. Chem., 2004. **279**(10): p. 8530-8.
4. Ninnemann, O., J. Jauniaux, and W. Frommer, *Identification of a high affinity NH₄⁺ transporter from plants*. EMBO J., 1994. **13**(15): p. 3464-3471.
5. Marini, A.M., et al., *A family of ammonium transporters in Saccharomyces cerevisiae*. Mol. Cell. Biol., 1997. **17**(8): p. 4282-4293.
6. Marini, A.M., et al., *Cloning and expression of the MEPI gene encoding an ammonium transporter in Saccharomyces cerevisiae*. EMBO J., 1994. **13**(15): p. 3456.
7. Gruswitz, F., et al., *Function of human Rh based on structure of RhCG at 2.1 Å*. Proc. Natl. Acad. Sci. USA, 2010. **107**(21): p. 9638-9643.
8. Bakouh, N., et al., *The challenge of understanding ammonium homeostasis and the role of the Rh glycoproteins*. Transfu. Clin. Biol. , 2006. **13**(1-2): p. 139-146.
9. Javelle, A., et al., *Structural and mechanistic aspects of Amt/Rh proteins*. J Struct Biol, 2007. **158**(3): p. 472-81.

10. Ludwig, U., *Ion transport versus gas conduction: function of AMT/Rh-type proteins*. Transfu. Clin. Biol. , 2006. **13**(1-2): p. 111-116.
11. Ludwig, U., B. Neuhauser, and M. Dynowski, *Molecular mechanisms of ammonium transport and accumulation in plants*. FEBS Lett., 2007. **581**(12): p. 2301-2308.
12. Planelles, G., *Ammonium homeostasis and human Rhesus glycoproteins*. Nephron Physiol., 2007. **105**(1): p. p11-p17.
13. Drepper, T., et al., *Role of GlnB and GlnK in ammonium control of both nitrogenase systems in the phototrophic bacterium Rhodobacter capsulatus*. Microbiology, 2003. **149**(Pt 8): p. 2203-12.
14. Javelle, A. and M. Merrick, *Complex formation between AmtB and GlnK: an ancestral role in prokaryotic nitrogen control*. Biochem. Soc. Trans., 2005. **33**(Pt 1): p. 170-2.
15. Khademi, S., et al., *Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 Å*. Science, 2004. **305**(5690): p. 1587-94.
16. Zheng, L., et al., *The mechanism of ammonia transport based on the crystal structure of AmtB of Escherichia coli*. Proc. Natl. Acad. Sci. USA, 2004. **101**(49): p. 17090-5.
17. Andrade, S.L.A., et al., *Crystal structure of the archaeal ammonium transporter Amt-1 from Archaeoglobus fulgidus*. Proc. Natl. Acad. Sci. USA, 2005. **102**(42): p. 14994-14999.
18. Soupene, E., et al., *Gas channels for NH₃: proteins from hyperthermophiles complement an Escherichia coli mutant*. J. Bacteriol., 2002. **184**(12): p. 3396-400.
19. Javelle, A., et al., *In vivo functional characterization of the Escherichia coli*

- ammonium channel AmtB: evidence for metabolic coupling of AmtB to glutamine synthetase*. Biochem. J., 2005. **390**(Pt 1): p. 215-22.
20. Javelle, A., et al., *Substrate binding, deprotonation, and selectivity at the periplasmic entrance of the Escherichia coli ammonia channel AmtB*. Proc. Natl. Acad. Sci. USA, 2008. **105**(13): p. 5040-5.
21. Andrade, S.L., et al., *Crystal structure of the archaeal ammonium transporter Amt-1 from Archaeoglobus fulgidus*. Proc. Natl. Acad. Sci. USA, 2005. **102**(42): p. 14994-9.
22. Lamoureux, G., M.L. Klein, and S. Berneche, *A stable water chain in the hydrophobic pore of the AmtB ammonium transporter*. Biophys. J., 2007. **92**(9): p. L82-4.
23. Wang, S., et al., *Ammonium transporters achieve charge transfer by fragmenting their substrate*. J. Am. Chem. Soc, 2012. **134**(25): p. 10419-10427.
24. Mayer, M. and U. Ludewig, *Role of AMT1; 1 in NH₄⁺ acquisition in Arabidopsis thaliana*. Plant Biol. , 2006. **8**(4): p. 522-528.
25. Mayer, M., et al., *Different transport mechanisms in plant and human AMT/Rh-type ammonium transporters*. J. Gen. Physiol., 2006. **127**(2): p. 133-44.
26. Fong, R.N., et al., *The W148L substitution in the Escherichia coli ammonium channel AmtB increases flux and indicates that the substrate is an ion*. Proc. Natl. Acad. Sci. USA, 2007. **104**(47): p. 18706-11.
27. Merrick, M.J. and R.A. Edwards, *Nitrogen Control in Bacteria*. Microbiol. Rev., 1995. **59**(4): p. 604-622.
28. Akgun, U. and S. Khademi, *Periplasmic vestibule plays an important role for*

- solute recruitment, selectivity, and gating in the Rh/Amt/MEP superfamily.* Proc. Natl. Acad. Sci. USA, 2011. **108**(10): p. 3970-3975.
29. Wang, J., et al., *Molecular Dynamics Simulations on the Mechanism of Transporting Methylamine and Ammonia by Ammonium Transporter AmtB.* J. Phys. Chem. B 2010. **114**(46): p. 15172-15179.
30. Nygaard, T.P., et al., *Substrate Recognition in the Escherichia coli Ammonia Channel AmtB: A QM/MM Investigation.* J. Phys. Chem. B, 2010. **114**(36): p. 11859-11865.
31. Lin, Y., Z. Cao, and Y. Mo, *Functional role of Asp160 and the deprotonation mechanism of ammonium in the Escherichia coli ammonia channel protein AmtB.* J. Phys. Chem. B, 2009. **113**(14): p. 4922-9.
32. Yang, H., et al., *Detailed mechanism for AmtB conducting $\text{NH}_4^+/\text{NH}_3$: molecular dynamics simulations.* Biophys. J., 2007. **92**(3): p. 877-85.
33. Ishikita, H. and E.W. Knapp, *Protonation states of ammonia/ammonium in the hydrophobic pore of ammonia transporter protein AmtB.* J. Am. Chem. Soc., 2007. **129**(5): p. 1210-5.
34. Cao, Z., Y. Mo, and W. Thiel, *Deprotonation mechanism of NH_4^+ in the Escherichia coli ammonium transporter AmtB: insight from QM and QM/MM calculations.* Angew. Chem. Int. Ed. Engl., 2007. **46**(36): p. 6811-5.
35. Bostick, D.L. and C.L. Brooks, 3rd, *Deprotonation by dehydration: the origin of ammonium sensing in the AmtB channel.* PLoS. Comput. Biol., 2007. **3**(2): p. e22.
36. Bostick, D.L. and C.L. Brooks, 3rd, *On the equivalence point for ammonium (de)protonation during its transport through the AmtB channel.* Biophys. J., 2007. **92**(12): p. L103-5.

37. Nygaard, T.P., et al., *Ammonium recruitment and ammonia transport by E. coli ammonia channel AmtB*. Biophys. J., 2006. **91**(12): p. 4401-12.
38. Liu, Y. and X. Hu, *Molecular determinants for binding of ammonium ion in the ammonia transporter AmtB-A quantum chemical analysis*. J. Phys. Chem. A, 2006. **110**(4): p. 1375-81.
39. Lin, Y., Z. Cao, and Y. Mo, *Molecular dynamics simulations on the Escherichia coli ammonia channel protein AmtB: mechanism of ammonia/ammonium transport*. J. Am. Chem. Soc, 2006. **128**(33): p. 10876-84.
40. Soupene, E., et al., *Ammonia acquisition in enteric bacteria: physiological role of the ammonium/methylammonium transport B (AmtB) protein*. Proc. Natl. Acad. Sci. USA, 1998. **95**(12): p. 7030-4.
41. Javelle, A., et al., *In vivo functional characterization of the Escherichia coli ammonium channel AmtB: evidence for metabolic coupling of AmtB to glutamine synthetase*. Biochem. J, 2005. **390**: p. 215-222.
42. Javelle, A., et al., *An unusual twin-his arrangement in the pore of ammonia channels is essential for substrate conductance*. J. Biol. Chem., 2006. **281**(51): p. 39492-8.
43. Javelle, A., et al., *An unusual twin-his arrangement in the pore of ammonia channels is essential for substrate conductance*. J. Biol. Chem., 2006. **281**(51): p. 39492-8.
44. Van der Spoel, D., et al., *GROMACS: Fast, flexible, and free*. J. Comput. Chem., 2005. **26**(16): p. 1701-1718.
45. Lindahl, E., B. Hess, and D. van der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis*. J. Mol. Model., 2001. **7**(8): p.

- 306-317.
46. Berger, O., O. Edholm, and F. Jahnig, *Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature*. Biophys. J., 1997. **72**(5): p. 2002-2013.
 47. Marrink, S.J., et al., *Adhesion forces of lipids in a phospholipid membrane studied by molecular dynamics simulations*. Biophys. J., 1998. **74**(2): p. 931-943.
 48. Tieleman, D.P., H.J.C. Berendsen, and M.S.P. Sansom, *An alamethicin channel in a lipid bilayer: Molecular dynamics simulations*. Biophys. J., 1999. **76**(4): p. 1757-1769.
 49. Breneman, C.M. and K.B. Wiberg, *Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials - the Need for High Sampling Density in Formamide Conformational-Analysis*. J. Comput. Chem., 1990. **11**(3): p. 361-373.
 50. Berendsen, H.J.C., et al., *Molecular-Dynamics with Coupling to an External Bath*. J. Chem. Phys., 1984. **81**(8): p. 3684-3690.
 51. Darden, T., D. York, and L. Pedersen, *Particle Mesh Ewald - an $N \cdot \log(N)$ Method for Ewald Sums in Large Systems*. J. Chem. Phys., 1993. **98**(12): p. 10089-10092.
 52. Hess, B., et al., *LINCS: A linear constraint solver for molecular simulations*. J. Comput. Chem., 1997. **18**(12): p. 1463-1472.
 53. Wallace, A.C., R.A. Laskowski, and J.M. Thornton, *LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions*. Protein Eng., 1995. **8**(2): p. 127-134.
 54. Bjelkmar, P., et al., *Implementation of the CHARMM force field in*

- GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models.* J. Chem. Theory Comput. , 2010. **6**(2): p. 459-466.
55. Hess, B., et al., *GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation.* J. Chem. Theory Comput. , 2008. **4**(3): p. 435-447.
56. Zoete, V., et al., *SwissParam: A fast force field generation tool for small organic molecules.* J. Comput. Chem., 2011. **32**(11): p. 2359-2368.
57. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling.* J. Chem. Phys., 2007. **126**(1).
58. Nose', S., *A molecular dynamics method for simulations in the canonical ensemble.* Mol. Phys. , 1984. **52**(2): p. 255-268.
59. Parrinello, M. and A. Rahman, *Polymorphic transitions in single crystals: A new molecular dynamics method.* J. Appl. Phys. , 1981. **52**(12): p. 7182-7190.
60. Hub, J.S., B.L. de Groot, and D. van der Spoel, *g_wham-A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates.* J. Chem. Theory Comput. , 2011. **6**(12): p. 3713-3720.
61. Zhu, W.L., et al., *How does ammonium interact with aromatic groups? A density functional theory (DFT/B3LYP) investigation.* J. Phys. Chem. A, 2000. **104**(42): p. 9573-9580.
62. Akgun, U. and S. Khademi, *Periplasmic vestibule plays an important role for solute recruitment, selectivity, and gating in the Rh/Amt/MEP superfamily.* Proc. Natl. Acad. Sci. USA, 2011. **108**(10): p. 3970-3975.
63. Hall, J.A. and S. Kustu, *The pivotal twin histidines and aromatic triad of the*

Escherichia coli ammonium channel AmtB can be replaced. Proc. Natl. Acad. Sci. USA, 2011. **108**(32): p. 13270-13274.

发表论文

1. **Jinan Wang**, Tim Fulford, Qiang Shao, Arnaud Javelle, Huaiyu Yang, Weiliang Zhu, Mike Merrick, Ammonium Transport Proteins with Changes in One of the Conserved Pore Histidines Have Different Performance in Ammonia and Methylamine Conduction, **PLoS ONE**, 2013, 8(5): e62745
2. Huiying Chu, **Jinan Wang**(共同第一作者),Hujun Shen, Yongliang Yang, Weiliang Zhu, Guohui Li, Investigation of Family 18 Chitinases and Inhibitors by Computer-aided Approaches. **Current Drug Targets**, 2012,13(4):502-11.
3. **Jinan Wang**, Weiliang Zhu, Guohui Li, Ulrich H. E. Hansmann. Velocity-scaling optimized replica exchange molecular dynamics of proteins in a hybrid explicit/implicit solvent. **Journal of Chemical Physics**, 2011,135, 084115
4. **Jinan Wang**, Huaiyu Yang, Zili Zuo, Xiuhua Yan, Yong Wang, Xiaoming Luo, Hualiang Jiang, Kaixian Chen, Weiliang Zhu. Molecular Dynamics Simulations on the Mechanism of Transporting Methylamine and Ammonia by Ammonium Transporter AmtB, **The Journal of Physical Chemistry B**,2010, 114(46):15172-15179
5. Jianzhong Chen, **Jinan Wang**, Beisi Xu, Weiliang Zhu, Guohui Li. Insight into Mechanism of Small Molecule Inhibitors of the MDM2-p53 Interaction: Molecular Dynamics Simulation and Free Energy Analysis, **Journal of molecular graphics modeling**, 2011, 30, 46–53
6. Junming Lu, Bo Zhang, Qingming Deng, **Jinan Wang**,Yunxiang Lu,Weiliang Zhu.The Nature and Magnitude of Specific Halogen Bonds Between Iodo-perfluorobenzene and Heterocyclic Systems, **International Journal of Quantum Chemistry**, 2011,111, 2352–2358

7. Yanyan Mao, Yun Du, Xiaohui Cang, **Jinan Wang**, Zhuxi Chen, Huaiyu Yang, Hualiang Jiang, Binding Competition to the POPG Lipid Bilayer of Ca^{2+} , Mg^{2+} , Na^{+} , and K^{+} in Different Ion Mixtures and Biological Implication **The Journal of Physical Chemistry B** 2013, 117, 850–858
8. Zhengyan Zhang, Zhijian Xu, Zhuo Yang, Yingtao Liu, **Jinan Wang**, Qiang Shao, Shujin Li, Yunxiang Lu, Weiliang Zhu, The Stabilization Effect of Dielectric Constant and Acidic Amino Acids on Arginine–Arginine (Arg–Arg) Pairings: Database Survey and Computational Studies, **The Journal of Physical Chemistry B**, 2013, 117, 4827–4835
9. Qiang Shao, **Jinan Wang**, Weiliang Zhu, Addition of KI Salt into Urea Aqueous Solution Improves Its Denaturing Ability: A Molecular Dynamics Simulation Study, **Journal of Chemical Physics**, version submitted
10. Qiang Shao, **Jinan Wang**, Jiye Shi, Weiliang Zhu, The Universality of β -Hairpin Misfolding Indicated by Molecular Dynamics Simulations, **The Journal of Physical Chemistry B**, under review
11. **Jinan Wang**, Qiang Shao, Zhijian Xu, Yingtao Liu, Zhuo Yang, Kaixian Chen, Jiye Shi, Weiliang Zhu, Exploring Intrinsic or Substrate-induced Large-scale Protein Conformational Transition and the Associated Free Energy Landscape Using A Highly Efficient Computational Approach (in preparing)

致谢

时光飞逝，转眼间五年的博士生涯已接近尾声，值此论文完成之际，我要衷心感谢我的导师朱维良研究员。本论文工作的完成离不开朱老师的悉心指导和无私关怀！是朱老师的指导和帮助使我从刚进实验室的无知懵懂到现在具备了一些基本的科研素养，我的每一点进步都凝结着朱老师的心血，在这里向您表示衷心的感谢！在这 5 年里，无论是在课题还是生活上遇到困难时，朱老师总是想尽一切办法来帮我解决，在此特别感谢。朱老师严谨的科研作风，勤奋踏实的工作态度以及身体力行的科研精神永远是我学习的榜样，我为是朱老师的学生而感到荣幸。

在此要特别感谢邵强副研究员在课题方面给予的指导和论文写作方面的帮助，您细致耐心的指导使我受益匪浅，本论文的完成与邵老师的指导和帮助密不可分。

感谢蒋华良研究员、罗小民研究员、许叶春研究员、罗成研究员、于坤千研究员、阳怀宇副研究员、周芝萍老师等在学习和生活上对我的悉心指导和关怀帮助。

感谢卢运祥博士、石婷博士、李波博士、仓小慧博士、陈志博士、王雍博士、蔡海燕博士、王非博士、刘晓峰博士、李宁博士、徐志建博士、陈晓婕博士、梁中洁博士、杜芸博士、朱雪莲博士、毛艳艳博士、闫秀花师姐、卢军明、陆寅、白芳、杨卓、刘颖涛、陈筑熙、孔祥谦、刘娴、徐圆、李闪闪、熊修明、蔡婷婷、黄子福、刘丽丽、李婧、彭建龙、张鑫贲、、张勇、周元飞、朱建明、王玉兰、杨洋、陈照强等 DDDC 所有同学和工作人员在日常生活和科研工作中的热情帮助。

感谢同级同学芦静、沈倩诚、叶飞、欧阳斯盛、王媛媛和秦光荣的帮助和关心，我将终身难忘这一起走过的 5 年研究生生涯。

本论文的第二章工作是在大连化物所李国辉研究员课题组完成的,在此特别感谢李老师在课题上的指导和帮助,感谢山东交通学院的陈建中副教授在动力学模拟应用方面的交流和帮助。第四章的生物部分实验是英国Mike Merrick课题组完成的。在此向我课题的所有合作者表示由衷地感谢。

在此要特别感谢熊修明和王玉兰同学对本论文的校对。

我还要特别感谢研究生部林巍老师、张敏兰老师、王佩老师、胡仲良老师对我学习、科研和生活上的关心和帮助。

最后,我要深深感谢我的家人给我精神和物质上的无私支持和鼓励,谢谢你们给予我的一切,你们是我生活和前进的动力!

王进安

2013年5月于药物所