·“双清论坛”专题：理论化学家视角中的仪器创制·

理论与计算化学程序软件发展需求与资助模式思考
——基于美国自然科学基金委软件基础构架项目

李晓松*

（华盛顿大学化学系，西雅图 98124-6108）

[摘 要] 随着计算机计算能力和算法的高速发展，现有的计算软件已不能满足目前日益增长的需求。因此开发一种实用、高效的新型的计算软件结构就成为当务之急。该新型结构将具有支持多层理论方法的能力，同时可以兼顾实验和理论化学的结果，优化搜索引擎。新一代的计算化学软件必须具备高度的并行性、可扩展性、重复使用性和兼容性，并可实现大型社区驱动。本文基于美国自然科学基金委软件基础构架项目，阐述了计算化学软件发展需求和相关的资助模式，包括层列式的资助模式和资助范围。此外，本文还将从开发软件项目的特点等方面展开讨论，包括软件基础构架，软件培训和推广，以及其对软件开发项目可持续性的影响。

[关键词] 计算化学软件；持续创新的软件基础构架；软件设计和软件工程

计算科学的发展，很大程度上依赖于计算程序软件所提供的计算方法和分析手段。一款成功的计算软件的研制，需要从解决实际应用和自身的科研发展两个角度来选择对应的理论方法。计算软件的发展和相应的应用研究相辅相成、密不可分。通常一款成功的计算软件需要一个庞大应用研究团体的支持。而计算化学软件作为一个特殊的科研方向，是基于化学、物理、应用数学和计算机科学等多学科的交叉领域。因此，计算化学软件方向人才培养也具有其独特性。基于以上情况，美国国家自然科学基金委为计算化学软件的发展制定了一套完整的项目资助方案，称为可持续创新的软件基础架构（Software Infrastructure for Sustained Innovation [1]）。

目前，美国国家自然科学基金委多个学部已达成共识，共同承担对计算科学软件的资助基金。相关学部包括计算机和信息科学工程学部、生物学部、数理学部、教育和人力资源部、工程学部、地理科学部和社会科学部。各个学部对计算科学软件的发展均提出了不同的侧重点和资助方案要求，最终统一策划了以下三类资助方案：

1 科学软件要素，Scientific Software Elements

科学软件要素项目旨在资助在现有软件基础上开发的新的软件模块，或者是小型的独立软件。该类项目同时也资助新软件功能的早期设计原型和促进终端用户实践的模块。项目必须说明开发新软件要素的必要性，及其对科研发展的推动能力。

2 科学软件集成，Scientific Software Integration

科学软件集成项目旨在资助多领域的较大规模应用软件，用于解决某些科学和工程领域的常见问题。项目开发的软件必须具有可持续性，能为多个科研团体服务。

收稿日期：2017-10-10；修回日期：2017-10-29

* 通信作者，Email：xsli@uw.edu

3　科学软件创新中心,Scientific Software Innovation Institutes

科学软件创新中心旨在建立一个长期的支持科学软件发展的基地。此项目有两种资助方式:概念计划和中心建设。概念计划步骤旨在集成多个领域的研究团队,来论证各领域的软件需求和面临挑战。概念计划必须有对中心未来发展方向和发展模式的蓝图设计,要能反映出对中心建设步骤的投入和设计。只有成功的概念设计才有可能获得科学软件创新中心建设的资助。

所有科学软件项目必须注重软件结构的设计、持续发展性、适用性、易管理和软件模块的兼容性。辅助软件开发的环境,比如说源代码的贮藏和测试框架,都需要有针对于目标科学领域的设计。项目须综合考虑软件的推广方式,软件使用的培训方法和软件的社会服务职能。在适当情况下,美国国家自然科学基金委鼓励软件发展团队与工业界和政府合作。

科学软件开发项目,需要具体考虑以下主要问题和事项:

• 研发的科学软件主要解决的科研问题是什么? 如何促进某些科研领域的发展? 软件使用对象是谁? 如何在软件开发中融合科技创新?

• 对比现有的科学计算软件,包括商业和开源软件,阐述研发项目的特点和弥补的科学计算软件领域的空白。

• 讨论软件结构和软件工程,包括软件设计,开发,参考文件,测试,验证,软件发布,运用,与软件有关的终端用户的培训和客户服务,和软件评估。

• 讨论研发的科学软件如何从软件安全、可行度、重复性和可用性的角度来设计软件构架。

• 研发项目如何适应计算机硬件和软件的持续发展,尤其是在新型技术出现的时候。

• 讨论软件授权的具体模式,和选择该软件授权的原因。

• 项目计划要考虑如何和终端用户保持交流,吸取科研团队的意见来设计软件发展方向。

• 提供概念验证和里程碑的时间限。概念验证过程要有具体指标,包括软件用户的调查。这些指标将用来衡量软件的成功与否。

• 讨论软件开发和培训的结合方式,包括和软件使用直接相关的培训,和影响到其他领域的培训。

• 讨论软件和软件发展在项目结束后的持续性。

• 研发的软件如何能利用到现有的可兼容的软件模块和国家计算中心的资源。

这套资助方案的制定,旨在考虑计算化学软件的可持续发展。新一代软件基础架构将具有高度的适用性、可重复性和软件模块兼容性。在发展新型计算化学软件的同时,还要考虑其集成高性能计算、大规模并行处理、高速网络利用、大数据存储和处理等综合能力。此外,软件结构的设计还需考虑未来应用科学和计算方法的复杂性,软件开发人才和软件使用人才的培训也应统筹考虑。为了推动计算化学软件的创新和发展,美国国家自然科学基金委鼓励国内外高校与科研机构、政府、工业届等充分合作,在统一的软件授权下开发软件。

总之,计算化学软件的开发是一个长期、艰巨的工作,为保证其发展的可持续性,需要得到国家层面的稳定支持。纵观其发展历史,大部分的计算化学软件早期都是由科研小组在国家科学基金的资助下发展起来的。在此过程中,人才培养、人才队伍稳定及学科自身发展都得到了极大的推动和提升。特别需要指出的是,通过国家层面的稳定支持,经过持续的人力物力投入、协调和组织工作,可以促进各类人才、各研究组通力合作,是迅速推进设计先进、功能强大、独立完整的计算化学软件形成的重要手段。以上情况不仅适用于美国,也同样适用于中国的情况。中国目前有许多优秀的理论与计算化学家在从事各种计算方法的发展,因此,在国家层面上支持计算化学软件的建设就十分必要且刻不容缓[2]。对于计算化学软件建设持续有力的基金资助,将是中国理论与计算化学实现跨越式发展,引领国际前沿领域的有力支撑和保障。

参　考　文　献

[1] Software Infrastructure for Sustained Innovation. https://www.nsf.gov/pubs/2016/nsf16532/nsf16532.htm.

[2] 国家自然科学基金委员会, 中国科学院. 中国学科发展战略·理论与计算化学. 北京:科学出版社,2016, 714.

On the need and funding model for supporting theoretical and computational chemistry software development: a perspective based on the software infrastructure for sustained innovation program of the US National Science Foundation

Li Xiaosong

(*Department of Chemistry, University of Washington, Seattle* 98124-6108)

Abstract With recent advances in computational power and algorithmic efficiency, we cannot rely on incremental changes to existing computational chemistry software. A new research infrastructure is needed that smoothly incorporates multi-scale methods into a cohesive experimental and theoretical chemistry discovery, optimization and design engine. The next generation of computational chemistry software for treating realistic chemical systems with atomic resolution will need to be highly parallelized, extensible, reusable, interoperable and community-driven. To this end, this article aims to present a perspective on the need and related funding model for supporting computational software development based on the Software Infrastructure for Sustained Innovation Program of the US National Science Foundation. A tiered model and its respective funding scope are illustrated herein. Important considerations that focus on software infrastructure, education and dissemination, and their impact on the sustainability of a software development program are discussed in this article.

Key words computational chemistry software; software infrastructure for sustained innovation; software design and software engineering

·资料信息·

我国学者在 DNA 测序方法与技术上取得重要进展

在国家自然科学基金(项目批准号:21327808,21525521)等资助下,北京大学黄岩谊课题组日前在 DNA 测序方法与技术上取得重要进展,发展一种全新概念的测序方法—纠错编码测序法(简称 ECC),该方法采取一种独特的边合成边测序(SBS)策略,利用多轮测序过程中产生的简并序列间的信息冗余,大幅度增加了测序精度。研究成果于 2017 年 11 月 6 日发表在 *Nature Biotechnology*(《自然—生物技术》)期刊上。论文链接:http://www.nature.com/articles/nbt.3982。

序列信息的冗余来自黄岩谊团队新发展的"对偶碱基荧光发生"SBS 测序流程,该流程通过全新设计的特殊测序反应底物,对待测 DNA 序列进行三轮独立的 SBS 测序,继而产生三条互相正交的简并序列编码。这三条编码可互为校验,后续不但能够通过解码推导出真实碱基序列信息,而且具备对单轮测序错误位点的校正能力。这种编码和解码策略已被广泛应用在其它科学领域中,用于有效检测和纠正错误。此次黄岩谊团队在测序技术中首次引入冗余编码概念,通过和低错误率的荧光发生测序技术相结合,在实验室搭建的原理样机上获得了单端测序超过 200 碱基读长无错误的实验结果。在 ECC 测序中,黄岩谊团队首先从化学原理上对荧光发生测序技术中的荧光标记分子进行了结构优化,设计合成了具有不同波长、更优性能的测序底物分子,并对聚合酶参与的各阶段反应动力学进行了细致的测量和建模。在深入理解荧光发生测序化学反应速度、完成度、副反应等关键技术细节的基础上,构建了精确的测序信号失相模型并提出了次级延伸理论,并据此开发出算法软件对测序反应失相过程做出了合理简化使其具备实用性。

(供稿:化学科学部 庄乾坤 王勇 陈拥军)