

·“双清论坛”专题:理论化学家视角中的仪器创制·

关于自主知识产权计算化学软件发展的一些思考

闫亚明 包 鹏 史 强*

(中国科学院化学研究所,北京 100190)

[摘 要] 本文重点介绍笔者课题组关于发展具有我国自主知识产权的计算化学软件的一些思考,主要包括稳定的经费支持和人才培养,新开发的计算化学软件应具备的独创性优势等方面。

[关键词] 计算化学软件;稳定支持和人才培养;独创性优势

近年来,发展具有我国自主知识产权的计算化学软件的建议得到了广大理论计算化学同行和国家自然科学基金委(以下简称“基金委”)的重视。笔者认为,黎乐民先生等人关于建立“计算化学软件平台专项”的建议^[1]和基金委组织的双清论坛“理论化学家视角中的仪器创制”^[2]非常及时;我国近年来在理论与计算化学研究领域发展迅速,无论从研究队伍的规模、研究基础,还是相关领域经费投入的强度来看,发展国内计算化学软件平台不仅条件已经成熟,而且对于我国理论化学的长远发展十分必要^[1,2]。

我们很高兴地看到,“计算化学软件平台就是理论化学家从事科研工作的仪器”这一观点正逐渐得到基金管理部门和相关领域专家的普遍认同。正如双清论坛“理论化学家视角中的仪器创制”所指出的^[2],通用计算程序软件平台的建设是一项艰巨而复杂的系统工程。根据笔者课题组的工作经验和使用各种通用计算化学软件的体会,我们充分认识到以下几点的重要性:

(1) 长期稳定的经费支持和人才培养必不可少。以我们较熟悉的分子动力学模拟软件 Gromacs 和 NAMD 为例,Gromacs 软件包从 1991 年开始开发,在 2001 年 3.0 版本发布之后才得到广泛的重视。该软件包目前程序代码总共一百多万行,至今还在不断发展完善。其发展得到欧洲科学基金会(European Science Foundation, ERC)、荷兰与瑞典等国的研究机构,以及 NVIDIA 等公司的资助,根据代码行数粗略估计需要花费两千万美元^[3]。NAMD 主要是在 NIH 的长期支持下,由伊利诺伊大学并行计算实验室和 NIH

的理论与计算生物物理组合作开发^[4]。

一些大型的电子结构计算软件如 Gaussian 等,其不断发展和完善更是有几十年的历史。Gaussian 最早的版本 Gaussian 70 在 1970 年发布,只能用小基组进行 Hartree Fock 计算,总代码只有 1 万多行。随后,Gaussian 公司对其不断地改进,优化,增加新功能并开发配套的结果分析工具,先后共发布了 14 个版本,这才使其成为用户最多的电子结构计算软件之一^[5]。另一种适用于材料体系的电子结构计算软件 VASP,从 MIT 的 Mike Payne 开发的最初程序到功能强大的 VASP 5X 也是经历了长达 15 年的时间^[6]。因此,与仪器研制项目类似,这些大型软件的开发,离开长期稳定的支持是不可能的。

与稳定的经费支持相对应,核心骨干团队长期的专注和投入同样必不可少。不仅在计算软件发展的初期需要投入大量的精力,建立软件的整体框架和发展核心算法。随着新的理论方法和计算技术的出现,代码的维护和在新计算构架下重写和优化一样是十分繁重的任务。在人才培养方面,计算化学软件的开发不仅需要熟悉理论化学和当前的主流计算方法,还需要结合现代计算机技术的最新进展,才有可能开发出具有足够竞争力的计算化学软件。因此,如果期望在发展自主知识产权的计算化学软件方面有所作为,吸引和培养最优秀的人才开展相关领域的工作十分重要。

(2) 新开发的计算化学软件需要有独特的优势和竞争力,才能够快速发展,这既是挑战也是机遇。例如,Gromacs 在版本 3.0 后迅速得到大家的重视

收稿日期:2017-10-10;修回日期:2017-10-21

* 通信作者,E-mail:qshi@iccas.ac.cn

是因为首先在 Intel Pentium 构架上引入汇编优化,成为最快的分子动力学模拟软件^[7]。与同类计算程序相比,3—10 倍的计算速度优势,以及 Intel 平台相对低廉的价格,使得 Gromacs 的用户迅速增加。2000 年左右也是并行计算迅速发展的时代,程序的并行效率是决定计算速度的关键因素之一。相比于其它分子模拟软件的在原有代码基础上修改实现并行, NAMD 专门为分布式存储并行机开发;采用了空间分解算法,消息驱动调度,从而最大程度地提高并行效率^[8],使得其在大尺度并行模拟中占有重要优势,成为各大计算中心最重要的计算化学软件之一。

Gromacs 和 NAMD 的早期版本与当时主流的分子动力学模拟软件还有很大的差距。但是他们分别抓住了 Intel X86 平台以及大规模并行计算技术发展的机遇,从而取得了成功。目前,利用计算机硬件和软件技术的最新进展,如众核技术、机器学习等,新的计算化学软件开发应该有很多机会。当前基于图形处理器(GPU)加速计算的分子动力学模拟程序已经非常普遍。与之类似的是,基于 GPU 计算的电子结构软件 TeraChem。与其它电子结构软件的在原有 CPU 代码上修改实现 GPU 加速不同, TeraChem 的开发者首先开发了用 GPU 实现双电子积分、直接 SCF、解析梯度的算法,从而最充分地利用 GPU 加速计算^[9]。这些新的发展趋势值得我们在发展自主知识产权计算软件时注意。

(3) 优秀的计算化学软件需要一个友好的用户界面以及完备易用的计算结果分析工具。Gaussian 软件的使用者都会对其图形界面工具印象深刻, Gromacs 软件包提供了完善的结果分析工具,并支持用户开发各种扩展。NAMD 程序更是和优秀的生物分子可视化程序 VMD 紧密结合。这些对于吸引用户使用新开发的计算化学软件都非常重要。除了以上理论计算化学工作者熟悉的计算软件包之

外,一些用于材料和生物计算的商业软件如 Materials Studio、Shrödinger 等都在软件的界面和易用性上花了很大的功夫。这样有助于实验工作者使用这些软件,从而更好地发挥理论计算化学对相关学科的重要推动作用。

综上所述,开发一个强大的通用计算化学软件需要巨大的人力物力投入、合适的机遇,新开发的软件还需要具有突破性思想和独特的优势。目前,我国理论化学界的研究工作者已经发展出了许多重要的原创性计算方法,现代计算机硬件和软件方面很多新兴技术的出现也提供了新的发展机遇。因此,基金管理部门的重视和理论化学工作者的积极参与将成为我国原创计算化学软件开发的东风,相关工作的开展将会大大提升我国理论与计算化学的国际地位,并且促进青年人才的培养。

参 考 文 献

- [1] 黎乐民等. 中国科学发展战略:理论与计算化学. 北京:科学出版社,2016.
- [2] 高飞雪等. 第 181 期双清论坛“理论化学家视角中的仪器创制”在大连召开. <http://www.nsf.gov.cn/publish/portal0/tab434/info68850.htm>; 2017.
- [3] Gromacs website. <http://www.gromacs.org/>.
- [4] TCB website. <http://www.ks.uiuc.edu/Research/namd/>.
- [5] Gaussian website. <http://gaussian.com/>.
- [6] History of VASP. http://cms.mpi.univie.ac.at/vasp/vasp/History_VASP.html.
- [7] Lindahl E, Hess B, Spoel D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *Journal of molecular modeling*, 2001, 7(8): 306—317.
- [8] Nelson MT, Humphrey W, Gursoy A, et al. NAMD: a parallel, object-oriented molecular dynamics program. *International Journal of Supercomputer Applications and High Performance Computing*, 1996, 10(4): 251—268.
- [9] PetaChem website. <http://www.petachem.com/products.html>.

Some thoughts on the development of computational chemistry software with independent intellectual property

Yan Yaming Bao Peng Shi Qiang

(Institute of Chemistry Chinese, Academy of Sciences, Beijing 100190)

Abstract Some thoughts on the development of China's computational chemistry software with independent intellectual property rights are discussed, including the stable funding support and training of young scientists, as well as the unique features that a newly developed computational chemistry software should have.

Key words computational chemistry software; stable funding support and training of young scientists; original ideas and unique features