

Professor GU Xiaocheng and the Center for Bioinformatics at Peking University

罗静初*

Citation: [中国科学: 生命科学](#); doi: 10.1360/SSV-2021-0332

View online: <https://engine.scichina.com/doi/10.1360/SSV-2021-0332>

Published by the [《中国科学》杂志社](#)

Articles you may be interested in

[Celebrating 90 years of Peking University School of Life Sciences \(1925-2015\)](#)

SCIENCE CHINA Life Sciences **58**, 1055 (2015);

[Plasma parameter diagnosis using hydrogen emission spectra of a quartz-chamber 2.45 GHz ECRIS at Peking University](#)

SCIENCE CHINA Physics, Mechanics & Astronomy **61**, 045211 (2018);

[The Research Center for the History of Science and Technology Inaugurated at Nankai University](#)

Chinese Annals of History of Science and Technology **5**, 193 (2021);

[Progress of compact laser plasma accelerator in Peking University](#)

SCIENTIA SINICA Physica, Mechanica & Astronomica **43**, 1282 (2013);

[Vasera-screening of premature vascular aging and its correction in young persons: University Health Center experience](#)

BIO Web Conf. **22**, 01011 (2020);



顾孝诚教授与北京大学生物信息中心

罗静初*

北京大学生命科学学院, 北京大学生物信息中心, 北京 100871

* 联系人, E-mail: luojc@pku.edu.cn

收稿日期: 2022-03-10; 接受日期: 2022-05-01; 网络版发表日期: 2022-06-15

2012年, 北京的冬天似乎比以往来得更早一些. 11月3日, 西山香炉峰那斑斓的秋叶还没来得及飘落, 一场突如其来的秋雨和紧跟着悄然而至的初雪, 把古都美丽的秋天毫不留情地带走, 也送走了我们心目中一位可敬可爱的长者——顾孝诚教授(图1).

顾孝诚教授长者风范、为人师表、鞠躬尽瘁、死而后已. 她去世当晚, 大家难以抑制悲痛心情, 撰写了悼念文章, 构建了纪念网站.

顾孝诚教授1930年生于上海, 1948年就读于北京大学植物学系, 1952年毕业留校, 在北大整整工作六十年, 为北大生命科学学科建设、国家重点实验室创建、国家高技术研究发展计划(863计划)蛋白质工程项目实施等做出了重要贡献. 20世纪80年代, 顾孝诚教授和美籍华生物学家吴瑞教授一起, 通过CUSBEA项目, 选送了四百多名优秀生物学本科生去美国攻读博士学位, 其中不少已经成为生命科学领军人物, 有些已经回国.

1997年, 顾孝诚教授创建了北京大学生物信息中心(Center for Bioinformatics, CBI), 作为她多年的学生和同事, 我愿把当年顾孝诚教授创建CBI的点点滴滴从沉睡的记忆中梳理出来, 与大家特别是年轻读者共勉.

1 建立CBI的基础

20世纪70年代末, 我从北大校办工厂回到生物系,



图1 2010年, 顾孝诚教授八十岁生日

Figure 1 Professor GU Xiaocheng, on her 80th birthday party in 2010

与顾孝诚教授同在生物物理教研室. 1986年秋, 美国亚利桑那州立大学戴伟·蒙特(David Mount)教授应邀来北大举办基因工程讲习班, 带来了他编写的序列分析软件DM, 可在个人电脑PC机上运行. 讲习班结束后, 蒙特教授给了我一盒软盘, 除了程序DM外, 还存有核酸和蛋白质序列数据库. 从此, 我进入了计算机在分子生物学中应用领域.

1986年, 863计划实施, 顾孝诚教授任生物技术领域专家兼蛋白质工程主题组长. 1987年, 北京大学蛋白质工程和植物基因工程国家重点实验室成立, 顾孝诚教授任实验室主任. 我也有幸加入重点实验室, 帮助

引用格式: 罗静初. 顾孝诚教授与北京大学生物信息中心. 中国科学: 生命科学, 2022, 52
Luo J C. Professor GU Xiaocheng and the Center for Bioinformatics at Peking University (in Chinese). Sci Sin Vitae, 2022, 52, doi: [10.1360/SSV-2021-0332](https://doi.org/10.1360/SSV-2021-0332)

各课题组将计算机用于课题研究,特别是蛋白质结构分析、比较和预测。重点实验室购置了一台蛋白质三维结构显示分析专用设备,放置在北京大学图书馆机房。在顾孝诚教授的努力下,北大图书馆给了我们很大支持。

1991年初,顾孝诚教授访问英国伦敦癌症研究基金会(Imperial Cancer Research Fund, ICRF),回国后第一时间找到我,说ICRF有个生物大分子建模实验室,专门从事蛋白质结构分析和预测研究,打算从863申请经费,让我去那进修。

1991年夏,我去英国进行为期六个月的访问。刚到伦敦当晚,住在驻英使馆教育处。第二天,实验室主任迈克·斯登伯格(Michael Sternberg)开车接我。聊了几分钟后,笑着对我说:“你的英语还不错”。自知英语不行,就赶紧实话实说:“和顾老师相比,我的英语水平差得太远”。听到这里,迈克哈哈大笑,并告诉了我一个有趣的故事。原来,年初顾孝诚教授和另一位专家访问ICRF,迈克介绍实验室研究方向时,怕来访者听不懂,特意讲得很慢,一字一句,有时还重复一遍。没想到刚介绍完,顾孝诚教授以极其纯正的英国英语、极其标准的语速,提出了极其专业的学术问题。迈克听了大吃一惊,脸刷地一下红了。“To my big surprise, she talked beautifully with typical British English. I was in a shame!”顾孝诚教授幼年时,家里请英国人当家庭教师。她专业词汇丰富,听说读写能力极强。改革开放初期,国外专家(其中包括不少诺贝尔奖获得者)来访做报告时,好多单位专门请她做同声翻译。

这次进修访问,初步掌握了蛋白质结构分析、预测的基本方法,成功预测了癌胚抗原三维结构,发表了相关文章。回国后重点实验室购置了图形工作站和结构分析软件,和各课题组合作,进行了胰蛋白酶、金属硫蛋白、胰岛素等结构分析和预测。

为推动国内蛋白质工程、蛋白质结构预测和蛋白质分子设计等领域的发展,获863计划支持,1994年暑期,顾孝诚教授和Chris Sander博士在北京共同主持了“中欧蛋白质工程和分子设计”讨论班。Chris时任欧洲分子生物学实验室(European Molecular Biology Laboratory, EMBL)计算生物学课题组长,20世纪90年代初构建了蛋白质二级结构DSSP、蛋白质家族FSSP和同源蛋白分类HSSP等一系列数据库。他会说多国语

言,包括简单的中文,并给自己起了个中文名“柯三达”。会后,他特地来北大,考察了当时北大的网络环境和计算资源。作为教育网华北地区主节点,北大校园网已经建立,并接入国际互联网。

2 建立CBI的机遇

第二年,我和顾孝诚教授去英国巴斯参加“基因组、蛋白组和生物信息学”学术会议。与会期间,专门和Chris探讨了在北大建立生物信息中心的可能性。他建议我们申请加入欧洲分子生物学网络组织(European Molecular Biology Network, EMBNet),以得到该组织的帮助^[1]。

EMBNet成立于1988年,Chris是创始人之一,主要成员为欧洲各国从事计算机在分子生物学中应用的科研院所和高等院校,其宗旨是为国内用户提供分子生物学资源服务和培训。20世纪90年代中期,EMBNet成员达到28个,包括英国、德国、荷兰、瑞士、西班牙等西欧国家,以及波兰、斯洛伐克和匈牙利等东欧国家。

顾孝诚教授于1995~2004年任国际遗传工程和生物技术中心(International Center for Genetic Engineering and Biotechnology, ICGEB)科学顾问委员会委员和主任。该中心由联合国教科文组织创办,总部设在意大利。1996年9月,顾孝诚教授在ICGEB开会时,拜访了该中心蛋白质结构研究组Sandor Pongor博士,而Sandor恰任EMBNet主席。顾孝诚教授表达了希望加入EMBNet的意愿,Sandor十分赞同,建议我们在10月召开的EMBNet年会上提出申请。

10月初顾孝诚教授回国后,立即把这消息告诉了我,争取尽快加入EMBNet。按规定,申请加入EMBNet必须得到所在国政府部门批准。为此,顾孝诚教授带我一起到教育部科技司,向袁成琛司长汇报。不料那天袁司长外出开会,等了一上午也没等到。直到下午,工作人员才和袁司长联系上。袁司长对顾孝诚教授的工作和为人十分了解,二话没说,就批准了我们的申请报告。

拿到批件后,我们立即给Sandor发传真。按规定,凡申请加入EMBNet,必须由申请者到会做报告,出具政府部门批件,承诺为本国提供生物信息资源服务和人才培养。当时离EMBNet年会召开不到10天,申办签证已经来不及。出于对顾孝诚教授的信任,也基于

Chris在北大做了实地考察, 破例由Sandor代表我们发言, 与会成员投票通过。

为实现我们对EMBNet的承诺, 更好地为国内用户提供服务, 1997年3月, 依托蛋白质工程和植物基因工程重点实验室, 并得到北大计算中心兼教育网华北地区主节点张兴华主任、北大图书馆馆长兼中国高等教育文献保障信息系统主任戴龙基教授支持, 我们成立了北京大学生物信息中心。

EMBNet年会每年举行一次, 由成员单位轮流举办。1997年9月, 我赴意大利参加EMBNet年会, 结交了许多朋友, 如EMBOSS软件包^[2]开发者Peter Rice和Alan Bleasby、数据库检索系统SRS开发者Thure Etzold、荷兰EMBNet负责人Jack Leunissen等。

1998年4月, 我们举办了首届生物信息培训班, 邀请了五位EMBNet专家担任培训教员, 100多名学员参加培训。1998年10月, EMBNet十周年年会在英国召开, 应主席Peter Rice邀请, 播放了讲习班实况录像, 给与会者留下了深刻印象。

1999年4月, 我们举办了第二届生物信息讲习班, 十名EMBNet同事应邀做报告或讲课。参加这次讲习班的学员共200多名, 除国内学员外, 泰国和埃及各派一名代表参加。863计划生物领域专家赵国屏教授、国家教育和科研计算机网学术委员会李星主任在会上发表了热情洋溢的讲话。

除报告会外, 还举行了为期一周的EMBOSS和SRS两个软件包上机操作培训。中国科学院理论物理研究所郝柏林院士当年已65岁, 以学生身份全程参加了这次讲习班。之后不久, 撰写了“尽快组建国家级的生物医学信息中心”的院士建议, 刊登在《国家自然科学基金委员会简报》上。此后几年, 我们先后举办了多次生物信息讲习班、培训班, 为国内生命科学和生物技术领域广大用户进行了数据库检索、序列分析、蛋白质结构预测等基本的生物信息技术培训。2000~2005年, 郝柏林院士应聘任CBI兼职教授, 指导两位博士研究生进行禾本科植物基因组复制研究, 解决了国际上关于水稻中是否存在基因组水平复制的争端^[3]。

3 CBI成立初期的困难和顾孝诚教授的努力

3.1 计算资源问题

CBI成立初期, 首先面临的是计算资源问题。当时

蛋白质工程和植物基因工程重点实验室仅有两台用于蛋白质结构分析的SGI图形工作站, 根本无法满足核酸和蛋白质序列数据库检索、BLAST数据库搜索等基本生物信息服务的需求。为此, 顾孝诚教授多次找北大计算中心张兴华主任和图书馆戴龙基馆长求助。看到已退休多年的顾孝诚教授为CBI建设跑上跑下、亲力亲为, 张兴华主任和戴龙基馆长深受感动, 就把他们新购置的SUN服务器给我们使用, 安装在计算中心和图书馆机房。计算中心安排了李笑难和滕小玲等, 图书馆安排了胡杰和杜晓峰等技术人员负责硬件维护和系统管理。

3.2 数据资源问题

初步解决计算机硬件后, 接下来的问题是数据资源问题。虽然当时国内高等院校和科研单位都已联网, 但与国际互联网之间的带宽十分有限。教育网规定, 国内高等院校浏览国际网站产生的网络流量费用由用户承担, 很大程度上限制了高校师生充分利用国际生物信息中心的资源。为此, 顾孝诚教授又去找张兴华老师。经张老师与教育网李星教授商量, 在网络IP地址十分紧缺的情况下, 决定给我们分配十个专用IP地址, 用于从国际生物信息中心下载数据, 存放到我们的服务器上, 供国内高校和科研院所免费下载。

3.3 网络带宽问题

为缓解国内用户访问国外数据库网速慢的问题, 建立国外数据库镜像不失为一个立竿见影的办法。1999年底, 人类基因组计划构建的第一个数据库GDB负责人Jamie Cuticchia和技术主管朱伟民访问北大, 举办了人类基因组计划报告会和基因组数据库讲习班。朱伟民后任欧洲生物信息学研究所(European Bioinformatics Institute, EBI)数据库系统部门负责人, 曾多次访问CBI, 和顾孝诚教授商讨回国计划。为此, 顾孝诚教授专门带他去拜访国家自然科学基金委。2011年他全职回国, 受基金委委托, 和多位专家一起起草了建立国家生物信息中心的可行性报告。

基因组数据库讲习班结束后, 我们安装了GDB镜像。之后, 又安装了大鼠基因组数据库RGD和瑞士蛋白质分析专家系统ExPASy等镜像, 为国内用户提供快速高效的生物信息数据资源和软件工具服务。安装镜像, 不仅需要解决计算机系统、软件工具等技术问题,

还涉及知识产权、网络带宽等诸多因素, 绝不仅仅是“拷贝数据”那么简单。2003年, 在时任美国国家生物技术信息中心(National Center for Biotechnology Information, NCBI)主任David Lipman的支持下, 我们曾安装了生物医学全文检索系统PubMed Central(PMC), 后因国外出版商的反对而不得不终止。

3.4 经费问题

生物信息中心的建设, 还需要有经费支持。2000年, 赵国屏教授担任蛋白质工程主题专家组组长, 对我们的工作给予了很大支持, 经过努力, 促成专项经费向北大生物信息中心拨付150万专款, 作为我们的运行经费。2001年, 新一轮863计划启动, 李亦学任生物领域专家兼生物信息主题专家组组长。北大生物信息中心、中国科学院上海生命科学研究院生物信息中心和华大基因合作承担了“国家生物信息基地建设”项目。

4 CBI创建初期的工作

在顾孝诚教授的带领和指导下, 在CBI全体师生的努力下, 在EMBNET同事帮助下, 我们安装了数据库查询系统SRS, 序列相似性数据库搜索系统BLAST、序列分析软件包EMBOSS等常用生物信息分析平台^[4~6], 开发了基于桌面电脑和Linux系统的便携式数据库查询系统^[7], 并先后于陈竺院士在上海召开的基因组大会和郝柏林院士在厦门举办的“海峡两岸生物启发的理论科学研讨会”期间举办讲习班, 向与会人员推广该系统的安装和使用。

除了安装常用生物信息软件外, 我们也自主开发了生物信息网络实验室WebLab、生物信息网格ABC-Grid、基因组浏览器ABrowser、代谢通路分析系统KOBAS、基因结构显示系统GSDS等分享平台。构建了分泌蛋白数据库SPD、植物转录因子数据库PlantTFDB、叶片衰老数据库LSD、拟南芥激素数据库AHD、成瘾相关基因数据库KARG、自闭症基因知识库AutismKB等。

2003年非典爆发初期, 在NCBI鲍一明博士帮助下, 我们建立了抗击非典生物信息资源网站, 开展了SARS病毒生物信息学分析研究^[8]。四月底的一个周末, 顾孝诚教授来实验室看望坚持工作的同学。考虑

到通风问题, 就带大家一起去未名湖边讨论。走到离办公楼不远处的小山坡, 看见树丛下一个不显眼的木门, 门背后是个防空洞, 就风趣地告诉我们, 当年这里是燕京大学进步青年抗日联络点, 今天成了我们抗击非典的联络点。

CBI成立初期, 仅有两位课题组长、两位讲师和一位实验室管理人员。2004年, 魏丽萍教授从美国全职回国, 任中心主任, 从事儿童自闭症遗传机制研究。同年, 芝加哥大学龙漫远教授任教育部长江客座教授, 指导CBI研究生从事果蝇和水稻中新基因起源研究。

CBI是教育部最早批准成立的生物信息学博士点, 上述分析平台的开发和数据库的构建, 绝大部分由博士研究生完成, 仅1999~2002级研究生就有李兵、陈新、陈蕴佳、何坤、刘翟、吴昕、高歌、孔雷、吴建民、张勇、郭安源、孙颖等参与了上述工作。

5 北京基因组研究所生命与健康大数据中心建立

CBI自1997年成立, 至今已有20多年。回顾中心初创时期顾孝诚教授所做的不懈努力, 不禁对老一辈科学家为我国生物信息事业呕心沥血的精神深受感动。由于种种原因, 近几年来, 生物信息资源服务不再是CBI主要任务。

2015年, 中国科学院北京基因组研究所生命与健康大数据中心(BIG DataCenter, BIGD)建立, 接下了生物信息资源服务的接力棒^[9]。2017年, 鲍一明博士从NCBI全职回国, 担任大数据中心主任, 与章张、赵文明副主任一起带领年轻的团队, 发展完善了多组学数据资源体系, 2019年成为国家基因组科学数据中心; 新冠疫情爆发以来, 建立了新冠病毒信息库。

2019年, 北京基因组研究所加挂“国家生物信息中心”牌子, 已经成为国际上除NCBI和EBI之外的主要生物信息数据资源中心^[10], 初步实现了顾孝诚教授、郝柏林院士等老一辈科学家建立我国生物信息中心的愿望。2021年10月, 中国科学院北京基因组研究所(国家生物信息中心)成为EMBNET中国节点, 鲍一明博士任节点负责人。

为纪念顾孝诚教授对我国生物信息事业的贡献, 顾孝诚教授纪念网站已部署在国家生物信息中心国家基因组科学数据中心。

6 附表

顾孝诚教授生平:

1930.2 生于上海

1948~1952 北京大学植物系就读, 毕业后留校工作

1981~1989 任CUSBEA项目中方协调员

1985~1992 任北京大学生物系主任

1986~1992 任863蛋白质工程主题专家组长

1987~1994 任北京大学蛋白质工程和植物基因工程重点实验室主任

1994~2004 任北京大学蛋白质工程和植物基因工程重点实验室学术委员会主任

1995~2004 任国际遗传工程及生物技术中心(IC-GEB)学术委员会委员、主任

2012.11 因病在北京去世

北京大学生物信息中心大事年表:

1986 863计划开始实施

1987 北京大学蛋白质工程和植物基因工程重点实验室成立

1991 顾孝诚教授访问英国伦敦癌症研究基金会

1994 顾孝诚教授和Chris Sander在北京举办中欧蛋白质工程讲习班

1995 顾孝诚教授在英国参加国际会议期间与Chris Sander商讨建立CBI

1996 北京大学蛋白质工程和植物基因工程重点实验室加入EMBnet

1997 CBI成立, 依托单位为北京大学蛋白质工程和植物基因工程重点实验室

1998 CBI举办首届生物信息培训班

1999 CBI举办国际生物信息讲习班, 郝柏林院士以学生身份全程参加

1999 郝柏林院士提出“建议尽快组建国家级的生物医学信息中心”的院士建议

1999 CBI举办基因组数据库GDB讲习班, 朱伟民参加

2000 CBI获863专项经费资助

2001 CBI和上海生物信息中心、华大基因合作承担863“国家生物信息基地建设”项目

2004 魏丽萍教授任CBI主任, 龙漫远教授任CBI长江客座教授

相关网站:

顾孝诚教授纪念网站 <https://ngdc.cnbc.ac.cn/education/biography/guxc/>

CUSBEA项目和吴瑞基金会 <https://www.raywu-memorialfund.org/wp/>

北京大学生物信息中心 <http://cbi.pku.edu.cn/>

北京大学蛋白质工程和植物基因工程国家重点实验室 <http://www.pepge.pku.edu.cn/>

北京大学计算中心 <https://cc.pku.edu.cn/>

中国高等教育文献保障信息系统 <http://www.calis.edu.cn/>

国家教育和科研计算机网 <https://www.cernet.edu.cn/>

中国科学院北京基因组研究所生命与健康大数据中心 <https://ngdc.cnbc.ac.cn/>

中国科学院北京基因组研究所新冠病毒信息库 <https://ngdc.cnbc.ac.cn/ncov/>

美国国家生物信息中心 <https://www.ncbi.nlm.nih.gov/>

欧洲生物信息学研究所 <https://www.ebi.ac.uk/>

瑞士蛋白质分析专家系统ExPASy <https://www.expasy.org/>

欧洲分子生物学网络组织 <https://www.embnnet.org/>

国际遗传工程和生物技术中心 <https://www.icgeb.org/>

致谢 感谢北京大学蛋白质和植物基因研究国家重点实验室多年来的支持, 感谢朱玉贤院士对本文写作的鼓励, 感谢北京基因组研究所鲍一明研究员、上海生物信息中心赵国屏院士和李亦学研究员、北京大学生物信息中心高歌博士、郭安源博士和徐礼鸣博士对本文的修改意见。感谢北京基因组所章张研究员、降帅博士和邹东高工部署和维护顾孝诚教授纪念网站。

参考文献

- 1 Luo J C. EMBOSS and EMBnet (in Chinese). *Chin J Bioinf*, 2021, 19: 213–231 [罗静初. EMBOSS和EMBnet. *生物信息学*, 2021, 19: 213–231]
- 2 Luo J C. Application examples of EMBOSS sequence analysis program (in Chinese). *Chin J Bioinf*, 2021, 19: 1–25 [罗静初. EMBOSS软件包序列分析程序应用实例. *生物信息学*, 2021, 19: 1–25]
- 3 Paterson A H, Bowers J E, Van de Peer Y, et al. Ancient duplication of cereal genomes. *New Phytol*, 2005, 165: 658–661
- 4 Huang Y, Luo J C, Gu X C. The application and re-development of bioinformatics resources (in Chinese). *Chin High Technol Lett*, 1999, 1: 60–62+34 [黄弋, 罗静初, 顾孝诚. 生物信息资源的应用与二次开发. *高技术通讯*, 1999, 1: 60–62+34]
- 5 Li B, Luo J C, Pan W, et al. The development and utilization of molecular biology databases and related softwares (in Chinese). *Hereditas*, 1999, 21: 52–53 [李兵, 罗静初, 潘卫, 等. 分子生物学数据库及相关软件的开发利用. *遗传*, 1999, 21: 52–53]
- 6 Luo J. Bioinformatics service, education and research: the EMBnet and CBI. *Silico Biol*, 2002, 2: 173–177
- 7 Luo J C. 基于PC/Linux的分子生物信息数据库查询系统 (in Chinese). *Chin Sci Bull*, 2000, 45: 1006–1008 [罗静初. 基于PC/Linux的分子生物信息数据库查询系统. *科学通报*, 2000, 45: 1006–1008]
- 8 Chen Y J, Gao G, Bao Y M, et al. Initial analysis of complete genome sequences of SARS coronavirus (in Chinese). *J Genetics Genomics*, 2003, 30: 494–500 [陈蕴佳, 高歌, 鲍一明, 等. SARS冠状病毒全基因组序列初步分析. *遗传学报*, 2003, 30: 494–500]
- 9 Xue Y, Bao Y, Zhang Z, et al. Database resources of the National Genomics Data Center, China National Center for Bioinformation in 2021. *Nucleic Acids Res*, 2021, 49: D18–D28
- 10 Rigden D J, Fernández X M. The 2021 *Nucleic Acids Research* database issue and the online molecular biology database collection. *Nucleic Acids Res*, 2021, 49: D1–D9

Professor GU Xiaocheng and the Center for Bioinformatics at Peking University

LUO JingChu

College of Life Science and Center for Bioinformatics, Peking University, Beijing 100871

doi: [10.1360/SSV-2021-0332](https://doi.org/10.1360/SSV-2021-0332)