REVIEW

Computational opportunities for remote collaboration and capacity building afforded by Web 2.0 and cloud computing

Willy Wriggers · Wilma K. Olson · Cristobal G. dos Remedios

Abstract In this paper, we state our aims and aspirations for building a global network of likeminded people interested in developing and encouraging students in the field of computational biophysics (CB). Global capacity building efforts have uncovered local computational talent in virtually every community regardless of where the students reside. Our vision is to discover and encourage these aspiring investigators by suggesting ways that they and other "garage scientists" can participate in new science even if they have no access to sophisticated research infrastructure. We argue that participatory computing in the "cloud" is particularly suitable for CB and available to any budding computational biophysicist if he or she is provided with open-minded mentors who have the necessary skills and generosity. We recognize that there are barriers to the development of such remote collaborations, and we discuss possible pathways to overcome these barriers. We point out that this Special Issue of *Biophysical Reviews* provides a much-needed forum for the development of several specific applications of CB.

Keywords Computational biophysics · White paper · Cloud computing · Capacity building · Garage science · Global network

Special Issue: Computational Biophysics

W. Wriggers
Department of Physiology and Biophysics and Institute for Computational Biomedicine,
Weill Medical College of Cornell University,
1300 York Ave,
New York, NY 10065, USA

W. K. Olson
Department of Chemistry and Chemical Biology,
BioMaPS Institute for Quantitative Biology,
Wright-Rieman Laboratories, Rutgers University,
610 Taylor Road,
Piscataway, NJ 08854-8087, USA
e-mail: wilma.olson@rutgers.edu

C. G. dos Remedios
Discipline of Anatomy and Histology & School of Medical Sciences, Bosch Institute (F13), University of Sydney,
Sydney 2006, Australia
e-mail: cris.dosremedios@sydney.edu.au

Present Address:
W. Wriggers (✉)
D. E. Shaw Research,
120 West 45th Street,
New York, NY 10036, USA
e-mail: wriggers@biomachina.org

Introduction

We began with the aim of writing a white paper on how to attract new, global talent to the field of Computational Biophysics (CB). The goal of this initiative is to discover talent in the field of CB regardless of where the talent resides, and then to introduce that talent to willing mentors using current global communications and computing techniques. We also offer some experience, and a willingness to help the students and mentors form the proposed network.

We note that CB, just like its parent discipline biophysics, is inherently an interdisciplinary science with multiple branches spanning all levels of biological organization. We define CB broadly as the study of biologically significant problems using a range of quantitative methods from small to large scales, such as ab initio analysis of molecular structure, molecular dynamics and mechanics, molecular modeling, molecular graphics, and even data mining. These computational techniques can be applied to a range of macromolecules and macromolecular systems as well as to cells, entire organisms, and their populations.

CB as a term is not widely established, but CB techniques have been used extensively in the literature (e.g., Craddock et al. 2012; Anisimov and Cavasotto 2011; Qin et al. 2009;

Phillips et al. 2005, to mention just a few). There are even graduate courses specifically devoted to CB. A CB course outline (MIT BioEECS 2012) at Massachusetts Institute of Technology states: "No single approach fully characterizes the research that falls into this area as the methods we employ are often problem dependent. Nevertheless, in many cases numerical techniques provide a platform that is used to gain insights into difficult biological problems." There have also been meetings specifically devoted to CB (Sagot et al. 2009), and there are NIH-funded training programs that provide educational certificates issued in the field of structural and computational biophysics (Poole 2011).

The remainder of the paper fleshes out our proposal for building a global network of researchers interested in developing and encouraging students in CB. Background and mission describes the opportunities afforded by global communications and computing techniques. Marketing computational biophysics emphasizes the need of marketing (or branding) CB as a new discipline. Application examples shows two application areas, software development and data mining, which we argue are particularly suitable for capacity building efforts. Effective capacity building discusses selection mechanisms for recruiting talent. Regularly scheduled retreats or conferences and Funding sources cover organizational topics such as the planning of meetings and fundraising. Open questions concludes the white paper with a discussion of open questions.

Background and mission

Global communication technology and remote computational services promise to eliminate geographic barriers through participatory information sharing. These "Web 2.0" and "cloud computing" techniques enable unconventional scientific collaborations which no longer require a physical laboratory. At the same time, there is a desire by many independently supported "citizen scientists" (and traditional graduate students without ready access to a university-level research infrastructure) to participate in bona fide scientific discovery. This fledging movement was termed "garage science" in a recent series of articles in the journal Nature (Ledford 2010).

Garage science, citizen science, crowd sourcing, and similar fashionable terms are used to describe volunteer work that has been accepted practice in many fields for more than a century (Cook 2011). For example, amateur astronomers have long participated in scientific studies led by professional astrophysicists (Astronomical Society of the Pacific 2012). The new power of the internet can be exemplified by the success of crowd-sourcing projects such as Wikipedia, and the fact that the "garage" movement has now reached the biological sciences (Khatib et al. 2011).

We believe the convergence of communication services in the cloud with the desire by individuals to donate their private time (and possibly CPU cycles) to assist research, provides new and unique opportunities for computational advances and training in biophysics. It is likely there is a large and untapped target group of scholars who may be interested in learning about and contributing to biophysics, and who should be able and willing to take advantage of computational services that can now be provided (Brabham 2012).

Although there are still some open questions as to which types of activities may prove ultimately fruitful (see Open questions below), we feel that the time is right for large organizations to take action. The International Union for Pure and Applied Biophysics (IUPAB) is a major force behind this Special Issue of Biophysical Reviews devoted to CB. There is a realistic opportunity for IUPAB and the family of International Scientific Unions (ICSU) to contribute to the development of a global student–mentor network that can be a great benefit to the global scientific community. The students in these networks would be talented, quantitative scholars of all ages and from any country, who wish to add research expertise and credentials to their resume, and who want to know how they can participate in biophysical research. The role of the mentors would be to find an accessible way to expand their personnel base with remote students and research partners. Depending on the project, it may not be necessary that the mentors bring significant hardware resources to the table, so our focus here is on human interactions rather than technical questions or development of a distributed computing grid.

The main mission of this contribution is to explore the use of CB as a global participatory framework that can be accessed by graduate students, particularly those from developing countries. We would consider ourselves successful if our efforts catalyzed meaningful interactions between these students and appropriate mentors, who produce, in turn, new educational experiences and research breakthroughs.

Marketing computational biophysics

The marketing (or branding) of CB will be critical if we wish to attract both students and mentors. Due to its interdisciplinary nature, CB can be characterized as a life-style choice that emerges from a general scientific vocation in the hard mathematical and scientific disciplines. Initial attraction to CB is often less motivated by direct biological or medical questions than by an interest in computation and mathematical or theoretical questions. Consequently, there are some possible advantages and limitations of CB that may affect the motivation and satisfaction of the collaborating researchers.

Pros:

- CB is at the interface of many disciplines (computer science, statistics, applied math, biology, physics, and chemistry). It is very rewarding from a learning perspective since it touches upon a wide range of intellectually stimulating concepts.
- It is great fun to work on a computer (satisfaction of creative work, building tools, software, and problem solving).
- A computational career is less risky than an experimental career; for example, CB skills are readily transferrable to other disciplines.
- There are opportunities to engage in global interactions and travel.

Cons:

- There are more lucrative applications of computer technology, at Silicon Valley companies and in finance, for example. Then again, most scientists prefer satisfaction over personal wealth, and it is fun to learn about and contribute to science. Ultimately, volunteering time for science is a personal decision.
- In the biological sciences, purely dry modeling and simulation only rarely lead to significant discoveries. Frequently, CB researchers engage in collaborations with experimentalists by solving software engineering projects or by assisting experimental scientists in the interpretation of their data (Application examples). The high regard placed in biophysics on wet laboratory data collection can be a humbling experience for researchers from quantitative disciplines who are accustomed to a historically more dominant role of theory and computation.
- The quantitatively most inspiring research areas in CB are often well populated by theorists attracted to them, and they are consequently very competitive in terms of funding or positions. The market eventually forces many CB researchers to find their niche outside the original gamut of research interests.

The challenges facing CB are not uncommon in science and affect mainly those researchers with academic career aspirations, but they do not weigh in heavily if the project is primarily seen as an educational experience. In Open questions, we propose additional motivating factors for students and mentors that may make the experience worthwhile.

Application examples

The proposed remote collaborations should not necessarily require specific high-performance computing hardware. We

wish to empower outstanding graduate students and researchers based on personal abilities and interests, regardless of their geographic location and the specific infrastructure available to them. In the following, we describe two fields of CB where, in the absence of expensive laboratories with high-end infrastructure, skilled researchers can still make a significant contribution.

- Software development. Computer programs can be easily written on a personal computer or laptop, provided that there is a central repository for version sharing and control among team members (GNU Savannah 2012; Google 2012). One successful example of software development in a remote collaboration framework is described by Starosolski et al. (2012) in this issue. The paper describes specific Web 2.0 services employed by that group in the development of the Sculptor 3D modeling and visualization software illustrated in Fig. 1. The collaboration involves a core group of four contributors who regularly volunteer their free time. Due to the absence of organizational overhead, the operation of such a virtual laboratory can be very efficient (Rusu and Wriggers 2012; Rusu et al. 2012). The glue that holds the team together in this case is the mutual desire to continue with work that was incubated in a traditional laboratory. However, we believe that one can foster a similar sense of community among any researchers who share a common goal (see Regularly scheduled retreats or conferences). For example, the 3DNA suite of programs for the analysis, construction, and visualization of nucleic acid structures (Lu and Olson 2003, 2008) supports an online community of scholars interested in DNA/RNA structural bioinformatics through its forum pages (Lu 2012).
- Bioinformatics data mining. A wealth of data available from gene arrays and in proteomics databases invites mining approaches for developing new hypotheses and discovering new connections (Wang et al. 2003; Clarke et al. 2012). Such contributions depend heavily on good primary data (e.g., gene array raw data that have not been "normalized" or filtered). These data are now freely accessible over the internet. In the Gene Expression Omnibus (GEO) database (National Center for Biotechnology Information 2012), for example, expressed genes are assembled into relational databases that can be combined and re-analyzed for purposes not envisaged by the researchers who reported those data (Fig. 2). Proteomics databases are not nearly as well developed as gene databases, but they are arguably more important. Perhaps only about half the altered genes result in corresponding changes at the protein level. Insertions, deletions, splicing variants, and other mechanisms result in many more proteins than the available 20,000 or so genes in man. Proteomics
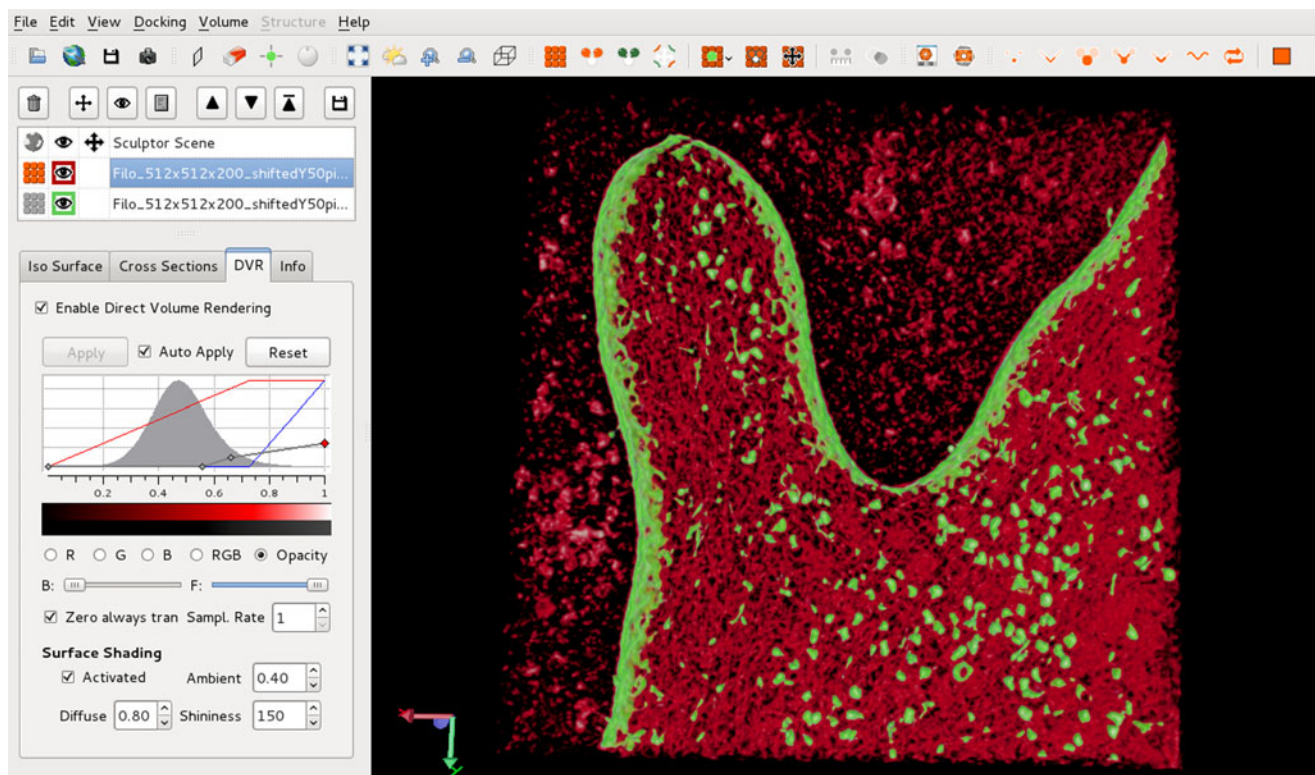
Fig. 1 Screenshot of the Sculptor visualization and modeling software. The image shows the volume rendering of actin filaments and the extracellular matrix (*red* densities) and of non-filamentous densities (membrane, ribosomes, and storage vesicles: *green*) of a 3D tomogram of a *Dictyostelium discoideum* filopodium (Rusu et al. 2012). The graphic was kindly created by Mirabela Rusu, one of the remotely collaborating authors of the Sculptor program

databases are similar to their gene "cousins" in that they produce the raw data, assembled data, and metadata. The raw data files are often obtained by mass spectrometry and, although they can be re-analyzed, it remains to be seen whether repositories of these data (Tranche Developers Group 2012) are viable (Poole 2011). At the next two levels, data are assembled by researchers into peptides and proteins (equivalent to nucleotide sequences) and developed into metadata. The proteomics equivalent of GEO is likely to move or take different forms, depending on future funding opportunities. In the meantime, there are ample data on intact known proteins, which can be freely downloaded through ExPASy (Swiss Institute of Bioinformatics 2012).

Effective capacity building

There are divergent aims and strategies for training the next generation of scientists. One must balance altruism [e.g., trying to help members of disadvantaged minority groups (MentorNet 2012) or students in the developing world] with the interest of mentors to be efficient with their time and to excel in their research; mentors are held accountable by their peers and need to be efficient with their mentorship investment.

Due to this tradeoff, we anticipate a need for a selection mechanism for admitting students to a remote mentorship program. A uniform set of standards should be applied in the selection. Those accepted into the program should have an understanding of calculus and matrix algebra, proficiency in at least one computer language, and a general knowledge of chemistry and physics. Thus, students in different countries might have completed different levels of schooling to reach this starting point. The application and selection process should follow standard professional practices, e.g., submission of official transcripts, personal statements, confidential letters of recommendation on behalf of the student, and involvement of mentors in the choice of students. We also realize that it is not possible to help everyone, so it may be better to choose a select few whom our mentors consider to have innate mathematical talent. The mentors will leave the program if their needs are ignored. Also, one does not do students a favor by providing encouragement that is not matched to their capabilities.

Fig. 2 Screenshot of the GEO dataset (GDS) cluster analysis window (an open visualization tool for displaying cluster heat maps), applied to GDS 3719 (National Center for Biotechnology Information 2012)
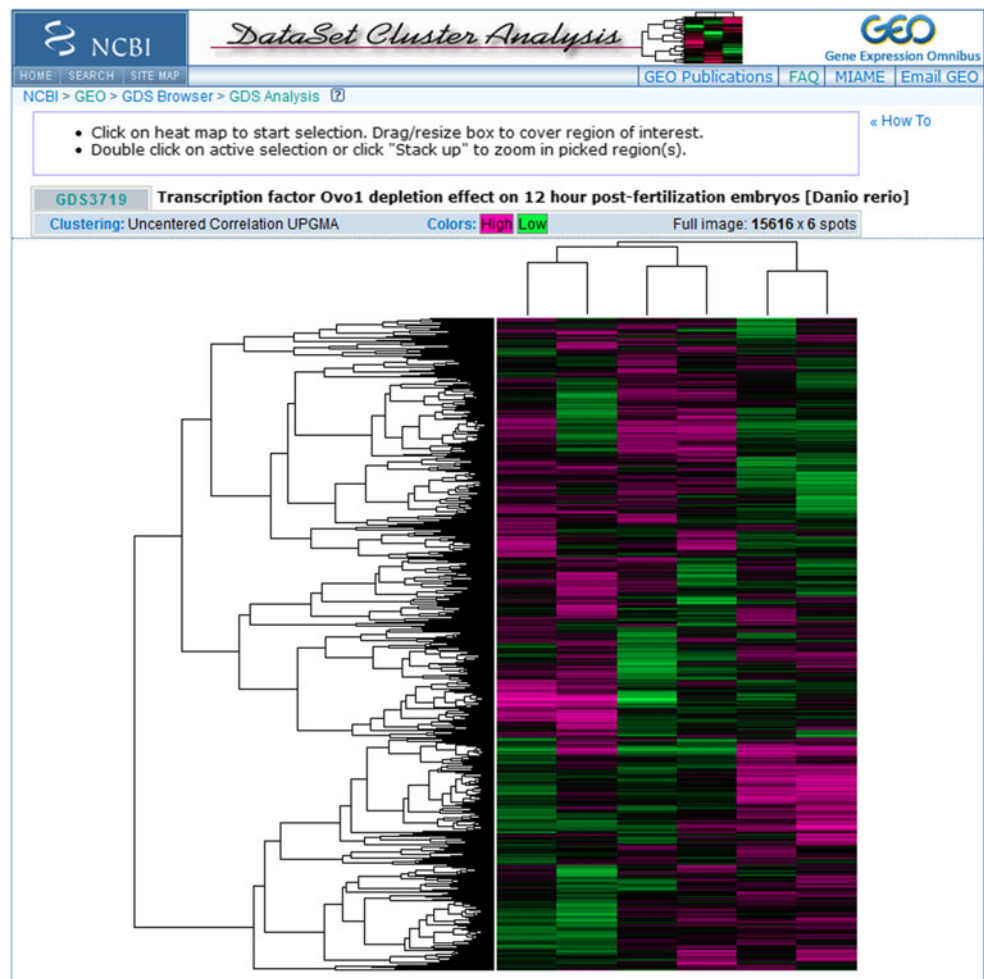
We realize that the late exposure of secondary education students to mathematics and physics in many countries suggests that a college level education might be necessary for the proposed program. On the other hand, the recent introduction of young African students to biophysics through a workshop experience in Latin America (Open questions) is tangible proof that our aims can be met.

Regularly scheduled retreats or conferences

All mentors and students should be invited to regular retreats or conferences to foster a sense of community. These could be attached to IUPAB-sponsored meetings and workshops, rather than waiting for the triennial IUPAB Congresses. Costs could be reduced by holding these events in less expensive venues around the world, but the safety of participants and the trustworthiness of local infrastructure must be a concern. The areas where potential mentors are located are often too expensive. Even in an inexpensive location in the US, a reasonable meeting can cost tens of thousands of dollars.

Funding sources

We have identified a number of possible funding strategies that could be explored in more detail:

– IUPAB regularly provides seed funding for educational, often very successful workshops that are co-funded in non-mainstream cities. ICSU may also be able to help in this regard.
– The US National Science Foundation has specific programs for fostering science exchange in a limited number of countries, although no CB programs have yet been envisioned.
– The US Howard Hughes Medical Institute (HHMI) has a wide variety of past and present international/educational initiatives, although none of the type proposed here. A program involving funded HHMI investigators as mentors might be attractive.
– The Human Frontiers Science Program (HFSP) currently offers no specific funding mechanism, but our goals match well with their international mission for collaborative research grants. Perhaps a new opportunity can be

created here, or HFSP investigators could use their funds for a mentorship program.

– Other foundations: the Agouron Institute (San Diego) and Burroughs-Wellcome (Research Triangle) support young US-based investigators working at the interfaces of biology and the physical/mathematical sciences. The Keck Foundation (California) limits funds to US institutions.

– Philanthropy: If asked, wealthy individuals with a science affinity might offer financial assistance. This approach probably would need to establish a 501(c)(3) non-profit in the US to establish tax-advantages for donors.

– Corporations: Companies in the US are willing to give to social causes as part of their corporate social responsibility programs. This requires 501 (c)(3) non-profit status. See MentorNet (2012) for a mentorship-based non-profit organization that caters to corporate donations.

– International humanitarian organizations such as the United Nations (UNDP or UNICEF capacity building programs) or the US Peace Corps may fund human empowerment efforts. Perhaps even religious institutions may be interested (e.g., the Vatican, which has a significant scientific program).

– It is important to note that any of the above funding opportunities may have restrictions and expectations that need to be taken into consideration.

Open questions

We are just starting to explore the opportunities of remote collaborations and much more work needs to be done. Some questions that remain open include:

– Computer hardware. It is unlikely that any remote scholars would bring significant computational resources to the table. There may be security risks involved in giving strangers access to local computer resources (almost certainly university policies would prohibit carte blanche access, although they may allow access via resident mentors). Computing in the cloud (see Amazon EC2 server; Amazon Web Services 2012) has recently emerged as a viable alternative to a local computing infrastructure, but regional access models and leasing costs still favor participants from the developing world. Perhaps the best starting strategy would be to provide students with laptops sponsored by the mentors or by donors. The projects in Application examples are relatively lean on hardware needs and suitable for laptop-based collaborations. If the need arises, mentors are of course free to provide additional access to their own Linux clusters and high-performance computing resources (provided that they can find a secure way to operate accounts), or, as more research laboratories gain

sufficient expertise, to lease computer resources in the cloud.

– Benefits to students. In addition to the education afforded by the outreach aspects of the program, there are specific benefits that may aid students (and other participants) in their future career development. In general, students who enter mentor programs have a lower drop-out rate in their local schools, report that they feel more confident, and express greater satisfaction with their education (MentorNet 2012). Also, letters of recommendation and petitions from established mentor-professors can help students meet their graduation requirements and frequently open the door to subsequent job opportunities. IUPAB cannot grant degrees, but it might enhance degree-granting programs by partnering local academic mentors with research mentors from other institutions. In South America, the Latin American Federation of Biophysical Societies (LAFeBS) found a way for students to achieve access to advanced facilities not available locally without the loss of continuous enrollment status at their home institution. For more details, students and mentors should contact Professor Marcelo Morales, the president of LAFeBS (LAFeBS 2012).

– Benefits to mentors. We assume that the student–mentor relationship would be the main driver of the scientific interactions, provided that the relationship is mutually beneficial. To make the investment of mentorship worthwhile to mentors, the participants of the program are expected to contribute to scientific work in the remote laboratories. The program thus has to walk a thin line between serving the motives of the students and those of the mentors. In terms of labor rights, the most often voiced criticism about Web 2.0 services and crowd sourcing is that expert knowledge is exerted by participants for relatively little reward from the entities they serve (Brabham 2012). In the scientific realm, this risk of exploitation is mitigated by the absence of a monetary profit motive. Also, mentors are expected to take a risk and invest most of their time in the relationship initially. There should be a contract and arbitration policies to help both sides meet their expectations. Also, many scientists (including the authors) believe it is appropriate for scientists to give something back to the scientific community, so our proposal can provide a conduit for such philanthropy.

– Unintended consequences. We acknowledge that, without appropriate employment opportunities, there may be no future for many scientists in their existing environment. The success of our proposed program could thus contribute to the "brain drain" problem faced by so many countries. This unintended outcome needs to be researched further. We need to understand what is stopping top talent from leaving. The increased outsourcing of technical jobs to places where employment costs are low, in combination

with continuing improvements in person-to-person communication, suggests that there may be employment opportunities in the future for trained computational biophysicists who work at remote sites in all parts of the world. Also, CB can provide an opportunity to allow overseas-trained computational biophysicists to return to their native country where they may become second-generation mentors. We need more information about the current and future employment of biophysicists. It may also be worth conducting a survey among potential candidates to find answers to what would motivate them to participate while staying in their home countries.

– Long-term sustenance. A successful program, as we envision it, would need dedicated leadership and staff support. The program should create and maintain a useful and attractive website. The leadership, which might initially take the form of an unpaid advisory board, would set the goals and policies of the program, identify and secure potential research mentors, establish criteria for selection of applicants, establish and sustain contacts with potential donors and corporate partners (Funding sources), and plan scientific meetings. The staff, who would be paid, would help to carry out these tasks, assuming that one can raise enough funding to provide salaries for the necessary staff support.

– Success stories. We acknowledge that developed countries might also be in need of capacity building; one could help members of under-represented minority groups in developed countries, or motivated "garage scientists" from more privileged backgrounds, to connect with mentors. However, it would be very desirable for our effort if one could obtain a positive outcome for young CB graduate students in specific developing geographic regions. For example, in 2010, the IUPAB raised funds to take nine young students from north, south, east, and central Africa to Latin America for a workshop on biophysics where they rubbed shoulders with over 200 like-minded graduate students (Morales and dos Remedios 2011). They were then introduced to research groups at the Federal University of Rio de Janeiro, Brazil, spending a week in the laboratories. Several of those students have returned to Latin America, and they include budding CB scientists. Clearly, this was just the beginning of a potentially much bigger effort, but it has demonstrated that a global mentor program could be not only valuable but also rewarding to all participants.

Conflict of interest None

References

AmazonWeb Services (2012) Amazon Elastic Compute Cloud. http://aws.amazon.com/ec2

Anisimov VM, Cavasotto CN (2011) Quantum mechanical binding free energy calculation for phosphopeptide inhibitors of the Lck SH2 domain. J Comp Chem 32:2254–2263

Astronomical Society of the Pacific (2012) Past Amateur Achievement Award winners. http://www.astrosociety.org/membership/awards/pastamateur.html

MIT BioEECS (2012) Computational Biophysics. http://www.eecs.mit.edu/bioeecs/CompBioPhy.html

Brabham DC (2012) The myth of amateur crowds: a critical discourse analysis of crowdsourcing coverage. Inform Comm Soc 15:394–410

Clarke L, Zheng-Bradley X, Smith R, Kulesha E, Xiao C, Toneva I, Vaughan B, Preuss D, Leinonen R, Shumway M, Sherry S, Flicek P, The 1000 Genomes Project Consortium (2012) The 1000 Genomes Project: data management and community access. Nat Meth 9:459–462

Cook G (2011) How crowdsourcing is changing science. The Boston Globe, Boston. (November 11 issue)

Craddock TJA, Tuszynski JA, Chopra D, Casey N, Goldstein LE, Hameroff SR, Tanzi RE (2012) The zinc dyshomeostasis hypothesis of Alzheimer's disease. PLoS One 7:e33552

Google (2012) Google Code. http://code.google.com

Khatib F, DiMaio F, Foldit Contenders Group, Foldit Void Crushers Group, Cooper S, Kazmierczyk M, Gilski M, Krzywda S, Zabranska H, Pichova I, Thompson J, Popovic Z, Jaskolski M, Baker D (2011) Crystal structure of a monomeric retroviral protease solved by protein folding game players. Nat Struct Mol Biol 18:1175–1177

LAFeBS (2012) Latin American Federation of Biophysical Societies. http://www.lafebs.org

Ledford H (2010) Garage biotech: life hackers. Nature 467:650–652

X.-L. Lu (2012) 3DNA: a suite of software programs for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. http://x3dna.org/

Lu X-J, Olson WK (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids Res 31:5108–5121

Lu X-J, Olson WK (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding, and visualization of three-dimensional nucleic-acid structures. Nat Protoc 3:1213–1227

MentorNet (2012) e-Mentoring for diversity in engineering and science. http://www.mentornet.net

Morales MM, dos Remedios CG (2011) Biophysical educational experiment: science and goodwill in Latin America and Africa. Biophys Rev 3:101–106

National Center for Biotechnology Information (2012) Gene Expression Omnibus. http://www.ncbi.nlm.nih.gov/geo

Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comp Chem 26:1781–1802

L. B. Poole (2011) Structural and computational biophysics training program. NIH research portfolio online reporting tools. http://report.nih.gov/ Project number 1T32GM095440-01

Qin Z, Kreplak L, Buehler MJ (2009) Hierarchical structure controls nanomechanical properties of vimentin intermediate filaments. PLoS One 4:e7294

Rusu M, Wriggers W (2012) Evolutionary bidirectional expansion for the annotation of alpha helices in cryo-electron microscopy reconstructions. J Struct Biol 177:410–419

Rusu M, Starosolski Z, Wahle M, Rigort A, Wriggers W (2012) Automated tracing of filaments in 3D electron tomography reconstructions using Sculptor and Situs. J Struct Biol 178:121–128

Sagot M-F, McKay BJM, Myers G (2009) ISMB/ECCB 2009 Stockholm. Bioinformatics 25:1570–1573

GNU Savannah (2012) Concurrent Versions System. http://savannah.nongnu.org/projects/cvs

Starosolski Z, Szczepanski M, Wahle M, Rusu M, Wriggers W (2012) Developing a denoising filter for electron microscopy and tomography data in the cloud. Biophys Rev doi:10.1007/s12551-012-0083-x

Swiss Institute of Bioinformatics (2012) SIB Bioinformatics Resource Portal. http://www.expasy.org

Tranche Developers Group (2012) Tranche Project file storage and dissemination software. https://trancheproject.org

Wang JTL, Wu CH, Wang PP (2003) Computational Biology and Genome Informatics. World Scientific, Singapore