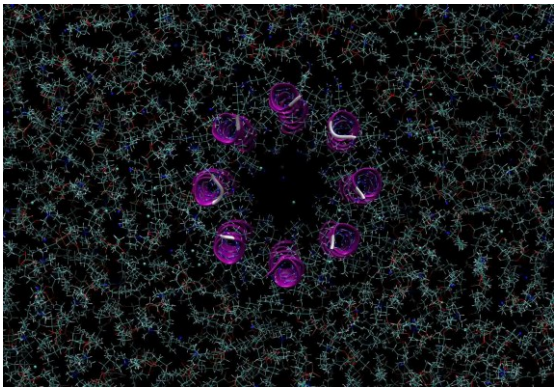
Institution: PHYESTA (Physics at Edinburgh and St Andrews)

Unit of Assessment: UoA 9 - Physics

Title of case study: Translational Biophysics – the IBM Blue Gene Application Demonstrator Portfolio

1. Summary of the impact



Impact: Economic gains, healthcare.
Since 2008 IBM has greatly expanded the division of its business that produces and markets High Performance Computing (HPC) products for the life sciences sector. This has in part been achieved by using new simulation methodologies, created by PHYESTA, to demonstrate the potential value of HPC to commercial enterprises in this sector. IBM is now a leader in HPC provision for the life science market.

Significance:
Between 2008 and 2013, IBM Pharmaceutical & Life Sciences Division has grown to more than 1500 employees, with per annum revenue in excess of US $1 billion.

Reach:
The revenue figure includes sales of Blue Gene hardware (IBM's flagship HPC product), software and other products to a variety of companies in the life sciences sector. These sales have improved those companies' ability to understand and design pharmaceutical and healthcare products.

Beneficiaries: IBM and its growing life science customer base. Healthcare companies, exemplified here by one developing antibody candidates for cancer screening. Healthcare patients (prospective).

Attribution: This work was led by Professor Jason Crain and supported by Professor Simon Bates.

2. Underpinning research

PHYESTA is a major contributor to research at the physics / life sciences interface. A great strength has been our coordinated use of computer simulation and experimental measurement to advance molecular models, capable of linking structure to function in complex and self-assembling systems.

The underpinning research began in 2000 [R1], and includes a seminal paper in 2002 overturning previous views of hydrophobic hydration in simple molecular systems such as mixtures of water and short-chain alcohols [R2]. In this work, isotopically-labelled neutron diffraction measurements were coupled with novel data analysis methods to show that even in simple alcohol-water mixtures a rich variety of self-assembled structures are formed. These previously unidentified structures were then shown to be the molecular cause of the strongly non-ideal thermodynamics shown by such mixtures. This work overturned the prevailing models (dating from 1945) in which ice-like structures were assumed to form around non-polar groups. Further work [R3] combined spectroscopic measurements, diffraction and advanced simulation. This led to improved thermodynamic models for the entropy of mixing in these simple systems and the demonstration that intermediate-range percolating molecular networks arise within particular ranges of concentration. This experimental and simulation work on the complex structure of simple molecular mixtures established an ideal benchmark for the development and evaluation of new simulation algorithms and code architectures, which could then be applied to more complex, biologically relevant molecules and materials.

Subsequent work moved decisively into the biological domain by addressing short proteins motifs in the form of peptide fragments [R4, R5]. Here the proper representation of hydrogen bonding and of the peptide bond in computer models was carefully explored and extensively tested against our

experimental data. Optimizing the computational treatment of these is a crucial step to addressing larger proteins and simulating their folding dynamics and assembly. Work on our peptide fragments (or 'minimal biomolecules') was rapidly accepted as an ideal testing ground for many aspects of biological self-assembly and rational structure prediction, which are amenable to both high-resolution experiment and massively parallel High Performance Computer (HPC) simulation. For this reason they formed the initial target demonstrators for life-science applications of IBM's new Blue Gene hardware.

Personnel:
Key PHYESTA researchers involved were Professor Jason Crain (1993-present), Professor Simon Bates (2003-2011) and Paul Tulip (PDRA 2005-2008).

3. References to the research

The quality of the work is best illustrated by R1, R2 and R3. *[Number of citations]*

[R1]	Dixit, S; Poon, WCK; Crain, J, *"Hydration of methanol in aqueous solutions: a Raman spectroscopic study",* Journal of Physics – Condensed Matter **12**, p. L323-L328 (2000) DoI: 10.1088/0953-8984/12/21/103, URL: tinyurl.com/mqf6n2q, *[66]*
[R2]	S Dixit, J Crain, J L Finney, A K Soper, *"Molecular segregation observed in a concentrated alcohol–water solution",* Nature **416**, p.829-832 (2002) DoI: 10.1038/416829a, URL: tinyurl.com/kxvvzum, *[392]*
[R3]	Dougan, L; Bates, SP; Hargreaves, R et al., *"Methanol-water solutions: A bi-percolating liquid mixture" ,* Journal of Chemical Physics **121**, p. 6456-6462 (2004) DoI: 10.1063/1.1789951, URL: tinyurl.com/n4sla3v, *[118]*
[R4]	Troitzsch, RZ; Tulip, PR; Crain, J, et al., *"A Simplified Model of Local Structure in Aqueous Proline Amino Acid Revealed by First-Principles Molecular Dynamics Simulations"* , Biophysical Journal **95**, P.5014-5020 (2008) DoI: 10.1529/biophysj.108.134916, tinyurl.com/lymd4rx, *[7]*
[R5]	Whitfield, TW; Martyna, GJ; Allison, S, and Crain J, *"Structure and hydrogen bonding in neat N-methylacetamide: Classical molecular dynamics and Raman spectroscopy studies of a liquid of peptidic fragments"*, Journal of Physical Chemistry **B110**, P.3624-3637 (2006) DoI: 10.1021/jp053140+, URL: tinyurl.com/mf3yy68, *[27]*

4. Details of the impact

Building directly on the above research, a well-publicised PHYESTA-IBM collaboration developed a portfolio of benchmarks for life science applications of HPC focussing on scalable molecular dynamics methods. These were developed initially for the first generation of IBM's Blue Gene supercomputers and subsequently formed the basis of a series of application demonstrators for life science problems. The demonstrator portfolio benchmarked the ability of Blue Gene systems to deliver enhancements in biomolecular simulation, and showed the accuracy of IBM's methodology by comparing simulations with a series of experiments on the liquid state structure of biomolecules including peptides. This joint work is recognized by IBM as contributing to sales in its life sciences business since 2008.

In early 2000, IBM began serious efforts to exploit the transformational opportunities that HPC, and specifically molecular simulation, then held for the healthcare industry. The then General Manager of IBM Life Sciences noted that it offered *"probably the greatest new market opening in IT that we see on the horizon." [text removed for publication]*. By 2002, IBM had spent over $150 million establishing partnerships and alliances, alongside $100 million in venture capital investments made to gain access to cutting-edge technologies via start-up acquisitions. *[text removed for publication]*. IBM's Blue Gene project began with a grand challenge declaration that the company would create a Petaflop supercomputer to address major scientific problems, with a particular focus on life sciences and biomolecular structure prediction. The life-sciences part of the project had two main business drivers: (1) to demonstrate to the pharmaceutical and biomedical sectors that rational structure prediction and rational design could be achieved via large-scale simulation methods, with accurate results when tested against high-resolution experimental data; and (2) to develop new simulation methods and algorithms for the massively parallel machine architectures that would be delivered by the Blue Gene hardware.

PHYESTA's contributions directly addressed IBM's requirement to improve the impact and uptake of high-performance computing in the life sciences, which was (and still is) seen as a key growth sector for the company. Our cutting-edge research and validated measurements were used by the company as a way to establish customer and in-house confidence in the application of its HPC product portfolio to high profile problems in life sciences such as protein structure and function. To achieve that confidence, IBM recognized the requirement to develop new computational methods, coupled with carefully coordinated experimental measurements (which IBM does not perform in-house). It recognized PHYESTA researchers as leaders in both fields and hence partners of choice. Professor Crain spent 2004 on sabbatical as a salaried staff member at IBM (TJ Watson Research Center, Yorktown Heights), and since then he has held a visiting position there (2005-present). His work with IBM included extensive neutron diffraction studies and optical spectroscopic measurements, as well as the incorporation of direct neutron diffraction structure factor evaluation from molecular dynamics trajectories. These advances formed a key part of definitive benchmarking exercises to test force field and sampling methods in highly disordered liquids, which demonstrated the capacity of IBM's HPC products to accurately model aqueous solutions of amino acids, small peptides and ultimately proteins. The specific methods developed for determining diffraction profiles directly from molecular dynamics trajectories were not only used in the demonstrator portfolio, but also form part of the data analysis suite in the OpenAtom software package released in 2009/2010. OpenAtom is a fine-grained parallel *ab initio* molecular dynamics code optimised for Blue Gene and other platforms.

An IBM team leader responsible for Blue Gene applications highlights the PHYESTA contributions to building the company's life science business, noting specifically that the interaction has *"spawned new technological and business opportunities in the life sciences"* and that engagement in the early phases of the Blue Gene project has *"helped to drive sales"* [F1]. IBM does not disclosed Blue Gene sales figures to life sciences companies even in aggregate. (This is unsurprising since large corporations might easily glean competitor information from such figures.) However, since the release of Blue Gene P (the second generation of machine in the Blue Gene series) in 2008, IBM's overall life science business, which includes Blue Gene alongside other products and services, has grown to achieve revenues of one billion dollars annually. A growing part of IBM's business model is to perform high-level computational research on behalf of client companies who lack the relevant expertise to do so themselves, using Blue Gene and other platforms within IBM. The company is now established as a prominent player in the application of IT infrastructure to healthcare challenges, and operates a stand-alone computational biology center (CBC) to interface to its life science business streams. Beyond the life sciences context, the Blue Gene platform has become IBM's premier HPC product line and, since 2008, a major business stream with high profile multi-rack installations at US National Labs and in Europe, 5 of which hold positions in the top 20 supercomputers in the world.

PHYESTA's work with IBM has led to improved algorithms for computer simulation, implementation of replica exchange molecular dynamics on Blue Gene, the development of scalable molecular dynamics modelling software implemented on Blue Gene, and new data analysis tools as well as experimental measurements against which to validate the simulations. This has enabled IBM to offer improved HPC solutions to customers in the business-critical target market of pharmaceuticals and life sciences. *[text removed for publication]*. PHYESTA's role in the enhancement of IBM's products and services has improved the ability of its client companies in the life sciences sector to efficiently pursue product discovery and development. This represents a significant secondary impact on these enterprises, and probably also tertiary impact on patient healthcare (although the latter is mainly prospective at this stage). Such knock-on impacts are hard to quantify, but we can give one exemplar of secondary impact because we were directly involved in it.

Soon after the initial demonstrator portfolio was in place, PHYESTA and IBM staff set up a joint project with a UK healthcare SME, Mologic. The shared goal was to develop a single assay platform for the detection of cancer biomarkers. PHYESTA and IBM used computational modelling approaches, as developed for the demonstrator portfolio, on Blue Gene hardware, to establish the structural basis behind the complex binding behaviour observed in one of Mologic's most promising antibody candidates, greatly refining their understanding of the molecular recognition mechanism. According to the supporting statement from Mologic's Chief Scientific Officer [F2] *"we had no way of taking this*

forward to challenge, substantiate or refine our hypothesis – an impasse which was, to say the least, frustrating, given the commercial importance of the antibody family in question". As a result of Mologic's joint work with PHYESTA and IBM, a single high-fidelity epitope was discovered, and is now under development by Mologic as a universal single-site hCG assay with particular applications in cancer screening. Although not yet at the stage to impact upon patients, this certainly represents an altered business strategy for Mologic, in which HPC simulation was introduced to solve a specific healthcare challenge that had not yielded to traditional discovery approaches. This exemplifies the type of uplift in business practice and technology that IBM Life Sciences is now offering to client companies across the sector. The work with Mologic was recently featured as an outstanding example of industry-academic collaboration in Drug Discovery World, a leading pharma development title [S1].

Our relationship with IBM has been used in the company's recruitment information and promotional videos, and has been highlighted by Scotland's First Minister as an "*exemplary University-Industry collaboration"* [S2]. The initial announcement of the joint effort between PHYESTA and IBM was heavily promoted, particularly in regard to the early phases of the Blue Gene programme and our work on HIV viral protein fragments simulations [S3]. The work was featured as one of IBM's "Icons of Progress" case studies in 2011, and is one of the IBM Centenary exemplars of University relationships [S4]. This campaign attracted considerable media attention from the BBC, Yahoo Finance, other news agencies [S5], as well as a wide range of pharma trade journals where IBM highlighted the importance of the PHYESTA work to the growth of HPC in the life sciences sector, focussing specifically on the need for stringent experimental validation of codes and algorithms as pioneered by PHYESTA [S6].

5. Sources to corroborate the impact

[F1]	Factual statement from IBM (Thomas J Watson Resarch Center) *Corroborates the link between Crain's research and impact with IBM research.*
[F2]	Factual statement from Chief Scientific Officer, Mologic *Corroborates the impact of the work on Mologic's corporate research and development programme.*
[S1]	Invited feature article in Drug Discovery World http://www.ddw-online.com/personalised-medicine/p149618-collaborative-research-leads-to-new-understanding-of-biomarkers-spring-12.html *Corroborates collaboration between PHYESTA, Mologic and IBM.*
[S2]	Statement from Scotland's First Minister quoted in the Scotsman http://www.scotsman.com/news/education/city-scientists-team-up-with-ibm-for-hiv-vaccine-research-1-1255169 *Corroborates collaboration with IBM and emphasises its importance to Scotland.*
[S3]	Initial promotion of collaboration through IBM press release http://www-03.ibm.com/press/uk/en/pressrelease/23799.wss *Corroborates collaboration with IBM, and contains quotes from Crain and IBM.*
[S4]	IBM Icons of Progress 2011 http://www-03.ibm.com/ibm/history/ibm100/uk/en/stories/university_relations.html *Corroborates partnership with University of Edinburgh*
[S5]	More media attention form BBC Scotland http://news.bbc.co.uk/1/hi/scotland/edinburgh_and_east/7330092.stm *Corroborates link to HIV treatment, and contains extensive quotes from Crain.*
[S6]	Pharma and other trade journals http://www.technologyreview.com/blog/editors/22046/ http://www.computing.co.uk/ctg/news/1843967/university-edinburgh-ibm-fight-hiv *Further corroboration of link to fighting HIV.*