Editorial

A Tribute to Phil Bourne—Scientist and Human

Cameron Mura [1],*, Emma Candelier [1] and Lei Xie [2]

[1] School of Data Science, University of Virginia, Charlottesville, VA 22903, USA
[2] Department of Computer Science, Hunter College, The City University of New York, New York, NY 10065, USA
* Correspondence: cmura@virginia.edu; Tel.: +1-434-249-3035

This Special Issue of *Biomolecules*, commissioned in honor of Dr. Philip E. Bourne, focuses on a new field of biomolecular data science. In this brief retrospective paper, we consider the arc of Phil's 40-year scientific and professional career, particularly as it relates to the origins of this new field.

Phil, as he is known to all—from students to university presidents and beyond—is the founding Dean of the School of Data Science (SDS) at the University of Virginia (UVA). He previously served as the first Associate Director for Data Science at the U.S. National Institutes of Health (NIH), where he led a novel *Big Data to Knowledge* initiative [1]. Prior to the NIH, Phil had a highly productive and impactful 20-year career at the University of California, San Diego (UCSD), with close ties to the San Diego Supercomputer Center and the Protein Data Bank (which he co-directed). At UCSD, Phil was also a Professor of Pharmacology, and ultimately an Associate Vice Chancellor.

This tribute, which accompanies an interview in this Special Issue, does not seek to delineate Phil's curriculum vitae or detail his many honors and achievements—e.g., serving as an early President of the International Society for Computational Biology and as the first Editor-in-Chief of *PLoS Computational Biology*—but rather to highlight the several ways in which Phil's contributions and leadership in multiple, disparate fields have coalesced as part of a new field of biomolecular data science. For details, note that a brief autobiographical account of Phil is available [2], as are his Wikipedia profile [3], his Ph.D. dissertation [4], and a list of the many scientists [5] whom Phil has trained, mentored and advised over the past four decades (this information is also available as a taxonomic tree [6], fittingly enough). Also, we would be remiss were we not to mention that one can learn what Phil, Monty Python, X-ray crystallography, and the county of Yorkshire, England all have in common by visiting ref [7]. Here, we intentionally intertwine the personal and the professional—as one can gather from even just brief interactions with him, Phil-the-human and Phil-the-scientist are refreshingly one and the same (Figure 1).

Currently a Professor of Biomedical Engineering and the Stephenson Dean of the School of Data Science at UVA, Phil spent much of his career exploring and helping *define* the intersection of biomolecules and computation—as a practicing scientist and as a leader [8] in academia, in open-access academic publishing [9], in the broader open-science movement [10,11], and in conjunction with government and industry (Phil's role as an associate vice chancellor at UCSD concerned "innovation and industrial alliances"). Over the span of Phil's four-decade career, our knowledge of biomolecular structures, dynamics, functions and evolution (in both health and disease) has rapidly advanced, often exponentially. *What enabled this?* The staggering advances were enabled, in no small part, by Phil's highly collaborative and foundational work, where three pervasive themes have been: (i) a **structural approach** to biological systems, including knowing when to be reductionist and when not to be; (ii) the development and application of core **computational methodologies**; and (iii) **multidisciplinarity**, to an extreme.

Figure 1. Phil's life in science started (**a**) very hands on, progressed to (**b**) mentoring, and then he finally (**c**) took off for the (**d**) White House with his family. While he's been a scientist for over 40 years, Phil's been an avid cyclist for even longer; at UVA, he's a founding member of the Hells Administrators (https://www.youtube.com/watch?v=ZgtNp1ditzE (accessed on 5 January 2023)).

To elaborate these three points—structure, computation, multidisciplinarity—we note that from the start of his career, first in small-molecule crystallography [12] and then in (very) large-molecule structural biology [13], Phil embraced the *key role of three-dimensional structure* [2] as an information-rich bridge between a biomolecule's sequence and its function. (Phil's *Structural Bioinformatics* text is a mainstay on many researchers' bookshelves [14].) As regards point (ii), a hallmark of Phil's research programs over the years has been the development and application of *computational methodologies & resources*, including state-of-the-art databases (most notably the Protein Data Bank [15])

and associated data standardization, dictionary and exchange approaches, such as the macromolecular crystallographic information file (mmCIF) [16]. Along the way, Phil and his teams created data standards and interoperable tools that were freely disseminated, before this was appreciated and accepted as scientific best practice, and they developed algorithms and software, such as the widely used combinatorial extension (CE) method for 3D structure alignment [17] and a novel approach to using "sequence order-independent profile–profile alignment" to examine protein functional sites across vast evolutionary distances [18]. Finally, as regards point (iii), computational biology and related areas are well-understood to be *highly* interdisciplinary [19], and here we simply reiterate that Phil was a pioneer in these fields from their inception (before they were 'a thing'). As an extreme example that is specific to Phil, not many scientists have both published research on "ancient shifts in trace metal geochemistry" [20] and written a book on Unix [21]!

In addition to foundational 'basic research' advances, Phil's work and its applications have had significant impact across a vast array of biological and biomedical domains, including early-stage drug discovery [22], molecular evolution [23], immunology [24], and more—resulting in over 350 papers, several books, and nearly 75,000 citations of his work [25]. In recent years, Phil's attention has turned to considering what is possible at the junction of data science and structural biology [26,27]; notably, Phil's receipt of Microsoft's *Jim Gray Award for eScience* (2010) foretold his move into this area, as this award cited his "*groundbreaking accomplishments in data–intensive science*". All throughout these career milestones, Phil has been unwavering in his support of public service in government and academia, in open scholarship, in research best practices [28], and in the professional development of all who have crossed his path, from students to peers to colleagues. Indeed, as regards professional development, many readers are likely familiar with the *Ten Simple Rules* (TSR) series that Phil conceived of and initiated 20 years ago. There are now well over 1000 rules [29], covering everything from strategically forging one's career path in academia, government and industry [30,31], to winning a Nobel Prize [32], to focused guides on leveraging Git/GitHub [33], to avoiding and resolving conflicts with your colleagues [34]. The full collection of TSRs, which is freely available at ref. [35] and organized by topical areas/categories (*Career development*, *Education & mentoring*, etc.), is a testament to how Phil empowers scientists to more effectively navigate the world of very-human scientific activities (papers, talks, careers) that begin where the data-collection and number-crunching end.

Those who have worked with Phil have likely noticed that a pronounced trait in his approach to biosciences, and now data science, is that it is expansive and forward-looking, with a healthy dose of irreverence and provocation [36]—in a word, *visionary*. Phil's interests in recent years have converged upon "biomedical data sciences", which can be viewed as a natural evolution (and synthesis) of bioinformatics, computational biology, structural biology, biophysics, systems biology, and other allied fields [36]. In a real sense, the intense multidisciplinarity of Phil's career foreshadowed a field such as biomedical data science. This Special Issue honors Phil by trying to capture his vision as it relates to biomolecules—how this vision arose and what it can encompass, as expressed in a collection of original research papers, perspectives and reviews. We hope that the breadth and depth of the contributions in this Special Issue convey the spirit of Phil's vision.

Finally, as we honor Phil in this Special Issue, recognizing his role today as the Dean of the UVA School of Data Science—the first of its kind in the nation—we close by noting that Phil's vision of biomedical data science can be mapped to four core elements of data science: *Systems*, *Analysis*, *Design* and *Value*. For example, *Systems*, in our context of biomolecular data science, relates to the underlying infrastructure, such as data structures, ontologies, software libraries and tools, that enables discovery. With respect to biomolecules, *Analysis* has been largely dominated by machine learning approaches such as deep learning, for which robust systems and frameworks to access and efficiently utilize training data are critical (e.g., [37]). *Design*, which can refer to human–computer interaction, visualization and so on, has played a vital role throughout the history of structural and computational

biology, and now biomolecular data science. Finally, the *Value* element seeks to optimize the benefit of research for those it serves, from society at large to local communities; here, clear links exist between drug and therapeutic development, health disparities research, and other realms at the heart of biomolecular and biomedical data sciences.

The papers in this Special Issue exemplify what a field of biomolecular data sciences can represent, as a fitting tribute to someone who has moved the field forward via his own work and by his steadfast support of many research communities, biomolecular and beyond. In keeping with Phil's mantra, '*Onwards!*' . . .

Author Contributions: All authors contributed to the writing and editing of the text. All authors have read and agreed to the published version of the manuscript.

References

1. Margolis, R.; Derr, L.; Dunn, M.; Huerta, M.; Larkin, J.; Sheehan, J.; Guyer, M.; Green, E.D. The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: Capitalizing on biomedical big data. *J. Am. Med. Inform. Assoc.* **2014**, *21*, 957–958. [CrossRef] [PubMed]
2. Bourne, P.E. Life Is Three-dimensional, and It Begins with Molecules. *PLoS Biol.* **2017**, *15*, e2002041. [CrossRef] [PubMed]
3. Philip Bourne–Wikipedia. Available online: https://en.wikipedia.org/wiki/Philip_Bourne (accessed on 5 January 2023).
4. Bourne, P.E. Crystal Structure Analyses: Metal Complexes of Biological Interest and the Stereochemistry of Substituted Phenylbicyclo[2,2,2]octanes. Ph.D. Thesis, Flinders University of South Australia, Adelaid, Australia, 1979. Available online: https://www.worldcat.org/title/708090572 (accessed on 5 January 2023).
5. Bourne, P.E. Bourne Lab Alumni. Available online: https://tinyurl.com/BourneLabAlumni (accessed on 16 November 2022).
6. Computational Biology Family Tree—Philip, E. Bouine. Available online: https://academictree.org/compbio/tree.php?pid=157978 (accessed on 16 November 2022).
7. Nicholls, A.; Bourne, P.E.; Banquet, H. *Macromolecular Crystallography Computing School*; IUCr: Bellingham, Washington, USA, 1996; Available online: https://tinyurl.com/IUCrBanquetHumour (accessed on 5 January 2023).
8. Bourne, P.E. Ten Simple Rules for Good Leadership. *PLoS Comput. Biol.* **2022**, *18*, e1010133. [CrossRef] [PubMed]
9. McKiernan, E.C.; Bourne, P.E.; Brown, C.T.; Buck, S.; Kenall, A.; Lin, J.; McDougall, D.; Nosek, B.A.; Ram, K.; Soderberg, C.K.; et al. How Open Science Helps Researchers Succeed. *Elife* **2016**, *5*. [CrossRef] [PubMed]
10. Bourne, P.E.; Bonazzi, V.; Brand, A.; Carroll, B.; Foster, I.; Guha, R.V.; Hanisch, R.; Keller, S.A.; Kennedy, M.L.; Kirkpatrick, C.; et al. Playing Catch-up in Building an Open Research Commons. *Science* **2022**, *377*, 256–258. [CrossRef]
11. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef]
12. Bourne, P.E.; Taylor, M.R. The Structure of Aqua[3-ethoxy-2-oxobutyraldehyde bis(thiosemicarbazonato)]zinc(II). *Acta Cryst. Sect. B* **1980**, *36*, 2143–2145. [CrossRef]
13. Clegg, G.A.; Stansfield, R.F.; Bourne, P.E.; Harrison, P.M. Helix Packing and Subunit Conformation in Horse Spleen Apoferritin. *Nature* **1980**, *288*, 298–300. [CrossRef]
14. Bourne, P.E.; Gu, J. *Structural Bioinformatics*, 2nd ed.; Wiley-Blackwell: Hoboken, NJ, USA, 2009; p. 1096.
15. Burley, S.K.; Berman, H.M.; Duarte, J.M.; Feng, Z.; Flatt, J.W.; Hudson, B.P.; Lowe, R.; Peisach, E.; Piehl, D.W.; Rose, Y.; et al. Protein Data Bank: A Comprehensive Review of 3D Structure Holdings and Worldwide Utilization by Researchers, Educators, and Students. *Biomolecules* **2022**, *12*, 1425. [CrossRef]
16. Bourne, P.E.; Berman, H.M.; McMahon, B.; Watenpaugh, K.D.; Westbrook, J.D.; Fitzgerald, P. Macromolecular Crystallographic Information File. In *Methods in Enzymology*; Academic Press: Cambridge, MA, USA, 1997; Volume 277, pp. 571–590.
17. Shindyalov, I.N.; Bourne, P.E. Protein Structure Alignment by Incremental Combinatorial Extension (CE) of the Optimal Path. *Protein Eng.* **1998**, *11*, 739–747. [CrossRef]
18. Xie, L.; Bourne, P.E. Detecting Evolutionary Relationships Across Existing Fold Space, Using Sequence Order-independent Profile-profile Alignments. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 5441–5446. [CrossRef]
19. Eddy, S.R. "Antedisciplinary" Science. *PLoS Comput. Biol.* **2005**, *1*, e6. [CrossRef]
20. Dupont, C.L.; Yang, S.; Palenik, B.; Bourne, P.E. Modern Proteomes Contain Putative Imprints of Ancient Shifts in Trace Metal Geochemistry. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 17822–17827. [CrossRef]
21. Bourne, P.E.; Holstein, R.; McMullen, J. *UNIX for OpenVMS Users*, 3rd ed.; Digital Press: Oxford, UK, 2003; p. 562.

22. Kinnings, S.L.; Liu, N.; Buchmeier, N.; Tonge, P.J.; Xie, L.; Bourne, P.E. Drug Discovery Using Chemical Systems Biology: Repositioning the Safe Medicine Comtan to Treat Multi-drug and Extensively Drug Resistant Tuberculosis. *PLoS Comput. Biol.* **2009**, *5*, e1000423. [CrossRef]

23. Scheeff, E.D.; Bourne, P.E. Structural Evolution of the Protein Kinase-like Superfamily. *PLoS Comput. Biol.* **2005**, *1*, e49. [CrossRef]

24. Ponomarenko, J.V.; Bourne, P.E. Antibody-protein Interactions: Benchmark Datasets and Prediction Tools Evaluation. *BMC Struct. Biol.* **2007**, *7*, 64. [CrossRef]

25. Philip, E. Bourne—Google Scholar. Available online: https://tinyurl.com/BourneGoogleScholar (accessed on 16 November 2022).

26. Mura, C.; Draizen, E.J.; Bourne, P.E. Structural Biology Meets Data Science: Does Anything Change? *Curr. Opin. Struct. Biol.* **2018**, *52*, 95–102. [CrossRef]

27. Bourne, P.E.; Draizen, E.J.; Mura, C. The Curse of the Protein Ribbon Diagram. *PLoS Biol.* **2022**, *20*, e3001901. [CrossRef]

28. Borgman, C.L.; Bourne, P.E. Why It Takes a Village to Manage and Share Data. *Harv. Data Sci. Rev.* **2022**, *4*. [CrossRef]

29. Bourne, P.E.; Lewitter, F.; Markel, S.; Papin, J.A. One Thousand Simple Rules. *PLoS Comput. Biol.* **2018**, *14*, e1006670. [CrossRef]

30. Searls, D.B. Ten Simple Rules for Choosing Between Industry and Academia. *PLoS Comput. Biol.* **2009**, *5*, e1000388. [CrossRef] [PubMed]

31. Bourne, P.E. Ten Simple Rules in Considering a Career in Academia Versus Government. *PLoS Comput. Biol.* **2017**, *13*, e1005729. [CrossRef] [PubMed]

32. Roberts, R.J. Ten Simple Rules to Win a Nobel Prize. *PLoS Comput. Biol.* **2015**, *11*, e1004084. [CrossRef] [PubMed]

33. Perez-Riverol, Y.; Gatto, L.; Wang, R.; Sachsenberg, T.; Uszkoreit, J.; Leprevost Fda, V.; Fufezan, C.; Ternent, T.; Eglen, S.J.; Katz, D.S.; et al. Ten Simple Rules for Taking Advantage of Git and GitHub. *PLoS Comput. Biol.* **2016**, *12*, e1004947. [CrossRef]

34. Lewitter, F.; Bourne, P.E.; Attwood, T.K. Ten Simple Rules for Avoiding and Resolving Conflicts with Your Colleagues. *PLoS Comput. Biol.* **2019**, *15*, e1006708. [CrossRef]

35. Ten Simple Rules Collection. Available online: https://collections.plos.org/collection/ten-simple-rules/ (accessed on 5 January 2023).

36. Bourne, P.E. Is "Bioinformatics" Dead? *PLoS Biol.* **2021**, *19*, e3001165. [CrossRef]

37. Draizen, E.J.; Murillo, L.F.R.; Readey, J.; Mura, C.; Bourne, P.E. PROP3D: A Flexible, Python-based Platform for Protein Structural Properties and Biophysical Data in Machine Learning. *bioRxiv* **2023**, Submitted. [CrossRef]