In Silico Veritas

Amanda S. Barnard*

CSIRO Materials Science and Engineering, 343 Royal Parade, Parkville 3052, Victoria, Australia

This Perspective is contributed by a recipient of the 2014 *ACS Nano* Lectureship Awards, presented at the XII International Conference on Nanostructured Materials (NANO 2014) in July 2014. The *ACS Nano* Lectureships honor the contributions of scientists whose work has significantly impacted the fields of nanoscience and nanotechnology.

ABSTRACT Innovations in computational nanoscience have traditionally come in conjunction with experimental innovations, but uncertainty often surrounds the trustworthiness of *in silico* studies. While the accuracy of simulations has been improving every year, considerably less attention has focused on dealing with increasing complexity, which may be the source of concern. Creating more realistic virtual experiments (without sacrificing theoretical and numerical accuracy) remains challenging, particularly when we are confronted with the polydispersity characteristic of *extra silico* samples. Fortunately, there are various theoretical methods that can be used in conjunction with first-principles simulations, not the least of which are the statistical tools and techniques promised by the emerging fields of materials informatics and data-driven sciences.

Scientists are traditionally among the early adopters of most new technologies, and with the advent of computers, it was not long before scientists began converting their theories and equations into "ones and zeros". The first computational studies of materials began to emerge in the 1970s, but it was perhaps a decade later before we could claim the birth of a new scientific domain, around the same time that nanoscience and nanotechnology were gaining momentum. The timing was fortuitous. Unlike computational predecessors in chemistry (physics, biology, and materials science), computational nanoscience is unique in that it evolved alongside its experimental counterpart and, for better or worse, shared in (and contributed to) the early years of uncertainty and serendipity of the field as a whole.

Beyond the excitement that accompanied the common struggle to understand matter and phenomena at the nanoscale, we were blessed with time and length scales that were mutually accessible. The early years also coincided with the convergence of increased resolution in experimental nanoscience, with massively parallel high-performance computing, and collaborations flourished. Given this environment, it was perhaps inevitable that we initially found ourselves attempting to recover some of the familiarity of conventional laboratories in our virtual domain—recreating established and experimental methods in cyberspace.

The term *in silico* is not truly Latin, but is acknowledged as originating from work by the mathematician Pedro Miramontes in 1989[1] to refer to experiments performed in or on a computer, in analogy to the Latin phrases *in vivo*, *in vitro*, and *in situ*. Although popularized for describing virtual screening experiments, it is now accepted that this expression encompasses a variety of well-established simulation methods, length/time scales, and underlying theories, all of which conceptually follow the *Input = Understanding of the problem/Output = Data* paradigm. The outcomes of *in silico* studies contribute to the body of understanding when the information contained in the result is greater than the information required to describe the problem in the first place, the same as in any other area of science.

However, even though computational chemistry, computational physics, and computational materials science have been contributing to knowledge in the field for decades, it is my impression from discussions with colleagues all over the world that these contributions are still regarded with some skepticism. This is perplexing, since *in silico* studies can be exactly repeated in any virtual laboratory using the same codes

* Address correspondence to amanda.barnard@csiro.au.

10.1021/nn502808y

In this Perspective, I will explore the concept of *truth* in computational nanoscience and identify some of the challenges that are different than those of computational chemistry and computational physics.

and parameters, whereas the reproducibility of experimental studies is hampered by variability in samples, equipment, and the laboratory conditions. Poorly calibrated equipment is just as damaging to results as poorly converged simulations, and yet they are viewed differently. Even among computational scientists, we discuss "validating" a model against an experiment, but "explaining" an experiment with a model, the implication being that the experiment is already "valid". I think it is time we took an objective look at this issue, and ask ourselves—is it justified?

In this Perspective, I will explore the concept of *truth* in computational nanoscience and identify some of the challenges that are different than those of computational chemistry and computational physics. Drawing from some of my own research on metal oxide nanoparticles, which I am perhaps more qualified to reframe and to criticize than the work of others, I will attempt to describe some of the ways that theoretical and computational nanoscience is striving to have impact, and to contribute more effectively to some of the grand challenges facing science and society today.

More Than a Matter of Trust. When thinking about "truth" in terms of computer simulations, there are two definitions that come to mind: accurate and realistic. Experimental methods probe entire samples of nanoparticles that have a distribution of sizes, shapes, and defects, and so provide an approximate solution to an exact problem because the results are an average over all of the different structural variations within a sample. In contrast, individual computer simulations in the field apply physicochemical theories to individual structures to provide exact solutions, but due to the inherent complexity in realistic systems, they are usually restricted to approximations of the problems. This gives the impression that realism is sacrificed at the altar of accuracy, and in the case of *ab initio* simulations, this is a fair judgment. Alternatively, there are other simulation methods that deliberately coarse-grain over the fine details so that the system can become more experimentally realistic. Both approaches are valid, and focusing on one aspect at the expense of the other is usually necessary to make a unique contribution to the field. One of the greatest advantages of collaborations between conventional experiments and "virtual experiments" is that they provide complementary information, rather than simply trying to replicate one another.

However, accuracy and realism are not mutually exclusive, and one of the great challenges to computational nanoscientists is how to achieve both, effectively and simultaneously. One approach is to use a hierarchy of established methods,[2] building on a foundation of *ab initio* electronic structure simulations, and applying increasing amounts of structural, chemical, and environmental realism as possible (or required). To illustrate this, let us follow a case study of titanium dioxide photocatalysts with which I have been involved over the past 10 years, where humble beginnings were systematically refined to deliver insight regarding their use in commercial sunscreens. This case study is conceptually represented in Figure 1.

At the beginning of this study, there was already considerable literature on the structure and properties of titania, both experimentally and computationally,[3] and the issue of size-dependent selectivity between the anatase and rutile polymorphs had already been considered.[4] The previous work had drawn on classical simulations, so the first step was to increase the accuracy of the input parameters by using higher order *ab initio* simulations.[5] These parameters largely consisted of a collection of surface properties, and as both anatase and rutile are crystalline materials, it was necessary to employ a more sophisticated multivariate phenomenological model[6] that treated the nanoparticles as faceted polyhedra.[7] This actually represented a considerable innovation at the time, as previous similar studies were content to approximate nanoparticles as spheres. As a result, a more detailed relationship between the size, shape, and solid phase was uncovered, and it was shown that, depending on the particle shape, the characteristic anatase-to-rutile phase transformation could be expected anywhere from ∼8 to ∼18 nm in average diameter.[8] While this was theoretically and computationally satisfying, in truth, it merely explained the source of the error bars in related experiments, which were conducted on samples with considerable structural polydispersivity.[9]

At the time, the principal criticism of this work was a lack of realism. The *ab initio* simulations were conducted at $T \approx 0$ K, *in vacuo*, and the surfaces were reconstructed but assumed to be "clean". While being accepted approximations among computational scientists (who often aim for exact solutions to approximate problems), the inconsistency with experimental systems was undoubtedly a source of skepticism. Therefore, using the same flavor of *ab initio* simulations, a much larger set of parameters was generated, which explicitly included different types of surface chemistry, characteristic of different types of local chemical environments.[10] This facilitated a new investigation of

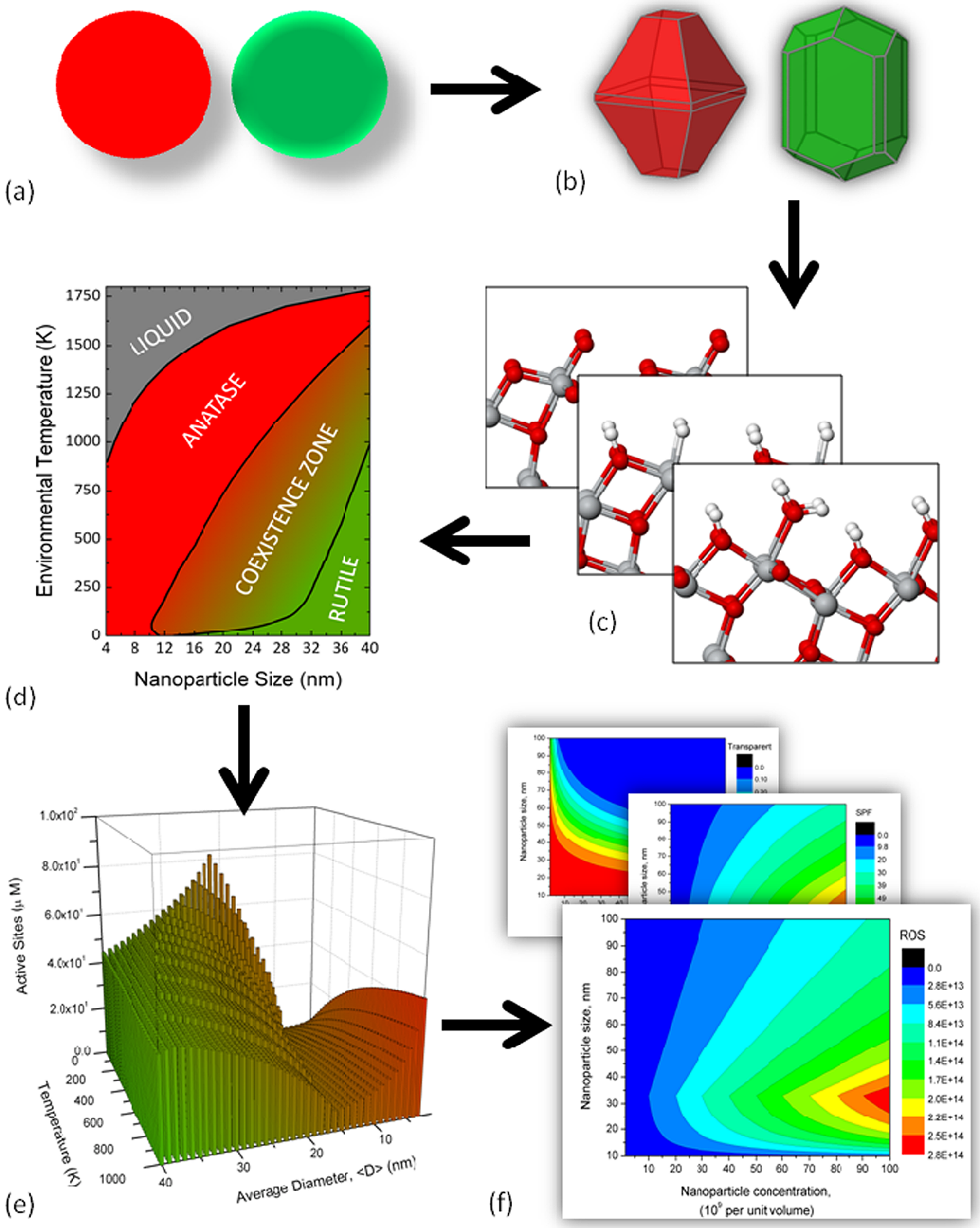

Figure 1. Conceptual representation of the progression of the six year case study on titania photocatalysts: beginning with a comparison of (a) "spherical" nanoparticles of anatase (red) and rutile (green) and (b) faceted polyhedral of the same. Reprinted with permission from ref 7. Copyright 2004 American Physical Society. (c) The introduction of diverse surface chemistry. Reprinted with permission from refs 10 and 11. Copyright 2005 Elsevier, American Chemical Society, respectively. (d) The encapsulation of and extension of this information in a size–temperature phase diagram. Reprinted from ref 13. Copyright 2008 American Chemical Society. (e) Mapping this diagram to the number density of surface cations, to create a structure/property map of maximal photocatalytic activity. Reprinted with permission from ref 14. Copyright 2011 Royal Society of Chemistry. (f) The combination of the photocatalytic structure/property map with complementary maps of the associated optical properties, for the assessment of titania nanoparticles in sunscreens. Reprinted with permission from ref 18. Copyright 2010 Nature Publishing Group.

the relationship between particle shape and chemical environment, and suggested ways that selective passivation could be used in the context of shape engineering.[11] Moreover, by exploiting the applicability and versatility of *ab initio* thermodynamics,[12] it was possible to extend the new parameters to include the impact of variations in the surrounding temperature and the supersaturation of the chemical reservoir. This led to the development of the first *environmentally sensitive* phase diagram of nanoscale titania, which mapped the equilibrium shape and phase as a function of size and temperature,[13] and located the conditions for a "coexistence zone" where phase transformations would be accompanied by surface reactions.

> One of the greatest advantages of collaborations between conventional experiments and "virtual experiments" is that they provide complementary information, rather than simply trying to replicate one another.

This combination of techniques is now referred to as *thermodynamic cartography*, and over the years, it has been used to map the relationship between the size and shape of any inorganic nanostructure as a function of "environmental parameters", such as temperature, pressure, or even humidity, based on a limited set of material-specific input parameters that can be obtained accurately. However, the greatest benefit comes in the application of the predictions contained in these maps, as a basis for more versatile structure/property relationships. In this case study, the environmentally sensitive phase diagram underpinned the development of a structure/property map of the photocatalytic activity, characterized by the density of cationic surface sites where reactive oxygen species (ROS) can be generated.[14] This can only be done when the materials in question are crystalline, the shapes are polyhedra, and the fractional area and atomic structure of each facet are known at all points in the phase diagram, as it was in this case.

Structure/property maps of this type are useful in guiding decision making, identifying the most (physically or economically) efficient options, and anticipating how stable it is likely to be under a range of different conditions. However, since some properties of nanomaterials are undesirable (such as the generation of ROS[15,16]) it is also possible to incorporate these predictions into technology assessment activities.[17] A fitting example is the final part of this case study, where the additional realism in the simulations assisted in the assessment of titania nanoparticles in sunscreen.[18] The desirable optical properties of metal oxide nanoparticles can make sunscreen more aesthetically pleasing to consumers (thereby increasing adoption), and offer a higher sun protection factor (SPF), but the photocatalytic generation of ROS at the surfaces could have undesirable side effects.[19,20] It has been estimated that 4000−6000 t of sunscreen wash off swimmers annually in oceans worldwide,[21] and so understanding the potential hazards posed by the nanoparticles when they are released from the topical matrix can assist in product design. The results of this case study suggested that (among the properties considered) there are certain nanoparticle sizes and concentrations that manufacturers should avoid to mitigate this risk, leaving many other size/concentration combinations that can still deliver a sustainable product.

> If we want to describe realistic nanomaterials and phenomena accurately *in silico*, then we must confront the need to include more complete spectra of all possible states of the systems, and treat the problem probabilistically.

This is a simple example, with scope for further advancement, but it serves to illustrate the point. A discrete set of accurate results can be combined with suitable descriptions of more realistic conditions, to contribute in a meaningful way to real world problems.

Diversity and Complexity. An assumption that is preserved throughout this example is one that has its roots in chemistry: that all of the nanoparticles are the same. This is understandable, since one of the great advantages in the field of chemistry is the opportunity to purify samples, but is it justifiable? In a sample of water, it is effectively possible to achieve $\sim$100% concentration of H_2O molecules, and all H_2O molecules are identical (instantaneous vibrational configurations aside). This is not the case in nanoscience, even if chemical purity could be achieved. In a sample of nanoparticles, the structures are not identical. Considerable effort has been directed to producing "monodispersed" nanoparticle systems, but in almost all cases (and certainly all economically feasible cases), structural polydispersivity persists. Omitting this complexity diminishes the realism, and could certainly be considered a failure of the case study described above. A more optimistic perspective to take, however, is that it has left room for improvement.

Embracing this complexity does, however, present a problem for computational nanoscience, which has gained acceptance on the promise that predictions could be extrapolated based on accurate simulations of a given "model" structure in the ground state. For many methods, and many researchers, this has been our distinguishing feature, our unique value proposition. Remember that the vast majority of researchers have been striving to achieve greater accuracy. However, the conditions under which any molecule, material, or nanostructure will actually be in the ground state are very rare (*in vacuo*, at low temperature, and entirely isolated from any external fields). All of the realism described above will perturb the system, so the true ground state of the system will only be achieved instantaneously, if at all. This issue is well-known to those working with classical molecular dynamics, for example.

If we want to describe realistic nanomaterials and phenomena accurately *in silico*, then we must confront the need to include more complete spectra of all possible states of the systems, and treat the problem probabilistically.[22] Most characterization experiments do not measure the relative stability of different morphologies, they infer it from the observed populations of different morphologies. Moreover, most devices will not operate under equilibrium conditions, so performance is not contingent on what is most stable, only what is most probable. Perhaps it is this lack of diversity, and the rarity of computational studies that successfully combine accuracy and realism, that fuels mistrust of many computational results.

Welcoming the Next Generation. Dealing with the complexity of ensembles of diverse nanostructures may seem overwhelming, so we are fortunate that a new paradigm is emerging, to redefine the meaning of *in silico*. Data-driven science is now maturing beyond established fields such as genomics and bioinformatics, and material informatics is providing new tools and new ways of thinking about solids state physics, materials science, and inorganic chemistry. The U.S. Materials Genome Initiative (MGI), launched in June 2011 by President Barack Obama, is an exemplary materials informatics platform and similar programs are beginning across the world. These programs represent new opportunities, but like any new research domain, there will be a period of adjustment as researchers embrace the new perspective: *Input = Data/Output = Better understanding of the problem*.

For nanoscience this formula is just what we need. The statistical methods inherent in data-driven science will provide the ideal tools to study the collective behavior of entire ensembles of polydisperse nanostructures, characteristic of *extra silico* samples. We will be able to reproduce the outputs generated by conventional experiments more closely, and to quantify the roles of distributions in determining properties and performance. We can apply these tools to simulation data collected combinatorially, in a return to our origins;[1] or to experimental data generated with high-throughput methods, extending our scope, encouraging deeper collaboration, and enriching the *in silico* domain.

As we enter this new era of *in silico* science, it is important to keep all of these issues in perspective. Just as traditional simulation-based modeling suffered from widespread skepticism until the body of benchmarks, standards, and validation procedures gained strength, so too materials informatics—and, one can only speculate, a future subfield of materials informatics, such as "nanoinformatics"—will need time to grow. By treating ensembles of structures that combine the theoretical/numerical accuracy with chemical/environmental realism, we will be working in the right direction to overcome this barrier, and come one step closer to providing exact solutions to exact problems. This is our new frontier.

Conflict of Interest: The authors declare no competing financial interest.

Acknowledgment. The author would like to thank I. Snook, G. Smith, L. Curtiss, H. Xu, E. Ōsawa, J. Louviere, and M. Per for useful discussions related to the ideas and research discussed in this Perspective, and the members of the Virtual Nanoscience Laboratory for critical reading and feedback.

REFERENCES AND NOTES

1. Miramontes, P. *Un Modelo de Autómata Celular para la Evolución de los Ácidos Nucleicos [A Cellular Automaton Model for the Evolution of Nucleic Acids]*. Tesis de Doctorado en Matemáticas. Universidad Nacional Autónoma de México (UNAM), 1992.
2. Barnard, A. S. Modelling of Nanoparticles: Approaches to Morphology and Evolution. *Rep. Prog. Phys.* **2010**, *73*, 086502.
3. Diebold, U. The Surface Science of Titanium Dioxide. *Surf. Sci. Rep.* **2003**, *48*, 53–229.
4. Zhang, H.; Banfield, J. F. Thermodynamic Analysis of Phase Stability of Nanocrystalline Titania. *J. Mater. Chem.* **1998**, *8*, 2073–2076.
5. Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for *Ab Initio* Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B* **2006**, *54*, 11169.
6. Barnard, A. S.; Zapol, P. A Model for the Phase Stability of Arbitrary Nanoparticles as a Function of Size and Shape. *J. Chem. Phys.* **2004**, *121*, 4276–4283.
7. Barnard, A. S.; Zapol, P. Effects of Particle Morphology and Surface Hydrogenation on the Phase Stability TiO2 at the Nanoscale. *Phys. Rev. B* **2004**, *70*, 235403.
8. Barnard, A. S. Thermodynamic Cartography and Structure/Property Mapping of Potential Nanohazards. In *Towards Efficient Designing of Safe Nanomaterials*; Leszczynski, J., Puzyn, T., Eds.; Royal Society Chemistry: Cambridge, 2012; pp 186−219.
9. Ranade, M. R.; Navrotsky, A.; Zhang, H. Z.; Banfield, J. F.; Elder, S. H.; Zaban, A.; Borse, P. H.; Kulkarni, S. K.; Doran, G. S.; Whitfield, H. J. Energetics of Nanocrystalline TiO_2. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6476–6481.
10. Barnard, A. S.; Zapol, P.; Curtiss, L. A. Anatase and Rutile Surfaces with Adsorbates Representative of Acidic and Basic Conditions. *Surf. Sci.* **2005**, *582*, 173–188.
11. Barnard, A. S.; Curtiss, L. A. Prediction of TiO_2 Nanoparticle Phase and Shape Transitions Controlled by Surface Chemistry. *Nano Lett.* **2005**, *5*, 1261–1266.
12. Reuter, K.; Stamp, C.; Scheffler, M. *Ab Initio* Atomistic Thermodynamics and Statistical Mechanics of Surface Properties and Functions. In *Handbook of Materials Modeling*; Yip, S., Ed.; Springer: The Netherlands, 2005; pp 149−194.
13. Barnard, A. S.; Xu, H. An Environmentally Sensitive Phase Map of Titania Nanocrystals. *ACS Nano* **2008**, *2*, 2237–2242.
14. Barnard, A. S. Predicting the Photocatalytic Activity or Potential Free Radical Toxicity of Nanoscale Titania. *Energy Environ. Sci.* **2011**, *4*, 439–443.
15. Nel, A.; Xia, T.; Mädler, L.; Li, N. Toxic Potential of Materials at the Nanolevel. *Science* **2006**, *311*, 622–627.
16. Barnard, A. S. How Can *ab Initio* Simulations Address Risks in Nanotech? *Nat. Nanotechnol.* **2009**, *4*, 332–335.
17. Jahnel, J.; Fleischer, T.; Seitz, S. B. Risk Assessment of Nanomaterials and Nanoproducts—Adaptation of Traditional Approaches. *J. Phys.: Conf. Ser.* **2013**, *429*, 012063.
18. Barnard, A. S. One-to-One Comparison of Sunscreen Efficacy, Aesthetics and Potential Nanotoxicity. *Nat. Nanotechnol.* **2010**, *5*, 271–274.
19. Wiseman, H.; Halliwell, B. Damage to DNA by Reactive Oxygen and Nitrogen Species: Role in Inflammatory Disease and Progression to Cancer. *Biochem. J.* **1996**, *313*, 17–29.
20. Nanomaterial Case Studies: Nanoscale Titanium Dioxide in Water Treatment and in Topical Sunscreen, National Center for Environmental Assessment—RTP Division, Office of Research and Development, U. S.

Environmental Protection Agency, EPA/600/R-09/057F, 2010.

21. Danovaro, R.; Bongiorni, L.; Corinaldesi, C.; Giovannelli, D.; Damiani, E.; Astolfi, P.; Greci, L.; Pusceddu, A. Sunscreens Cause Coral Bleaching by Promoting Viral Infections. *Environ. Health Perspect.* **2008**, *116*, 441–447.
22. Barnard, A. S. Clarifying Stability, Probability and Population in Nanoparticle Ensembles. *Nanoscale* **2014**. DOI: 10.1039/C4NR01504E.