

# researchtrends

## IN THIS ISSUE:

PAGE 2

The value of bibliometric measures



## Adding perspective to US higher education rankings

The Carnegie Classification meets an important need in US higher education. Without ranking colleges or universities, it compares a broad range of criteria, and publishes data that enable researchers, policymakers and even prospective students to perform further qualitative analysis of US higher education institutions.

PAGE 3 - 4

Research trends



## Is nanoscale research slowing down?

The compound growth rate of papers published in nanoscience and nanotechnology increased 16% between 1996 and 2006. This growth now seems to be slowing down. Éric Archambault, President of Science-Metrix, examines the data to find out why.

PAGE 4 - 5

Country trends



## United States research output continues to decline

The US has long led the global knowledge economy, but the last decade has seen its dominant position weakening. What is causing this decline?

PAGE 5 - 6

Expert opinion



## The h-index and its variants: which works best?

Numerous variants of the h-index have been developed since Jorge Hirsch first proposed the h-index in 2005. Yet, while increasingly refined bibliometric tools can only be a good thing, are so many indices necessary and valuable, or just confusing?

PAGE 7

Why did you cite...?



## Why was I TopCited...?

In this section, we ask authors about what motivates citation patterns. This issue we talk to four authors whose papers were highly cited to discover the ingredients of success.

Welcome to issue 5 of Research Trends, which focuses on the United States. The US remains the world's dominant scientific nation, but research has demonstrated that its share of scientific articles published in peer-reviewed journals has been in decline over the past decade. We examine what might be causing this decline and how it could be reversed.

We also look at the Carnegie Classification, which compares – rather than ranks – US higher education institutions. Based on a broad range of criteria, the data produced by the classification is useful for researchers and policymakers in carrying out further qualitative analysis.

In a stand-alone article, Dr. Lutz Bornmann expands on research he conducted into the h-index and nine of the variants that have been developed since 2005. How valuable are they?

If you would like to comment on any of the topics covered, please use our [feedback](#) facility.

Kind regards,

The Research Trends Editorial Board

## Did you know?

### Researchers are surprised by their own findings

Discovery is the watchword of scholarly research, and new observations serve to propel scientific knowledge ever forward. Yet the lay observer might be interested to learn just how often the authors of research papers are surprised by their own findings. A search of the Scopus database for articles containing the word root 'surpris\*' in the title, abstract or keywords retrieves almost 6,000 articles published in 2006, more than 0.5% of all articles that year. In fact, looking at this proportion over the last decade, authors appear to be about 3% more surprised by their own research every year!

A disciplinary breakdown of the 'surprising' articles identified in 2006 shows that the greatest proportion are published in Economics, Econometrics and Finance (1.25% of articles), while the least are found in the Energy and Engineering fields (0.11% and 0.17% respectively). Authors in the Agricultural and Biological Sciences are surprised at the average rate of 0.51% of articles.

Researchers should take heart, as one recent example of such a 'surprising' research finding was reported by Andrew Fire and Craig Mello in 1998 [1] and led to the 2006 Nobel Prize in Physiology or Medicine. It has been cited almost 3,400 times to date.

[1] Fire, A. (1998) "Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*", *Nature*, Vol. 391, No. 6669, pp. 806-811.

## The value of bibliometric measures



# Adding perspective to US higher education rankings

In the United States, higher educational establishments are ranked by several organizations, including publishers, universities and other research institutions. However, in recent years, the *US News and World Report* rankings in particular have come under criticism from the academic community for their failure to accurately represent educational excellence. They often rely on subjective assessments of reputation, and emphasize inputs and resources over what happens on campus.

Many colleges have, as a result, chosen to boycott the ranking schemes. They argue that the work of their institutions is too complex, varied and ephemeral to be reduced to simple output measures, says Alexander McCormick, previously Senior Scholar at the Carnegie Foundation for the Advancement of Teaching and now Director of the National Survey of Student Engagement at Indiana University School of Education (1). Yet according to him, efforts to improve higher education depend on comparative assessment data.

### Classification without bias

Without some form of objective assessment of what an institution has to offer, researchers, policymakers, institutional personnel and even prospective students and their parents are left with little to go on when comparing institutes of higher education.

In this respect, the Carnegie Classification meets their needs. Without producing a ranking table of US colleges, the Carnegie Classification compares a broad range of criteria which seek to answer the questions:

1. What is taught?
2. To whom is it taught?
3. In what setting does this teaching take place?

The data produced provide a useful starting point for researchers, policymakers and institutional personnel to perform further qualitative analysis.

### Updated criteria add nuance

The Carnegie Classification was developed in 1970 to support research on US higher education, and that continues to be its primary purpose. It classifies all accredited degree-granting colleges and universities in the United States. Carnegie reports were issued in 1973, 1976, 1987, 1994, 2000 and 2005/06.

In response to the growing need for more flexible and representative comparative data, the Carnegie Foundation changed its assessment criteria in 2005, updating its basic analysis and adding a further five all-inclusive classifications: undergraduate instructional program, graduate instructional program, enrolment profile, undergraduate profile, and size and setting. At the same time, it added an elective classification – community engagement – and is piloting another voluntary criterion – undergraduate education – in 2008 (2).

While the basic classification may continue to serve as a key point of reference and analytic tool for many users, McCormick believes the updated classifications will add much-needed texture and nuance.

“Colleges and universities are no longer characterized on the basis of a single view of what they do,” he explains. “For example, a research university’s ‘portrait’ will capture not only its commitment to graduate education, but also the nature of its undergraduate program, the characteristics of its undergraduates, the relative size of undergraduate and graduate populations and the absolute size and residential character of the campus.” (2)

He continues: “As the simple, mutually exclusive terminology of the traditional classification gives way to a richer multidimensional framework, our conception of institutional similarities and differences will necessarily become more nuanced. We believe the added flexibility [of the new classifications], and the responsibility that goes with it, will enhance research and policy development, as well as campus conversations about institutional priorities and distinctiveness.” (2)

For more information on the Carnegie Classifications or to look up an institution, please click [here](#).

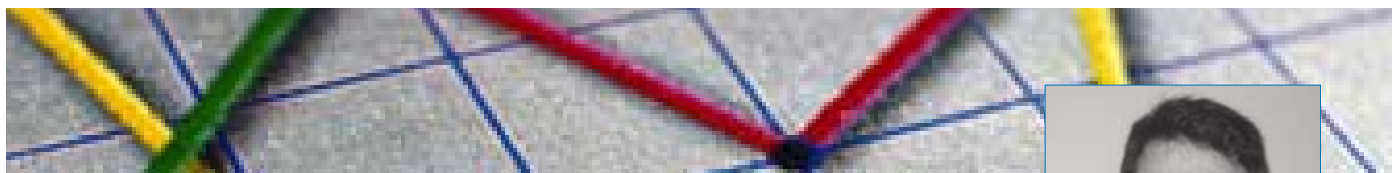
#### References:

(1) McCormick, A. (2007), “First, Do No Harm”, *Carnegie Perspectives*

(2) McCormick, A. (2007), “A New Set of Lenses for Looking at Colleges and Universities”, *Carnegie Perspectives*

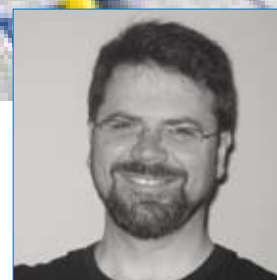
See also McCormick, A. (2007), “Hidden in Plain View”, *Inside Higher Ed*.

Research trends



# Is nanoscale research slowing down?

Éric Archambault, President Science-Metrix



The number of papers published in nanoscience and nanotechnology has increased rapidly in the last 11 years. The compound annual growth rate (CAGR) from 1996 to 2006 was 16% and papers indexed in Scopus rose from 16,000 to 64,000. Calculating growth is never easy, as there is seldom one estimation method that gives the whole picture. For instance, we know the CAGR value is influenced by the growth in the coverage of Scopus itself. It is also necessary to consider that science may be growing and that nanoscience could be riding on that wave. Consequently, it is customary in bibliometrics to examine how a field is evolving as a percentage of the database used, in this case Scopus. One can see that in 1996, about 1.5% of papers indexed in Scopus were about nanoscience and this has increased to 4.2% of the database's contents in 2006, close to a threefold increase (see Figure 1).

### Slowing growth

In light of these data, one can safely say that nanoscience research grew rapidly in the last decade. However, growth appears to be slowing in both absolute and relative terms. Bearing in mind that not all papers had been included in the database for 1996 when these data were calculated, it is more relevant to examine the curve that presents data in percentage terms. This curve is starting to look s-shaped, which suggests that growth has started to slow down. How can we explain this phenomenon? We will present four hypotheses, the most parsimonious first, followed by those of increasing complexity.

The first hypothesis is that this slowdown is only a random variation along the exponential growth path observed. In addition to obtaining a strong R-Square value, it is indeed a requirement to accept that an exponential growth curve (or exponential regression curve) offers a robust model that observed data points be distributed randomly on each side of the curve. Consequently, it is possible that in the future, additional data points will show that a slowdown had not actually occurred – it was merely the normal occurrence of yearly variation.

### From exponential to linear growth

The three remaining hypotheses are compatible with the observation that nanoscience research has started to look like an s-shaped curve, a type of growth process observed in biological but also in scientific, technological and social systems. This is not really surprising, as a system can never grow indefinitely. There are also limits to growth, as the means necessary to produce anything are always finite. For instance,

there are only a certain number of researchers who can work on nanotechnology and once they are all mobilized and have learnt everything they can to perform and publish research in this field efficiently, their rate of publication will inevitably stabilize. For this reason, the second hypothesis is simply that nanotechnology is maturing, and that future growth will be more linear than exponential.

A third hypothesis would be linked with the idea that the field may be momentarily experiencing a slowdown and that a further boom cycle is

forthcoming. If this model is appropriate, it would be akin to what German researcher Ulrich Schmoch calls "double-boom

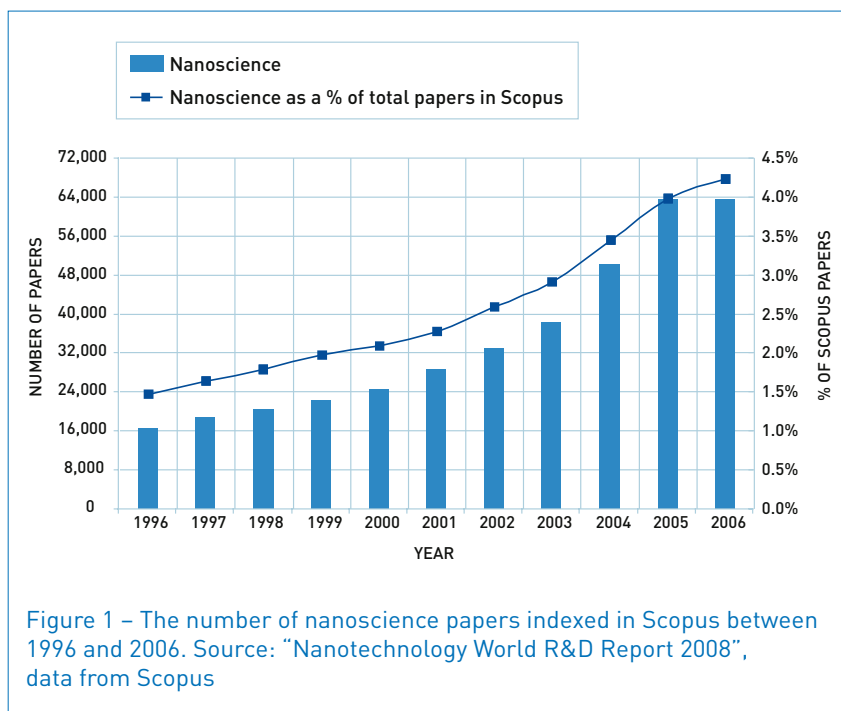


Figure 1 – The number of nanoscience papers indexed in Scopus between 1996 and 2006. Source: "Nanotechnology World R&D Report 2008", data from Scopus

Continued from page 3

cycles" (1). Schmoch argues that several domains, such as robotics and immobilized enzymes, evolved through an initial period of great patenting activity, followed by a slowdown and a less publicized follow-up cycle of high-frequency patenting. It is therefore possible that nanotechnology would be undergoing a slowdown after an initial boom in publishing and patenting and that a second boom would be forthcoming.

**Obliteration by incorporation**

The fourth hypothesis is that the use of nanoscale R&D is increasingly incorporated into mainstream S&T, which means that it is no longer being specifically mentioned as often or as prominently by researchers in their scientific publications and patents. This would mean that nanotechnology is undergoing a process analogous to "obliteration by incorporation". This concept, described by Robert Merton (2), suggests a process by which the origin of an idea is forgotten due to prolonged use, as it enters the mainstream language of academic disciplines, and

is thereafter no longer linked with its originators. An analogous situation in the development of a scientific discipline is provided by developments in the field of genomics. In this field, it was common in the 1990s to mention the use of the polymerase chain reaction (PCR) method in the title and the abstract of papers, but this was subsequently obliterated by incorporation. Indeed, it is now considered obvious that if one carries out gene sequencing, the PCR method is used. The same may be true of nanotechnology today and researchers may not mention that they are working at the nanoscale level as it may be deemed obvious to knowledgeable practitioners.

For further information about these reports, please [click here](#).

References:

- (1) Schmoch, U. (2007) "Double-boom cycles and the comeback of science-push and market-pull", *Research Policy*, Vol. 36, issue 7, pp. 1000-1015.
- (2) Merton, R.K. (1949) *Social Theory and Social Structure*. New York: Free Press.

Country trends



# United States research output continues to decline

An extensive body of research has consistently demonstrated that the US share of scientific articles published in peer-reviewed journals has been in decline over the last decade (see Figure 1). This has typically been ascribed to the effect of the developing knowledge economies of **China** and the four Asian Tiger nations, Taiwan, Singapore, Hong Kong and South Korea, and has not been considered a policy concern (1). However, since the 1990s the absolute number of articles published by US-based researchers has plateaued (see Figure 2).

This flattening of scholarly output has been confirmed by the "Science and Engineering Indicators" (SEI) 2008 (2), published in January by the US National Science Board. This biennial report contrasts this finding with strong annual growth in research funding in the US over the same period, from US\$200 billion in 1997 to around US\$340 billion (or 2.6% of GDP) in 2006. A companion policy statement, "Research and Development: Essential Foundation for U.S. Competitiveness in a Global Economy" (3), nevertheless calls for a "strong national response" by further increasing the level of US government funding for basic research.

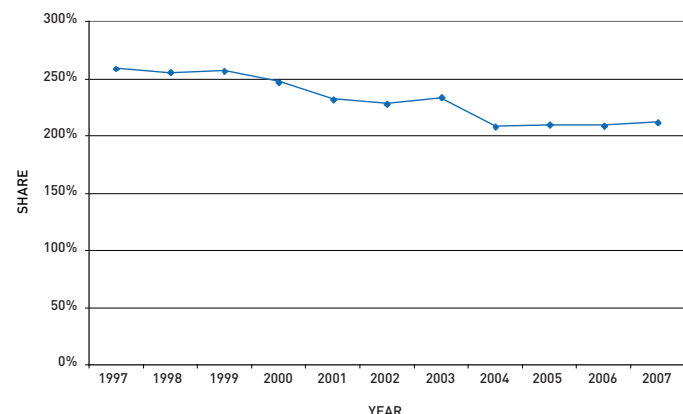


Figure 1 – Share of world articles published by US researchers, 1997–2007. Source: Scopus

Continued from page 4

Despite these trends in article output, the SEI 2008 report demonstrates that the US continues to produce the best-cited research in the world, as indicated by its dominant share of articles in the top 1% of cited articles across all fields. This finding is borne out by comparing the **h-index** of the US with those of selected world regions (see Figure 3).

By any measure, the US remains the world's dominant scientific nation. The question facing government policymakers in the age of knowledge-based economies is: for how much longer?

References:

[1] Hill, D., Rapoport, A.I., Lehming, R.F., and Bell, R.K. (2007) "Changing U.S. output of scientific articles: 1988–2003", National Science Foundation special report.

[2] "Science and Engineering Indicators 2008", National Science Board report.

[3] "Research and Development: Essential Foundation for U.S. Competitiveness in a Global Economy", National Science Board report.

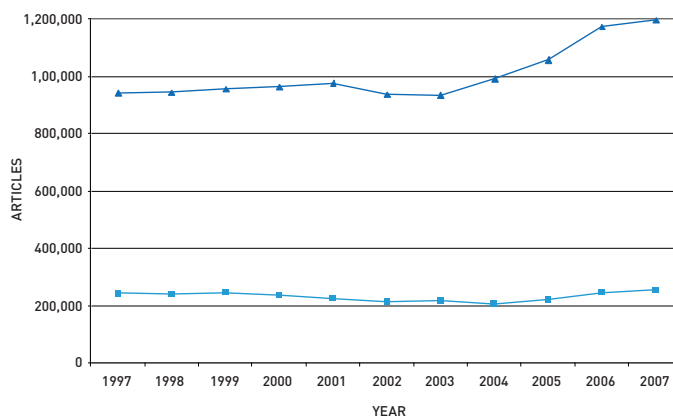


Figure 2 – Number of articles published by US researchers (light blue) versus world (dark blue), 1997–2007. Source: Scopus

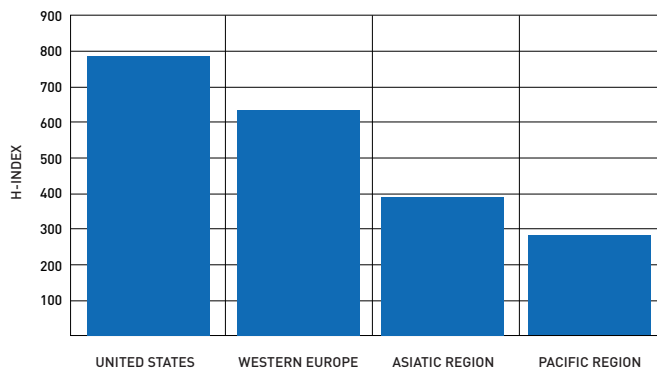


Figure 3 – H-index of US versus selected global regions. Here, the h-index defines the number of documents published in the period 1996–2006 that receive the same or greater number of citations during the same period. Source: [SCImago SJR – SCImago Journal & Country Rank](#)

## Expert opinion



# The h-index and its variants: which works best?

Dr. Lutz Bornmann

The h-index was originally proposed by Jorge Hirsch in 2005 to quantify the scientific output of an individual researcher. It was conceived as an improvement on previous indices, which tended to focus on the impact of the journals in which the researcher had published, and

so assumed that the author's performance was equivalent to the journal's average. If a scientist's publications are ranked in order of the number of lifetime citations they have received, the h-index is the highest number, h, of their papers that have each received at least h citations.

Continued from page 5

### Room for improvement

The **h-index** quickly gained widespread popularity. This is largely due to the fact that it is conceptually simple, easy to calculate and gives a robust estimate of the broad impact of a scientist's cumulative research, explains Dr. Lutz Bornmann, post-doctoral researcher active in bibliometrics, scientometrics and peer-review research at ETH Zurich, the Swiss Federal Institute of Technology [1].

However, the h-index has received some criticism, most notably:

- It is not influenced by citations beyond what is required for entry to the h-defining class. This means that it is insensitive to one or several highly cited papers in a scientist's paper set, which are the papers that are primarily responsible for a scientist's reputation.
- It is highly dependent on the length of a scientist's career, meaning only scientists with similar years of service can be compared fairly.
- A scientist's h-index can only rise (with time), or remain the same. It can never go down, and so cannot indicate periods of inactivity, retirement or even death.

### Quantity versus impact

1. Quantity of the productive core: the h, g and h(2) indices and the m-quotient describe the productive core of a scientist's output and show the number of papers in the core.
2. Impact of the productive core: the a, m, r, ar and h<sub>w</sub> indices show the impact of papers in the core. This is closer to peer-assessment results.

Variants of the h-index that have been developed in an attempt to solve one or more of its perceived shortcomings include the m-quotient, **g-index**, h(2)-index, a-index, m-index, r-index, ar-index and h<sub>w</sub>-index. Hirsch himself proposed the m-quotient, which divides the h-index by the number of years a scientist has been active, thereby addressing the problem of longer careers correlating with higher h scores.

### Two types of index

For Bornmann, the value of a bibliometric index lies in how closely it predicts the results of peer assessment. In a paper published in the *Journal of the American Society for Information Science and Technology*, of which he was co-author [1], he analyzed nine indices to find out whether any has improved upon the original h-index, with particular focus on their ability to accurately predict peer assessment.

He discovered that there are two basic types of index: those that better represent the quantity of the productive core (defined as the papers that fall into the h-defining class), and those that better represent the impact of the productive core (see sidebar). In a further study to validate these findings, Bornmann tested his results against 693 applicants to the Long-Term Fellowship program of the European Molecular Biology Organization, Heidelberg, Germany. The study confirmed these two basic types.

This is useful, as the indices that better represent the impact of the productive core agree with the opinions of the applicants' peers. "The results of both studies indicate that there is an empirical incremental contribution associated with some of the h-index variants that have been proposed up to now; that is, with the variants that depict the impact of the papers in the productive core," Bornmann says.

### A balanced approach

Several researchers in the field have suggested that bibliometricians would do well to use several indices when assessing a scientist's output and impact. Bornmann agrees, adding that the results of his research indicate that the best way to combine the

different indices is to ensure that one chooses an index from each category.

"After analysis of all indices, matching the results to the real results of peer assessment, using two indices – one to measure output and another to measure impact – is the closest to peer-assessment results," he explains. "In the future, we definitely need fewer h-index variants and more studies that test their empirical application with data sets from different fields."

Dr. Bornmann's full paper can be found [here](#).

Reference:

[1] Bornmann, L., Mutz, R., and Daniel, H.D. (2008) "Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine". *Journal of the American Society for Information Science and Technology*, Vol. 59, No. 5, pp. 830-837.



## Why did you cite...?



## Why was I TopCited...?

Scopus recently introduced a new service, **Scopus TopCited**, which presents the most-cited articles in 26 different subject areas worldwide.

However, while TopCited can tell us *who* was highly cited, it does not tell us *why* the paper was so popular. To shed some light on what makes a successful paper, we asked four TopCited authors why they think their paper had such an impact.

Dr. David Spergel's paper, "Three-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Implications for cosmology", published in the *Astrophysical Journal* (1), received 504 citations in the subject area Earth and Planetary Science between 2004–2008.

Dr. Spergel believes his paper was highly cited because it "discusses new results from a major scientific mission and the implications for our understanding of cosmology. Also, it includes the basic cosmological parameters, which are widely used in physics and astronomy."

Prof. Younan Xia and Dr. Dan Li's paper, "Electrospinning of nanofibers: Reinventing the wheel?", published in *Advanced Materials* (2), was cited by 379 other papers within Materials Science from 2004 to 2008.

Professor Xia thinks the main reason his paper was highly cited is because it reviews the fundamentals and presents recent advances in the field of electrospinning, serving as a useful reference paper.

He adds, "The recognition of electrospinning by the research community is due to the fact that it is simple, inexpensive and convenient to use; it is applicable to many kinds of materials; and nanofibers, with complex secondary structures, can be readily prepared with modified setups."

Drs. Paul Emsley and Kevin Cowtona's paper, "Coot: Model-building tools for molecular graphics", published in *Acta Crystallographica Section D: Biological Crystallography* (3), was cited by 1,419 papers in the subject area Physics and Astronomy between 2004–2008.

Dr. Emsley explains that his paper was highly cited because it discusses a useful program for structural biologists. The impact of the paper can be attributed to the fact that the program is so widely used. He says, "The use of interactive graphs and plots, particularly for model validation, was novel and provides a simpler environment than had been available before."

On citation habits in general, he adds that some popular journals have started to omit citations in supplementary materials where the experimental details are more fully described. "This badly bites into the citation count for works such as ours and is something of which I, and I presume Dr. Cowtona, disapprove."

Dr. Søren Brunak's paper, "Improved prediction of signal peptides: SignalP 3.0", published in *Journal of Molecular Biology* (4), was cited by 1,242 papers in the subject area Immunology and Microbiology in the time period 2004–2008.

Dr. Brunak says his paper was highly cited for two reasons. First, his tool for signal peptide prediction, SignalP, is unique in that it produces a higher than expected quality of predictions. Second, **SignalP** is easy to use and freely available on the Internet.

#### References:

[1] Spergel, D.N. (2007), "Three-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Implications for cosmology", *Astrophysical Journal*, Supplement Series, Vol. 170, Issue 2, pp. 377–408.

[2] Li, D., Xia, Y.N. (2004), "Electrospinning of nanofibers: Reinventing the wheel?", *Advanced Materials*, Vol. 16, Issue 14, pp. 1151–1170.

[3] Cowtona, K. and Emsley, P. (2004), "Coot: Model-building tools for molecular graphics", *Acta Crystallographica Section D: Biological Crystallography*, Vol. 60, Issue 12 I, pp. 2126–2132.

[4] Brunak, S. (2004), "Improved prediction of signal peptides: SignalP 3.0", *Journal of Molecular Biology*, Vol. 340, Issue 4, pp. 783–795.