Theory-Changing Breakthroughs in Science: The Impact of Research Teamwork on Scientific Discoveries

Jos J. Winnink
Center for Science and Technology Studies (CWTS), Leiden University, Wassenaarseweg 62A, Leiden 2333 AL, The Netherlands; Leiden University Dual PhD Center The Hague, P.O. Box 13228, The Hague 2501 EE, The Netherlands; NL Patent Office, A Division of Netherlands Enterprise Agency, P.O. Box 10366, The Hague 2501 HJ, The Netherlands. E-mail: winninkjj@cwts.leidenuniv.nl

Robert J. W. Tijssen
Center for Science and Technology Studies (CWTS), Leiden University, Wassenaarseweg 62A, Leiden 2333 AL, The Netherlands and DST-NRF Center of Excellence in Scientometrics and STI Policy, Stellenbosch University (South Africa), Stellenbosch, Western Cape, South Africa. E-mail: tijssen@cwts.leidenuniv.nl

Anthony F. J. van Raan
Center for Science and Technology Studies (CWTS), Leiden University, Wassenaarseweg 62A, Leiden 2333 AL, The Netherlands. E-mail: vanraan@cwts.leidenuniv.nl

We have developed and tested an evidence-based method for early-stage identification of scientific discoveries. Scholarly publications are analyzed to track and trace breakthrough processes as well as their impact on world science. The focus in this study is on the incremental discovery of the ubiquitin-mediated proteolytic system in the late 1970s by a small international team of collaborating researchers. Analysis of their groundbreaking research articles, all produced within a relatively short period of time, and the network of citing articles shows the cumulative effects of the intense collaboration within a small group of researchers working on the same subject. Using bibliographic data from the Web of Science database and the PATSTAT patents database in combination with expert opinions shows that these discoveries accumulated into a new technology. These first findings suggest that potential breakthrough discoveries can be identified at a relatively early stage by careful analysis of publication and citation patterns.

Introduction

The Context of Scientific Breakthroughs

Can scientific breakthroughs be identified within a few years after the fact? What kind of sudden changes and distinctive structural developments do these events leave behind in the research literature? We study the observable effects of these breakthroughs on the research community. Our bibliometric approach relies on systematic large-scale searches within the world-wide scholarly and patent literature. In most cases, the moment in time when a breakthrough occurred cannot be precisely pinpointed—not even in retrospect. External observers and analysts, that is, researchers who are not experts in the field under study, have to resort to indirect evidence and proxies. Assembling information and time-series data from bibliographic databases such as the Web of Science[1] (WoS) or SCOPUS,[2] enables nonexpert analysts to identify significant short-term changes in the general pattern of knowledge creation processes and communication patterns in the field of science within 2 years after publication of key research articles. By adopting this method, our early-detection approach differs from work carried out by others; Arbesman (2010), for instance, focuses on the ease with which scientific discoveries can be made; a Delphi approach applied to Japanese-German cooperation is presented in Breiner, Cuhls, and Grupp (1994); C. Chen, Y. Chen, Horowitz, Hou, Liu, and Pellegrino (2009) model transformative discoveries focus on connections across structural holes in network representations of scholarly knowledge in scientific discovery; the focus of Julius,

Received October 6, 2014; revised December 27, 2014; accepted October 29, 2014

 • Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.23505

[1]Web of Science is a product of Thomson Reuters.
[2]SCOPUS is a product of Elsevier B.V.

Berkoff, Strack, Krasovec, and Bender (1977) is on using expert knowledge to judge, at an early stage, whether a discovery is expected to be of importance for new technology that will effect research and development (R&D) in the pharmaceutical sector; Leydesdorff and Rafols (2011) use geographical and cognitive diffusion to analyze emerging topics; Martin (1995) introduces a model of a foresight process to help identify scientific developments of strategic importance; Small (1977) answers the question, "Does the citation picture of a specialty's development match the perception of the specialists?"; Ponomarev, Williams, Hacket, Schnell, and Haak (2014a) use forecasting models to predict future citation patterns of individual articles to determine whether these articles are expected to exceed a high citation threshold after 5 years; in Ponomarev, Lawton, Williams, and Schnell (2014b), the usability of interdisciplinary in the subject categories, and geographical diversity, is studied to determine whether these measures can be used to identify potential breakthrough articles.

Discoveries, breakthroughs, and subsequent shifts in scientific "paradigms" are characteristic of revolutionary science according to Kuhn, who distinguishes evolutionary "normal" and "revolutionary" science (Kuhn, 1962). These structural shifts are marked by the gradual introduction and application of new theories, insights, and concepts and their increasing dominance, and thus structurally changing the "cognitive" fabric of a research domain (Andersen, Barker, & Chen, 2006). However, science does not always progress in such a neat linear and cumulative way. Changes in theoretical frameworks proliferate slowly as the relevant scientific research community needs to be convinced to alter its views and approaches. Radically novel approaches, new information, and breakthrough findings, which are incompatible with the current dominant theoretical framework and beliefs within a research field, may suddenly appear on the scene and revolutionize the cognitive structure of this field of research within a brief span of time. Wray (2011, p. 202) states: "According to Kuhn's mature view, a new theory is developed in a field in an effort to account for an anomaly that the accepted theory was unfit to account for." Wang, Song, and Barabási (2013, p. 127) conclude: "Paradigm-changing discoveries have notoriously limited impact."

There is no generally accepted description, let alone a universal definition, of the term *breakthrough* that can count on full support throughout the international scientific community. We rely on Hollingsworth (2008, p. 317) who states: "A major breakthrough or discovery is a finding or process, often preceded by numerous small advances, which leads to a new way of thinking about a problem." Building on this definition, we adopt Koshland's "Cha-Cha-Cha" theory of scientific discovery, which distinguishes three general types of scientific discoveries (Koshland, 2007, p. 761). He argues: "In looking back on centuries of scientific discoveries, however, a pattern emerges which suggests that they fall into three categories—Charge, Challenge, and Chance —." Where *charge* discoveries can be considered normal science in the Kuhnian sense; the *challenge* discoveries and *chance* discoveries are examples of revolutionary science marked by sudden changes in knowledge creation processes. Science progresses by building on the achievements of other scholars, as is expressed by the metaphor "If I have seen further, it is by standing on the shoulders of giants."[3] But does this general metaphor apply to scientific discoveries of the challenge type? According to Koshland (p. 761): "Challenge discoveries are a response to an accumulation of facts or concepts that are unexplained by or incongruous with the scientific theories of the time." We use the term "theory-changing breakthroughs" for discoveries of this type given that they lead to alterations in theoretical concepts.

In some cases, the challenge lies within a small closely knit, but geographically dispersed, group of researchers facing the same puzzling experimental results. As the research within this team progresses, their publications will cite one another's work. Several incremental "minor" discoveries are made and published. The accumulation of their efforts, the teamwork discovery, will only be recognized and publicized by these "insiders." Their resulting (joint) research publications will therefore at first receive only a modest number of "within-team" citations. The further diffusion of the ensuing new theoretical insights and the proliferation of their applications will result in a growing citation network. We hypothesize that citation patterns of such revolutionary teamwork-based challenge discoveries differ from "evolutionary" breakthroughs that are congruent with existing theories. We expect a sociocognitively dense, but low-level, citation network within the first couple of years after the discovery, followed by a sudden increase in citation numbers across the wider scientific community. This assumption is put to the test in our case study of a Nobel Prize–winning discovery, which is introduced in the next section.

Several studies focus on research teams and their role in the production of new knowledge. Weinberg (1970) focuses on the role of big team science; the focus in Guimerà, Uzzi, Spiro, and Amaral (2005) is on mechanisms of team assembly. International scientific collaboration is discussed in Luukkonen, Persson, and Sivertsen (1992); Jones, Wuchty, and Uzzi (2008) focus on multiuniversity research teams. Wuchty, Jones, and Uzzi (2007) discuss the growing dominance of scientific research teams in knowledge production. Darden and Maull (1977) focus on the role that "interfield theories" play in the production of new knowledge; Upham, Rosenkopf, and Ungar (2010) discuss the advantages and disadvantages of "schools of thought" in relation to the creation of new knowledge; the relation between collaboration and creativity is analyzed in Uzzi and Spiro (2005); Uzzi, Mukherjee, Stringer, and Jones (2013) discuss the role of "boundary spanning ideas" and especially

[3]This metaphor is ascribed to Isaac Newton, but has to be attributed to Bernard of Chartres and was first recorded in the 12th century (Merton, 1965, p. 37).

DOI: 10.1002/asi

"atypical combinations." In this study, we focus on the question, "Is there a cognitive reason for this, as we know now, fruitful collaboration?"

Discovery of the Ubiquitin-Mediated Proteolytic System

The judgment of scientific experts in the specific field and on argumentation provided by other knowledgeable reviewers, in this case the Nobel Prize committee, may be relied upon as a means of judging and contextualizing the importance of a scientific discovery. In 2004, the Nobel Prize for Chemistry was awarded to Aaron Ciechanover, Avram Hershko (both scientists at the Technion, the Israel Institute of Technology at Haifa), and Irwin Rose (Fox Chase Cancer Center, Philadelphia, USA) for the discovery of the ubiquitin-mediated proteolytic system, which is crucial for the proper functioning of eukaryotic[4] cells. Nath and Shadan (2009, p. 421) illustrate the function and importance of the ubiquitin-mediated proteolytic system in the following way:

> The destruction of proteins is as important as their synthesis for the maintenance of protein homeostasis in cells. In eukaryotes, the ubiquitin–proteasome system is responsible for most of this protein degradation: the small protein ubiquitin acts as a death warrant, tagging and targeting other proteins to the large proteolytic chamber of the proteasome. (. . .) It is now known that ubiquitin-mediated destruction plays a crucial part in cell-cycle regulation, DNA repair, cell growth and immune function, as well as in hormone-mediated signalling in plants. (. . .) Given the central role of the ubiquitin system in diverse cellular processes, it is not surprising that its dysfunction contributes to cancer and to neurodegenerative and immunological disorders. An understanding of the ubiquitin system is therefore important in devising treatments for such diseases (. . .)

According to the Nobel Prize Committee (2004, p. 3), "The breakthrough came in 1980. It was described in two papers that were both communicated on 10 December 1979 to the journal *Proceedings of the National Academy of Sciences* of the USA." It continues: "The unravelling of the Ubiquitin proteolytic system is not an exception to the rule that scientific discoveries are based on findings of others and that it can take a long period between the first preliminary findings and the breakthrough discovery."

The Ciechanover (2009, p. 1) article concurred with Koshland's view of challenge discoveries:

> (. . .) Could the Ubiquitin system have been discovered earlier? Possibly yes. (. . .) As we now know, the system was not discovered by chance but rather by challenge—mostly as a natural response to developments in the field (. . .) A new system and concept(s) were needed to explain all of these new findings and assumptions, gathering them under a unifying umbrella. (. . .)

[4]Eukaryotes have cells in which the genetic material is DNA in the form of chromosomes and is contained within a distinct nucleus. Eukaryotes include all living organisms other than the eubacteria and archaebacteria.

Hypothesis, Research Questions, Data, and Method

Hypothesis and Research Questions

The main hypothesis we want to test using the discovery of the ubiquitin system as a test bed is "the observable changes in the sociocognitive structure, that are visible as changes in the citation network of relevant bibliographic information, enable the detection at early stage of a scientific discovery that results in an identifiable shift in theoretical concepts." This hypothesis leads to the following research questions:

- How does the metaphor "science progresses by standing on the shoulders of others" manifest itself in this discovery?
- What changes in the sociocognitive structure are visible in the period around a theory-changing breakthrough?
- Does the citation network of relevant publications support the idea of a challenge breakthrough by showing only modest numbers of citations for the three landmark publications?
- Is it possible to perceive a gradual increase in the multidisciplinarity of the citing publications?
- Can publications that are not directly manifest, but that play an important role in the developments following the discovery, be identified in the citation network around the landmark publications?
- Does the citation network show a "clustering" of publications from a small group of scientists over a short period?
- Does the constraint we impose in using brief windows of time prevent the identification at an early stage of a theory-changing breakthrough?

Data

As data sources, we used the online version of Thomson Reuters's WoS database as well as the Center for Science and Technology Studies's (CWTS's) in-house version for the scholarly publications, and the European Patent Office's (EPO's) Worldwide Patents Statistics database (PATSTAT), the October 2012 edition, for the patent publications. All documents covered by the WoS up to the first quarter of 2013 and all patent documents published up to August 2012 were included in our search for relevant publications. In addition, we referred to the original "paper" publications for those publications not covered by the WoS to reveal the citation information needed in our research.

Applying "ubiquitin" as a search topic to the WoS resulted in 28,778 relevant research publications of the document types "article," "letter," and "proceedings paper" spanning the period 1975–2013. Searching PATSTAT for patent publications related to "ubiquitin" resulted in 1,522 documents belonging to 682 distinct DOCDB[5] patent families that were applied for in the period 1985–2012.

[5]DOCDB patent families are sets of equivalent patent publications referring to the same patent invention.

DOI: 10.1002/asi

Methods

We start with an analysis of the evolution of the research field by looking at the trends with regard to the numbers of scholarly and patent publications. CWTS uses 250 distinct subject categories based on the subject categories Thomson Reuters uses in the WoS database. Reclassification of all documents is carried out to reflect the most actual version of the classification scheme. In this context, a group of documents is called “monodisciplinary” if all documents belong to one and the same subject category, and “multidisciplinary” if multiple subject categories are assigned to the publications in the group. We look at the evolution of the multidisciplinarity of the publications in this research field by analyzing the evolution of the number of distinct subject categories assigned to the publications, and we conclude with the analysis of the citation networks centered on the breakthrough articles.

Results

Publication Trends

Figure 1 shows the trend for scholarly publications as well as for patent publications. Several significant and visible changes in the first few years can be seen after 1980, when the breakthrough articles were published. A significant increase is clearly visible starting in 1983. Combining both figures, we see that this increase marks a pivotal moment, after which a continuously increasing number of scholarly publications can be found. Following this pivotal moment in 1983, we notice the appearance of patent publication from 1988 onward, indicating the gradual evolution of the scientific ideas into technology. The first patent publication found in this study was applied for in 1985. The long-term trend for ubiquitin-related scholarly publications shows an exponential growth rate until it starts to deviate around 2003.

The trends shown in Figure 1 also illustrate that the ubiquitin field matured into a substantial research area around 13 years after the main discovery. This moment is marked by the steep rise from 1990 to 1991. The number of patent families shows a significant and large increase in the period 1998–2000. This increase gives the impression of a switch to a new or improved technology, resulting in a higher level of patent applications.

Multidisciplinarity of the Research Field

As a measure of multidisciplinarity, we counted the number of different CWTS subject categories that correspond to the WoS-indexed publication output per publication year. As ubiquitin-related research becomes more dispersed across related science fields, the number of publications in journals assigned to more and different subject categories will increase. This process reflects the pervasive spread of a new paradigm across the sciences, as well as the wider use of the knowledge on the ubiquitin system. Figure 2 shows an overall fairly linear increase in the number of CWTS subject categories, totaling more than 100 in recent years. Deviations from this linear trend are visible at a more detailed level. From 1980 until 1985, the number of subject categories stayed more or less at the same level. In 1987, a jump took place and the number reached a plateau and stayed at this level until the steep rise in 1991. From 1991 onward, a more or less linear growth is visible.

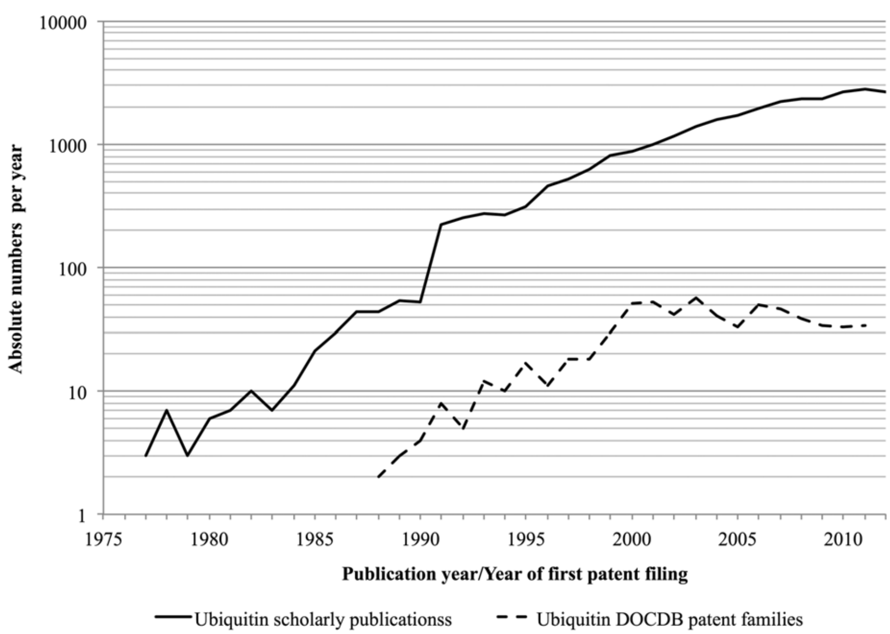

FIG. 1. Numbers of scholarly publications (articles, letters, and proceedings papers), patent applications (DOCDB patent families) per year (1975–2012).

DOI: 10.1002/asi

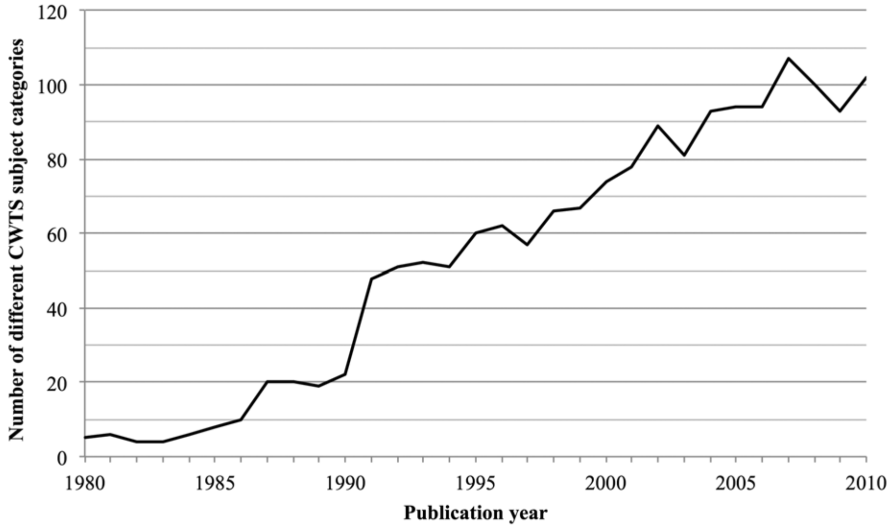

FIG. 2. Number of different CWTS subject categories assigned within a year to ubiquitin-related publications (1980–2010).

Citation Network

Network analysis. Aaron Ciechanover (Ciechanover, 2009, p. 2) writes as follows: "(. . .) but it is nevertheless possible to point to a single article, or a handful of articles, that laid the foundation to the discovery. This is the case with the Ubiquitin system, where the pioneering findings were described in a single article published in this journal (. . .)." The publication referred to is Ciechanover, Hod, and Hershko (1978) that was published almost 2 years before the two publications Ciechanover, Heller, Elias, Haas, and Hershko (1980a) and Hershko, Ciechanover, Heller, Haas, and Rose (1980). According to the Nobel Prize Committee, these latter two publications together describe the paradigm-involving breakthrough. We refer to these three publications as the "landmark publications."

To analyze the role the breakthrough articles played in the development of the ubiquitin field, we constructed an extended citation network consisting of publications directly citing the landmark publications and of publications cited by these three publications. In constructing this network, only publications containing original research are included; reviews, editorials, corrections, and the like were therefore omitted. The citation network contains 174 publications and spans the period 1951–1986. To find the publications that, based on citation relations, were the most influential, we included all 1,023 citation relations between documents within this network. For every node in this network, the "indegree" was calculated and used as a measure of importance of the node. Indegree indicates the number of times a publication is cited by publications within this time-constrained network, and it is therefore less than the overall number of times a publication is cited according to the WoS database.

The publications are divided into two groups. One group consists of publications written by authors and coauthors of the Technion Institute and Fox Chase Cancer Center (Technion–Fox Chase group), and a second group consists of publications by authors not in the Technion–Fox Chase group. We present key figures for this network in Table 1. The indegree divides the publications into two distinct groups. There are five publications in this network, all from the Technion–Fox Chase group, with an indegree above 30. All other publications have an indegree of 15 or less, meaning that these publications play a less prominent role in ubiquitin research. The distribution of the publications within this network with respect to their indegrees is given in Figure 3. The five publications with the highest indegree stand out noticeably and are clearly marked. Two of these five publications, Ciechanover, Elias, Heller, Ferber, and Hershko (1980b) and Wilkinson, Urban, and Haas (1980), became visible as concentration points in the extended citation network owing to the fact that they were published shortly after the Nobel Prize–winning publications.

TABLE 1. Key figures in the citation network for the period 1951–1986 with Ciechanover et al. (1978), Ciechanover et al. (1980a), and Hershko et al. (1980) as focus points.

Number of nodes	174
Number of vertices (links) within this network between the nodes	1,023
Publication	**Indegree**
Hershko et al. (1980)	83
Ciechanover et al. (1980a)	53
Wilkinson et al. (1980)	42
Ciechanover et al. (1978)	39
Ciechanover et al. (1980b)	33

Citation network of the ubiquitin discovery. To investigate the direct and indirect cognitive relations between these three publications, we created a citation network for the

DOI: 10.1002/asi

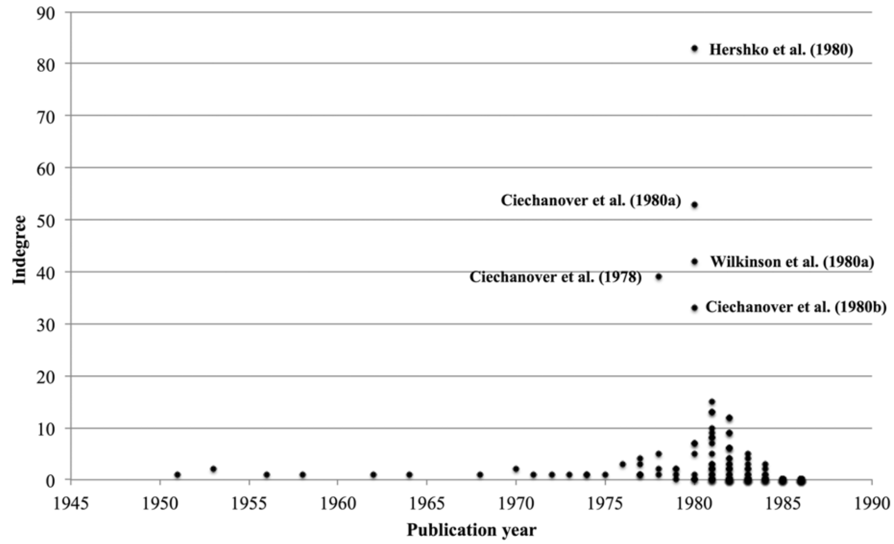

FIG. 3. Indegree for all publications within the citation network (1951–1986). The dots represent individual publications. The labeled dots represent the five publications with an indegree above 30.

discovery of ubiquitin (Figure 4). The network is based on the citation links for the landmark publications. The opaque disks present these three publications. Open circles represent publications with an author or coauthor from the Technion Institute. Non-Technion publications are represented by the name of the first author and the year of publication. The publications referenced by these three publications are also included. Continuous lines indicate citation relations between documents authored or coauthored by researchers from the Technion Institute.

Dashed lines in this figure indicate references to non-Technion publications. The citation relations indicate that the research was directly based on scientific knowledge published from 1951 onward. The publication in 1951 is Lowry, Rosebrough, Farr, and Randall (1951), which is a well-known example of a very highly cited (>300,000) publication (Moed, 2005, p. 82).

Figure 5 shows an annotated detail of Figure 4 and is based on the publications from the two institutes up to the discovery publications. Authors and coauthors of the publications, their affiliations, and citation relations are shown. The octagons represent affiliations, ellipses represent authors, and disks publications.

The affiliations and the authors are placed on the left-hand side (organizational structure); the right-hand side (time-dependent citation structure) contains the publications with the time dimension on the horizontal axis. The opaque ellipsis represents the three Nobel Prize laureates; the opaque disks represent the landmark publications. These two figures clearly show that the Technion Institute is at the center of the discovery of the ubiquitin proteolityc system. The scientific knowledge in the landmark publications comes from publications that are closer on a timeline to one another when approaching 1980; the citation network becomes denser on the time dimension.

The sociocognitive bridge: the Technion–Fox Chase research team. Cooperation between the Technion Institute and the Fox Chase Cancer Center (Ciechanover, 2014) was sought by the researchers at the Technion Institute to solve the problem of the involvement of ATP[6] (energy) in the ubiquitin process; Dr. Rose's group offered fundamental help in solving this issue.

Ciechanover, Herskho, Rose, Haas, and Wilkinson formed the core of this research team. The first publication related to ubiquitin in which at least one author of the core research team from each institute is involved is Rose, Warms, and Hershko (1979); the last publication is Hershko, Heller, Eytan, Kaklij, and Rose (1984). As the team members share information, it is to be expected that they cite their publications before other researchers do. Figure 6 shows, for the period 1979–1981, publications with citation relations with the landmark publications; publications from authors within the core group (within-group publications) and those from authors outside (out-group publications) are clearly marked. The date a manuscript was submitted to the publisher is used to place the publications on a timeline.

Knowledge profiles based on WoS subject categories assigned to the publications of the Nobel Prize laureates before the collaboration was established are presented in Figure 7; overlap and complementarity in subject categories are clearly visible.

Citation Characteristics of the Landmark Publications

Which publication is considered a highly cited paper? How do the documents presented compare when taking the number of times they are cited as a measure?

[6]Adenosine triphosphate (ATP) is a molecule that acts as an energy source within cells (footnote added by the authors).

DOI: 10.1002/asi

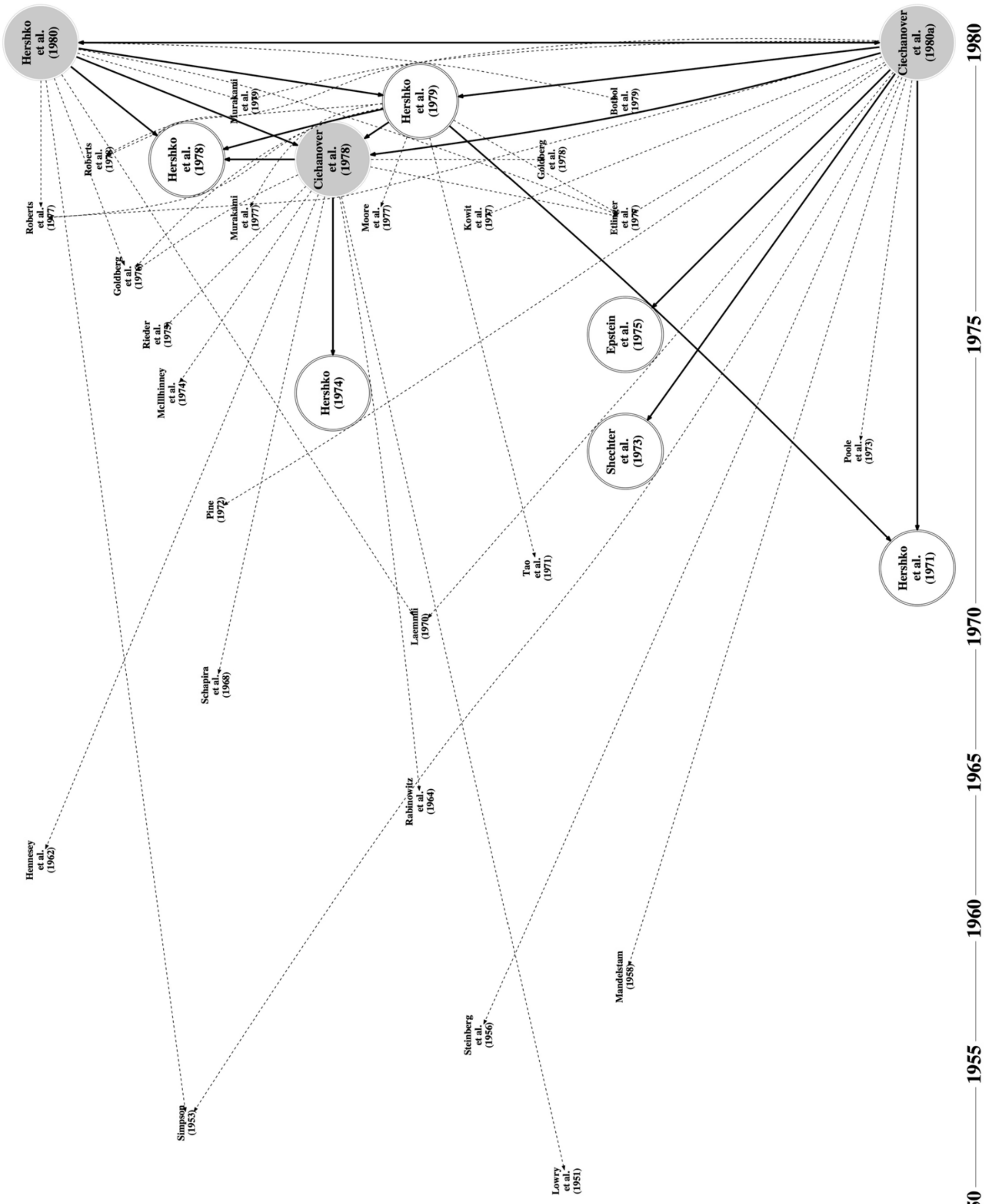

FIG. 4. Citation network of the ubiquitin discovery based on publications cited by the landmark publications.

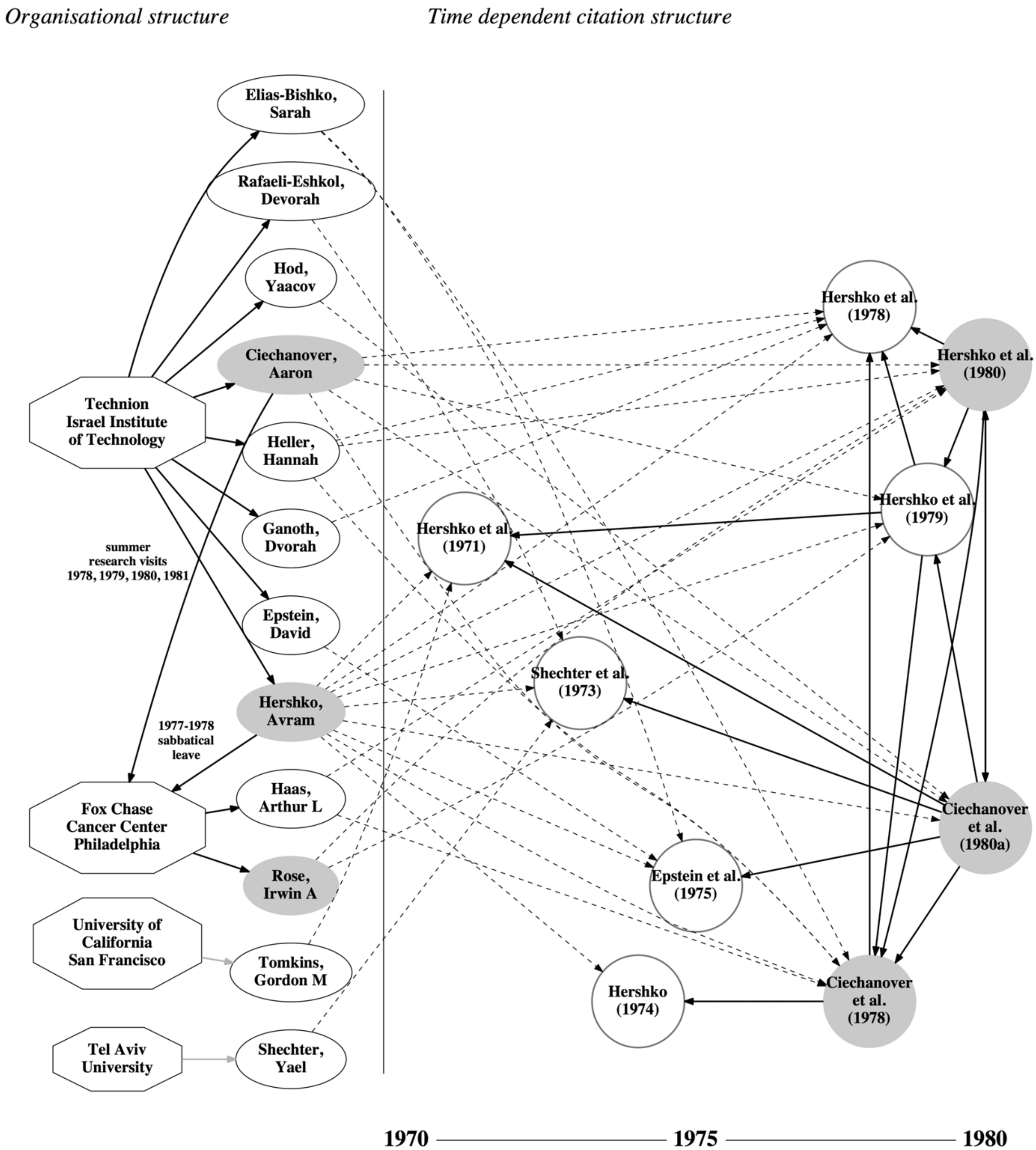

FIG. 5. Organizational-temporal anatomy of the discovery of ubiquitin. Shown are publications, (co) authors, research institutions, and interrelations. The opaque (gray) ellipses present the Nobel Prize laureates, and the opaque (gray) disks the landmark publications. Solid arrows represent affiliation-author relations and citation relations; dashed arrows represent the (co) author-publication relations.

Table 2 gives an overview. As the table shows, the core documents in the ubiquitin case show modest citation behavior. Redner (2005, p. 52) defines, arbitrarily he admits, a "discovery paper" as having at least 500 citations. Using this criterion, none of the ubiquitin-related papers can be considered a "discovery paper." The WoS subject categories "Biochemistry & Molecular Biophysics" and "Biology" are assigned to the four publications Ciechanover et al. (1980a, 1980b), Hershko et al. (1980), and Wilkinson et al. (1980). Citation distributions were calculated for all publications of

DOI: 10.1002/asi

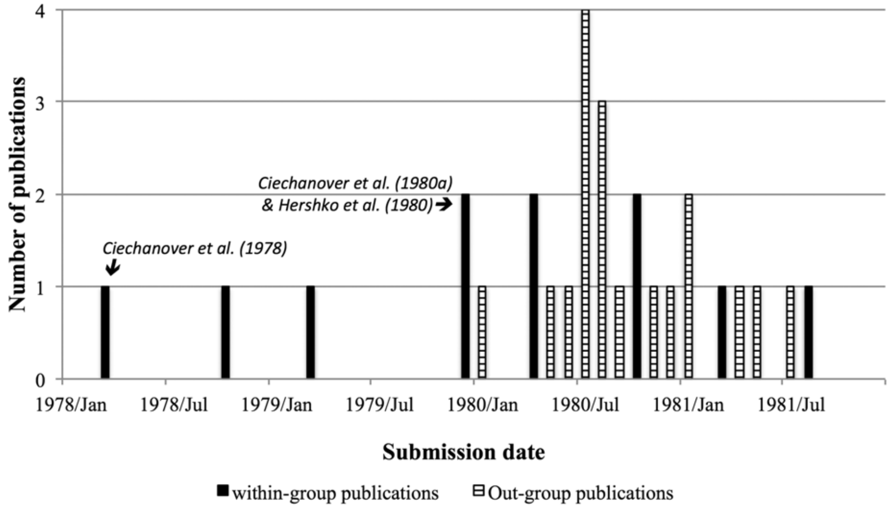

FIG. 6. Within- and out-group publications connected by citation relations with the landmark publications (1978–1981).

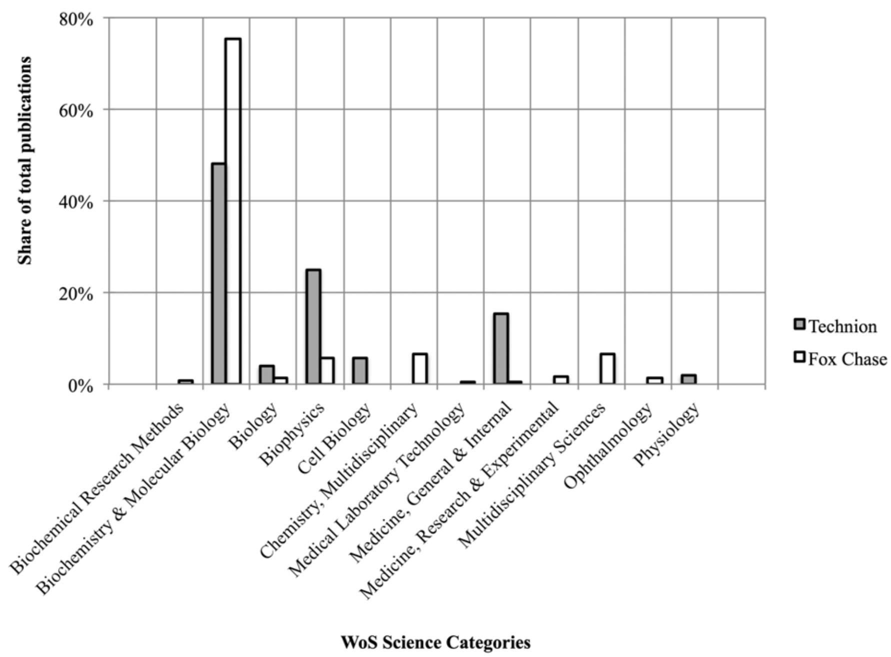

FIG. 7. Knowledge profiles of the Technion Institute (Ciechanover and Hershko) and Fox Chase Cancer Center (Rose) based on the WoS subject categories of publications up to 1978 covered in the WoS database.

type "article" and "letter" published in 1980 to which these subject categories were assigned; the four publications were excluded from the calculations. We present the results in Table 3. All four publications are among the highest cited publications published in 1980 for the two WoS subject categories.

The role of Ciechanover et al. (1978). The citation history for Ciechanover et al. (1978) started in 1979 with a number of group self-citations by authors and coauthors from the Technion Institute. It seems that the real value of the discovery presented in this publication was not recognized at this time outside the inner circle of researchers working on the same subject. Ciechanover et al. (1978) was published in *Biochemical and Biophysical Research Communications* (BBRC). Compared with all publications published in 1978 in this journal, it has proven to be one of the most cited. The distribution and the relative position of the article within this distribution are presented in Figure 8. It has received fewer citations than Ciechanover et al. (1980a) and Hershko et al. (1980). These latter two publications were published in the *Proceedings of the National Academy*

DOI: 10.1002/asi

TABLE 2. Key citation figures for the “core” document shown in Figure 5. Hershko (1974) and Hershko, Heller, Ganoth, and Ciechanover (1978) are conference papers for which no citation information is available.

Publication	Cumulative number of citations			
	≤12 months	≤24 months	≤36 months	Until September 2013
Hershko and Tomkins (1971)	7	25	31	294
Shechter, Rafaeli-Eshkol, and Hershko (1973)	1	4	6	17
Hershko (1974)	.	.	.	.
Epstein, Elias-Bishko, and Hershko (1975)	4	15	22	116
Ciechanover et al. (1978)	0	4	16	283
Hershko et al. (1978)	.	.	.	.
Hershko et al. (1979)	6	18	30	137
Ciechanover et al. (1980a)	4	22	42	337
Ciechanover et al. (1980b)	1	16	31	159
Hershko et al. (1980)	7	30	59	467
Wilkinson et al. (1980)	3	26	51	250

TABLE 3. Based on the number of times cited the position of the four “core” publications from 1980 compared to the publications in the “reference set” is shown.

Publication	≤2 years	Up to 2013
Ciechanover et al. (1980a)	Top 2%	Top 1%
Ciechanover et al. (1980b)	Top 4%	Top 5%
Hershko et al. (1980)	Top 1%	Top 1%
Wilkinson et al. (1980)	Top 1%	Top 2%

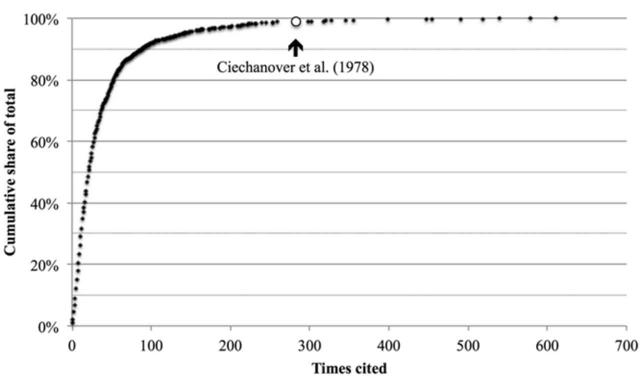

FIG. 8. Cumulative distribution of the number of publications in BBRC published in 1978 versus the number of times cited (1978–2013).

of Sciences (PNAS) and are among the top 6% most cited PNAS publications published in 1980.

Discussion

The aggregate-level trend data depicted in Figure 1 show several development stages in the number of scholarly publications, notably in 1984–1985, 1990–1991, and 2008–2009. These upswings in the time series suggest that previous scientific discoveries caused sudden increases in research activity and a concomitant rise in the number of publications. The type of the discovery and its impact on R&D processes worldwide cannot, however, be determined from these trend data.

For the patent data, we notice that the applications start to appear in 1985 after the rise in the number of scholarly publications in 1984–1985. This rise also coincides with the sudden increase in the number of CWTS subject categories (Figure 2). This indicates that the scientific knowledge about ubiquitin became mature enough to be used in patent applications. The level switch in the number of patent application that we notice in 1999–2000 is an indication that new or improved usage of the knowledge was found to be the basis for further technological developments.

In Figure 1, which focuses on the period when the breakthrough discovery was made, a steep rise is noticeable in the period 1983–1987. The evolution of the number of patent applications, the time delay with the upswing of scholarly publications, and the simultaneous level change in multidisciplinarity (Figure 2) indicate that the scientific knowledge became sufficiently well developed to be transformed into patentable technological applications (inventions).

The analysis of the citation network (Figure 5) shows that researchers from the Technion–Fox Chase group played a central role in ubiquitin development. The discovery of the ubiquitin-mediated protolytic system was, at first, not recognized outside the inner circle of the research group around Ciechanover and Hershko. Hershko spent a sabbatical leave at the Fox Chance Cancer Center in Philadelphia in 1977–1978; Ciechanover visited this institute in the summers of 1978–1981 (Ciechanover, 2004). These stays and visits, and also the publications by members of the Technion–Fox Chase research team, signal the importance of this collaboration for Ubiquitin research.

Figure 4 and especially Figure 5 show that the three landmark publications are related by citation links, and that they also share a group of authors and coauthors, of which most are affiliated to the Technion–Fox Chase group. These authors form a research group based in two institutes that originated at the Technion Institute. Various aspects of scientific research groups are discussed in Guimerà et al. (2005), Jones et al. (2008), Stokols, Hall, Taylor, and Moser (2008), Uzzi and Spiro (2005), Uzzi et al. (2013), Weinberg (1970), and Wuchty et al. (2007)

Of the nine publications of interest (Table 1), five were published in the period 1978–1980. Within the triadic network formed by Ciechanover et al. (1978, 1980a) and Hershko et al. (1980), a fourth publication (Hershko, Ciechanover, & Rose, 1979) seems to act as an information bridge between these three publications.

In the long run, all the three landmark publications receive a respectable, but not excessive, number of citations. Ciechanover et al. (1978) start receiving a modest number of citations in the second year after publication. From year 2 on, the number of citations to Ciechanover et al. (1980a) and Hershko et al. (1980) is relatively high, but not extreme. Based on the citations received, the 1978 article at first went almost unnoticed compared to the two 1980 articles. This is especially true if the article is compared with another publication (Hershko et al., 1979) that received much more attention. There are several plausible explanations for this behavior, for instance (a) the research field is becoming more mature and starting to realize the importance of the discovery and (b) the use of different journals BBRC (Ciechanover et al., 1978) versus PNAS (Ciechanover et al., 1980a; Hershko et al., 1979, 1980) with, partially, different audiences, to name just two of the possibilities.

The citation network for all publications directly cited by the three landmark publications or citing them in the period 1951–1986 shows two publications, Wilkinson et al. (1980) and Ciechanover et al. (1980b), that also seem to play an important role. The role of the findings presented in these two publications is mentioned in Hershko and Ciechanover (1982, p. 350). These two publications describe, in combination, the definitive identification of the component known as APF-1 as ubiquitin.

The knowledge profiles (Figure 7) largely overlap, but show some differences. The Technion team seems to be more focused on medicine, biophysics, and physiology. The conclusion arises that a cognitive bridge was created between two groups with at least partial complementary knowledge and that this combination facilitated the disentangling of the nature of the ubiquitin system. This makes it plausible that the collaboration between the two groups of authors introduced new combined theoretical insights that played a significant role in the further development of the original discovery; Ciechanover (2014) supports this view.

Other information sources and related bibliographic data have to be examined to determine the nature and background of the breakthrough, in order to differentiate the roles the various key publications played. The discovery is a product in a wider field of research on protein degradation. We note that the discovery of the Ubiquitin-mediated proteolytic system emerged from research executed by a small group of interacting scientists, as far as we can tell from bibliographic information. From Figures 4 and 5, we conclude that the closer we approach the year 1980, the denser the network of publications linked by citation relations becomes.

The evolution of networks of researchers and institutions becoming active in this research field can be taken to reflect the diffusion of the new theoretical concepts within the scientific community.

In this study, our goal was to answer the following research questions:

How does the metaphor "Science progresses by standing on the shoulders of others" manifest itself in this discovery?

Our research shows that, in the period around the breakthrough discovery, a small group of intensively cooperating researchers were endeavoring to attain the breakthrough discovery through bootstrapping. At a later stage, new researchers gradually enter the research field and "stand on the shoulders" of the achievements of the group of original researchers.

Is there a cognitive reason for this, as we know now, fruitful collaboration?

According to Ciechanover (2014), the collaboration between the two research groups was instigated by the fact that the research group at the Fox Chase Cancer Center was needed to help in elucidating the role of ATP in the Ubiquitin-mediated proteolytic system.

What changes in the "sociocognitive" structure of the discovery of ubiquitin are visible in the period around a theory-changing breakthrough?

The sociocognitive structure of the discovery of ubiquitin shows that the very small original research group enlarged as the result of a sabbatical visit of Herskho, as well as several visits by Ciechanover. The start of these visits marks the moment publications from both groups become part of the same citation network that led to the landmark publications for which the Nobel Prize was awarded.

Does the citation network of relevant publications support the idea of a challenge breakthrough by showing only modest numbers of citations for the three landmark publications?

The landmark publications are among the most cited publications in the field, but the number of citations is modest; at most, 30 within 36 months. The citation network shows several other publications with a comparable number of citations. The breakthrough discovery did not lead to an explosion of research activity, but proliferated in subsequent iterative advances.

Is a gradual increase in the multidisciplinarity of the citing publications noticeable?

A linear increase in multidisciplinarity is visible (Figure 2) over a long period. For the period 1980–1984, it remained stable at a modest level. Instances of abrupt increases that might be the result of "follow-up" breakthroughs are also visible.

Can publications that are not directly manifest, but that play an important role in the developments following the discovery, be identified in the citation network around the landmark publications?

Analyses of the citation network up to 1983 revealed two publications, published in 1980 shortly after the two 1980 landmark publications, which also seem to play a central role in the further development of research on the ubiquitin-mediated proteolytic system (Hershko & Ciechanover, 1982).

Does the citation network show a "clustering" of publications from a small group of scientists over a short period?

A "clustering" can be perceived of publications from a small group of collaborating authors in a relative short period around the theory-changing breakthrough discovery (Figures 4 and 5). Publications from this group of collaborating authors show a within-group advantage (Figure 6).

Does the constraint we impose in using short time windows prevent the identification at an early stage of a theory-changing breakthrough?

If we consider publications for the period 1978–1982, structural changes in the citation network around the breakthrough moment are manifest. The articles describing the ubiquitin discovery received a modest number of citations. This modest number of citations is in line with Wang et al. (2013, p. 127), "Paradigm-changing discoveries have notoriously limited impact," if we interpret "impact" as "citations received."

Concluding Remarks

This study finds that there is bibliometric evidence supporting the idea of a *challenge* breakthrough. The breakthrough discovery of the ubiquitin-mediated proteolytic system is the result of several incremental discoveries distributed over a number of publications that are have citation relations. The individual publications are not exceptional in terms of the number of citations they received and are by a small group of collaborating scientists. Five publications seem to form the "heart" of the discovery and constitute the basis for further development. These publications stand out in terms of indegree (i.e., number of citations received).

Another conclusion is that "standing on the shoulders of others" does not adequately describe the progress of science at the "micro" level of this individual theory-changing discovery (i.e., *challenge* discovery), given that the discovery is the result of the work of a small group of closely collaborating scientists. This group of scientists was, in some respects, at least temporarily, cognitively isolated as the result of reluctance in the scientific community to accept new theories. Wang et al. (2013, p. 127) cites Kuhn (1962) as "(. . .) precisely because the more a discovery deviates from the current paradigm, the longer it takes to be appreciated by the community (. . .)." Bibliographic information shows a within-group advantage (Figure 6) and, furthermore, that the scientific knowledge in the landmark publications comes from publications that are closer on a timeline to each other when approaching 1980.

Analysis of the publications and their citation relations in this study shows effects on a timescale that might be appropriate for use as early stage (i.e., within 12–24 months) identification of a theory-changing breakthrough. It needs to be proven that our hypothesis

> observable changes in the socio-cognitive structure that are visible as changes in the citation network of relevant bibliographic information, enable the detection at early stage, within 24 months, of a potential scientific discovery that results in an identifiable shift in theoretical concepts

holds in general. Redner (2005, p. 52) defines a discovery article as having at least 500 citations, Wang et al. (2013) points to the fact that publications with exceptional long-term impact appear to be the hardest to recognize on the basis of their early citation patterns, van Noorden, Maher, and Nuzzo (2014) find that none of the articles announcing some of the world's most famous discoveries come anywhere close to ranking among the 100 most highly cited publications of all time, and Ioannnidis, Boyack, Small, Sorenson, and Klavans (2014) find that authors conclude that their most cited papers are the most important and that the chart-topping work was evolutionary and not revolutionary. To address this issue, this study deviated from "simple" citation analysis by incorporating in the analysis other information related to the publications. Careful analyses of citations at the early stage revealed the inception of a core research group. The members of this research group were responsible for all citations during the first 2 years after publication of the discovery. From the moment the "breakthrough" discovery is recognized by other scientists, "out-group citations" become increasingly dominant.

A follow-up study should focus on operationalizing a method to identify from bibliographic data a "core research group." Furthermore, this study should provide threshold values for the within-group citations to out-group citations ratio that would hint at a theory-changing discovery. An identified core research group in combination with the value for the within-group citations to out-group citations ratio should help in generalizing the results of this case study and answer the question, "To what extent is the identification of potential theory-changing breakthroughs possible by searching for core research groups in early-stage citation networks?"

DOI: 10.1002/asi

Acknowledgments

We sincerely thank Professor Aaron Ciechanover for his willingness to answer questions posed and pointing to additional relevant literature. We also thank the reviewers for their comments and suggestions.

References

Andersen, H., Barker, P., & Chen, X. (2006). The cognitive structure of scientific revolutions. New York: Cambridge University Press.

Arbesman, S. (2010). Quantifying the ease of scientific discovery. Scientometrics, 86(2), 245–250.

Breiner, S., Cuhls, K., & Grupp, H. (1994). Technology foresight using a Delphi approach—A Japanese-German cooperation. R & D Management, 24(2), 141–153.

Chen, C., Chen, Y., Horowitz, M., Hou, H., Liu, Z., & Pellegrino, D. (2009). Towards an explanatory and computational theory of scientific discovery. Journal of Informetrics, 3(3), 191–209.

Ciechanover, A. (2004) Aaron Ciechanover—Biographical.Nobelprize.org. Nobel Media AB 2014. Retrieved from http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2004/ciechanover-cv.html (accessed on 29 January 2015).

Ciechanover, A. (2009). Tracing the history of the ubiquitin proteolytic system: The pioneering article. Biochemical and Biophysical Research Communications, 387(1), 1–10.

Ciechanover, A. (2014). Personal communication (unpublished).

Ciechanover, A., Hod, Y., & Hershko, A. (1978). A heat-stable polypeptide component of an ATP-dependent proteolytic system from reticulocytes. Biochemical and Biophysical Research Communications, 81(4), 1100–1105.

Ciechanover, A., Heller, H., Elias, S., Haas, A. L. L., & Hershko, A. (1980a). ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation. Proceedings of the National Academy of Sciences, 77(3), 1365–1368.

Ciechanover, A., Elias, S., Heller, H., Ferber, S., & Hershko, A. (1980b). Characterization of the heat-stable polypeptide of the ATP-dependent proteolytic system from reticulocytes. Journal of Biological Chemistry, 255(16), 7525–7528.

Darden, L., & Maull, N. (1977). Interfield theories. Philosophy of Science, 44(1), 43–64.

Epstein, D., Elias-Bishko, S., & Hershko, A. (1975). Requirement for protein synthesis in the regulation of protein breakdown in cultured hepatoma cells. Biochemistry, 14(23), 5199–5204.

Guimerà, R., Uzzi, B., Spiro, J., & Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. Science, 308(5722), 697–702.

Hershko, A. (1974). Regulation of gene expression in eukaryotic cells. In M. Harris & B. Thompson (Eds.), Fogarty International Center proceedings, number 25 (pp. 85–96). Washington, DC: U.S. Printing Office.

Hershko, A., & Ciechanover, A. (1982). Mechanisms of intracellular protein breakdown. Annual Review of Biochemistry, 51(1), 335–364.

Hershko, A., & Tomkins, G. M. (1971). Studies on the degradation of tyrosine aminotransferase in hepatoma cells in culture: Influence of the composition of the medium and adenosine triphosphate dependence. Journal of Biological Chemistry, 246(3), 710–714.

Hershko, A., Heller, H., Ganoth, D., & Ciechanover, A. (1978). Mode of degradation of abnormal globin chains in rabbit reticulocytes. In H. L. Segal & D. J. Doyle (Eds.), Proceedings of a symposium on protein turnover and lysosome function (pp. 149–169). New York: Academic Press.

Hershko, A., Ciechanover, A., & Rose, I. A. (1979). Resolution of the ATP-dependent proteolytic system from reticulocytes: A component that interacts with ATP. Proceedings of the National Academy of Sciences, 76(7), 3107–3110.

Hershko, A., Ciechanover, A., Heller, H., Haas, A. L., & Rose, I. A. (1980). Proposed role of ATP in protein breakdown: Conjugation of protein with multiple chains of the polypeptide of ATP-dependent proteolysis. Proceedings of the National Academy of Sciences, 77(4), 1783–1786.

Hershko, A., Heller, H., Eytan, E., Kaklij, G., & Rose, I. A. (1984). Role of the alpha-amino group of protein in ubiquitin-mediated protein breakdown. Proceedings of the National Academy of Sciences, 81(22), 7021–7025.

Hollingsworth, J. R. (2008). Scientific discoveries: An institutionalist and path-dependent perspective. In C. Hannaway (Ed.), Biomedicine in the twentieth century: *Practices, Policies, and Politics* (pp. 317–353). Bethesda, MD: National Institutes of Health.

Ioannnidis, J. P., Boyack, K. W., Small, H., Sorenson, A. A., & Klavans, R. (2014). Is your most cited work your best? Nature, 514(7524), 561–562.

Jones, B. F., Wuchty, S., & Uzzi, B. (2008). Multi-university research teams: Shifting impact, geography, and stratification in science. Science, 322(5905), 1259–1262.

Julius, M., Berkoff, C. E., Strack, A. E., Krasovec, F., & Bender, A. D. (1977). A very early warning system for the rapid identification and transfer of new technology. Journal of the American Society for Information Science, 28(3), 170–174.

Koshland, D. E. (2007). The cha-cha-cha theory of scientific discovery. Science, 317(5839), 761–762.

Kuhn, T. S. (1962). The structure of scientific revolutions. Chicago: The University of Chicago Press.

Leydesdorff, L., & Rafols, I. (2011). Local emergence and global diffusion of research technologies: An exploration of patterns of network formation. Journal of the American Society for Information Science and Technology, 62(5), 846–860.

Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randall, R. J. (1951). Protein measurement with the folin phenol reagent. Journal of Biological Chemistry, 193(1), 265–275.

Luukkonen, T., Persson, O., & Sivertsen, G. (1992). Understanding patterns of international scientific collaboration. Science, Technology, & Human Values, 17(1), 101–126.

Martin, B. R. (1995). Foresight in science and technology. Technology Analysis & Strategic Management, 7(2), 139–168.

Merton, R. K. (1965). On the shoulders of giants—A Shandean postscript. Chicago: The University of Chicago Press.

Moed, H. F. (2005). Citation analysis in research evaluation. Information science and knowledge management. Dordrecht, The Netherlands: Springer Netherlands.

Nath, D., & Shadan, S. (2009). The ubiquitin system. Nature, 458(7237), 421.

Nobel Prize Committee. (2004). The Nobel Prize in Chemistry 2004—Advanced information. Technical report, Nobelprize.org.

Ponomarev, I. V., Williams, D. E., Hackett, C. J., Schnell, J. D., & Haak, L. L. (2014a). Predicting highly cited papers: A method for early detection of candidate breakthroughs. Technological Forecasting and Social Change, 81(January 2014), 49–55.

Ponomarev, I. V., Lawton, B. K., Williams, D. E., & Schnell, J. D. (2014b). Breakthrough paper indicator 2.0: Can geographical diversity and interdisciplinarity improve the accuracy of outstanding papers prediction? Scientometrics, 100(3), 755–765.

Redner, S. (2005). Citation statistics from 110 years of Physical Review. Physics Today, 58(6), 49–54.

Rose, I., Warms, J., & Hershko, A. (1979). A high molecular weight protease in liver cytosol. Journal of Biological Chemistry, 254(17), 8135–8138.

Shechter, Y., Rafaeli-Eshkol, D., & Hershko, A. (1973). Influence of protease inhibitors and energy metabolism on intracellular protein breakdown in starving *Escherichia coli*. Biochemical and Biophysical Research Communications, 54(4), 1518–1524.

Small, H.G. (1977). Co-citation model of a scientific specialty—Longitudinal study of collagen research. Social Studies of Science, 7(2), 139–166.

Stokols, D., Hall, K. L., Taylor, B. K., & Moser, R. P. (2008). The science of team science: Overview of the field and introduction to the supplement. American Journal of Preventive Medicine, 35(2 Suppl.), S77–S89.

Upham, S., Rosenkopf, L., & Ungar, L. (2010). Positioning knowledge: Schools of thought and new knowledge creation. Scientometrics, 83(2), 555–581.

Uzzi, B., & Spiro, J. (2005). Collaboration and creativity: The small world problem. American Journal of Sociology, 111(2), 447–504.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. Science, 342(6157), 468–472.

van Noorden, R., Maher, B., & Nuzzo, R. (2014). The top 100 papers—Nature explores the most-cited research of all time. Nature, 514(7524), 550–553.

Wang, D., Song, C., & Barabási, A.-L. (2013). Quantifying long-term scientific impact. Science, 342(6154), 127–132.

Weinberg, A. M. (1970). Scientific teams and scientific laboratories. Daedalus, 99(4), 1056–1075.

Wilkinson, K. D., Urban, M. K., & Haas, A. L. (1980). Ubiquitin is the ATP-dependent proteolysis factor I of rabbit reticulocytes. Journal of Biological Chemistry, 255(16), 7529–7532.

Wray, K.B. (2011). Kuhn's evolutionary social epistemology. Cambridge, UK; New York: Cambridge University Press.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. Science, 316(5827), 1036–1039.

DOI: 10.1002/asi