Discovering Paradigm Shift Patterns in Biomedical Abstracts: Application to Neurodegenerative Diseases

Frédérique Lisacek[a,b] Christine Chichester[a] Aaron Kaplan [c] Ágnes Sándor [c]

[a] Geneva Bioinformatics (GeneBio) S.A., Geneva,, [b] Swiss Institute of Bioinformatics, Geneva,
[c]Xerox Research Centre Europe, Meylan, France

Abstract

Millions of facts are stored within the biological literature. Most of these facts represent small advances in the knowledge on an established theory, but a small fraction offer new insight into a biological phenomenon. We propose a method based on computational linguistic tools for distinguishing these facts (extraction) and exposing knowledge that may be important in future developments (prediction). The method is based on finding linguistic cues indicating that the authors of biological articles have identified a problem with, or a break from, conventional knowledge.

Introduction

In reviewing the literature on a given biological process, a researcher's goal is to integrate the available information into a model of the process. For a topic that is the subject of ongoing research, this model will necessarily be incomplete and perhaps inconsistent. These weaknesses in the model indicate important directions for further research. In fact, discovery usually follows a chaotic path. Hypotheses are set, challenged and often reformulated before a reliable and stable representation of the investigated phenomenon is reached (Lisacek, 2003).

We propose a methodology for rapidly arriving at a synthetic view of the literature on a particular research problem, by automatically identifying research reports that contain key elements in the field. The methodology involves searching for linguistic cues that indicate actual or potential breakthroughs. A "potential breakthrough" indicates a weakness in the current model, e.g. observations that appear contradictory, competition among multiple hypotheses, unexpected results, or new ideas that must be integrated into the model. The following are examples of the kinds of text we wish to find:

- First appearance of evidence:

  Altogether, these data provide the first evidence of the participation of proteasomes in the control of mammalian mitochondrial inheritance and suggest a new role of the ubiquitin-proteasome pathway in mammalian fertilization. (PMID: 12606393)

- Emerging trend:

  Growing evidence indicates that aldehydic products of lipid peroxidation play an important role in the pathophysiology of neurodegenerative disorders such as Parkinson's disease. (PMID: 12644268)

- Contradiction of conventional knowledge:

  In contrast with previous hypotheses, compact plaques form before significant deposition of diffuse A beta, suggesting that different mechanisms are involved in the deposition of diffuse amyloid and the aggregation into plaques. (PMID: 12697936)

- Identification of controversy, debate, contradiction:

  Presenilin 1 (PS1) regulates beta-catenin stability; however, published data regarding the direction of the effect are contradictory. (PMID: 11606587)

Our concern for following the evolution of hypotheses and trends in the literature through identifying the described changes is comparable to the phenomena characterised in a recent paper as "paradigm shifts" (Storrie and Nilsson, 2002). These shifts express a significant gap between accepted (or expected) knowledge and new knowledge described (observed).

We will henceforth use the term "paradigm shift" as shorthand to denote phenomena of change and potential change as described above. This includes problems whose eventual resolution may be a breakthrough, as well as actual breakthroughs that have already occurred.

Synthesizing research literature is a problem that has already received a great deal of attention. Much of the work has concentrated on integrating information from as many documents as possible, by automatic means. A significant number of applications have been developed to identify gene or protein names and symbols in biomedical texts (Tanabe and Wilbur, 2002, Yu et al., 2002, Proux et al., 1998, Krauthammer et al., 2000). Once assembled as networks, these enormous collections of data allow hypotheses to be generated and

tested using all pre-existing data. Alternative strategies for knowledge discovery, based on data mining combined with literature analysis, have also produced promising results. In most cases, sequence homology and/or gene expression profiling together with manual literature analysis help interpreting results (Blaschke et al., 2001, Chaussabel and Sher, 2002, Raychaudhuri and Altman, 2003). In contrast, our approach is to seek connections within the literature by using relatively few articles selected for their novel, controversial, or unexpected content.

Put broadly, the problem addressed by our method is to search a large collection of documents for the few that meet our needs. In this sense, it is an information retrieval (IR) problem, but the techniques used by classical IR systems turn out not to be applicable. Most IR systems, following the example of Salton (1971), are based on "bag of words" techniques that have been shown to be useful for finding documents about a particular topic. However, they are less so for finding paradigm shift documents. As we have seen above, paradigm shifts are described using a wide range of expressions. Clearly, it would be infeasible to use entire expressions as query terms in a bag-of-words approach.

It is thus necessary to look not merely for individual words or sequences of words, but for more complex ways of combining words. Information extraction (IE) systems in the tradition of the MUC competitions, e.g. (Hobbes et al., 1996), typically provide a regular expression formalism in which the user can describe patterns of sentences that should be identified. This has been shown to work well for some concepts whose range of expressions in language is relatively restricted, for example, sentences about joint business ventures (Hobbs et al.,1996). But this sort of pattern formalism would take a prohibitively large number of patterns to achieve acceptable levels of precision and recall for describing the wide range of expressions used to convey the notion of paradigm shift.

Since the problem solved by our system is to find expressions of new ideas, it is somewhat related to the problem of "novelty detection" (Zhang et al. 02), in which the goal is to decide whether or not a given sentence or document contains information that was not already expressed in previous documents, and to the problem of "first story detection" (Allan et al. 00), in which the object is to find sets of newspaper articles that describe the same event, in order to identify the earliest report of each event. The approach in these applications is to compare content among candidate sentences or documents, whereas our approach is to look for explicit linguistic markers of novelty (and of importance) in a single document. Our approach is thus perhaps not applicable to as wide a range of document types, but since it is a less difficult problem, it makes more useful levels of performance attainable in areas where it is applicable.

Our method is based on the detection of cue phrases that indicate a specific aspect of argumentative structure. A similar task is undertaken in (Teufel 1998, Teufel and Moens 2002), in support of automatic summarization of scientific articles. The authors aim to extract cue phrases designated as "meta-comments". However, Teufel's method is limited to finding contiguous sequences of words, which the authors acknowledge as a weakness, since the expressions that could be considered meta-comments are in fact as varied in vocabulary and structure as the expressions that indicate paradigm shifts. Consistent with this approach, (Mizuta and Collier, 2004) have suggested "zone identification" (ZI) as a solution to organise factual information in biology papers. In fact, we address a sub-problem of the ZI problem.

The remainder of this article is structured as follows. In Section 2, we describe some prototype software we have developed for automatically finding "paradigm shift" articles on a given subject matter. In Section 3, we present the results of several preliminary tests we have performed to evaluate the system at various levels, including both tests of the software itself, and evaluations of the methodology of which the software is one element. Section 4 concludes with some notes on the current and future evolution of our system.

Method

The aim of the linguistic processing performed by the system is to present a selection of articles that describe a "paradigm shift" in the understanding of a given topic. We divide this task into two subtasks:

1. Identify articles that describe paradigm shifts
2. Identify articles about the topic

Identifying descriptions of paradigm shifts

The concept of paradigm shift as we have defined it is complex, and there are no fixed, conventional expressions for it. This is illustrated by the sentences quoted in the introduction-they all convey the idea we are looking for, but in each case the vocabulary and the linguistic structures are different. The variety seen in these examples is but the tip of the iceberg: in fact, it is rare to find two documents that describe paradigm shifts using the same expression. We have therefore developed a pattern formalism that allows us to describe, in a relatively concise way, this variety of expressions.

Consider the following three sentence fragments, in which the expressions conveying the notion of paradigm shift are in bold type:

> ***Further research*** *on the zinc* ***paradox*** *in AD* ***is needed****. . .*

> ***Information*** *regarding the nature of its participation in this process* ***remains controversial*** *and* ***unclear****.*

> ***Despite recent advances*** *. . . mechanisms of aggregate formation have* ***proved challenging to study****.*

Although these fragments have no vocabulary in common, there is a common pattern. Three basic notions are represented in all of them, namely the notions of [TIME], [IDEA] and [CONTRAST]:

> *Further[TIME] research[IDEA] on the zinc paradox[CONTRAST] in AD is needed . . .*

> *Information[IDEA] regarding the nature of its participation in this process remains[TIME] controversial[CONTRAST] . . .*

> *Despite recent[TIME] advances[IDEA] ... mechanisms of aggregate formation have proved challenging[CONTRAST] to study.*

We consider that the complex concept of paradigm shift is expressed in the above sentences via the composition of the constituent notions of [TIME], [IDEA], and [CONTRAST]. These constituent notions, unlike the complex concept of paradigm shift, tend to be contained in the meanings of single words. We can thus build a system that identifies (some) paradigm shift sentences by compiling a vocabulary list for each of the three constituent notions, and then searching for sentences in which words expressing the three constituent notions are combined. Naturally, the combination of [TIME], [IDEA], and [CONTRAST] is just one example; many other decompositions of the concept of paradigm shift are possible. Nevertheless, we believe that the number of possible decompositions is relatively small, particularly when compared with the enormous number of possible expressions. In the current version of our system, which is the result of a corpus study, we use a total of seven constituent notions, accompanied by rules indicating which of the possible combinations of these notions indicates a paradigm shift. With the seven constituent notions are associated a total of some 600 words, which were selected by hand, with some automatic support for proposing candidates.

To identify paradigm shifts, sentences in which an appropriate set of constituent notions is combined must be detected. Mere co-occurrence in the same sentence is not sufficient. Consider the following sentence:

> *The* ***results****[IDEA] show a* ***long-lasting****[TIME] neuroprotection induced by activation . . . as early as three days prior to induction of an . . . epileptic* ***challenge****[CONTRAST].*

The three words indicated belong to the word lists we have compiled for the respective constituent notions, yet the sentence does not express a paradigm shift. This can be explained by the fact that the words expressing the constituent notions do not constitute a coherent linguistic expression, unlike in the earlier sentences.

We therefore use a parser to identify the syntactic relationships between words that express the constituent notions. The syntactic constraints we impose do not require any particular syntactic structure; they merely require that the constituent notions be connected to each other in the sentence's dependency graph.

The system is implemented in the XIP (Xerox Incremental Parser) environment. (Aït et al., 2002). XIP is a broad-coverage syntactic dependency parser developed at Xerox Research Centre Europe. It identifies syntactically related pairs of words in sentences, and we have provided it with our word lists for the seven constituent notions so that it can label them appropriately. For example, in the sentence

> *Further research on the zinc paradox in AD is needed.*

the parser identifies a number of pairs of words as being related, among which are the following pairs composed of words from our vocabulary lists for the notions of [TIME], [IDEA], and [CONTRAST]:

> (research[IDEA],Further[TIME])

> (research[IDEA],paradox[CONTRAST])

The words 'research' and 'Further' are considered to be syntactically related in this sentence because 'Further' is an adjective modifying the noun 'research,' and the words 'research' and 'paradox' are considered related because 'paradox' is the head of a prepositional phrase that modifies 'research.' This sentence is identified as expressing a paradigm shift because our rules indicate that any sentence containing a [TIME, IDEA] pair and an [IDEA, CONTRAST] pair should be flagged.

Ranking by subject relevance

The previous section explained how we find documents that describe a paradigm shift. Since the goal is to find documents that describe a paradigm shift in a particular subject area, it remains to find documents relevant to the subject area. As detailed below, documents are retrieved from PubMed, via a simple boolean keyword search, but the relevant abstracts are diluted with too many false positives. We therefore implemented a simple ranking system based on the bag-of-words model. Using a set of keywords related to the subject matter, and the system ranks abstracts according to the concentration of these keywords in the abstract. The final result of the system is thus a list of abstracts that contain paradigm shift expressions, ranked in order of the concentration of topic keywords in the abstract. To help increase the accuracy of the ranking, after the input of an initial set of subject matter keywords, the system can suggest additional keywords, based on frequency of co-occurrence with the initial set.

Results

We are currently testing a prototype system built as described in the previous section. Further development, including significant refinement of the selection criteria, is ongoing, but we have performed some preliminary evaluations on the prototype.

As a test case for evaluating the system, we chose the topic of neurodegenerative diseases. In the past decade, protein aggregation, the hallmark of neurodegenerative diseases, resulting from defects in protein degradation is the substance of the current model. Therefore, in order to obtain a document set with high coverage of relevant articles, we used the PubMed query `"neurodegen*" OR ("protein" AND ("degrad*" OR "aggreg*"))`.

Direct evaluation of linguistic processing

We have evaluated the performance of the concept-matching and ranking modules on a test corpus consisting of articles that were published in the past six months, and therefore weren't included in our development corpus. With the original query string (see above) and this new date restriction, we retrieved some 3300 abstracts from PubMed. Our system detected at least one putative paradigm shift sentence in 175 of these. We applied the ranking algorithm using a predetermined set of 30 subject matter keywords, and discarded abstracts that received a score of zero (which indicates that they contained none of the subject ranking keywords and thus are unlikely to be relevant to the topic of neurodegeneration). We then evaluated the remaining 131 abstracts by hand, for two criteria: the degree to which we felt they described a paradigm shift, and the degree to which they were relevant to the subject of neurodegeneration. Both evaluations were done on a four-point scale. The results for the quality of the detected paradigm shifts were the following:

Score	No. of abstracts	% of abstracts
Low	4	3%
Medium	32	24%
Medium-High	41	31%
High	54	41%

We see that the vast majority of the abstracts selected are ones that contain relevant information concerning a change in research direction.

Figure 1 shows the performance of the ranking module. Note that less relevant abstracts tend to be found towards the end of the ranked list, as desired.

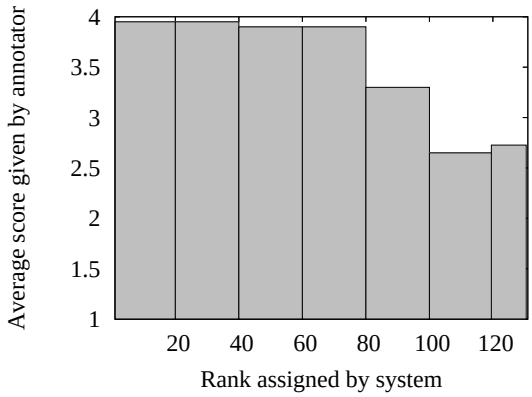

Figure 1: Effectiveness of ranking.

On an earlier test set that was a subset of the original development corpus, we performed an evaluation to determine the contribution of syntactic constraints to our concept matching approach. We constructed a set of rules that are analogous to those described in Section 2.1 except that they consider merely co-occurrence in the same sentence rather than syntactic connectedness; that is, any two words occurring in the same sentence were taken to constitute a word pair. This relaxed rule set necessarily selects all sentences selected by the original rule set, plus potentially some others. In a test set of 55 sentences not selected by the original rules, 17 were selected by the relaxed rules. All 17 were false positives, i.e. none of them described paradigm shifts. It is thus clear that the syntactic constraints significantly improve precision.

Evaluation of the methodology

We assume that the knowledge described in documents containing paradigm shift sentences is pivotal in understanding and explaining a biological phenomenon. Sentence selection requires no a priori knowledge on the relative importance of biological entities. Quite the opposite, it actually contributes to weighing the role of entities involved. Traditionally, this information is found in good review articles. So, in order to substantiate our assumption, we tried to determine whether the limited selection of paradigm shift documents is sufficient to cover common knowledge on the topic described in reviews.

We comprehensively read a few articles reviewing neurodegeneration at different points in time (Tran and Miller, 1999, Maccioni et al., 2001, Bossy-Wetzel et al, 2004). Then we compared the description of protein roles as described between reviews and as described through our selection of documents. In this first study, entities are restricted to proteins but ongoing work also targets ligand, tissue type, etc. Furthermore, proteins not described in reviews but cited in paradigm shift abstracts were tracked in recent articles to evaluate the pertinence of their role in the current understanding of neurodegeneration.

Finally we compared our protein set with those mapped in the KEGG interaction network (`http://www.genome.jp/kegg/pathway/hsa/hsa01510.html`) generated with sequence homology and gene expression profiling data as described in (Limviphuvadh et al., 2004).

Molecular actors of neurodegeneration from reviews Table 1 features proteins identified as involved in neurodegeneration. Column 2 lists proteins occurring in reviews published over three years ago. Column 3 lists proteins occurring in the most recent review, and finally Column 4 lists proteins occurring in selected abstracts containing paradigm shift sentences. Except for the Pael-receptor and Septin, all cited proteins were matched. At this stage, we do not consider protein- protein interactions since we simply wish to check the soundness of our selection of relevant molecular actors. The context in which proteins are referred to is the next obvious targeted information. Table 2 shows examples of selected paradigm shift sentences for articles cited in Table 1.

Proteins that have gained further interest A number of documents selected with paradigm shift sentences and published over two years ago describe proteins that are not even included in the most recent review (Bossy-Wetzel et al., 2004). We therefore tracked in PubMed other recent articles mentioning these proteins in relation to neurodegeneration and could confirm the pertinence of their role in the current understanding of neurodegeneration. Two examples shown in Table 3, support this view. Another five proteins not shown here have been identified.

Molecular actors of neurodegeneration predicted in silico A collection of proteins is described in the KEGG interaction network (`http://www.genome.jp/kegg/pathway/hsa/hsa01510.html`). In particular, the network highlights those proteins that are involved in several neurodegenerative diseases. For some of those proteins (incidentally, not listed in the above reviews), we could also identify their multiple disease implication with paradigm shift sentences. Table 4 shows these sentences for two of the proteins found.

These results are preliminary and need to be refined. As mentioned earlier, we aim at gathering more contextual details and focus on molecular actors in a broader sense than simply listing proteins involved. To begin with, protein posttranslational modifications as functional modulators are a priority (Chichester et al., 2003).

Conclusion

In this article we described a novel approach to mining biomedical literature for relevant information indicating directions for future research. Instead of extracting and processing factual information in scientific abstracts, we use discourse structure cues. We hypothesize that relevant papers are the ones that refer to contradictions, inconsistencies, and gaps in existing models by pointing them out, suggesting possible solutions or yielding solutions. We have developed a rule-based NLP system that, using syntactic parsing and a new concept-matching module, extracts such Medline abstracts with high precision. The results are ranked according to their relevance to the desired subject matter. To evaluate the method, we used the NLP system to select Medline articles about neurodegenerative diseases from a given span of time, and compiled a list of proteins that, according to these articles, are involved in neurodegeneration. Some of these proteins were overlooked in survey articles contemporary with the articles we processed, but appear in surveys published in subsequent years. This indicates that our method is successful at identifying important new developments before they become widely known.

We continue to refine the criteria used for identifying abstracts that describe paradigm shifts. In particular, while the prototype described in this article can only detect single sentences that describe paradigm shifts, we are working on detecting abstracts in which the detection of a paradigm shift requires combining infor-

mation from multiple sentences.

Acknowledgements:
We are very grateful to Jean-Pierre Chanod and Michel Gastaldo for their unconditional support of this work. We also thank Eric Cheminot and Patrick Brechbiehl for their valuable technical assistance.

Address for Correspondence:
Frédérique Lisacek
frederique.lisacek@genebio.com

References

[1] Aït-Mokhtar,S., Chanod J.-P., Roux, C. (2002) Robustness beyond shallowness: incremental dependency parsing. *Special issue of the NLE Journal*.

[2] Allan, J., Lavrenko V., Jin H. (2000) First Story Detection *In TDT Is Hard. Proc. Ninth international conference on Information and knowledge management*, pp 374–381.

[3] Blaschke, C., Oliveros J. C., Valencia A. (2001) Mining functional information associated with expression arrays. *Funct Integr Genomics*, **1**256–68.

[4] Bossy-Wetzel, E, Schwarzenbacher R, Lipton SA. (2004) Molecular pathways to neurodegeneration. *Nat Med*. **10** Suppl: S2-9.

[5] Chaussabel, D., Sher A. (2002) Mining microarray expression data by literature profiling. *Genome Biol*, **3**RESEARCH0055.

[6] Chichester, C., Nikitin F., Ravarini J-C., Lisacek F. (2003) Consistency checks for characterizing protein forms. *Computational Biology and Chemistry* **27**, 29-35.

[7] Hobbs,J., Appelt, D., Bear, J., Israel, D., Kameyama, M., Stickel, M., Tyson, M. (1996) FASTUS: A cascaded finite-state transducer for extracting information from natural-language text. *Finite State Devices for Natural Language Processing*, MIT Press.

[8] Limviphuvadh, V., Tanaka, S., Goto, S., Ueda, K., Kanehisa, M. (2004) Analysis of protein interaction networks in neurodegenerative disorders. *Genome Inform. Workshop*, **P125**, 1-2.

[9] Lisacek, F. (2003) Shaping biological knowledge *Pharmacogenomics*, **4**5-8.

[10] Maccioni, R.B, Munoz JP, Barbeito L. (2001) The molecular bases of Alzheimer's disease and other neurodegenerative disorders. *Arch Med Res.*, **32**367-81.

[11] Mizuta, Y., Collier N. (2004) Zone identification in biology articles as a basis for information extraction. *Proceedings of the International Workshop on Natural Language Processing in Biomedicine and its Applications (JNLPBA)*.

[12] Proux, D., Rechenmann, F., Julliard, L., Pillet, V.V., Jacq, B. (1998) Detecting gene symbols and names in biological texts: A first step toward pertinent information extraction. *Genome Inform. Ser. Workshop Genome Inform.*, **9**, 72-80.

[13] Raychaudhuri, S., Altman R. B. (2003) A literature-based method for assessing the functional coherence of a gene group. *Bioinformatics*, **19**, 396–401.

[14] Salton, G., (1971). The SMART Information Retrieval System. Prentice Hall, Englewood Cliffs, NJ.

[15] Storrie, B, Nilsson T. (2002) The Golgi apparatus: balancing new with old. *Traffic*, **3**, 521-9.

[16] Tanabe, L. and Wilbur, W.J. (2002) Tagging gene and protein names in biomedical text. *Bioinformatics*, **18**, 1124-1132.

[17] Teufel, S.(1998) Meta-discourse markers and problem-structuring in scientific articles. *Workshop on Discourse Structure and Discourse Markers.*

[18] Teufel, S., Moens, M.(2002) Summarizing scientific articles: Experiments with relevance and rhetorical status. *Computational Linguistics*, **28**, 409-445.

[19] Tran, P.B, Miller RJ. (1999) Aggregates in neurodegenerative disease: crowds and power? *Trends Neurosci.*, **22**,194-7.

[20] Yu, H., Hatzivassiloglou, V., Friedman, C., Rzhetsky, A., Wilbur, W.J. (2002) Automatic extraction of gene and protein synonyms from Medline and journal articles. *Proc. AMIA Symp*, 919-923.

[21] Zhang, Y, Callan, J, Minka, T (2002). Novelty and redundancy detection in adaptive filtering. *Proc. ACM SIGIR*,81-88.

Disease	**Proteins collected from 2000-2001 reviews**	**Proteins collected from 2004 review (Wetzel et al., 2004)**	**Proteins collected from paradigm shift sentences (reference)**
AD	Amyloid precursor protein (APP)	APP	APP: used as keyword
AD	tau	tau	tau: used as keyword
AD	presenilin	presenilin	presenilin: used as keyword
AD	γ-secretase	γ-secretase	γ-secretase (Pitsi, D. et al., 2002)
AD/PD	apoplipoprotein-E (apoE)	apoplipoprotein-E (apoE)	apoE (Trieu, V.N. and Uckun, F.M., 2000) (Mulder M., et al., 2001) (Glockner F., et al., 2002)
AD		apolipoprotein-J	Clusterin (Debure, L., et al., 2003)
AD	GSK3-β		GSK3-β (Kaytor, M.D. and Orr, H.T.,2002)
PD	α-synuclein	α-synuclein	α-synuclein: used as keyword
PD	parkin	parkin	Parkin: used as keyword
AD/PD	NMDA receptors	NMDA receptors	NMDA receptors (Grima, G., et al., 2001)
PD		synphilin	synphillin-1 (Engelender, S., et al., 1999)
PD		Pael-receptor	Not found
PD		Septin	Not found
PD		PTEN-induced kinase 1 (PINK 1)	PINK 1 (Mayo, M.W., et al., 2002) (Gary, D.S. and Mattson, M.P., 2002)
PD		DnaJ protein homolog 1 (HSP40)	HSP40 (Farinha, C.M., et al. 2002) (Boillee, S., et al., 2002)
PD		DJ-1	DJ-1 (Bonifati V., et al., 2003)
HD	huntingtin	huntingtin	huntingtin: used as keyword
HD		cAMP-response element-binding protein (CREB)/ CREB binding protein (CBP)	CREB/CBP (Takahashi, N., et al., 2002)
HD		PSD-95	PSD-95 (Olsen, O. and Bredt D.S., 2003)
HD		Htt-associated protein (HAP1)	HAP1 (Li Y., et al., 2001)
ALS	Cu/Zn SOD1	Cu/Zn SOD1	SOD1: used as keyword

Table 1: Published reviews cite proteins that play a key role in neurodegenerative diseases. Protein names manually extracted from reviews are shown in column 2 and 3 in correspondence with diseases listed in column 1 (AD: Alzheimer's, PD: Parkinson's, HD: Huntington's, ALS: Amyotrophic Lateral Sclerosis). Whenever these protein names were found in paradigm shift abstracts, the corresponding abstract is cited. Well-known characteristic proteins were used as keywords for ranking purposes.

Selected articles as cited in Table 1	Paradigm shift sentence(s)
Trieu, V.N. and Uckun, F.M., 2000	To our knowledge, the results presented herein provide the first experimental evidence that brain microglia-macrophage recruitment-accumulation is affected by apoE deficiency
Glockner F., et al., 2002	A number of studies investigating apoE expression and protein levels in AD brain in correlation to its genetic polymorphism has yielded controversial results.
Debure, L., et al., 2003	This activity might reconcile the opposite purported influences of clusterin on cell survival and explain how clusterin can be causally involved in neurodegeneration
Kaytor, M.D. and Orr, H.T.,2002	How GSK3 acts in this regard is still open to debate, but it may involve both extracellular and nuclear apoptotic activities.
Grima, G., et al., 2001	Glial-derived arginine, the nitric oxide precursor, protects neurons from NMDA-induced excitotoxicity. — This potential neuroprotective mechanism may lead to an alternative approach for the treatment of neurodegenerative diseases involving excitotoxic processes, such as ischemia.
Engelender, S., et al., 1999	It is unknown how alpha-synuclein contributes to the cellular and biochemical mechanisms of PD, and its normal functions and biochemical properties are poorly understood. — We identified a novel interacting protein, which we term synphilin -1 (encoded by the gene SNCAIP).
Boillee, S., et al., 2002	Early defect in the expression of mouse sperm DNAJ 1, a member of the DNAJ-heat shock protein 40 chaperone protein family, in the spinal cord of the wobbler mouse, a murine model of motoneuronal degeneration. — These results provide the first example of an association between transcriptional repression of a chaperone protein and a neurodegenerative process.)
Bonifati V., et al., 2003	The function of the DJ -1 protein remains unknown, but evidence suggests its involvement in the oxidative stress response. — Elucidating the physiological role of DJ -1 protein may promote understanding of the mechanisms of brain neuronal maintenance and pathogenesis of Parkinson's disease.
Olsen, O. and Bredt D.S., 2003	Surprisingly, and in contrast to previously published studies, we failed to detect GMP binding to the MAGUKs postsynaptic density -95 (PSD -95) and CASK.

Table 2: Example sentences of the cited references in Table 1

S100A6 protein	PMID: 12601007 (2003) Paradigm shift abstract	These results show that co -localization of annexin 11 and S100A6 at the nuclear envelope may be regulated either by elevation of intracellular Ca2+ or by cell cycle progression and provide the first evidence that these proteins may associate in vivo.
	PMID: 15590066 (2004) in most recent articles	"...may suggest that S100A6 plays a role in AD neuropathology."
Immunophilin FKBP	PMID: 12526984 (2002) Paradigm shift abstract	The immunophilin receptors implicated in generating the neurotrophic effects of FK506 and rapamycin (RM) are unknown (...) Our studies also show for the first time that the immunophilin FKBP is present in the human brain and that its levels are altered in the brain of patients with neurodegenerative diseases.
	PMID: 15514157 (2004) in most recent articles	These studies show that FKBP recruitment can protect cultured neurons from cell death triggered by toxic Aβ aggregates.

Table 3: Confirmation of the role of two proteins first detected with paradigm shift abstracts.

	caspase8	**Glyceraldehyde-3-phosphate dehydrogenase**
AD	In an attempt to reconcile these two hypotheses, we investigated APP processing during apoptosis and found that APP is processed by the cell death proteases caspase -6 and -8 (PMID: 10409650)	Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is a well-studied glycolytic enzyme that plays a key role in energy metabolism [...] In recent years, that view has changed since GAPDH is now known to contribute to a number of diverse cellular functions unrelated to glycolysis [...] Most recently, it has been shown that GAPDH is a target for deprenyl related compounds (Carlile et al., 2000; Kragten et al., 1998) and may contribute to the neuroprotection offered by those compounds. (PMID: 11205159)
HD	This novel non-receptor-mediated pathway for activating caspase -8 might contribute to neuronal death in Huntington disease (PMID: 11788820)	There is growing evidence that truncated forms of mutant Htt in the nucleus influence gene transcription by binding to proteins such as CREB binding protein (CBP) response element binding protein binding protein, N-COR, glyceraldehyde -3-phosphate dehydrogenase, and p53. (PMID: 11466559)

Table 4: Sentences that highlight the multiple disease implications of some proteins in neurodegeneration (AD: Alzheimer's, HD: Huntington's) with correspondence with KEGG data.