Case Report ■

Generating Hypotheses by Discovering Implicit Associations in the Literature: A Case Report of a Search for New Potential Therapeutic Uses for Thalidomide

MARC WEEBER, PHD, REIN VOS, MD, PHD, HENNY KLEIN, PHD,
LOLKJE T. W. DE JONG-VAN DEN BERG, PHD, ALAN R. ARONSON, PHD,
GRIETJE MOLEMA, PHD

A b s t r a c t   The availability of scientific bibliographies through online databases provides a rich source of information for scientists to support their research. However, the risk of this pervasive availability is that an individual researcher may fail to find relevant information that is outside the direct scope of interest. Following Swanson's *ABC* model of disjoint but complementary structures in the biomedical literature, we have developed a discovery support tool to systematically analyze the scientific literature in order to generate novel and plausible hypotheses. In this case report, we employ the system to find potentially new target diseases for the drug thalidomide. We find solid bibliographic evidence suggesting that thalidomide might be useful for treating acute pancreatitis, chronic hepatitis C, *Helicobacter pylori*-induced gastritis, and myasthenia gravis. However, experimental and clinical evaluation is needed to validate these hypotheses and to assess the trade-off between therapeutic benefits and toxicities.

Researchers practice science in highly specialized intellectual environments and communities. It takes years of training to contribute new discoveries to the continually growing body of scientific knowledge. The staggering amount of available online knowledge and information, concerning both the personal field of expertise and developments in other disciplines, may overwhelm the individual researcher. Moreover, the latter, relevant, information may be overlooked. Swanson has phrased this particular situation as one of "complementary but disjoint structures within the literature of science."[1]

Affiliations of the authors: Department of Social Pharmacy and Pharmacoepidemiology, Groningen University Institute for Drug Exploration, The Netherlands (MW, HK, LTWdJ-vdB); Lister Hill National Center for Biomedical Communications, National Library of Medicine, Bethesda, Maryland (MW, ARA); HEALTH Research Institute, Department of Health Ethics and Philosophy, University of Maastricht, The Netherlands (RV); Department of Pathology and Laboratory Medicine, Medical Biology Section, Groningen University Institute for Drug Exploration, The Netherlands (GM).

Assume that one set of literature discusses the argument that *A* relates to *B*, while a separate literature provides a discussion on how *B* relates to *C*. These sets may have no articles in common and are therefore disjoint, yet the arguments are complementary because, when they are combined, there is the inference that *A* relates to *C*. Using titles from the bibliographic database MEDLINE, Swanson has made several discoveries in biomedicine by connecting disconnected sets of literature according to this simple *ABC* model.[2–4] Importantly, the first two literature-based hypotheses of the therapeutic effects of fish oil for patients with Raynaud's disease and the role of magnesium deficiency in migraine were corroborated clinically.[5] For drug discovery, this model is similar to Vos' model,[6] in which discovery is the rapprochement of a drug profile (*AB*) to a disease profile (*BC*).

Following Swanson's ideas of connecting disconnected sets of literature, we developed a literature-based scientific discovery support tool that applies advanced Natural Language Processing (NLP) techniques to MEDLINE citations.[7,8] This tool provides the means to efficiently analyze huge amounts of textual data and to generate new hypotheses in the biomedical domain. The hypotheses can then be evaluated bibliographically by studying computer-selected MEDLINE titles and abstracts.

In this case report we have generated new knowledge in the domain of drug discovery by finding new hypotheti-

cal therapeutic applications of the drug thalidomide. Between 1959 and 1961, thalidomide (α-N-phtalimido-glutarimide) was a popular over-the-counter sedative. Devastating teratogenic effects led to its withdrawal from the market. In recent years, however, interest in thalidomide has intensified based on its reported immunomodulatory and anti-inflammatory properties.[9–12] In 1998, the Food and Drug Administration (FDA) approved thalidomide for the indication of erythema nodosum leprosum, an inflammatory manifestation of leprosy.[13,14] Additionally, thalidomide seems to have beneficial effects on ulcers and wasting associated with HIV infection.[11,13,14] The observation that new knowledge can be generated on a relatively old drug and that research is still ongoing led to our hypothesis that there may be new, not yet discovered applications for thalidomide.

Materials and Methods

Concept-based Discovery Support Tool

In earlier publications, we reported on the development of a concept-based discovery support system.[7] Using this system, we were able to simulate, or rediscover, two of Swanson's discoveries (i.e., the association between the fish oil and Raynaud's disease and the association between magnesium deficiency and migraine).[8] These papers provide more details of the system. The most salient feature of the system is that it is "concept-based." This means that it applies NLP techniques to identify biomedical concepts[15] in PubMed titles and abstracts at <http://www. ncbi.nlm.nih.gov/PubMed/>. The concepts originate from the Unified Medical Language System (UMLS) Metathesaurus,[16] the largest biomedical thesaurus to date. The use of concepts instead of text words has several advantages. First, different expressions such as variants or synonyms collapse to one concept. For instance, IL-12, IL12, interleukin 12, CLMF, cytotoxic lymphocyte maturation factor, and natural killer cell stimulatory factor all refer to the same concept: *Interleukin-12*. Second, finding meaningful strings of multiple words (phrases, compounds) is nontrivial in NLP. By using concepts from the Metathesaurus, only biomedically relevant (multiword) strings are used, whereas the rest are discarded as noise. The most important reason to use UMLS concepts, however, is the semantic information that is added to them by human experts. All concepts have been assigned to one or more semantic categories or types. There is a total of 134 categories, including "Disease or Syndrome," "Gene or Genome," and "Amino Acid, Peptide, or Protein." The concept *Thalidomide*, for instance, has been assigned the semantic types "Organic Chemical," "Pharmacologic Substance," and "Hazardous or Poisonous Substance." In our discovery procedure we use this semantic information as a filter to reduce the size of the search space.

The Discovery Process

We have divided the discovery process into two steps: the generation of a hypothesis and its subsequent evaluation. In the hypothesis generation phase, only one pathway (represented by *B* in the *ABC* model) is pursued because of the potentially huge search space. Once a hypothesis that *A* relates to *C* has been generated, a search in a more restricted space may uncover additional pathways (*B*) between *A* and *C* and thus strengthen (or reject) the hypothesis.

Experimental Setting

Two authors of this paper (MW and GM), an information scientist and a pharmacologist/immunologist, respectively, performed the actual discoveries and formulated the proposed hypotheses. MW was the main developer of the system and used it to retrospectively (re)discover Swanson's most famous discoveries.[8] GM was involved as the prototypical user of a discovery support system. As a researcher with a background in pharmacology and immunology, she acknowledged the overflow of information in current biomedical research practices and was interested in putting the system to a real-life test. GM's pharmacological domain knowledge of thalidomide was of a general kind. She knew the recently discovered effects of thalidomide on TNFα mRNA degradation and its role in suppressing pro-inflammatory processes. GM's immunology expertise is broad, but she has no specific background in any of the mentioned diseases.

The search processes and the interaction with the discovery support system were mostly done in collaboration between MW and GM. It consisted of several one-hour sessions during a two-week period. Several search parameters and cut-off settings were based on pragmatic issues such as reducing the list of possibilities to a manageable amount (not too long) without using too much of a special focus (not too short).

Generating Hypotheses

We started with thalidomide as *A* in the *ABC* model (Fig. 1). The discovery tool downloaded PubMed titles and abstracts mentioning thalidomide (and its variants and synonyms) on July 27, 2000 and mapped the natural language texts to UMLS concepts. Subsequently, we applied our semantic filter. We selected only UMLS concepts classified as "Immunologic Factor" from sentences mentioning *Thalidomide* because we hypothesized that we may find new therapeutic applications through the immunologic actions of the drug. The system provided a rank-ordered list of immunologic concepts as output. The discovery tool has been developed in such a way that the user also can view each of the immunologic concepts in the original context. Based on background knowledge and the provided bibliographic information, we selected promising concepts that are the *B*s in the *ABC* model (see Fig. 1). "Promising" was defined on two grounds. First, the concepts were ordered by frequency. The higher the frequency, the more likely it is that this factor is related to thalidomide. Also, it provides us a substantial amount of possible textual "evidence" of this putative relation. Second, the expert used her knowledge to look for promising novel pathways.
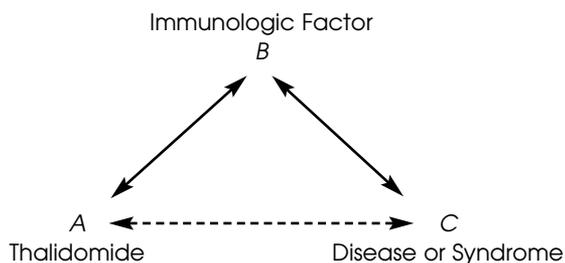
Immunologic Factor
B

A ← - - - - - - - - - - - - - - - → *C*
Thalidomide   Disease or Syndrome

F i g u r e 1. The *ABC* discovery model. Starting with thalidomide (*A*), immunologic factors (*B*) known to be affected by the drug can be identified in the PubMed database when applying the discovery support tool. In the second step, diseases (*C*) for which these factors are known to be important are identified. The relationships *AB* and *BC* (solid lines) are known and reported in the literature, the relationship *AC* (dashed line) is unknown to date and is putatively a new discovery.

The discovery tool downloaded the PubMed citations on the most promising *B*s and identified the available concepts. Again, a semantic filter was applied. This time, we were interested in diseases: only concepts classified as "Disease or Syndrome" were presented, representing the *C* concepts in the *ABC* model (see Fig. 1). The system presented the sentences in which the *A* and *B* concepts and the sentences in which the *B* and *C* concepts co-occurred in a juxtaposed manner to facilitate the user to make the inference that thalidomide (*A*) may be effective in disease *C* through immunologic pathway *B*.

Filtering Process

As the list of diseases was very long, we adopted the following strategy to reduce the number of hypotheses. First, we selected diseases of which the literature had more than two immunologic *B* concepts in common with the thalidomide literature. Lack of co-occurring immunologic concepts represented a lack of useful immunologic pathways and therefore was not considered interesting. Subsequently, we executed automated PubMed queries on thalidomide combined ("AND" in PubMed) with every disease *C*. Of note, our tool searched only in titles and abstracts, thus ignoring Medical Subject Headings (MeSH), and *AB* and *BC* relations were identified as existing only if the concepts co-occurred in one sentence. This means that our search was more restricted (and more precise) than a normal PubMed search because by default PubMed searches in all available fields. If the number of hits on a default PubMed query was larger than three, we discarded the finding based on the assumption that the relation between *A* and *C* was known and hence did not represent a new discovery. These cut-off values were mainly based on our aim to reduce human workload. With the current thresholds, about one hundred diseases had to be assessed by the expert. Using the support system's contextual presentation of the bibliographic "evidence," this was a manageable task. When other cut-off values are used, the list will either contain too many diseases or not represent novel possible therapeutic applications.

For the remainder of the diseases, we studied the output of the discovery system on the putative pathway between thalidomide and disease *C*, using the extracted sentence, the complete PubMed citation, and the full text paper, if available. At this stage, we had generated a list of diseases for which we had found some bibliographic indications that thalidomide might be an effective pharmacologic agent.

Evaluating Hypotheses

In the evaluation process we tried to find additional bibliographic and other evidence for the putative pathways between thalidomide and the listed diseases. The first step was to download and analyze citations on each of the diseases (*C*; see Fig. 1) found in the hypothesis generation phase. From these citations the discovery system selected immunologic factors (*B*) that co-occurred in sentences with this disease. The *B* concepts also had to co-occur in the thalidomide literature in sentences that contained the concept *Thalidomide*. The system listed these concepts and provided the juxtaposed *AB* and *BC* sentences for human expert assessment. This process strengthened some hypotheses while rejecting others.

Other Databases

In addition to PubMed, we also queried other databases available at Groningen University: Biological Abstracts (from 1990 to 2000/06), CINAHL—Nursing & Allied Health (from 1982 to 2000/06), EMBASE—Excerpta Medica (from 1989 to 2000/06), and Current Contents (from 1997 to 2000/07/28). Additionally, we queried the Internet through Altavisa at <http://www.altavista.com> and Google at <http://www.google.com>.

Results

Finding Immunologic Factors

The discovery system downloaded 1,366 citations that included (a variant of) the concept *Thalidomide* in either title or abstract. These variants were "thalidomide," "sedoval," "synovir," and "kevadon." In the sentences in which *Thalidomide* occurred, 3,860 different concepts co-occurred. Only 82 of them were assigned the semantic type "Immunologic factor," a reduction of 98%. Table 1 shows the 31 concepts that had a frequency of occurrence higher than two. A lower frequency indicated that the concepts were hardly mentioned in association with thalidomide and were not investigated further. Concepts such as *Antigens, Antibody*, or *Chemotactic factors* are too general to be useful. "General" corresponds loosely with the position of the concept in the thesaurus hierarchy. The higher a concept is located in the hierarchical tree, the more general it is. Because we are interested in specific and testable hypotheses, more precise concepts are preferred.

The high frequency of the concept *Tumor necrosis factor* (TNFα), illustrated thalidomide's well-known characteristic of inhibition of TNFα production via increased

TNFα mRNA degradation. The search also provided the concepts *Interleukin-12* and *Interleukin-10* (IL-10) (Table 2). The domain expert almost instantly selected these concepts for further exploration. Because of her immunologic knowledge about the role of different cytokines in the T-helper cell differentiation, she appreciated thalidomide's pharmacological potential in this immunologic process.

Thalidomide has strong inhibitory effects on mononuclear cell production of IL-12[9,17] and a stimulatory effect on IL-10 production.[10] IL-12 is a pleitropic cytokine that favors the differentiation of T-helper 0 (Th0) cells into T-helper 1 (Th1) cells; hence the cellular response against cells expressing auto-antigens.[18] The subsequent induced production of IFNγ by T-cells and natural killer (NK) cells, furthermore, enhances general immune cell activation, growth, and differentiation.[19] Based on this bibliographic information, schematically depicted in Figure 2, we selected the concept *Interleukin-12* as the immunologic pathway that may result in discovering new applications for thalidomide.

Finding Diseases through Interleukin-12

The discovery system downloaded and analyzed the 3,846 MEDLINE citations in which the *B* concept *Interleukin-12* occurred 16,262 times. In the sentences with IL-12, 5,707 different other concepts occurred, of which 420 had the semantic type "Disease or Syndrome," a 93% reduction. During the filtering process, as described in the Methods section, we found that for some cases the information indicated that the pharmacologic pathway would exacerbate instead of alleviate the disease. We also found NLP errors in the mapping process. The number "pi," for instance was incorrectly mapped to the disease *Pulmonary valve insufficiency*. Additionally, some disease concepts were considered too general (*Critical Illness*), whereas others were too specific. For instance, the retrieved concept *Salmonella infections* did not have any PubMed citations in common with thalidomide. However, the PubMed query "salmonella AND thalidomide" returned nine citations, indicating that thalidomide has received interest in a salmonella context.

Interestingly, in the filtering process we were able to strengthen the hypothesis put forward in one citation that multiple sclerosis (MS) might benefit from thalidomide. The reported putative pathway concerned thalidomide's inhibitory effects on TNFα synthesis, thereby preventing acute exacerbations.[22] In the underlying study, we found bibliographic evidence for the importance of the IL-12 in this disease.[21,22] We also found that a shift in Th1/Th2 balance might be desirable to improve disease status.[23]

By using the described filtering techniques we were able to obtain a list of 12 diseases that may benefit from thalidomide treatment (see Table 2). In other words, we generated 12 new hypotheses for thalidomide use.

Table 1 ■ Immunologic Factors Identified by the Discovery Support Tool that Co-occur in Sentences with Thalidomide

Concept	Frequency
Tumor necrosis factor	312
Interleukin-2	58
Adjuvants, immunologic	40
Cytokines	34
Antigens, CD4	15
Lymphocyte antigen CD69	14
Interleukin-12	13
Phytohemagglutinins	12
Antigens, CD5	11
Antigens	11
Antibodies	11
Antigens, CD8	10
Integrins	8
Interleukin-6	7
Interleukin-10	7
Antigens, CD3	6
Receptors, interleukin-2	6
Purified protein derivative of tuberculin	6
IgM	6
Antithymoglobulin	5
Granulocyte-macrophage colony-stimulating factor	5
Interleukin-4	5
Antibodies, monoclonal	5
Interleukin-1	4
Interferon type II	4
Antibodies, anti-idiotypic	4
Chemotactic factors	3
Interleukin-8	3
POA	3
IgG	3
HLA antigens	3

*These factors are potentially of interest because they may be affected by thalidomide. Presented factors occurred with a frequency > 2.

Evaluating Hypotheses

The second column of Table 2 provides the number of PubMed citations that the discovery system analyzed for each of the 12 diseases. For some of these diseases, the bibliographic evidence was not strong because there was either not enough information or the information available was contradictory. For other cases we assumed that there was too much "circumstantial bibliographic evidence" to claim a new discovery (i.e., in diseases closely related to the identified diseases for which thalidomide had already been investigated extensively). For instance, at the time no link in the literature indicated that patients with atherosclerosis may benefit from thalidomide. However, this disease is often associated with systemic lupus erythematosus, a disease for which thalidomide treatment has been investigated. Similarly, atherosclerosis is characterized by angiogenesis (i.e.,

Table 2 ■ Twelve Diseases for Which Thalidomide May Be Used Therapeutically

Disease	No. of MEDLINE Citations
Acute pancreatitis	8,446
Atherosclerosis	15,973
Brucellosis	4,282
Chickenpox	6,304
Chronic hepatitis C	1,023
Graves' disease	5,353
H. pylori-induced gastritis	7,108
Myasthenia gravis	6,231
Pulmonary fibrosis	1,652
Purpura	7,660
Sialadenitis	706
Sjögren's syndrome	5,262

*These uses are not yet reported in the biomedical literature (PubMed) and are considered hypotheses that warrant further bibliographic investigation. The second column provides the number of citations to be analyzed.

new blood vessel formation from pre-existing ones), a complex cellular phenomenon for which thalidomide's regulatory properties are widely acknowledged. In a similar manner we judged that there was too much circumstantial evidence for thalidomide application in Sjögren's syndrome and the closely related disease sialadenitis. Indeed, a search in the NIH clinical trial database at <http://ClinicalTrials.gov> in January 2001 showed that a phase II clinical trial was being conducted on thalidomide in patients with Sjögren's syndrome.

Diseases Discovered as Potential Targets for Thalidomide Treatment

Based on the discovery support system's output (sentences with co-occurring relevant topics, PubMed abstracts, and sometimes full papers available through publishers' websites) we compiled a list of four diseases for which we hypothesize that thalidomide may be an effective treatment (Table 3).

Chronic Hepatitis C

Chronic hepatitis C (CHC) is an inflammatory disease of the liver caused by the hepatitis C virus. The Th1/Th2 cytokine balance is important in persistence of infection and liver injury in CHC.[24] Progressive liver injury in CHC is associated with upregulation of intrahepatic Th1 cytokines (IFNγ, IL-2),[25] downregulation of Th2 cytokines,[26] and the induction of IL-12.[27] Besides a pronounced production of the Th1-associated cytokines, circulating levels of the pro-inflammatory cytokine TNFα also correlated with the degree of inflammation in the liver.[28,29]

Myasthenia Gravis

Myasthenia gravis (MG) is an organ-specific autoimmune disease, putatively aimed at the nicotine acetylcholine receptor (AChR). MG afflicts the neuromuscular junctions.[30] The pathophysiologic background of MG is complex and at present mostly unknown. Aberrant production of, among others, TNFα, IL-10, and IL-12, after antigen challenge is considered to play a role in the pathology.[31,32] In addition, a dysbalance in Th1/Th2 cell activation has been implicated in the disease.[33]

H. Pylori-Induced Gastritis

Gastritis is an inflammation of the lining of the stomach that can be caused, for example, by prolonged irritation from the use of non steroidal anti-inflammatory drugs and infection with the *H. pylori* bacteria. Most MEDLINE abstracts concerned H. pylori-induced gastritis. In a mouse model, *H. pylori*-induced mucosal inflammation was shown to be Th1-mediated, with disease exacerbation in IL-4, but not IFNγ, gene-deficient mice.[34] A failure to promote Th2 relative to Th1 responses would impede resolution of the infection and promote chronic inflammation. Furthermore, treatment of mice with IL-12 or anti-IFNγ antibodies, respectively, decreased the severity of gastritis, while adoptive transfer of Th1 cell lines significantly exacerbated the inflammation.[35]
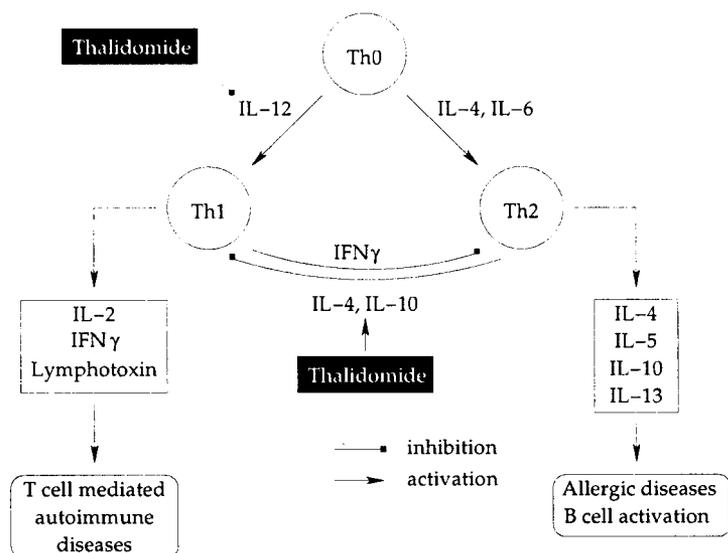

Figure 2. Role of cytokines in Th1/Th2 differentiation and in the regulation of immunopathological manifestation. Thalidomide inhibits the production of IL-12 and induces the production of IL-10, among others, and may therefore strongly affect T cell mediated immune responses. Adapted from reference 33, ch. XXVI, p. 322.

Acute Pancreatitis

In acute pancreatitis (AP), serum levels of IL-1, IL-6, and especially TNFα have been demonstrated to be elevated.[36,37] Additionally, administration of the Th2 type cytokine IL-10 exerted a protective effect in experimental AP.[38,39] The role of IL-12 in AP has not been investigated intensively, but the few citations available suggest that increased amounts of IL-12p40 in AP patients may be responsible for their increased susceptibility to infection.[40] This finding suggests a differentiation toward a Th1 type of immune reaction in patients with AP. PubMed provided one reference that co-cited thalidomide and pancreatitis.[41] In this case, however, pancreatitis was supposed to be a previously unreported, side effect of thalidomide use.

Searching Other Databases

EMBASE provided one citation with regard to chronic hepatitis C and thalidomide.[42] In this poster presentation, it was observed that in patients with multiple myeloma who were treated with thalidomide, two responders had a concomitant improvement in CHC. The authors did not provide any explanation of this observation. EMBASE also included one citation on thalidomide treatment of myasthenia gravis, a Brazilian conference paper in Portuguese.[43] A web search resulted in one interesting page that described new potential treatments for thalidomide: "We also are looking at thalidomide in Alzheimer's disease and myasthenia gravis."[44] EMBASE, furthermore, provided several new references with regard to pancreatitis and thalidomide. Interestingly, most of these references were also indexed in PubMed. It turned out that the difference in index terms caused the positive identification by EMBASE. These four citations seemed to suggest that pancreatitis is a side effect of thalidomide; however, closer inspection revealed that pancreatitis was put forward as a side effect caused by corticosteroids rather than by thalidomide.

In summary, only a little information is available about a relation between thalidomide and chronic hepatitis C and myasthenia gravis. We have found strong bibliographic evidence that suggests a mechanistic (immunologic) point of view on thalidomide and chronic hepatitis C and myasthenia gravis that is novel and has not been reported yet. Furthermore, we have not found any direct bibliographic indication that *H. pylori*-induced gastritis and pancreatitis may benefit from thalidomide. Therefore, these diseases represent truly novel potential therapeutic targets for thalidomide. All of these diseases have been related to a dysbalance in Th1/Th2 immunologic responses favoring Th1 reactions. However, we also found some references on the opposite situation, i.e., the production of excessive Th2 cytokines in the disease of interest. For example, a small number of studies indicated that the Th1/Th2 balance in CHC tends to favor Th2 responses.[45,46] Whether the observed elevated Th2 cytokine levels represented a systemic response or are a result of increased local production within the

Table 3 ■ Four Diseases for Which Thalidimide May Be Used*

Disease
Acute pancreatitis
Chronic hepatitis C
H. pylori-induced gastritis
Myasthenia gravis

*These diseases have not been associated with thalidomide before, but we uncovered solid bibliographic evidence of an immunologic pathway through Th1/Th2 differentiation which may be affected by thalidomide.

liver[46] remains unknown at present. Furthermore, it should be taken into account that disease initiation and subsequent progression may be differentially controlled by Th1 and/or Th2 specific cytokines and that patients can become prone to other infections when a Th1 response is therapeutically eradicated. Only experimental data about the effects of thalidomide will provide the final evidence whether the diseases identified using our discovery support tool truly benefit from the drug.

Discussion

We have reported on the application of a computer discovery support tool to identify complementary but disjoint structures within the literature of medical life sciences regarding disease profiles that may benefit from thalidomide. We argued that thalidomide shifts the Th1/Th2 balance toward Th2 in human immune response. In theory, autoimmune diseases that are characterized by a Th1 differentiation may benefit from thalidomide use. Because of the complexity of the processes in the immune system, it was difficult to grasp the molecular changes induced by thalidomide when applied to the large variety of diseases known to date. However, the literature-based discovery support system successfully facilitated the formation of hypotheses regarding some of thalidomide's mechanistic pathways. The current investigations provided only bibliographic evidence for positive pharmacologic effects of thalidomide in the diseases identified. Experimental and clinical evaluation of therapeutic benefits versus toxicities should shed light on their potential use for the treatment of various pro-inflammatory diseases that have not been studied in this context yet.

As we executed our discovery task in the end of July 2000, we now have two more years of medical research and publications to evaluate our proposed hypotheses. The application of thalidomide as a potential treatment for chronic hepatitis C has been a subject of recent discussion.[47] Additionally, it is interesting to observe that one of the reasons of using INFα for treating hepatitis C is the inhibition of INFγ, which is also one of the pharmacologic effects of thalidomide. To complicate matters, however, clinical trials are also being conducted to use INFγ for IFNα nonresponders (see <http://ClinicalTrials.gov>). This is a further illustration of the

human immune system's complexity and also suggests careful pre-clinical and clinical testing to truly evaluate the generated hypotheses.

New knowledge about existing drugs is generated continuously. Famous examples of new indications for existing drugs include acetylsalicylic acid (aspirin) as prophylaxis for myocardial infarction and colorectal cancer,[48] minoxidil for male pattern baldness,[49] and sildenafil (Viagra) for erectile dysfunction.[50] The discovery of these new applications originated from clinical observation, theoretical reasoning, and serendipity. With our discovery support system, we have extended the process of drug discovery by generating hypotheses in a systematic manner by combining existing knowledge. This aspect of knowledge re-use may result in a more rapid identification of other potential applications for drugs as well as identification of patterns of side effects, for example.

Working with our tool is more complicated than executing a general PubMed search. It is an intellectually intensive process in which a domain expert has to continually evaluate hypotheses to reduce the long initial list to strong, testable ones. An example of the indispensability of expert knowledge concerns the discovery of myasthenia gravis as target disease for thalidomide. During the analyses, one PubMed citation was found that co-mentioned thalidomide and experimental myasthenia gravis.[51] This study, conducted in Lewis rats, did not show any pharmacologic effect of thalidomide. A likely reason for this is the fact that Lewis rats respond in a typical Th1 fashion after immunologic challenge.[52] In a system in which the Th1 response prevails, pharmacologic interference at this level may not be sufficient to obtain an effect. The expert knowledge enabled us not to identify this citation as a refutation of our hypothesis.

This example of the human expert's indispensability is typical to our view of literature-based discovery. Computational tools assist human domain experts in studying the literature and deriving and evaluating novel hypotheses. Using tools such as our system may lead to more focused laboratory experiments and may make new and fruitful connections between scientific disciplines. The scientific environment of biomedical research is characterized by a significant increase of data and information from highly automated experiments and online databases. Intelligent and computer-assisted approaches to sift through this information will become necessary tools on the researcher's bench in the near future.

References ∎

1. Swanson DR. Complementary structures in disjoint science literatures. In Bookstein A, Chiaramella Y, Salton G, Raghavan VV (eds): Proceedings of the 14th Annual International ACM/SIGIR Conference. New York, ACM Press, 1991. pp 280–289.
2. Swanson DR. Fish oil, Raynaud's syndrome, and undiscovered public knowledge. Perspect Biol Med 1986;30(1):7–18.
3. Swanson DR. Migraine and magnesium: Eleven neglected connections. Perspect Biol Med 1988;31(4):526–557.
4. Swanson DR, Smalheiser NR. An interactive system for finding complementary literatures: A stimulus to scientific discovery. Artif Intell 1997;91(2):183--203.
5. Smalheiser NR, Swanson DR. Using ARROWSMITH: A computer-assisted approach to formulating and assessing scientific hypotheses. Comput Methods Programs Biomed 1998; 57:149–153.
6. Vos R. Drugs Looking for Diseases. Dordrecht, The Netherlands, Kluwer Academic Publishers, 1991.
7. Weeber M, Klein H, Aronson AR, Mork JG, et al. Text-based discovery in biomedicine: The architecture of the DAD-system. In Overhage JM (ed): Proceedings of the AMIA Annual Fall Symposium. 2000. Philadelphia, Hanley & Belfus, 2000. pp 903–907.
8. Weeber M, Vos R, Klein H, de Jong-van den Berg LTW. Using concepts in literature-based discovery: Simulating Swanson's Raynaud–fish oil and migraine–magnesium discoveries. J Am Soc Inf Sci Tech 2001;52(7):548–557.
9. Calabrese L, Fleischer AB. Thalidomide: Current and potential clinical applications. Am J Med 2000;108:487–495.
10. Corral LG, Haslett PA, Muller GW, Chen R, et al. Differential cytokine modulation and T cell activation by two distinct classes of thalidomide analogues that are potent inhibitors of TNF-alpha. J Immunol 1999;163:380–386.
11. Marriott JB, Muller G, Dalgleish AG. Thalidomide as an emerging immunotherapeutic agent. Immunol Today 1999; 20(12):538–540.
12. Sands BE, Podolsky DK. New life in a sleeper: Thalidomide and Crohn's disease. Gastroenterology 1999;117(6):1485–1488.
13. Powell RJ. New roles for thalidomide. BMJ 1996;313:377–378.
14. Raje N, Anderson K. Thalidomide—A revival story. N Engl J Med 1999;341(21):1606–1609.
15. Aronson AR. Effective mapping of biomedical text to the UMLS metathesaurus: The MetaMap Program. In Bakken S (ed): Proceedings of the AMIA Annual Fall Symposium. 2001, Philadelphia, Hanley & Belfus, 2001, pp 17–21.
16. McCray AT, Miller RA. Making the conceptual connections: the Unified Medical Language System (UMLS) after a decade of research and development. J Am Med Inform Assoc 1998; 4(6):484–500.
17. Moller DR, Wysocka M, Greenlee BM, Ma X, et al. Inhibition of IL-12 production by thalidomide. J Immunol 1997;159:5157–5161.
18. McHugh SM, Rifkin IR, Deighton J, Wilson AB, et al. The immunosuppressive drug thalidomide induces T helper cell type 2 (Th2) and concomitantly inhibits Th1 cytokine production in mitogen- and antigen-stimulated human peripheral blood mononuclear cell cultures. Clin Exp Immunol 1995; 99(2):160–167.
19. Callard RE, Gearing A. The Cytokine Factsbook. London, Academic Press, 1994.
20. Sastry PS. Inhibition of TNF-alpha synthesis with thalidomide for prevention of acute exacerbations and altering the natural history of multiple sclerosis. Med Hypoth 1999;53(1):76–77.
21. Drulovic J, Mostarica-Stojkovic M, Levic Z, Stojsavljevic N, Pravica V, Mesaros S. Interleukin-12 and tumor necrosis factor-alpha levels in cerebrospinal fluid of multiple sclerosis patients. J Neurol Sci 1997;147(2):145–150.
22. Inoue A, Koh C-S, Yamazaki M, Yahikozawa H, et al. Suppressive effect on Theiler's murine encephalomyelitis virus-induced demyelinating disease by the administration of anti-IL-12 antibody. J Immunol 1998;161(10):5586–5593.
23. Zang YC, Yang D, Hong J, Tejada-Simon MV, et al. Immunoregulation and blocking antibodies induced by interferon beta treatment in MS. Neurology 2000;55(3):397–404.

24. McHutchison JG, Giannelli G, Nyberg L, Blatt LM, et al. A pilot study of daily subcutaneous interleukin-10 in patients with chronic hepatitis C infection. J Interferon Cytokine Res 1999;19(11):1265–1270.

25. Shields PL, Morland CM, Salmon M, Qin S, Hubscher SG, Adams DH. Chemokine and chemokine receptor interactions provide a mechanism for selective T cell recruitment to specific liver compartments within hepatitis C-infected liver. J Immunol 1999;163(11):6236–6243.

26. Napoli J, Bishop GA, McGuinness PH, Painter DM, McCaughan GW. Progressive liver injury in chronic hepatitis C infection correlates with increased intrahepatic expression of Th1-associated cytokines. Hepatology 1996;24(4):759–765.

27. Quiroga JA, Martin J, Navas S, Carreno V. Induction of interleukin-12 production in chronic hepatitis C virus infection correlates with the hepatocellular damage. J Infect Dis 1998; 178(1):247–251.

28. Neuman MG, Benhamou JP, Martinot M, Boyer N, et al. Predictors of sustained response to alpha interferon therapy in chronic hepatitis C Clin Biochem 1999;32(7):537–545.

29. Fabris C, Soardo G, Falleti E, Toniutto P, et al. Relationship among hepatic inflammatory changes, circulating levels of cytokines, and response to IFN-gamma in chronic hepatitis C. J. Interferon Cytokine Res 1998;18(9):705–709.

30. Balasa B, Sarvetnick N. Is pathogenic humoral autoimmunity a Th1 response? Lessons from myasthenia gravis. Immunol Today 2000;21:19–23.

31. Huang WX, Huang P, Fredrikson S, Pirskanen R, Hillert J. Decreased mRNA expression of TNF-alpha and IL-10 in non-stimulated peripheral blood mononuclear cells in myasthenia gravis. Eur J Neurol 2000;7(2):195–202.

32. Matusevicius D, Navikas V, Palasik W, Pirskanen R, Fredrikson S, Link H. Tumor necrosis factor-alpha, lymphotoxin, interleukin (IL)-6, IL-10, IL- 12 and perforin mRNA expression in mononuclear cells in response to acetylcholine receptor is augmented in myasthenia gravis. J Neuroimmunol 1996;71(1–2):191–198.

33. Theze J. The Cytokine Network and Immune Function. Oxford, Oxford University Press, 1999.

34. Smythies LE, Waites KB, Lindsey JR, Harris PR, Ghiara P, Smith PD. *Helicobacter pylori*-induced mucosal inflammation is Th1 mediated and exacerbated in IL-4, but not IFN-gamma, gene-deficient mice. J Immunol 2000;165(2):1022–1029.

35. Blanchard TG, Czinn SJ. Immunological determinants that may affect the *Helicobacter pylori* cancer risk. Aliment Pharmacol Ther 1998;12(Suppl. #1):83–90.

36. Blanchard JA II, Barve S, Joshi-Barve S, Talwalker R, Gates LK Jr. Cytokine production by CAPAN-1 and CAPAN-2 cell lines. Dig Dis Sci 2000;45(5):927–932.

37. McKay CJ, Gallagher G, Brooks B, Imrie CW, Baxter JN. Increased monocyte cytokine production in association with systemic complications in acute pancreatitis. Br J Surg 1996; 83(7):919–923.

38. van Laethem JL, Eskinazi R, Louis H, Rickaert F, Robberecht P, Deviere J. Multisystemic production of interleukin 10 limits the severity of acute pancreatitis in mice. Gut 1998;43(3): 408–413.

39. Rongione AJ, Kusske AM, Kwan K, Ashley SW, Reber HA, McFadden DW. Interleukin 10 reduces the severity of acute pancreatitis in rats. Gastroenterology 1997;112(3):960–967.

40. Pezzilli R, Miniero R, Cappelletti O, Barakat B. Behavior of serum interleukin 12 in human acute pancreatitis. Pancreas 1999;18(3):247–251.

41. Mehta P, Kedar A, Graham-Pole J, Skoda-Smith S, Wingard JR. Thalidomide in children undergoing bone marrow transplantation: Series at a single institution and review of the literature. Pediatrics 1999;103(4):e33.

42. Durie BGM, Stepan DE. Efficacy of low dose thalidomide (T) in multiple myeloma. Blood 1999;94(10 Suppl 1 Part 1):316a.

43. Lamartine de Assis J. The treatment of acquired myasthenia gravis in clinical neurology of the Clinical Hospital of the Faculty of Medicine of the University of Sao Paulo. Rev Bras Neurol 1994;30(3):62–71.

44. Brown KS. New uses for thalidomide yielding valuable lessons. Scientist 1997;11(3):1–8.

45. Kobayashi K, Ishii M, Igarashi T, Satoh T, et al. Profiles of cytokines produced by CD4-positive T lymphocytes stimulated by anti-CD3 antibody in patients with chronic hepatitis C. J Gastroenterol 1998;33(4):500–507.

46. Reiser M, Marousis CG, Nelson DR, Lauer G, et al. Serum interleukin 4 and interleukin 10 levels in patients with chronic hepatitis C virus infection. J Hepatol 1997;26:471–478.

47. Raufman JP, Lamps LW. Thalidomide-induced normalization of serum ALT levels in a patient with hepatitis C. Am J Gastroenterol 2001;96(11):3209–11.

48. Giovannucci E, Egan KM, Hunter DJ, Stampfer MJ, et al. Aspirin and the Risk of Colorectal Cancer in Women. N Engl J Med 1995;333(10):609–614.

49. Price VH. Treatment of hair loss. N Engl J Med 1999; 341(13):964–973.

50. Charatan F. First pill for male impotence approved in US. Br. Med. J. 1998;316:1111.

51. Crain E, McIntosh KR, Gordon G, Pestronk A, Drachman DB. The effect of thalidomide on experimental autoimmune myasthenia gravis. J Autoimmun 1989;2(2):197–202.

52. Hylkema MN, van der Deen M, Pater JM, Kampinga J, Nieuwenhuis P, Groen H. Single expression of CD45RC and RT6 in correlation with T-helper 1 and T-helper 2 cytokine patterns in the rat. Cell Immunol 2000;199(2):89–96.