Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Literature-related discovery: Potential treatments and preventatives for SARS☆

Ronald N. Kostoff ⁎,1

Georgia Institute of Technology, School of Public Policy, USA

A R T I C L E   I N F O

Article history:
Received 2 November 2010
Received in revised form 26 February 2011
Accepted 26 March 2011
Available online 20 April 2011

Keywords:
Severe acute respiratory syndrome (SARS)
Literature-related-discovery
Literature-based discovery
Text mining
Radical discovery
Radical innovation
Infectious disease

A B S T R A C T

Literature-related discovery (LRD) is the linking of two or more previously disjoint concepts in order to produce novel, interesting, plausible, and intelligible connections (i.e., potential discovery). LRD has been used to identify potential treatments or preventative actions for challenging medical problems, among myriad other applications. Severe acute respiratory syndrome (SARS) was the first pandemic of the 21st century. SARS was eventually controlled through increased hygienic measures (e.g., face mask protection, frequent hand washing, living quarter disinfection), travel restrictions, and quarantine. According to recent reviews of SARS, none of the drugs that were used during the pandemic worked.

For the present paper, SARS was selected as the first application of LRD to an infectious disease. The main goal of this research was to identify non-drug non-surgical treatments that would 1) prevent the occurrence, or 2) reduce the progression rate, or 3) stop/reverse the progression of SARS. The MeSH taxonomy of Medline was used to restrict potential discoveries to selected semantic classes, and to identify potential discoveries efficiently. To enhance the volume of potential discovery, databases were used in addition to Medline. These included the Science Citation Index (SCI) and, in contrast to previous work, a full text database. Because of the richness of the full text, ‘surgical’ queries were developed that targeted the exact types of potential discovery of interest while eliminating clutter more efficiently.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

1.1. Overview and background

Literature-related discovery (LRD) is the linking of two or more previously disjoint concepts in order to produce novel, interesting, plausible, and intelligible connections (i.e., potential discovery). The open discovery systems (ODS) component of LRD starts with an unsolved problem, and generates solutions to that problem through potential discovery. ODS LRD has been used to identify potential treatments or preventative actions for challenging medical problems, among myriad other applications. The closed discovery systems (CDS) component of LRD starts with an unsolved problem and a potential solution, and generates potential mechanisms that link the solution to the problem. ODS is the only approach that will be used for the present study.

Two points should be emphasized before proceeding further. First, linking of disjoint literatures is a necessary but not sufficient condition for discovery. There needs to be value-added in the novel concept(s) that results. Second, while the term ‘potential discovery’ is used in this paper and throughout the LRD literature, ‘hypothesis’ is more accurate. What results from these LRD ‘discovery’ studies are hypotheses that have to be tested in the laboratory/field before they can be properly termed ‘discoveries’.

☆ Disclaimer: The views in this paper are solely those of the author, and do not necessarily represent the views of the MITRE Corporation.

⁎ Tel.: +1 571 527 0693.
E-mail address: ronald.kostoff@pubpolicy.gatech.edu.

1 The MITRE Corporation (Retired).

Technological Forecasting & Social Change 78 (2011) 1164–1173
© 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.techfore.2011.03.022
A 2009 review paper by the author showed that, while a number of LRD published papers claimed to have generated potential discovery, essentially none of these claims could be validated [1]. The only published LRD potential discovery claims that could be validated as credible hypotheses were in a journal Special Issue devoted to LRD (e.g., [2,3]). The four medical papers in this Special Issue describe the application of ODS LRD to four chronic diseases: Raynaud’s Phenomenon (RP), cataracts, Parkinson’s Disease (PD), and Multiple Sclerosis (MS).

The present paper presents a comprehensive approach to systematic acceleration of potential discovery and innovation, and demonstrates the generation of large amounts of potential discovery for prevention/treatment of an infectious disease: severe acute respiratory syndrome (SARS). The general issues of potential discovery and innovation in the LRD context are discussed in the first paper of the Special Issue [4], and the general methodology for this discovery approach was shown in the second paper of the Special Issue [5]. The SARS biomedical background has been published in a detailed review article, and the interested reader is referred to that article [6].

The present paper provides an overview of the etiology and challenges of SARS, then presents a retrieval and analysis of the core SARS literature and literatures related directly to the core SARS literature (e.g., immune system component literatures). These related literatures might contain the seeds of potential discovery (treatments and preventive measures) for SARS, and some examples of potential discovery are presented. In contrast to previous work, this paper includes full text analysis for the related literatures. This provided a substantial increase in the volume of potential discovery retrieved. Also, examples of interesting but non-discovery (i.e., potential innovation) concepts from the core SARS literature are presented, since they have practical value in their own right.

The four previous medical papers in the LRD Special Issue also included potential discovery from indirectly-related literatures. The indirectly-related literatures for the present infectious disease proof-of-principle demonstration were not examined. The amount of potential discovery retrieved from the directly related literature alone (including the full text directly related literature) is voluminous. A major challenge is to select those combinations of potential discoveries that provide the maximum synergy. Until these potential discoveries have been exploited, there is little practical need of going to indirectly related literatures to search for yet more potential discovery. In the practical situation, even the potential innovation as defined above has not been exploited, and accelerating these potential innovations should be the first order of business [55–59].

1.2. Purpose of study

SARS is a contagious disease that resulted in the hospitalization of about 8000 people world-wide in 2002–2003, and resulted in the deaths of about 800 people. According to the author’s interpretation of reviews of the pandemic, none of the drugs worked (e.g., “Despite an extensive literature reporting on SARS treatments, it was not possible to determine whether treatments benefited patients during the SARS outbreak. Some may have been harmful.”) [60]. Those who recovered did so by natural means; their immune systems were sufficiently strong to contain the viral attack. Many were aided by public health interventions (e.g., face mask protection, frequent hand washing, living quarter disinfection, travel restrictions, and quarantine) as well.

The subject of SARS was selected for study because of its pandemic nature, and its apparent intractability to all drug treatments. The main goal of this study was to identify non-drug non-surgical treatments that would 1) prevent or delay the onset, or 2) reduce the progression rate, or 3) stop/reverse the progression, of SARS. For much of the study, Medline was used as the data source, and the MeSH taxonomy of Medline was used to restrict potential discoveries to selected semantic classes.

The second goal was to generate large amounts of potential discovery in more than an order of magnitude less time than required for the RP study. To enhance the volume of potential discovery, a full text database was used in contrast to previous work. Because of the richness of the full text, ‘surgical’ queries were developed that targeted the exact types of potential discovery of interest while eliminating clutter more efficiently. The ‘surgical’ nature of these queries compensated for the additional ‘noise’ characteristic of the more voluminous full text. However, since the full text database (Science Direct-SD) did not have an associated MeSH taxonomy, the MeSH taxonomy headings were essentially used as text phrases to restrict SARS treatment and prevention potential discoveries to selected substances. The Science Citation Index (SCI) was also used to search for discovery, and again the MeSH taxonomy headings were used as text phrases to restrict potential discoveries to selected substances. Approximately four times as many records were retrieved from Medline when MeSH terms were included in the query compared to using only terms in the title or Abstract, due to the greater choice of potential discovery substances. This means the SCI or SD queries as presently constituted will retrieve only about 25% of the records that are possible using MeSH to define the substance pool.

Before the specific approach and results are described, the medical issues for SARS that served as targets for the discovery search query will be summarized.

1.3. SARS medical issues

The first pandemic of the 21st century was the outbreak of SARS caused by the SARS coronavirus (SARS-CoV). As far as is known, this outbreak was not due to the deliberate release of the SARS-CoV, but rather was a naturally occurring event. The appearance of SARS seems to have involved: 1) a zoonotic origin for SARS-CoV (e.g., horseshoe bats and/or Chiroptera as one wildlife reservoir [71]); 2) transmission to intermediate hosts (e.g., civet cats, raccoon dogs [8]); 3) human contact with these intermediate hosts in Southern China [Guangdong Province, Fall 2002] and subsequent cross-species transmission of the coronavirus to humans [8]; 4) transmission of the virus through both non-hospital personal contact and hospital staff contact [9]:
and, 5) global transmission of the virus via travelers from affected regions in Asia to other countries. SARS was eventually controlled through increased hygienic measures (e.g., face mask protection, frequent hand washing, living quarter disinfection), travel restrictions, and quarantine.

A number of recent reviews have focused on different components of the above SARS etiology, with the central focus of 1) identifying common characteristics of those who succumbed to the disease in order to 2) develop treatment targets for future outbreaks. A careful reading of these reviews shows that the humoral and cellular components of the adaptive immune system of those who succumbed were deficient on presentation and deteriorated thereafter. There is controversy about whether adequate antiviral interferons were generated during the innate response, but there is common agreement that the switch from innate to adaptive immunity was defective [10–23].

Specifically, those who succumbed from SARS-CoV infection tended to have the following characteristics, based on diagnostic measurements made at the time of presentation with symptoms:

- Older age
- Male gender
- Presence of comorbidities
- Elevated LDH or C-reactive protein/high initial lactate dehydrogenase level
- Higher initial viral load of SARS coronavirus
- Elevated neutrophil count/neutrophilia
- High levels of chemokines CXCl10 and IL-18
- High levels of IL-6, IL-8, and MCP-1
- Significant increase in the TH2 cytokines IL-4, IL-5, IL-10
- Increased levels of IP-10, MIG and IL-8
- Lower levels/deficient anti-SARS spike antibody production
- Lymphopenia/low counts of CD4 and CD8 at presentation
- Reduced ACE2 expression
- Reduced levels of IL-12p70 and TNF-alpha (relative to positive outcome)
- Positive RT-PCR on nasopharyngeal aspirate samples
- Elevated pulse rate
- Raised serum albumin
- Raised serum creatinine phosphokinase (CPK) levels
- Increased serum creatine kinase, and urea
- Deviated ISG and immunoglobulin gene expression levels

These characteristics served as the targets to be improved by potential discoveries.

2. Approach

Fig. 2 in the second paper of the Special Issue (Methodology) [5] is a flow chart that outlines the steps used in the present study. These steps include: database selection; query development to identify the core SARS literature; retrieval of the core SARS literature; analysis of the retrieved core SARS literature to identify the main generic themes (biomedical roadblocks); query development for, and retrieval of, literatures that represent the main generic themes; subtraction of the SARS core literature from the literatures that represent the main generic themes; intersection of the net retrieved literatures for the main generic themes with the literature of desired solution types (e.g., non-drug substances); searches for potential discovery candidates in the net retrieved literature for the given classes of substances; and validation of potential discovery candidates as potential discovery.

2.1. Database selection

Three databases were selected for source material. Medline is very comprehensive in the coverage of medical issues. The SCI overlaps Medline strongly in coverage of the major medical journals and biology journals, but in addition covers technical disciplines well beyond the medical sphere. The SCI also allows citing papers, references, and papers that share common references to be retrieved. For both these databases, titles, Abstracts, and keywords were searched. In addition, it was desired to ascertain the increase in potential discovery possible from searching full text. The SD database was used for this purpose.

2.2. Query development to identify core SARS literature

The core SARS literature is defined as those documents that the research/medical community would associate unambiguously with SARS. An iterative relevance feedback approach was used, and produced the following query: “(Severe-Acute-Respiratory-Syndrome OR SARS-Virus OR (SARS AND (coronavirus OR infect* OR virus* OR viral OR epidemic* OR epidemiology OR antibodies OR antibody OR vaccine* OR influenza OR pandemic* OR outbreak* OR syndrome)) OR SARS-patient* OR SARS-transmission OR SARS-CoV)”. The query was inserted in the three databases to retrieve the core SARS literature.
2.3. Identification of generic biomedical themes from core SARS literature

Multiple grouping approaches were used to identify the main generic themes of the core SARS literature: document clustering, auto-correlation mapping of phrases, and factor matrix analysis of phrases, using the Vantage Point software package [24]. These grouping techniques were applied to retrievals from the Medline and SCI databases, and no significant differences were observed.

2.4. Query development for, and retrieval of, literatures that represent the main generic biomedical themes

This step represents the main methodological advance over the author's previous LRD studies. Whereas previous LRD approaches used Boolean combinations of important biomedical terms (e.g., protein AND aggregation) in the query for retrieving generic biomedical literatures, the present approach used functional proximity queries (e.g., Inhibit within X words of Viral within X words of Entry). This type of query allowed 'surgical' targeting of relevant records, and for full text in particular was necessary to filter out the large numbers of irrelevant records that would occur with a Boolean query.

The query has two components: \((T \text{ NOT } S) \text{ AND } C\). \(T\) represents the technical query terms of interest, and \(S\) represents the technical query terms defining the SARS core literature (shown in Section 2.2). \(T \text{ NOT } S\) will retrieve those records of interest in the total database, excluding records found in the SARS core literature. \(C\) mainly represents the substances/behaviors that are in the potential discovery pool. As an example, if non-drugs were the only types of discovery of interest, \(C\) would be the pool of all non-drug records. Thus, all records in Medline, for example, that addressed medicinal plants would be included in \(C\). \(C\) also includes the journals whose articles tend to focus on such substances/behaviors, and thus would expand the substance pool for potential discovery beyond a simple listing of substances.

The specific steps employed in determining \(T\) are as follows. As stated previously, multiple grouping approaches were used to identify the main generic themes of the core SARS literature. The main problem was deciding which hierarchical level of grouping to use for the query. Initially, groupings at the lowest level of detail (e.g., CD4, CD8, IL1, IL2, IFN-gamma, etc.) were examined. However, a detailed examination of the SARS literature showed inconsistencies in the desired directions of some of these items, based on clinical observations and data. This led to examination of functional groupings at a higher level of aggregation (e.g., inhibit viral replication, enhance humoral immunity, improve cell-mediated immunity, increase Th1, etc.), which could accommodate different directions at the lowest level of detail while achieving the targets represented by the higher level of aggregation.

To retrieve the directly related literature from which potential discovery would be extracted, this higher level functional query was applied to the search engines of three databases: SCI, Medline, and SD. The first two of these databases provide Abstracts as the major text source, and SD provides full text. Since the SD analysis took place a few months after the SCI and Medline analyses, the initial query was modified slightly to exploit the SD search engine features. Appendix 1 of [54] contains the full SCI and Medline queries, and Appendix 2 of [54] contains the full SD query. The unique features of each are explained in these appendices.

There were two types of fundamental terms in the full query (Type I and Type II), and examples from the more abbreviated SD query will be presented here. Type 1 focused on inhibiting viral entry to healthy cells and/or inhibiting their replication, whereas Type 2 focused on improving immune system component performance. These two generic categories were identified from analysis of the retrieved SARS core literature. Factor analysis and document clustering of this core literature identified these key thrust areas, and further phrase frequency analysis of the retrieved SARS core literature identified the biomedical term ('viral entry')--functional term ('inhibit') combinations eventually selected for the query.

In the proximity form of the query used below, 'A PRE/5 B' is a precedence relation, and means that term A precedes term B, with a spacing ranging from zero (adjacency) to five words. 'A W/15 B' is a proximity relation, and means that term A is within 15 words of term B. This proximity form of the query (as contrasted to the Boolean form used in prior LRD studies) provided highly 'relevant' retrievals, where 'relevant' is defined as any article that contains a potential discovery or innovation candidate. Ref [54] contains a more detailed discussion of 'relevance' in the present context.

**Type 1**

\[
(((\text{inhibit}^* \text{ PRE/5 “virus entry”}) \text{ OR } (\text{inhibit}^* \text{ PRE/5 “viral entry”}) \text{ OR } (\text{inhibit}^* \text{ W/15 replicat}^* \text{ W/15 virus}) \text{ OR } (\text{inhibit}^* \text{ W/15 replicat}^* \text{ W/15 viral})) \text{ AND } (\text{potent PRE/15 “antiviral activity”}))
\]

**Type 2**

\[
(((\text{enhanc}^* \text{ PRE/5 “humoral immun”}) \text{ AND } (\text{enhanc}^* \text{ PRE/5 “humoral response”}) \text{ AND } ((\text{enhanc}^* \text{ PRE/5 “antibod” response”}) \text{ OR } (\text{enhanc}^* \text{ PRE/5 “antibod” production”}) \text{ OR } (\text{enhanc}^* \text{ PRE/5 “virus neutraliz”}) \text{ OR } (\text{enhanc}^* \text{ PRE/5 “viral neutraliz”})) \text{ OR } ((\text{enhanc}^* \text{ PRE/5 “cellular immun”}) \text{ AND } (\text{enhanc}^* \text{ PRE/5 “cell mediated immun”}) \text{ AND } ((\text{enhanc}^* \text{ PRE/5 CD8}) \text{ OR } (\text{enhanc}^* \text{ PRE/5 “t cell response”}) \text{ OR } (\text{enhanc}^* \text{ PRE/5 “t cell immune response”}) \text{ OR } ((\text{enhanc}^* \text{ PRE/5 “innate immun”}) \text{ AND } (\text{enhanc}^* \text{ PRE/5 “innate antiviral”}) \text{ AND } ((\text{enhanc}^* \text{ PRE/5 “antiviral activity”}) \text{ OR } (\text{enhanc}^* \text{ PRE/5 “antiviral response”}))))\]

When the Type 1 query shown above was applied to full text, reasonable numbers of relevant articles were retrieved. When the Type 1 query was applied to Abstracts, the articles retrieved were highly relevant, but the numbers retrieved (as will be shown) were miniscule. To obtain more articles when searching the Abstracts, a modified Type 1 query was generated. This modified form changed the ‘AND’ (in the query above) to an ‘OR’, a much less restrictive condition. This resulted in an order of magnitude more retrieved and relevant articles (when searching Abstracts), although still small compared to the retrievals obtained from searching full text.
Type 2 queries focused on enhancing the performance of the different components of the immune system. There were a number of variants of the Type 2 query shown above that were examined. Appendix 1 of Ref. [54] shows the form of the query used to generate the SCI results. The Type 2 entry under the T component is for ‘Improv’*, but the full query added blocks that replaced ‘Improv’* with Enhanc* or Induc* or Stimulat* or Increas* or Activat* or Regulat*. Obviously, other such terms could be identified and used as well, but these terms were deemed adequate for the present proof-of-principle demonstration, which shows the feasibility of the methodology.

The query examples provided above were for the SD database, which has intrinsic proximity query capabilities in addition to Boolean query capabilities. The SCI search engine does not have adjacency/proximity search capability presently. The only search capabilities are Boolean/co-occurrence in a selected field or among all fields (e.g., A SAME B, or A AND B). To overcome this deficiency, the author developed an algorithm that would provide such capability [25]. In the algorithm, SCI stopwords are used to set spacing between terms. Thus, a query term of the form [improv*-of-humoral-immun*] will retrieve those records containing variants of improv* that precede variants of “humoral-immun”* with one word intercalated. The intercalated word could be any word, not just the stopword used in the query. As can be seen from the terms in Appendix 1 of Ref. [54], the actual query used did not go beyond precedence spacings of two words, but a larger production-oriented study could use greater spacings between the terms of interest. This would retrieve far more records, and lead to more candidate potential discoveries.

In Appendix 1 of Ref. [54], the S block of terms represents the core SARS records, and its inclusion as a negation expression insures the records retrieved are disjoint from the core SARS literature.

2.5. Intersection of the net retrieved literatures for the main generic themes with the literature of desired solution types

In Appendix 1 of Ref. [54], the blocks listed under C are the types of substances/behaviors considered for discovery. They consist of records from non-drug journals as shown, and records that contain non-drug substances. The value of including the journals is that their records could include substances/behaviors not identified in the substances/behaviors blocks (i.e., pre-specified lists of substances/behaviors). Obviously, many more substances/behaviors could be added to the list in a more comprehensive production-oriented study.

The records retrieved by the T NOT S query were intersected with the records retrieved by the C query, to yield records that identified non-drug substances/approaches that would produce immune system changes in the desired directions. These records were inspected visually, and sample results are presented in the next section.

2.6. Validation of potential discovery candidates as potential discovery

The purpose of this study is to identify potential discovery and innovation. For this purpose, ‘relevant’ is interpreted as any article that contains a potential discovery or innovation candidate. What is potential discovery? It is the linking of two or more literature concepts (that have not been linked previously in the literature; i.e., disjoint) in order to produce novel, interesting, plausible, and intelligible connections. A potential discovery candidate is an interesting linkage that has to be vetted against prior knowledge to validate disjointness. In practice, a major roadblock is defining ‘prior knowledge’, and in particular the databases that will be used to represent ‘prior knowledge’ for the vetting process and how these databases will be interpreted.

To make the problem tractable, only the main source databases are selected for the ‘prior knowledge’ determination. In the present case, the same three major sources that were used in previous LRD medical studies have been selected for the ‘prior knowledge’ determination: SCI; Medline; and patent database as represented by Derwent Innovation Index (DII). For each candidate potential discovery, a query was developed that intersected the candidate potential discovery (e.g., curcumin) with the SARS core literature query (severe acute respiratory syndrome OR ....), and the query was inserted into each of the three validation databases. If no prior records were retrieved, the concept was viewed as a potential discovery. If prior records were identified, the concept could still be viewed as a potential innovation if the judgment was made that the concept could be developed at a more accelerated pace.

3. Results

3.1. Numbers of retrievals from different source databases

3.1.1. SCI

Application of the full query to the SCI/SSCI database (1989–2008, Articles and Reviews) yielded 662 records. The records were sampled for ‘relevancy’ to discovery or innovation, using the definitions of relevancy as discussed previously. Approximately 85% were judged to be ‘relevant’ (potential discovery candidates). In addition, the 7000 most recent papers in the SCI that cited the 662 records were retrieved. These citing papers covered approximately the seven year period 2002–2008, and about 50% were judged to be relevant. In the previous LRD studies on medical topics, before the ability to retrieve all the papers that cite an initial retrieval became available in the SCI, only spot checks could be done of papers that cited potential discovery candidates. Not only are papers that cite potential discovery candidates good potential discovery candidates themselves, but the author’s preliminary (unpublished) experiments which show many different types of citation linkages (e.g., papers that share references) to potential discovery candidates will identify good potential discovery candidates.
3.1.2. Medline

Application of the query to the Medline database yielded 1149 records. While the version of the Medline database used (through the Web of Knowledge search engine) goes back to 1950, it effectively started in about 1975, when Abstracts were introduced. The records retrieved were sampled for relevancy to discovery or innovation, using the definitions of relevancy as discussed previously. Approximately 80% were judged to be 'relevant' (potential discovery or innovation candidates). There is much overlap between Medline and the SCI.

3.1.3. Science Direct

Application of the appropriate query from Appendix 1 of Ref. [54] to the Science Direct database (1999–2008) yielded different types of results, depending on which field was searched, and some of these findings are reported in Table 1. The retrieval results among SCI, Medline, and Science Direct are not comparable due to the different journal coverage of each database. Specific examples of potential discovery from each of these databases will be shown later in the present Results section.

In Table 1, the Science Direct results are for the three components of the Type 1 query. The first row (INHIBIT*ACETABULARIA) reflects the intersection of the terms for the 'inhibit' group (e.g., "(inhibit*-virus-entry OR inhibit*-viral-entry OR inhibit*-of-virus-entry OR inhibit*-of-viral-entry OR inhibit*-of-of-virus-entry OR inhibit*-of-of-viral-entry OR (inhibit*-replication SAME (virus OR viral))) OR (inhibit*-of-replication SAME (virus OR viral))) and the group of substances and behaviors that starts with 'Acetabularia' (e.g., Acetabularia OR Achlya OR Acupressure OR Acupuncture OR Algae OR Alkaloid* OR Allium OR Angiosperms OR Anthocerotophyta OR Anthocyanin...). Listed in Appendix 2 of Ref. [54]. The second and third rows substitute the Euglenida and Plankton groups for the Acetabularia group.

For the first row, the second column (UNMOD QUERY — ABSREC) contains the number of records retrieved when the full text query was applied to the Abstract field. The full text query was designed to provide 'surgical' targeting of key phrase relations in the full text, contains the intersections of a number of proximal relations, and is a rather restrictive condition. Applied to the full text, the query is very effective, but it is far too restrictive for the Abstract field, as the results show. Very few records are retrieved, and they are all relevant (Column 3). Essentially no records were retrieved in this column for any of the Type 2 queries, since the Type 2 queries are even more restrictive (more intersecting terms) than the Type 1 query.

Column 4 (MOD. QUERY – ABSREC) contains records retrieved using the modified Type 1 query applied to the Abstract field. All occurrences of 'AND' were replaced with 'OR'. On average, about an order of magnitude more records were retrieved with the modified query compared to the full text query, but the numbers are still small. The relevance of the retrieved records is still quite high. For the Type 2 query components (not shown), the Column 4 retrievals were about half as much for the 'enhance' group, about twice as much for the 'induce' group, about the same for the 'stimulate' group, and about the same for the 'increase' group. There were overlaps among the groups, and search engine limitations did not allow estimation of the degree of overlap.

Column 5 is the relevance percentage of the Column 4 retrievals. It is highest for the 'inhibit' group, is about 60% for the 'enhance' group, and is about 50% for the other three (Type 2) groups.

Column 6 (UNMOD. QUERY — FLTXREC) represents the retrievals for the full text query of Appendix 2 applied to the full text. They are almost an order of magnitude larger than those of Column 4, and the relevance percentage (Column 7) is almost the same as that for the Abstracts in Column 5. The percentage is highest for the 'inhibit' group, is about 80% for the 'enhance' group, and is about 50% for the other three (Type 2) groups.

Column 8 (FT/ABS-NORM) is the ratio of full text retrievals to Abstract retrievals (normalized to the total numbers of full text and Abstract records in the database) using the restricted form of the full text query for each, and Column 9 is the same ratio where the less restricted form of the query was used to search the Abstracts.

The bottom line is that many potential discovery candidates have been retrieved. Many more are possible: increasing the substance base (through more non-drug items or by inclusion of drugs) would probably increase the number of candidates by an

| Query                  | Unmod. query | Mod. query | Unmod. query | FLTX | FT/ABS | Mod |
|------------------------|--------------|------------|--------------|------|-------|-----|
|                        | ABS REC | % REL | ABS REC | % REL | FLTX REC | % REL | ABS REC | % REL | NORM |
| INHIBIT* ACETABULARIA  | 1      | 100   | 31     | 90    | 170   | 90   | 172.3   | 5.56 |
| INHIBIT* EUGLENIDA     | 5      | 100   | 60     | 90    | 243   | 90   | 49.26   | 4.1  |
| INHIBIT* PLANKTON      | 7      | 100   | 28     | 90    | 350   | 90   | 50.68   | 12.67|
| ENHANC* ACETABULARIA   | 0      | 14    | 60     | 90    | 90    | 80   | INF     | 6.52 |
| ENHANC* EUGLENIDA      | 0      | 29    | 70     | 129   | 80    | INF   | 4.51    |
| ENHANC* PLANKTON       | 0      | 18    | 40     | 170   | 80    | INF   | 9.57    |
| INDUC* ACETABULARIA    | 0      | 56    | 30     | 464   | 50    | INF   | 8.4     |
| INDUC* EUGLENIDA       | 0      | 101   | 40     | 749   | 60    | INF   | 7.52    |
| INDUC* PLANKTON        | 1      | 79    | 70     | 944   | 40    | INF   | 12.11   |
| STIMULAT* ACETABULARIA | 0      | 37    | 50     | 142   | 60    | INF   | 3.89    |
| STIMULAT* EUGLENIDA    | 0      | 49    | 60     | 235   | 50    | INF   | 4.86    |
| STIMULAT* PLANKTON     | 0      | 34    | 80     | 265   | 60    | INF   | 7.9     |
| INCREASES* ACETABULARIA| 0      | 35    | 40     | 192   | 40    | INF   | 5.56    |
| INCREASES* EUGLENIDA   | 0      | 58    | 40     | 303   | 50    | INF   | 5.29    |
| INCREASES* PLANKTON    | 0      | 44    | 40     | 291   | 60    | INF   | 6.7     |
order of magnitude; increasing the number of terms in the query would enhance the retrieval; relaxing the proximity conditions would increase the retrieval; relaxing the intersection requirements would increase the retrieval; and adding further types of citation linkages would increase the retrieval. A major factor in the high relevance fractions achieved is the form of the query terms; they were not used in any previous LRD studies, but will become a fixture in future studies.

3.2. Examples of retrievals from different literatures

This remainder of this section contains representative examples of potential discovery from literatures related directly to the core SARS literature. Before proceeding to analyses, a few illustrative examples from the core SARS literature restricted to semantic classes will be presented. While these are not discovery, they nevertheless reflect the types of impact that the non-drug approaches could potentially have for delaying or preventing the onset of SARS. In addition, as will be discussed later, some of these core concepts are prime candidates for innovation.

For example, “Aurinetricarboxylic acid (ATA) has been shown to inhibit the replication of viruses from several different families, including … the coronavirus causing severe acute respiratory syndrome. Vaccinia virus replication is significantly abrogated upon ATA treatment, which is associated with the inhibition of early viral gene transcription. This inhibitory effect may be attributed to two findings. First, ATA blocks the phosphorylation of extracellular signal-regulated kinase 1/2, an event shown to be essential for vaccinia virus replication. Second, ATA inhibits the phophatase activity of the viral enzyme H1L, which is required to initiate viral transcription. Thus, ATA inhibits vaccinia virus replication by targeting both cellular and viral factors essential for the early stage of replication” [26].

As another example, “we identified that three widely used Chinese medicinal herbs of the family Polygonaceae inhibited the interaction of SARS-CoV S protein and ACE2. The IC50 values for Radix et Rhizoma Rhei (the root tubers of Rheum officinale Baill.), Radix Polygoni multiflori (the root tubers of Polygonum multiflorum Thunb.), and Caulis Polygom multiflori (the vines of P. multiflorum Thunb.) ranged from 1 to 10 [Lg/ml]. Emodin, an anthraquinone compound derived from genus Rheum and Polygonon, significantly blocked the S protein and ACE2 interaction in a dose-dependent manner. It also inhibited the infectivity of S protein-pseudotyped retrovirus to Vero E6 cells. These findings suggested that emodin may be considered as a potential lead therapeutic agent in the treatment of SARS” [27].

3.2.1. Non-drug concepts in the core SARS literature

More detailed descriptions of the following potential innovations and discoveries can be found in Ref. [54]. Cimicifuga rhizoma, Meliae cortex, Coptidis rhizoma, Phellodendron cortex and Sophora subprostrata radix decreased the MHV production and the intracellular viral RNA and protein expression, and could be potential candidates for new anti-coronavirus drugs [28]; Betulinic acid and savinin were competitive inhibitors of SARS-CoV 3CL protease [29]; quercetin-3-beta-galactoside was identified as an inhibitor of the protease [(3CL(pro)) [30]; alpha,beta-unsaturated peptidomimetics, anilides, metal-conjugated compounds, boronic acids, quinolinocarboxylate derivatives, thiophenecarboxylates, phthalhydrazide-substituted ketoglutamine analogs, isatin and natural products have been identified as potent inhibitors of the SARS-CoV main protease [31]; tannic acid and 3-isothaeflavin-3-gallate were found to be inhibitive. These two compounds belong to a group of natural polyphenols found in tea, and only theaeflavin-3,3’-digallate (TF3) was found to be a 3CLPro inhibitor [32].

3.2.2. Non-drug potential discovery concepts in the SCI retrieved by query

GLF[fermentation product of Chinese medicinal mushroom] up-regulated the cell-mediated immune response related cytokines (IL-2, IFN-gamma, and TNF-alpha) expression in different lymphoid tissues [33]; Jacalin, an antigen-specific lectin from jackfruit seeds, has been shown to induce mitogenic responses and to block infection by HIV-1 in CD4(+ ) T lymphocytes [34]; sulforaphane significantly downregulated the serum levels of proinflammatory cytokines such as IL-1 beta, IL-6, TNF-alpha, and GM-CSF during metastasis [35]; immunomodulatory activity of methanolic extract of M. koenigii leaves was evaluated on humoral and cell mediated immune response to ovalbumin, and the extract holds promise as immunomodulatory agent, which acts by stimulating humoral immunity and phagocytic function. [36]; the aqueous extract of T. cordifolia was found to enhance phagocytosis in vitro. The aqueous and ethanolic extracts also induced an increase in antibody production in vivo. [37]; oral intake of the fucoidan might take the protective effects through direct inhibition of viral replication and stimulation of both innate and adaptive immune defense functions. [38]; Atractylodes macrocephala Koidz (AMK) markedly stimulated lymphocyte proliferation, antibody production, and cytokine secretion in mouse splenocytes, showing the ability to induce the preferential stimulation of Th1 type, rather than Th2 type T lymphocytes [39]; a potent anti-influenza virus activity was discovered in summer leaves of Japanese wasabi [(Wasabia japonica)], inhibiting influenza virus replication regardless of the hemagglutinin antigen type. [40]; purification of an antiviral peptide from seeds of Sorghum bicolor L strongly inhibited the replication of herpes simplex virus type I (HSV-1) [41].

3.2.3. Non-drug potential discovery concepts in the SCI retrieved by citing papers

Combined treatment with pidotimod and red ginseng acidic polysaccharide has an immunostimulatory effect in a synergistic manner on antibody response to challenge with lipopolysaccharide and sheep red blood cells without toxic changes [42]; Myrica rubra leaf ethanol extract showed anti-influenza virus activity irrespective of the hemagglutinin antigen type in the influenza virus type A (H1N1), its subtype (H3N2), and type B [43]; oral intake of L. paracasei NCC2461 by aged mice enhanced the specific adaptive immune response to in vivo antigenic challenge without altering other cellular and humoral immune responses [44];
methanol extract of Asarum sieboldii inhibited the H5N1 influenza viruses from the infected cells [45]; Caffeoyl Glycoside from the roots of Picrorhiza scrophulariiflora (Kutki) stimulated cell proliferation of splenocytes and peritoneal macrophages, enhanced the cytotoxicity of natural killer (NK) cells, increased CD4 and CD8 cell populations, and has immunomodulatory activity by regulating expression of Th1 and Th2 related cytokines [46].

3.2.4. Vaccine adjuvants — Medline

Korean mistletoe lectin C increased influenza-specific antibodies with dominant IgG1 subclass in serum, IgG in genital secretions and IgA in saliva, and significantly enhanced influenza-specific lymphocyte proliferation and cytotoxic activity in spleens and in mediastinal lymph nodes. When KML-C was used as a mucosal adjuvant, mice were completely protected from mortality after the challenge with a homologous (H1N1) mouse-adapted influenza virus [47]; seeds of Cochinchina momordica numerically increased the antibody levels, suggesting potential to improve the immune responses by use as an adjuvant [48]; Ag85B of mycobacteria, which cross-reacts among mycobacteria species, elicits helper T-cell type 1 (Th1) immune responses as a novel adjuvant [49]; probiotic Bacillus cereus var. toyoi-treated piglets showed a significantly lower frequency of CD8(high)/CD3+ T cells and CD8(low)/CD3+ T cells and a significant higher CD4+/CD8+ ratio [50].

3.2.5. Non-drug potential discovery concepts in Science Direct

Different components of kefir have an in vivo role as oral biotherapeutic substances capable of stimulating immune cells of the innate immune system, to down-regulate the Th2 immune phenotype or to promote cell-mediated immune responses against tumors and also against intracellular pathogenic infections [51]; Lentinar possesses antiviral activity due to an induction of interferon-γ production; enhances host resistance against infections with bacteria as well as fungi, parasites, and viruses, including the agent of AIDS, and reduces the toxicity of AZT [52]; a polyphenol rich extract (CYSTUS052) from the Mediterranean plant Cistus incanus exerts a potent anti-influenza virus activity in A549 or MDCK cell cultures infected with prototype avian and human influenza strains of different subtypes [53].

Because the purpose of the SARS study was to demonstrate an approach, and not necessarily to be comprehensive, a number of shortcuts were taken. Not all possible semantic categories for potential discoveries were identified, only the most obvious. Relatively few terms were selected for the queries; many more were available. Not all retrieved records were examined; only enough to demonstrate the quality of results. The potential expansion to indirectly related literatures using both text linking and citation linking described previously was not done. Thus, the results obtained should be viewed as the tip of a very large iceberg.

4. Discussion and conclusions

In previous medical applications of LRD, the discovery approach was to cluster the disease literature into groups of disease characteristics, generate combinatorials of intra-group characteristics (essentially synonyms), construct a discovery query from these combinatorials, and apply the query to non-drug substances/behaviors. In the present application, a somewhat different approach was taken. Since the characteristics of those who had succumbed to SARS tended to be sub-optimal performance of different immune system components, the query combined the immune system components at a higher aggregation level with functional terminology designed to improve this performance (e.g., enhance humoral immunity). Such a query structure allowed a large number of potential discovery candidates to be identified. As the previous paragraph shows, much more is possible in terms of potential discovery volume.

The picture from the handful of potential discoveries reported in this paper (and the hundreds of additional potential discoveries possible with a properly resourced study) is a synergy of lifestyle/dietary practices that could be interpreted as anti-SARS. Along with non-discovery items such as Aurintricarboxylic acid (ATA), Emodin (an anthraquinone compound derived from genus Rheum and Polygonum), Cimicifuga rhiza, Melia cortex, Coptidis rhiza, Phellodendron cortex, Sophora subprostrata radix, Betulinic acid, savinian-type diterpenoids and lignoids, quercetin-3-beta-galacto side, d.alpha,beta-unsaturated peptidomimetics, anilides, metal-conjugated compounds, boronic acids, quinolinecarboxylate derivatives, thiophenecarboxylates, phthalhydrazide-substituted ketoglutamine analogs, isatin, tannic acid, and 3-isothioflavin-3-gallate (TF2B) are potential discovery items such as Ganoderma lucidum, Jacalin, Sulfuraphane, methanolic extract of M. koenigi leaves, Tinospora cordifolia, Fucoidan, Atractylodes macrocephala Koidz, Wasabia japonica, seeds of Sorghum bicolor L, pidotimod and red ginseng acid polysaccharide (RGAP), Myrcura rubra leaf ethanol extract, L. paracasei NCC2461, methanol extract of Asarum sieboldii, Caffeoyl Glycoside, and adjuvants Korean mistletoe lectin C, Cochinchina momordica seed, Ag85B, probiotic Bacillus cereus var. toyoi, Kefir, Lentina, polyphenol rich extract (CYSTUS052) from the Mediterranean plant Cistus incanus. As stated above, more laboratory tests and field trials would have to be done on all these items to insure that they are anti-SARS and safe, but these preliminary literature-based results offer some promise of what is possible.

There is a major disconnect between the absence of therapies presently or potentially available on all the major medical Web sites (and in SARS mainstream journal review papers), and the potential therapies suggested by what has already been demonstrated in the core SARS literature, much less what this study has generated from the related literatures. Few medical Web sites even mention any of the approaches shown in the SARS core results section. The core literature and related literature potential discoveries and innovations have the potential to evolve into mainline medical treatments.
Acknowledgment

The research described in this paper was supported by MITRE Corporation Internal Research funds.

References

[1] R.N. Kostoff, J.A. Block, J.A. Solka, M.B. Briggs, R.L. Rushenberg, J.A. Stump, D. Johnson, J.R. Wyatt, Literature-related discovery, ARIST 43 (2009) 241–285.
[2] R.N. Kostoff, M.B. Briggs, Literature-related discovery (LRD): potential treatments for Parkinson's Disease, Technol. Forecast. Soc. Change 75 (2) (2008) 226–238.
[3] R.N. Kostoff, M.B. Briggs, T. Lyons, Literature-related discovery (LRD): Potential treatments for Multiple Sclerosis, Technol. Forecast. Soc. Change 75 (2) (2008) 229–255.
[4] R.N. Kostoff, M.B. Briggs, J.L. Solka, R.L. Rushenberg, Literature-related discovery (LRD): introduction and background, Technol. Forecast. Soc. Change 75 (2) (2008) 165–185.
[5] R.N. Kostoff, M.B. Briggs, J.L. Solka, R.L. Rushenberg (LRD): methodology, Technol. Forecast. Soc. Change 75 (2) (2008) 186–202.
[6] R.N. Kostoff, The highly cited SARS research literature, Crit. Rev. Microbiol. 36 (4) (2010) 299–317.
[7] D. Janies, F. Habib, B. Alexandrov, A. Hill, D. Pol, Evolution of genomics, host shifts and the geographic spread of SARS-CoV and related coronaviruses, Cladistics 24 (2) (2008) 111–130.
[8] Y.J. Feng, G.F. Gao, Towards our understanding of SARS-CoV, an emerging and devastating but quickly conquered virus, Comp. Immunol. Microbiol. Infect. Dis. 30 (5–6) (2007) 309–327.
[9] Z.B. Zhang, The outbreak pattern of SARS cases in China as revealed by a mathematical model, Ecological Model. 204 (3–4) (2007) 420–426.
[10] Y. Guo, C. Korteweg, Pathogenesis of Severe Acute Respiratory Syndrome, Am. J. Pathol. 170 (4) (2007) 1136–1147.
[11] V. Thiel, F. Weber, Interferon and cytokine responses to SARS-coronavirus infection, Cytokine Growth Factor Rev. 19 (2008) 121–132.
[12] Y. Guo, C. Korteweg, M.A. McNutt, J. Gu, Pathogenetic mechanisms of Severe Acute Respiratory Syndrome, Virus Res. 133 (2008) 4–12.
[13] Y. Zhang, J. Li, Y. Zhan, L. Wu, X. Yu, W. Zhang, L. Ye, S. Xu, R. Sun, Y. Wang, J. Lou, Analysis of serum cytokines in patients with Severe Acute Respiratory Syndrome, Infect. Immun. 72 (8) (2004) 4410–4415.
[14] M. Baba, B.Y. Ma, M. Nonaka, Y. Matsuishi, M. Hirano, N. Nakamura, N. Kawasaki, T. Kawasaki, Glycosylation-dependent interaction of Jacalin with the N-glycans present on HIV-1 gp120, Virology 299 (2) (2003) 369–379.
[15] L.L. Chen, J. Li, C. Luo, H. Liu, W.J. Xu, G. Chen, O.W. Liew, W.L. Zhu, C.M. Puah, X. Shen, H.L. Jiang, Binding interaction of quereetin-3-beta-galactoside and its synthetic derivatives with SARS-CoV 3CL(pro), Phytomedicine 15 (3) (2008) 202–208.
[16] M.S. Ranjith, A.J.A. Ranjitsingh, S.G. Shankar, G.S. Vijayalaksmi, K. Deepa, H.S. Siddhu, Enhanced phagocytosis and antibody production by Tinospora cordifolia (TF3), Evid. Based Complement. Altern. Med. 2 (2) (2005) 209–215.
[17] S.D. Chen, M.C. Hishe, M.T. Chiu, Y.S. Lai, Y.H. Cheng, Effects of fermentation products of Ganoderma lucidum on growth performance and immunocompetence in weaning pigs, Arch. Anim. Nutr. 62 (1) (2008) 22–32.
[18] M. Baba, B.Y. Ma, M. Nonaka, Y. Matsuishi, M. Hirano, N. Nakamura, N. Kawasaki, N. Kawasaki, T. Kawasaki, Glycosylation-dependent interaction of Jacalin with CD45 induces Th lymphocyte activation and Th1/Th2 cytokine secretion, J. Leukoc. Biol. 81 (4) (2007) 186–193.
[19] S.D. Chen, M.C. Hishe, M.T. Chiu, Y.S. Lai, Y.H. Cheng, Effects of fermentation products of Ganoderma lucidum on growth performance and immunocompetence in weaning pigs, Arch. Anim. Nutr. 62 (1) (2008) 22–32.
[20] M. Baba, B.Y. Ma, M. Nonaka, Y. Matsuishi, M. Hirano, N. Nakamura, N. Kawasaki, N. Kawasaki, T. Kawasaki, Glycosylation-dependent interaction of Jacalin with CD45 induces Th lymphocyte activation and Th1/Th2 cytokine secretion, J. Leukoc. Biol. 81 (4) (2007) 186–193.
42. Du, XF, Jiang, CZ, Wu, CF, Won, EK, Cheung, SY. Synergistic immunostimulatory effect of pidotimod and red ginseng acidic polysaccharide on humoral immunity of immunosuppressed mice. Pharmazie. 63:12(2008). 849. 904–908.

43. K. Mochida, Anti-influenza virus activity of myrica rubra leaf ethanol extract evaluated using madino-darby canine kidney (MDCK) cells, Biosci. Biotechnol. Biochem. 72 (11) (2008) 3018–3020.

44. K. Vidal, J. Benyacoub, M. Moser, J. Sanchez-Garcia, P. Serrant, I. Segura-Roggero, G. Reuteler, S. Blum, Effect of lactobacillus paracasei NCC2461 on antigen-specific t-cell mediated immune responses in aged mice, Rejuvenation Res. 11 (5) (2008) 957–964.

45. J.H. Lee, S.M. Kwon, S.H. Seo, Y.S. Park, Y.B. Kim, S.K. Kim, H.D. Paik, Screening of a natural feed additive having anti-viral activity against Influenza A/H5N1, Korean J. Food Sci. Anim. Resour. 28 (4) (2008) 512–516.

46. S. Zeng, D.C. Wang, Y.G. Cao, N. An, F.Q. Zeng, C.T. Han, Y. Song, X.M. Deng, Immunopotentiation of Caffeoyl Glycoside from Picrorhiza scrophulariiflora on activation and cytokines secretion of immunocyte in vitro, Int. Immunopharmacol. 8 (12) (2008) 1707–1712.

47. Z.I. Rajput, C.W. Xiao, S.H. Hu, A.G. Anjoo, N.M. Soomro, Improvement of the efficacy of influenza vaccination (H5N1) in chicken by using extract of Cochinchina momordica seed (ECMS). J. Zhejiang Univ. Sci. B 8 (5) (2007) 331–337.

48. Yuya Shimizu, Katsuo Karamatsu, Akihiro Matsubara, Yasuhiro Yasutomi, Innovation of vaccine adjuvants, Nippon Rinsho 10 (2008) 1915–1921.

49. P. Schierack, L.H. Wieler, D. Taras, V. Herwig, B. Tachu, A. Hlinak, M.F.G. Schmidt, L. Scharek, Bacillus cereus var. toyoi enhanced systemic immune response in piglets, Vet. Immunol. Immunopathol. 118 (1-2) (2007) 1–11.

50. S.K. Song, Z. Moldoveanu, H.H. Nguyen, E.H. Kim, K.Y. Choi, J.B. Kim, J. Mestecky, Intranasal immunization with influenza virus and Korean mistletoe lectin C (KML-C) induces heterosubtypic immunity in mice, Vaccine 25 (34) (2007) 6359–6366.

51. G. Vinderola, G. Perdigon, J. Duarte, D. Thangavel, E. Farnsworth, C. Matar, Effects of kefir fractions on innate immunity, Immunobiology 211 (3) (2006) 149–156.

52. W. Andlauer, P. Furst, Nutraceuticals: a piece of history, present status and outlook, Food Res. Int. 35 (2-3) (2002) 171–176.

53. T. Van der Valk, M.M.H. Chappin, G.W. Gijsbers, Evaluating innovation networks in emerging technologies, Technol. Forecast. Soc. Change 78 (2) (2011) 346–357.

54. D. Thorleuchter, D. Van den Poel, A. Prinzie, Mining ideas from textual information, Expert Syst. Appl. 37 (10) (2010) 7182–7188.

55. D. Thorleuchter, D. Van den Poel, A. Prinzie, A compared R&D-based and patent-based cross impact analysis for identifying relationships between technologies, Technol. Forecast. Soc. Change. 77 (7) (2010) 1037–1050.

56. P.C. Lee, H.N. Su, Quantitative mapping of scientific research—the case of electrical conducting polymer nanocomposite, Technol. Forecast. Soc. Change 78 (1) (2011) 132–151.

57. L.J. Stockman, R. Bellamy, P. Garner, SARS: systematic review of treatment effects, PLoS Med. 3 (9) (2006) e43.