World Patent Information 41 (2015) 11–14

Contents lists available at ScienceDirect

World Patent Information

journal homepage: www.elsevier.com/locate/worpatin

CBK searching (chemistry-biology-keyword): Performing cross-discipline collaborative searches

Kimberly Miller [a, *], Seth Mendelson [b]

[a] Novartis Institutes for Biomedical Research (NIBR), Cambridge, MA, USA
[b] Eisai Inc., Andover, MA, USA

ARTICLE INFO

Article history:
Available online 16 February 2015

Keywords:
Chemical structure
Sequence
Markush
Amino acid
Cross discipline searching
ADC

ABSTRACT

A debate over whether a search is a chemistry or a biology search is becoming increasingly more frequent within search teams. The line between non-text chemistry and biology searching has become blurred in recent years. The appearance of "non-natural" amino acids and modified peptides, for example, has given rise to the questions: Is it a structure? Is it a sequence? Could it be both? It may not be a question of one or the other, but perhaps working with a colleague to bring the best possible answer to the client. Several examples clarify the line and present guidance when beginning a search. Consider the size of the molecule, how many amino acids are present and what the client needs in order to best cover all bases and create a more comprehensive search report.

1. Introduction

There are similarities in chemical structure searching and biology sequence searching. A chemist may begin a structure search by drawing the specific compound or Markush and running the query through specially designed databases, e.g.: Marpat, the Merged Markush System (MMS), Reaxys etc. Search criteria may be discussed with the client surrounding R-groups definitions. A biologist may begin a search by carrying out a sequence search, using sequence databases and alignment tools (BLAST, Smith-Waterman, etc.). Search criteria is often discussed with the client concerning percent homology. However, this is where the processes of structure searching and sequence searching begin to diverge.

In most cases, it is clear which type of analyst should carry out a search. Typically if the request involves amino acids, genes, or antibodies, it is forwarded to a biologist. Similarly, if a request concerns heteroaryls or polyethylene it lands with a chemist. This is not always the case now. There may be a need to collaborate with a colleague or cross-discipline train to create a more complete search report. It may not be productive or efficient to cross-discipline train, therefore collaboration will be explored.

There are many advantages to collaborating. Chemists and biologists consider their work to be very different from each other. Most searchers conduct either chemistry searches or biology searches, but very few routinely perform both. An example of this is the way a chemist would refine keywords for a biology search compared to how a biologist may refine keywords for a chemistry search. A chemist is familiar with benzene generically described as aryl while a biologist may not be. Conversely, a biologist may know that Kearns-Sayre syndrome is a neuromuscular disease while a chemist may not. A simple keyword search can become more in-depth by the addition of terms provided by a colleague.

2. Discussion

2.1. Convergence of chemistry and biology

An antibody-drug conjugate (ADC) is an example of where chemistry and biology meet. ADCs are large molecules consisting of an antibody linked to a drug [1]. If the focus of the ADC is on the antibody, the search would be carried out by a biologist. If the interest surrounds the linker or the payload, a chemist would be better suited to conduct the search (Fig. 1). Although ADCs involve both chemistry and biology, the elements are often searched independently. Linkers are peptides, but are commonly searched as chemical entities. For example, "val-cit" is a valine-citruline linker and "phy-ly" is a phenylalanine-lysine linker [2]. A general rule is 1–3 amino acids chains (mono-peptides, di-peptides or

* Corresponding author.
E-mail address: kimberly.miller@novartis.com (K. Miller).

http://dx.doi.org/10.1016/j.wpi.2015.01.006
0172-2190/

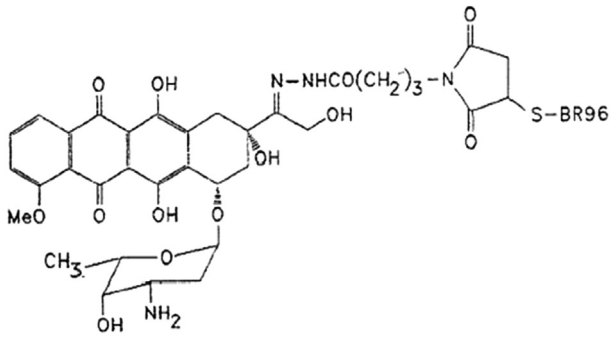

Fig. 1. US20020192223.

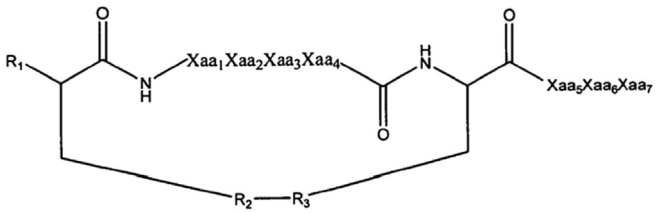

Fig. 2. Claim 1 WO2011120071.

tri-peptides) or 1–8 nucleotides are considered low molecular weight (LMW) compounds. In addition, proteins with more than 30 amino acids or more than 9 nucleotides are considered sequences. Compounds that fall into the 4-30 amino acids range may be classified as both a chemical structure and a sequence [3]. This is where collaboration may be called for.

A scenario where it may not be as obvious could be a request, as Fig. 2 shows, where at first glance it appears to be a Markush structure [4]. There are specific amides and generic R-groups and "X" variables. However, a closer review will show the "X" variables are defined as amino acids, e.g.: Xaa1 is L-tyrosine, L-phenylalanine or L-tryptophan, providing a possible sequence. This type of hybrid Markush leads to the question "is it a chemistry search or is it a biology search or both?"

Larger LMW compounds, such as macrocycles may fit into this gray area. Examples are molecules such as certain antibacterials, cyclosporine and cyclic peptides. Fig. 3 illustrates a cyclic peptide Markush and one specific example [5]. At first glance, the Markush claim appears to be chemical. The R-groups may be atoms such as Hydrogen, Nitrogen and Carbon. However, when a lengthy chain, R_9, is inserted, e.g.: $-(CH_2)_x$-NHC(O)-$(CH_2)_z$-C(O)NH-$(CH_2)_y$, where x, y, and z may be 1 to 5, the ring becomes a cyclic peptide. One now sees a potential sequence, Ac-Nle-cyclo(Glu-Orn-D-Phe-Arg-Dab)-NH_2 to search.

2.1.1. Use of multiple search techniques in a specific example

A specific cyclic peptide, Pitocin, can be used to illustrate the possible need for both chemistry and biology searching for one request to be comprehensive. Pitocin is a smooth muscle contractor typically used during childbirth to enhance labor contractions and is characterized as both a chemical structure and a sequence (Fig. 4). Depending on how the request is interpreted, different approaches may be taken. A request for information on Pitocin may lead a chemist to perform a search centered on the structure in databases such as Marpat and Merged Markush Service (MMS). In contrast, a biologist, knowing this is a hormone, may begin with a sequence search or a keyword search in an appropriate database. It can be shown that carrying out all three of these types of searching elements may be the best approach to create the most comprehensive report.

A reference interview with the client may reveal that the real request is information pertaining to "the use of Pitocin to treat autism". When beginning the search, a first step may be a keyword search using the CAS Registry number for Pitocin along with synonyms for Pitocin and combining these with terms related to autism in CAPlus:

s ((50–56-6/rn) or (pitocin or oxytocin or oxytocic …)) and (autism or autistic …)

Although the above search will provide 169 hits, it may be worth going a bit further and conducting a sub-structure search on Pitocin and combining those results with terms related to autism. Again, relevant answers will be obtained. Upon comparison of the answers obtained via the structure search and the answers obtained via the keyword search, it is shown in Table 1 that a number of unique answers are retrieved from the structure search. Some of these references may describe specific types of compounds and may not have been retrieved by performing a keyword search alone:

A keyword and a chemical structure search have now both been performed to obtain answers from two different methods. An

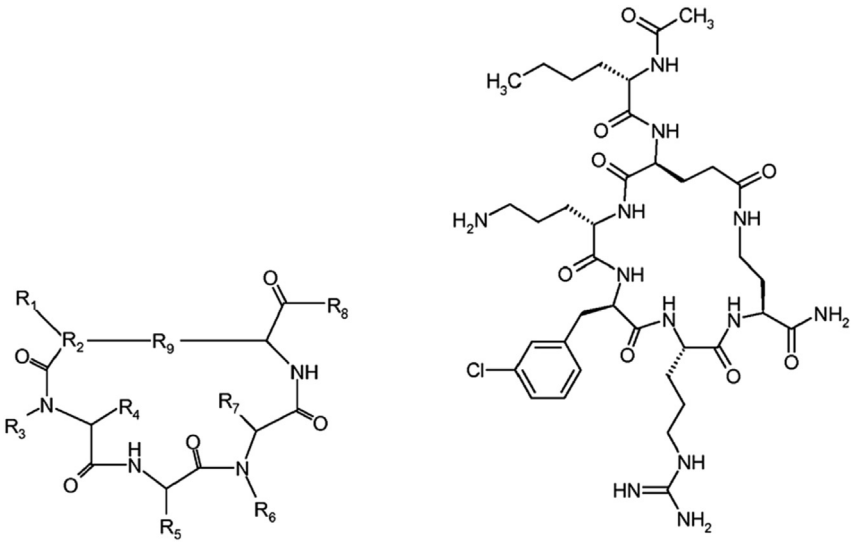

Fig. 3. Markush claim and example WO2011063366.