



patSnap

A guide to
searching and
using Markush
structures

For most people who read patents, Markush structures are fabulous things.

Your mind glazes over upon finding one, casting it and the accompanying Claim 1 texts into intellectual property oblivion. Whomever the Markush structure was formulated for, it wasn't you. But when questions relating to Markush are asked, the dangerously poor understanding of these enigmas is exposed.

But, for chemists, there are many advantages to using Markush structures. They primarily define all the compounds that have been claimed in a patent. Therefore, a reader is able to get a complete picture of, for example, every drug candidate utilised for a particular medical condition in a clinical trial. A Markush structure in Claim 1 of a patent would serve as legal notice to any reader that certain molecules covered within are "*off-limits*"—any attempts to make these compounds in a commercial setting would likely result in an infringement of that patent. A reader may also use Markush structures to distinguish compounds that have actually been synthesised by the patent holders from those that have not—so-called "*prophetic*" compounds, such as bioisosteres.

In this paper, we will identify the components that make up the Markush structure and how you can exploit the value they possess.

What is a Markush structure and what does it do?

For most people who read patents, Markush structures are fabulous things. Your mind glazes over upon finding one, casting it and the accompanying Claim 1 texts into intellectual property oblivion. Whomever the Markush structure was formulated for, it wasn't you. But when questions relating to Markush are asked, the dangerously poor understanding of these enigmas is exposed.

But, for chemists, there are many advantages to using Markush structures. They primarily define all the compounds that have been claimed in a patent. Therefore, a reader is able to get a complete picture of, for example, every drug candidate utilised for a particular medical condition in a clinical trial. A Markush structure in Claim 1 of a patent would serve as legal notice to any reader that certain molecules covered within are "*off-limits*"—any attempts to make these compounds in a commercial setting would likely result in an infringement of that patent. A reader may also use Markush structures to distinguish compounds that have actually been synthesised by the patent holders from those that have not—so-called "*prophetic*" compounds, such as bioisosteres.

In this paper, we will identify the components that make up the Markush structure and how you can exploit the value they possess.

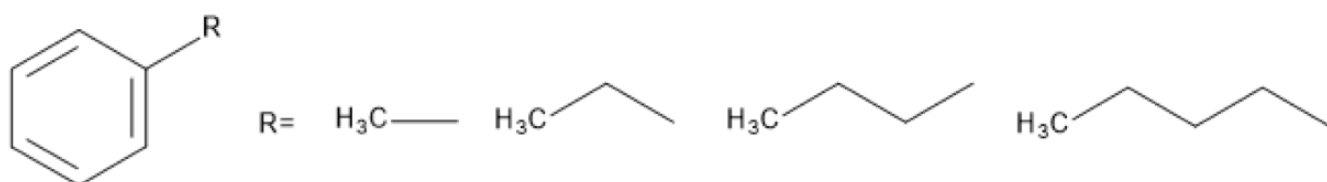


Fig. (1) A simple representation of a Markush structure and enumeration

Figure 1 shows a simple example of a Markush structure, where an aryl ring is monosubstituted with a single R-group substitution that can be methyl, ethyl, propyl or butyl. Therefore, instead of drawing four separate molecules, the single Markush structure represents all of them. Such representations are very useful in patents where more complicated structures are present.

Figure 2 shows a more complicated and drug-like example. Here, we have both R-groups and X-groups. In Markush structures, X-group substituents tend to denote hetero atoms, while everything else is denoted by R-groups. In this example, where R=CH₃ and X=OH, aspirin (O-acetylsalicylic acid) is identified from the Markush structure. Other R and X-group variations of the Markush structure will provide differing aspirin derivatives.

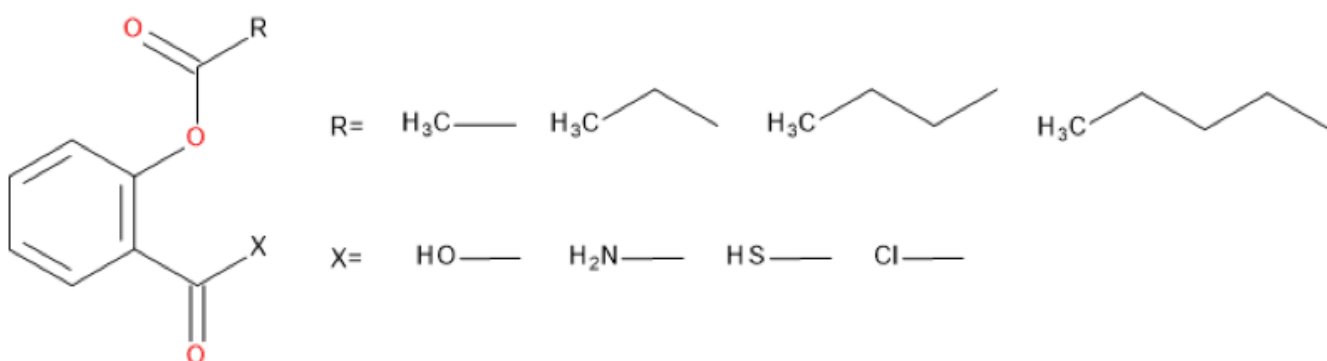


Fig. (2) Markush structure showing aspirin and derivatives

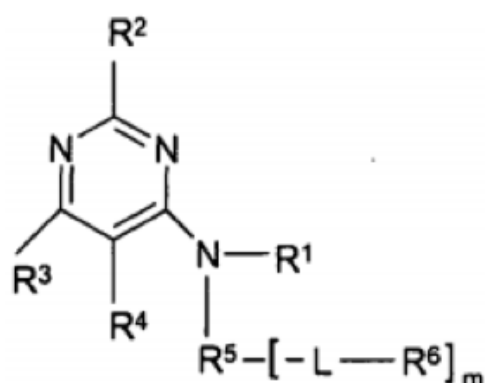
Ultimately, Markush structures will mean different things to different people. A patent agent or lawyer will look at the Markush structure to ascertain whether it has legal integrity, with respect to all compounds claimed, and may use it to narrow down claims specific to molecules of legal interest. A scientist will utilise the Markush structure to seek patentability space, so they can synthesise molecules without infringing any other patents. The advantages of using Markush structures are, in general, biased towards providing the patent holder with a commercial advantage. Compounds can be easily hidden in a Markush structure by claiming thousands of compounds from it without identifying the specific compound of interest. Patent authors can remain deliberately vague about the specific use for a set of claimed compounds in a Markush structure or, alternatively, make a broad set of claims by use of the Markush structure to maximise the coverage of protection. Furthermore, authors of chemical patents can even protect compounds they have not actually made. So-called "prophetic" compounds, such as bioisosteres, can be claimed as biologically active and protected within a patent.

What is the problem with Markush structures? Fear and loathing

Fear? What fear? This fear:

Claims

1. Compounds having the general formula (I)



wherein

R¹, R² and R⁴ are each -H

R³ is a substituted phenyl

R⁵ is a substituted or unsubstituted phenyl

-H, linear or branched substituted or unsubstituted C₁-C₈ alkyl, substituted or unsubstituted aryl, substituted or unsubstituted heteroaryl, substituted or unsubstituted heterocyclyl, substituted pyrrolidinyl, substituted or unsubstituted C₃-C₈ cycloalkyl, disubstituted cyclohexyl, cyclopentyl, substituted or unsubstituted C₅-C₁₂ bicycloalkyl, substituted or unsubstituted adamantyl, -(CH₂)_q-group, wherein q is an integer from 1 to 3, under the proviso, if R⁶ is selected to be a methylene chain -(CH₂)_q-group, R¹⁷ or R¹⁹ are selected to be a methylene chain -(CH₂)_s-group, wherein s is an integer from 1 to 3 or a -(CH₂)_t-A-group, t is an integer from 1 to 3 and A is selected from O or N, respectively, and R⁶ and R¹⁷ or R⁶ and R¹⁹ form together a 5 to 8 membered ring system, or R⁶ represents -(CH₂)_p-Z, wherein p is an integer from 0 to 6 and Z is selected from the group comprising: substituted or unsubstituted aryl, substituted or unsubstituted heteroaryl, substituted or unsubstituted heterocyclyl, -N(R⁷R⁸), wherein R⁷ and R⁸ represent independently from each other -H, or linear or branched substituted or unsubstituted C₁-C₆ alkyl, or Z is selected from -(CR⁹R¹⁰R¹¹), wherein R⁹, R¹⁰ and R¹¹ are independently of each other selected from the group consisting of: -H, linear or branched substituted

Fig. (3) How to stop a scientist from reading your patent: a typical Markush in Claim 1

Figure 3 captures the fear associated with Markush structures perfectly well. There are parts of the Markush that are perfectly understandable, e.g., R₁, R₂ and R₄=H. However, understanding R₃ is confusing, while R₅ is indecipherable. Unfortunately, most Claim 1 Markush structures look similar—this is normal in patents. Understandably, anyone faced with this sort of complexity will not be inclined to utilise the information that is given in Claim 1. So, they relinquish the opportunity to potentially gain valuable intellectual property, legal and chemical information. There are also certain other “fears” that are separate from the Markush structure itself. The need to understand chemical structure drawing is a prerequisite for using any Markush application and this is not a skill possessed by everyone. ChemDraw, Biovia, Symyx, IsisDraw, ACD Labs and MarvinSketch are a few among many drawing packages that are being used today. They each possess a unique user-interface (UI), which must be mastered by the user before any application of the tool. This is time-consuming and very alienating. It is no surprise that Markush structures are seen as features to be feared, not utilised.

Loathing, simply put, is for the Markush structure searching tools that are currently out there. The three main types of searching tools to consider are Merged Markush Service (MMS), Marpat (CAS Registry) and Derwent World Patents Index (DWPI). These three databases form the foundation of current Markush structure searching. They each boast millions of patent records relating to Markush structures dating back to 1957. The main issue with these databases is that the procedure-heavy searching UI appeals only to those who are highly versed in the technical enumeration process of Markush structures, chemical databases and organic chemistry. Furthermore, there exists a very narrow scope of application for these databases. For example, questions such as “What is my freedom to operate with these compounds?”, “Are my drug candidates novel?”, “What other drug discovery patents are closely related to my patent?” and “Who are my pharmaceutical competitors?” cannot be answered using the current Markush searching databases.

What is the solution?

PatSnap has been studying the challenges faced by chemists so it can create better tools for them than any of the others out there. It's been looking at, for example, the R&D medicinal chemist and asking, "What does *this person do every day?*" We fully understand that each day a chemist will be:

1. Setting up synthetic chemical reactions
2. Doing chemical reaction work-ups
3. Submitting crude and pure NMR samples—and analysing the corresponding spectra
4. Running silica-gel column chromatography and prep/analytical HPLC purification
5. Establishing the logistics for meeting agreed research deadlines and
6. Presenting data and publishing papers

Nowhere in the list of daily duties is there time for learning cryptic software so a Markush searching tool can be utilised.

PatSnap has created a Markush tool that's all about simplicity and scope of use. Anyone from any discipline can learn and use the tool in minutes. We also provide a comprehensive toolkit to help users achieve a high degree of functional application, after completing Markush structure searching.

Expert-level chemical structure drawing skills are not required. A user can input the structure of interest as an IUPAC name, trivial name, research code, CAS No, SMILES keycode, JPEG image, or by good old-fashioned structure drawing.

Figure 4 shows the starting point of a Markush search, using benzylpenicillin. The two basic elements in this tool are (1) R-group definition/enumeration and (2) substituent attachment point definition.

- Markush structure drawing is **simple**
- Structure can be input *via* **chemical structure drawing, name, SMILES, CAS No or photo upload**
- Only **2 buttons** are required to create Markush structures (highlighted)

STEP 1.

This button allows you to **define** the R groups on the molecule and **enumerate** the position with fragments

STEP 2.

This button allows you choose attachment points in a fragment

Chemical Name	Structure	Chemical Formula	Synonym Name
BENZYLPENICILLIN		C19H18N2O4S	BENZYLPENICILLIN; penicillin G; benzyl...
benzylpenicillin methyl ester		C17H20N2O4S	benzylpenicillin methyl ester; schem1104...
potassium benzylpenicillin		C19H18KN2O4S	potassium benzylpenicillin; schem03407...
d903485		C33H36KN4O4S2	d903485; potassium benzylpenicillin(1-)...
PENICILLIN G BENZATHINE		C32H38N4O4S	PENICILLIN G BENZATHINE; e798; bicil...

Fig. (4) Markush searching for dummies: 2-steps to initiate a Markush structure search

STEP 1.

define the R groups on the molecule


- Simply delete bonds where you want Markush representation
- Designate R₁, R₂, R₃, R₄ positions by clicking with 





Fig. (5) Step 1

Step 1 uses the highlighted button to provide R-groups for Markush structure generation, as shown in Figure 5. The desired bonds from benzylpenicillin can be deleted and replaced with R-groups using. The R-groups are automatically numbered generating R1-R4 in this example.

STEP 2.

enumerate the position with fragments

- Draw the fragments and click-and-drag over them to define $R_1 - R_4$ with 
- Define the points of attachment in the fragments using 

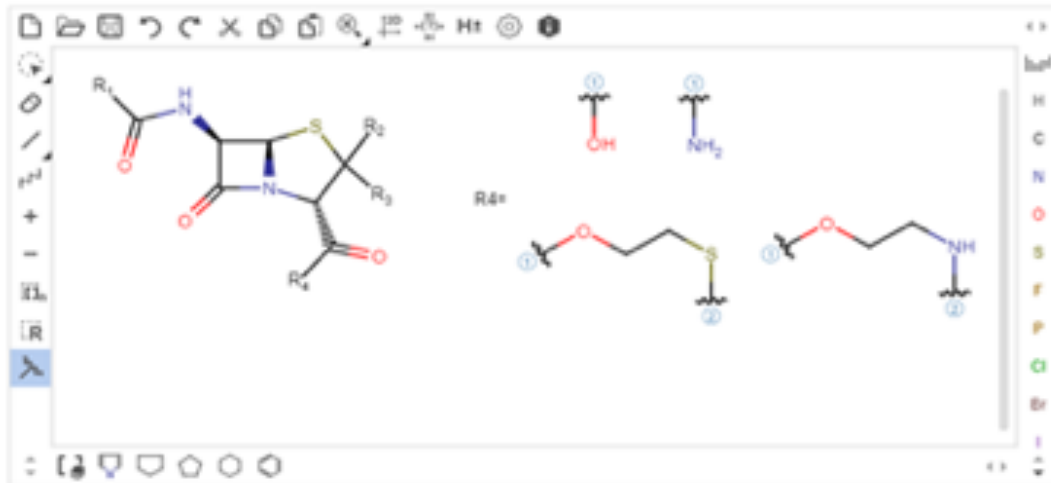


Fig. (6) Step 2

Step 2 involves drawing the R-group fragments which attach to the generated Markush structure in the preceding step. Figure 6 shows enumeration of the R4 position for the penicillin Markush structure. The fragments are then assigned attachment points using the highlighted button. Here, the -OH and -NH₂ fragments have only one point of attachment, while the -OCH₂CH₂S- and -OCH₂CH₂NH-fragments can be attached at 2 points. All the drawn fragments are highlighted, then is pressed to label them as R4, as shown in **Figure 7** (all other R-groups have been removed). The 2-step process is now complete and searching can commence.

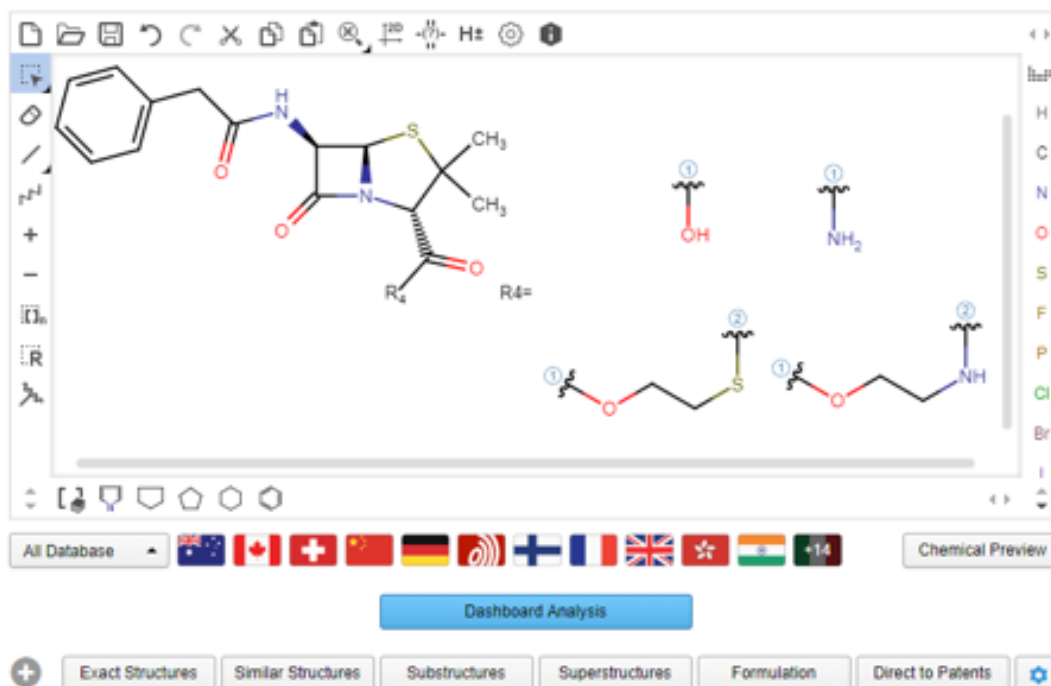


Fig. (7) The Markush structure request completed and ready for searching

The Markush structure search is specifically designed for the “*Similar Structures*” search option, **Figure 7**. A snapshot of the output obtained can be observed in **Figure 8**. This search has identified 5968 discrete molecules associated with the Markush structure—this list can be further refined using filters. The accuracy associated with the discrete compound records is based on a Similarity Score—with “1” being the closest to the Markush structure. Specific records can be selected and useful functions can be accessed to manipulate the selected records. One of the most useful functions is the Structure-to-Patents feature—this lets you identify which patents contain the molecules of interest from a particular Markush structure search.

LabBench allows up to 20000 compound structures to be stored, where further data manipulations can be made. Many commercial users of Markush searches want to know how many other patents are similar to the one they’ve filed. This can be easily visualised in PatSnap’s Chemscape, which provides a heat map corresponding to patent similarity. The PatSnap Chemical platform also offers an “*Analysis*” function, which provides a plethora of important information about selected patents, as shown in **Figure 9**. The large number of functional applications available after a Markush search is useful to commercial users.

Structure-to-Patents takes your selected structures and identifies all patents in which they appear

LabBench is a place to store up to 20000 compounds, which you can then search and manipulate

Analysis is a multi-tool data analysis platform, see Fig. (9)

Chemcape shows the chemical similarity between your patent and others

1-20 chemicals, 5968 chemicals in total

Tanimoto threshold 0.7 ↓ Similarity ↓ 20

zinc6617263
Absolute
Synonyms: zinc6617263
Similarity Score: 0.96
[Search with Structure](#) [Just Track it](#)

(2s,5r,6r)-3,3-dimethyl-7-oxo-6-[(phenylacetyl)amino]-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxamide
Absolute
Synonyms: (2s,5r,6r)-3,3-dimethyl-7-oxo-6-[(phenylacetyl)amino]-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxami... [Show all](#)
Similarity Score: 0.96
[Search with Structure](#) [Just Track it](#)

Select records you wish to analyse

Similarity Score shows the accuracy of the structure shown relative to the Markush search

Track exact compounds in other patents via email alerts

Fig. (8) Snapshot of output for a Markush structure search and post-search functions

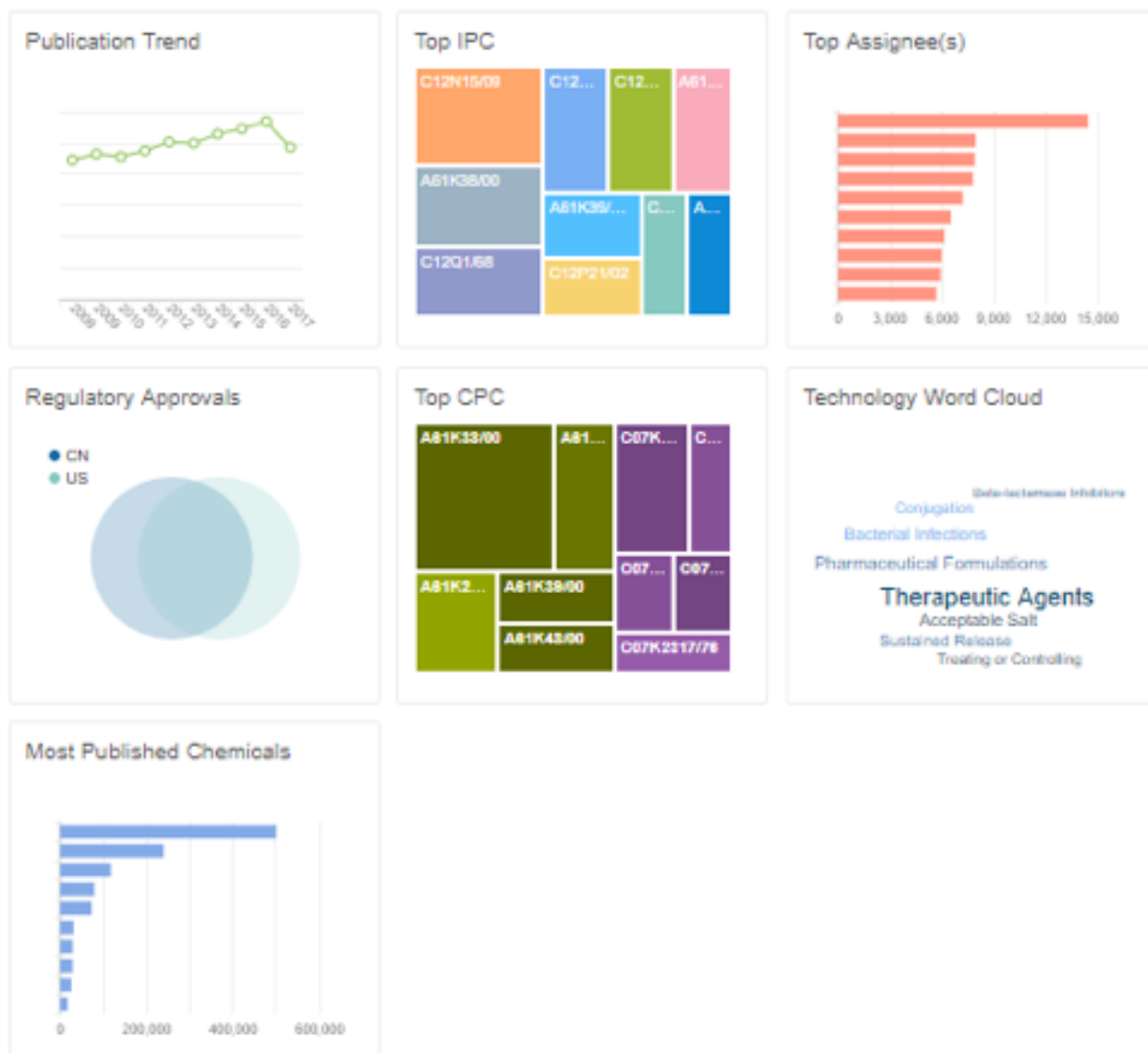
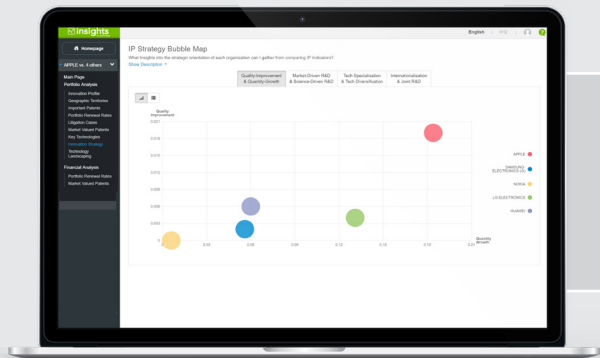


Fig. (9) What you can learn about your selected patents if you hit the Analysis button

Why use PatSnap Chemical for Markush structure searching?

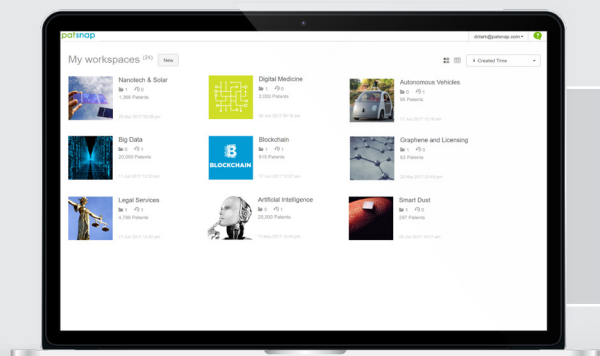
1. You can actually use it: it is not the multi-step, procedural nightmare that is other Markush searching databases. It is a simple 2-step process to define R-groups and enumerate them, before hitting the search button.
2. PatSnap is not competing with existing databases: The PatSnap Chemical Markush tool is completely unlike Marpat/MMS/DWPI databases and is an analysis-focused tool.
3. It is not just for chemists: PatSnap has created a Markush searching tool where the primary user type is everyone. Patent lawyers/agents, academics and pharmaceutical personnel who are not related to research chemistry can access simple Markush searching—via chemical structure and non-chemical structure methods.
4. Our focus is on customers: Our focus is on the invaluable scientific, commercial and legal information which can be extracted from Markush structure searches. PatSnap's Markush tool is designed for effective "*post-search analysis*", using features such as Chemscape, Workspace (for inter-departmental collaboration) and LabBench, among others.

The PatSnap platform gives your team the data it needs to make better innovation decisions.



R&D is a race against time
(and your competitors)

Where is your next million
dollar idea?



Collaborate to accelerate
your innovation rate

You've made it all the way. Now take your business all the way to the top by putting your new skills to practice on the PatSnap platform.

[Request a demo](#)



Where will you find your next opportunity?

About Chemical by PatSnap

Chemical by PatSnap lets you explore 114 million structures linked to over 120 million patents. These data sets are combined with REACH, US FDA, Chinese FDA and European EMA regulatory information, and clinical trial data. The platform combines innovation data with scientific information, in one easily searchable interface.

Find out more: patsnap.com