Elaine N. Cheeseman, Ph.D.

Science IP®; CAS

echeeseman@cas.org

Milestone for Markush Searching: Celebrating 30 Years of MARPAT®

259th ACS National Meeting & Exposition
Division of Chemistry & the Law
Spring, 2020

CAS®

A DIVISION OF THE
AMERICAN CHEMICAL SOCIETY

Agenda

- **CAS providing access to the world's chemical information**
- Why search for Markush structures?
- Trends in MARPAT
- Examples and Science IP
- Summary

As science evolves, information is a critical differentiator influencing business strategy and return on R&D investment

Accelerate Discovery

- Idea generation
- Repurpose technology
- Identify strategic partners

Decrease R&D Costs

- Avoid reinventing the wheel
- Efficient chemical sourcing and synthesis
- Keep up to date on new technologies

Gain Competitive Advantage

- Find white space
- Monitor competition
- Mitigate risk

CAS is a specialist in scientific information solutions

We provide products and services that power discovery to solve our world's biggest challenges by helping organizations plan, innovate, and protect their innovations

With over 110 years of experience, no one knows more about scientific information and related technology than CAS

UNPARALLELED
SCIENTIFIC CONTENT

SPECIALIZED
TECHNOLOGY

UNMATCHED
HUMAN EXPERTISE

CAS was the original crowdsourcing of chemical and scientific discoveries

- CAS, as a division of the ACS, has a long, rich history of providing access to the world's chemical information going back to 1907.
- The ACS' vision
Improving people's lives through the transforming power of chemistry

CAS providing access to the world's chemical information

- Early on, CAS' inverted names in print indices provided an overview of substances with the same main structural component
- CAS Registry was formed initially to aid in the indexing of substances in patents and literature
- Harry Morgan of CAS invented the Morgan algorithm which built on the work of Donald Gluck of DuPont for structure searching based on connection tables
 - Connection tables could be searched by computers
 - Became basis for CAS Registry®; still used today.
- MARPAT was build on the technology developed for CAS Registry.
 - W. Fisanick; US4642762; 1987; "Storage and Retrieval of Generic Chemical Structure Representations"

CAS scientists analyze, summarize and make scientific information accessible to colleagues worldwide

Source
Selection

Document
Indexing

Reaction
Indexing

Markush
Indexing

Authority
Processing

In the age of algorithms, human intellectual indexing by CAS's 600+ scientists is still critical in chemistry

Agenda

- CAS providing access to the world's chemical information
- **Why search for Markush structures?**
- Trends in MARPAT
- Examples and Science IP
- Summary

MARPAT – a source of Markush structures

- “Eugene Markush, a dye chemist, was the first person to successfully obtain a patent based on generic structures [filed] in 1923.” (Wikipedia)
- The patent was granted in 1924.

US1506316:

Claims: 95
1. The process for the manufacture of dyes which comprises coupling with a halogen-substituted pyrazolone, a diazotized unsulphonated material selected from the group consisting of aniline, homologues of aniline and halogen substitution products of aniline. 100

- The name MARPAT from **MAR**kush (or generic) structures from **PAT**ents

Why search for Markush structures?

- Patentability:
 - To retrieve patents in which your compounds could be covered by someone else's patent(s) by a Markush structure.
- Freedom to Operate:
 - To retrieve patents in which your compounds are covered by someone else's patent(s) by a Markush structure. Will you be infringing?
- Validity
 - To retrieve older patents (based on the date provided) which could be used to invalidate specific patent(s) by a Markush structure.

In addition to specific substances, Markush structures are important

- CAS REGISTRY provides access to specific substances
 - Reference information is available in CAplusSM.
 - Retrieved structures are searched for references in CAplus by their CAS RN[®]
- MARPAT is a source of Markush structures and references.
 - Unique references are usually retrieved in a MARPAT search.

CAS[®]

A DIVISION OF THE
AMERICAN CHEMICAL SOCIETY

MARPAT covers many types of generic structures

- MARPAT covers organic and organometallic compounds
 - Including coordination compounds, small polymers, monomers, & small biologics
- MARPAT does not cover
 - Alloys, metal oxides, inorganic salts, intermetallics, large polymers, & large biologics

MARPAT covers Markush structures in many technology areas

Classification Codes of MARPAT Records

BIO – biochemistry

- Radiation biochemistry
- Fermentation biochemistry
- Agrochemical bioregulators

ORG – organic

- Heterocycles
- Steroids
- Organometallics

MAC – macromolecular

- Dyes
- Monomers
- Surface-active agents

APP – applied

- Pharmaceuticals
- Cosmetic compositions
- Essential oils

PIA - Physical, Inorganic, & Analytical

- Radiation Chemistry
- Photochemistry
- Reproductive Processes

Celebrating 30 years of MARPAT

MARPAT became available to all STN users on September 30, 1990

	At Launch	Now
Records	21,000	> 525,000
Markush structures	65,000	> 1.25 million
Time coverage	Patents found in CAplus with the patent publication year of 1988 to the present	Patents found in CAplus with the patent publication year of 1988 to the present <ul style="list-style-type: none">▪ Selective coverage back to 1984 depending on the language and patent office▪ INPI data from 1961 to 1987
Content coverage	Markush structures in all 80 CA sections	Markush structures in all 80 CA sections
Updates	Biweekly	Daily

Agenda

- CAS providing access to the world's chemical information
- Why search for Markush structures?

- **Trends in MARPAT**

- Examples and Science IP
- Summary

MARPAT keeps growing

Source: CAlus on STN

Trends in MARPAT - priority countries over time

Source: CAlus on STN

Agenda

- CAS providing access to the world's chemical information
- Why search for Markush structures?
- Trends in MARPAT
- **Examples and Science IP**
- Summary

Patents can be huge

```
L9 ANSWER 1 OF 1 HCAPLUS COPYRIGHT 2020 ACS on STN PatentPak PDF
AN 2017:1405898 HCAPLUS Full-text<<LOGINID::20200102>>
DN 167:368379
TI Preparation of spironaphthalene oxathiadiaza tetracyclopentacosa
tetraenonedioxide derivatives and analogs for use as MCL-1
protein inhibitors
IN Brown, Sean P.; Bedke, David Karl; Degraffenreid, Michael R.; ...
PA Amgen Inc., USA
SO PCT Int. Appl., 2426pp. CODEN: PIXXD2
DT Patent
LA English
FAN.CNT 1
PPPI
PATENT NO.          KIND  DATE          LANGUAGE  PatentPak
-----
WO 2017147410      A1    20170831  English   PDF
PI
PATENT NO.          KIND  DATE          APPLICATION NO.  DATE
-----
WO 2017147410      A1    20170831      WO 2017-US19336  20170224
...
PRAI US 2016-62300013      P    20160225
OS CASREACT 167:368379;MARPAT 167:368379; CASFORMULTNS 2017:1405898
```

Statistics:

- 1096 synthetic examples
- Claim 1 in patent application – 7 pages
- 3026 indexed cpds in CAS Reg
- One (independent) claim indexed in MARPAT
- 6384 reactions indexed in CASREACT

CAS[®]

A DIVISION OF THE
AMERICAN CHEMICAL SOCIETY

Example: Elexacaftor

- “Elexacaftor/ivacaftor/tezacaftor, sold under the brand name Trikafta, is a combination medication used in those that have cystic fibrosis with a f508del mutation.” (Wikipedia)
- “Oct. 21, 2019 The U.S. Food and Drug Administration today approved Trikafta (elexacaftor/ivacaftor/tezacaftor), the first triple combination therapy available to treat patients with the most common cystic fibrosis mutation. ...” (www.fda.gov/news-events/press-announcements/fda-approves-new-breakthrough-therapy-cystic-fibrosis)

Absolute stereochemistry shown

Elexacaftor

CAS
A DIVISION OF THE
AMERICAN CHEMICAL SOCIETY

CAS Registry record for Elexacaftor

L5 ANSWER 1 OF 1 REGISTRY COPYRIGHT 2020 ACS on STN

RN 2216712-66-0 REGISTRY

ED Entered STN: 20 Apr 2018

CN 3-Pyridinecarboxamide, N-[(1,3-dimethyl-1H-pyrazol-4-yl)sulfonyl]-6-[3-(3,3,3-trifluoro-2,2-dimethylpropoxy)-1H-pyrazol-1-yl]-2-[(4S)-2,2,4-trimethyl-1-pyrrolidinyl]- (CA INDEX NAME)

...

CN Elexacaftor

CN N-(1,3-dimethyl-1H-pyrazole-4-sulfonyl)-6-[3-(3,3,3-trifluoro-2,2-dimethylpropoxy)-1H-pyrazol-1-yl]-2-[(4S)-2,2,4-trimethylpyrrolidin-1-yl]pyridine-3-carboxamide

CN VX 445

FS STEREOSEARCH

...

LC STN Files: ADISINSIGHT, CA, CAPLUS, CASFORMULTNS, CASREACT, CHEMCATS, EMBASE, IMSRESEARCH, IPA, MEDLINE, TOXCENTER, USPATFULL

****PROPERTY DATA AVAILABLE IN THE 'PROP' FORMAT****

13 REFERENCES IN FILE CA (1907 TO DATE)

16 REFERENCES IN FILE CAPLUS (1907 TO DATE)

Family structure search retrieves the specific compound, salts, stereoisomers, labelled cpds, mixtures, etc.

- A family search on this structure in CAS REGISTRY retrieved 16 substances

Absolute stereochemistry shown

Absolute stereochemistry shown

Absolute stereochemistry shown

Absolute stereochemistry shown

- 12 patents and 4 non-patents records in CAPLUS
- The oldest record is WO 2017177124 (basic).

Bibliographic information of WO2017177124

L8 HCAPLUS COPYRIGHT 2019 ACS on STN

[PatentPak PDF](#) | [PatentPak PDF+](#) | [PatentPak Interactive](#)

AN 2017:1635239 HCAPLUS [Full-text](#)<<LOGINID::20191204>>

DN 167:508112

ED Entered STN: 12 Oct 2017

TI Preparation of ivacaftor analogues containing silicon atoms as CFTR modulators

IN Munoz, Benito; Parks, Daniel; Bastos, Cecilia M.

PA Proteostasis Therapeutics, Inc., USA

...

PPPI

PATENT NO.	KIND	DATE	LANGUAGE	PatentPak
WO 2017177124	A1	20171012	English	PDF PDF+ Interactive

...

PPPI

PATENT NO.	KIND	DATE	LANGUAGE	PatentPak
WO 2017177124	A1	20171012	WO 2017-US26579	20170407
CA 3019663	A1	20171012	CA 2017-3019663	20170407
AR 108203	A1	20180725	AR 2017-100899	20170407

..

PRAI US 2016-62319439 P 20160407

WO2017-US26579 W 20170407

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)	
(19) World Intellectual Property Organization International Bureau	
(43) International Publication Date 12 October 2017 (12.10.2017)	(10) International Publication Number WO 2017/177124 A1

(51) International Patent Classification: C07D 215/56 (2006.01) A61K 31/095 (2006.01) C07F 7/08 (2006.01) A61P 11/00 (2006.01)	(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
(21) International Application Number: PCT/US2017/026579	
(22) International Filing Date: 7 April 2017 (07.04.2017)	
(25) Filing Language: English	
(26) Publication Language: English	
(30) Priority Data: 62/319,439 7 April 2016 (07.04.2016) US	
(71) Applicant: PROTEOSTASIS THERAPEUTICS, INC. [US/US]; 200 Technology Square, 4th Floor, Cambridge, MA 02139 (US).	(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, NG, SN, TD, TG).
(72) Inventors: MUNOZ, Benito; 377 Lowell Avenue, Newtonville, MA 02460 (US). PARKS, Daniel; 11a Orion Rd, Pepperell, MA 01463 (US). BASTOS, Cecilia, M.; 1 Bi-gelow Way, South Grafton, MA 01560 (US).	
(74) Agents: KAVANAUGH, Theresa, C. et al.; Goodwin Procter LLP, 100 Northern Avenue, Boston, MA 02210 (US).	Published: — with international search report (Art. 21(3))

CAS

A DIVISION OF THE
AMERICAN CHEMICAL SOCIETY

CAS indexing provides context to hit substances

CAS Hit indexing:

"IT 936727-05-8, VX-809 1152311-62-0, VX-661 1918143-53-9 1954716-30-3, VX 152 1954716-32-5, VX 440 1954716-50-7, GLPG 2665 2138326-34-6, FDL 169 2138326-53-9, GLPG 2851 2138328-20-6, GLPG 2737 2204245-48-5, VX 659 **2216712-66-0, VX 445**

RL: THU (Therapeutic use); BIOL (Biological study); USES (Uses) (co-drug; preparation of ivacaftor analogs containing silicon atoms as CFTR modulators)"

85 substances were indexed to this patent application in CAPlus

PatentPak® provides the location in the patent from where the substances was indexed

24. The method of any one of claims 18-23, wherein the subject is a human patient.
25. The method of any one of claims 18-24, further comprising administering at least one or two additional CFTR modulators.
26. The method of claim 25, wherein one additional CFTR modulator is selected from the group consisting of VX-152, VX-440, VX-445, VX-659, VX-809 (lumacaftor), VX-661, FDL169, GLPG2851, GLPG2665, GLPG2737, and GLPG2222.

MARPAT contains search features to focus your retrieval

Match level	Example of a hit
Atom – specifically (most narrow)	pyridine
Class – specifically or generically	pyridine, “any ring/heterocycle”
Any – specifically, generically, or with a text descriptor (most broad)	pyridine, “any ring/heterocycle”, or text – such as “optionally substituted”

The match level settings can be adjusted to retrieve the type and the approximate amount of retrieved references.

Approach 1: Match level “atom” on the pyridine nitrogen, “any” on the end groups

Structure query

Black circle = atom

Red circles = any

No circles = class

Closed substructure search
retrieved 7 records; 4 were not
found in CAS REGISTRY/HCAplus

The oldest:

Preparation of benzylnaphthalenes for the treatment of viral infection,
Kemia Inc., USA

PI WO 2004071426 A2 20040826 (basic)

PRAI US 2003-60446713 20030211 (three)

MSTR 1 Assembled

499: heterocycle ... (opt. substd.)

601: opt. substd.

725: heterocycle ... (opt. substd.)

1195: heterocycle ... (opt. substd.)

1479: opt. substd.

Patent location: claim 2...

CAS®

A DIVISION OF THE
AMERICAN CHEMICAL SOCIETY

Approach 2: Match level “class” on most; “any” on the end groups

Structure query:

Red circles = any

No circles = class

Closed substructure search retrieved 7 records; 3 were not found in the CAS REGISTRY/HCAplus or earlier MARPAT searches

Processes for preparing fused heterocyclic ion channel modulators,

Gilead Sciences, Inc., USA

US 20150225384 A1 20150813 (basic)

PRAI CN 2014-10050699 20140213 (three)

MSTR 15 Assembled

135: heterocycle ... (opt. substd. by 1 or more G4)

150: heterocycle ... (opt. substd.)

159: heterocycle ... (opt. substd. by 1 or more G5)

207: R

209: carbon chain ... (opt. substd. by 1 or more G6)

326: heterocycle ... (opt. subst'd. by 1 or more G38)

344: alkyl <containing 1-4 C>

345: alkyl <containing 1-4 C>

346: R

347: R

CAS

A DIVISION OF THE
AMERICAN CHEMICAL SOCIETY

MARPAT can be searched in a variety of ways

- Options demonstrated: closed substructure searches
 - Approach 1: Heteroatom(s) in the central heterocycle is atom; end groups any
 - Approach 2: Central part of the compound is class; end groups any (fewer)
 - Approach 1 and/or 2
- Other options: substructure searches
 - Allow the nodes to be open to any substitution (no additional fusion on rings)
 - Leave off end groups; open to any substitution at those positions.
- Requestors frequently want specified range of references (e.g. 50 – 100 references).
 - The match level can generally be adjusted to get the number of references in specific ranges.
- **What approach do you want?**

The MARPAT content can be searched on SciFinderⁿ

- Match levels are fixed to the defaults
 - Rings = match level atom
 - Chains = match level class
- Elexacaftor structure query in SciFinderⁿ
 - “Search Patent Markush”; “As drawn” (no additional substitution)
 - Retrieved 5 Markush structures from **3 references**
- Elexacaftor searched both the above approaches in Marpat on STNext retrieved 10 references
 - 7 Markush references not retrieved in the Markush searching on SciFinderⁿ
 - We recommend a search in STNext with adjusted match level settings.

MARPAT is available on SciFinderⁿ, cont'd

- An example of a MARPAT reference from SciFinderⁿ

WO2018107100

[View Reference Detail](#)

- An example of a MARPAT reference retrieved on STNext (Approach 1), but not on SciFinderⁿ

WO 2004071426

MSTR 1 Assembled

499: heterocycle ... (opt. substd.)

601: opt. substd.

725: heterocycle ... (opt. substd.)

1195: heterocycle ...(opt. substd.)

1479: opt. substd.

Patent location: claim 2...

The other Markush source, DWPIM, is now available on STNext

- DWPIM is produced by Derwent, then Thomson Reuters, now Clarivate
- Search in the same way as MARPAT (match level atom, class, any)
- Search with the STN query options and/or the Derwent superatoms
- Duplicates and be removed since it is on the same system (STNext)

Science IP[®], the official search service of CAS, delivers comprehensive IP search results and analysis on a per-project or retainer basis

- Team of search experts with **over 250 years of experience**
 - Proficient in searching chemistry, sequences, polymers, formulations, biomedical topics, etc
 - Advanced degrees in chemistry and/or biology.
- Searchers with industry experience that understand the why behind your search so they can help you **customize each project** to best meet your needs and budget
 - Iterative, consultative process that puts you in direct contact with the searcher
 - Free cost estimates to price project before work begins
- Fast turn-around to **keep your projects moving forward**

Contact Science IP: scienceip@cas.org or www.scienceip.org

Summary

- Markush structures are important for the purpose of patentability, FTO, and validity.
- In CAS' mandate is to provide access to all types of chemical information, MARPAT is a key source.
- MARPAT has been available on STN® for 30 years!
- Searching MARPAT can retrieve references which are not retrieved by searching only the exemplified compound source, CAS REGISTRY/HCAplus.
- MARPAT makes use of match level settings to focus in on the type of records and the amount which the requestor wants.
- The MARPAT indexing can be searched on SciFinderⁿ – but the match level settings are the default settings.
- The Science IP staff are very experienced in working with clients in scientific searches – including Markush searches.

Elaine N. Cheeseman, Ph.D.

Senior Searcher, Science IP

echeeseman@cas.org

www.cas.org

Connect to CAS at

CAS[®]

A DIVISION OF THE
AMERICAN CHEMICAL SOCIETY