

STUDIE VYUŽITÍ GENERICKÝCH (MARKUSHOVÝCH) VZORCŮ V CHEMICKÝCH PATENTECH

JAROSLAV ŠILHÁNEK, EVA BENEŠOVÁ
a PETR KAČER

Ústav organické technologie, VŠCHT Praha, Technická 5,
166 28 Praha 6
silhanek@vscht.cz

Došlo 12.9.16, přijato 18.10.16.

Klíčová slova: chemické patenty, Markushovy vzorce,
generické vzorce, chemické báze dat, SciFinder, Reaxys,
patentové rešerše

Obsah

1. Úvod
2. Volba testovacího souboru patentů pro porovnávání rozsahu excerpce sloučenin
3. Excerpce chemických sloučenin z patentů a jejich analýza
4. Možnosti zadání strukturních rešerší v režimu *Markush*
5. Závěr

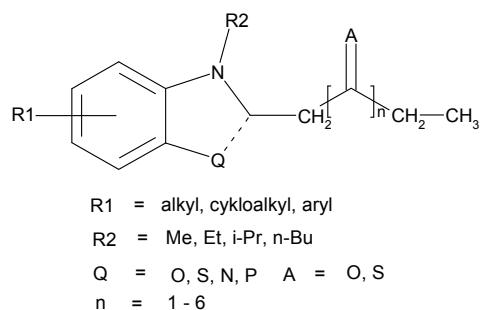
1. Úvod

Součástí chemických patentů, zvláště z oblasti farmaceutické chemie, je deklarace v patentu chráněných sloučenin obecným, přesněji generickým a v patentové terminologii Markushovým vzorcem (obr. 1). Tuto praxi úspěšně obhájil americký chemik Eugene A. Markush¹ v soudní při s americkým patentovým úřadem v r. 1924 (cit.²). I když je patent především právní dokument stvrzující právní nárok na duševní vlastnictví v něm obsažených skutečností, z hlediska věcného obsahu je i zdrojem vědeckých informací. V případě chemických sloučenin nás tedy zajímá nejenom struktura odvoditelná z generického vzorce v kombinaci s uvedenou sadou strukturních složek, ale samozřejmě i jak byla připravena a jaké má vlastnosti. Zájmy patentového specialisty a praktikujícího chemika se zde rozcházejí, ale z právního hlediska je rozhodující skutečnost, že sloučenina popsaná generickým vzorcem je explicitně uvedena v patentu. Z hlediska chemie lze ale položit otázku, jak, a zda vůbec byla připravena, když vlastnosti a identifikační údaje chybí. V první instanci se s otázkou, zda uznat či neuznat existenci chemické sloučeniny jako reálnou, setkávají excerpční zpracovávající patenty do nějaké chemické báze dat. Má se za rozumné, že

do bází dat jsou excerpovány jen sloučeniny, pro které jsou v patentu uvedeny syntetické postupy, nebo některé vlastnosti, např. jejich biologická aktivita. Pak ale nutně existuje pravděpodobnost, že při rešerších nenajdeme sloučeninu i velmi blízkou té, kterou hledáme, ale která je v patentu jen zahrnuta do generického vzorce.

Vývoj a postupné zdokonalování nástrojů pro práci s chemickými bázemi dat SciFinder a Reaxys vedl v nedávné době k nabídce strukturních rešerší v režimu, který je ve strukturních editorech označen jako *Markush Search*, nebo jen *Markush*, což by mohlo riziko neúspěšných rešerší alespoň částečně eliminovat. Bohužel funkce těchto strukturních rešerší není příliš detailně dokumentována. To nás vedlo k rozhodnutí pokusit se ozřejmit možnosti využití generických (Markushových) vzorců na praktických příkladech, a to jak z hlediska rozsahu a pravidel excerpce sloučenin, tak z hlediska využití strukturních vyhledávacích režimů *Markush* ve výše zmíněných chemických bázích dat.

Je ale nutné upozornit, že výše popsany problém není nový a existuje prakticky od zavedení generických vzorců do patentových dokumentů. To vedlo už v r. 1951 anglického chemika Montagu Hyamse k založení firmy s názvem DERWENT právě s cílem zpracovávat chemické patenty tak, aby byly zachovány generické vzorce jako zdroj informací. Nejdříve k tomu byla využívána technika dřevných štítků, kde jednotlivé pozice znamenaly individuální substituenty nebo jejich skupiny. To bylo později převedeno do počítačové podoby s daleko většími možnostmi a postupně rozšiřováno do dalších technických oborů. Nejdůležitější báze dat této společnosti je tzv. World Patent Index (WPI), přístupný na databázovém středisku STN International³. Patenty obsahující generické vzorce zpracovává i CAS, a to do samostatné báze MARPAT⁴, rovněž přístupné na STN International. Společnou nevýhodou těchto bází jsou jednak vysoké náklady na přístup a dále nutnost rozsáhlých a stále udržovaných praktických zkuše-



Obr. 1. Ilustrace generického Markushova vzorce v patentu

ností, což vede k tomu, že jsou využívány téměř výhradně specializovanými patentovými řešeršery. Nabídka strukturálních řešeršerů typu Markush v bázích dat SciFinder a Reaxys by tak mohla otevřít přístup k těmto informacím i pro praktikující chemiky a alespoň částečně překlenout určitou dvojkolejnost chemických řešeršerů.

2. Volba testovacího souboru patentů pro porovnávání rozsahu excerptce sloučenin

Rozšířená představa chemiků je, že z patentů jsou excerptovány chemické sloučeniny, které jsou uvedeny v příkladech, nebo pro které jsou v patentu konkrétní data. Již méně si chemici uvědomují, že z patentů musí být excerptovány také sloučeniny, které jsou specificky (např. plným vzorcem nebo systematickým názvem) uvedeny v patentových nárocích. Oficiální zásada pro excerptci sloučenin do bází Chemical Abstracts říká, že excerptovány

Tabulka I
Přehled analyzovaných patentů

Patent	SciFinder						Reaxys		
	All	Prep	Prophetic	Reactant/ Reagent	Reactions	Experi- mental conditions	All	Prep	Reagent/ Catalyst
WO 2015157005	6477	6179	5995	46	0	0	500	172	40
US 20130072468	2222	1503	1115	1175	6384	6384	1064	534	55
WO2014100533	917	611	13	451	7280	0	500	469	32
US 20130184248	586	543	6	248	3129	3129	574	477	49
EP 1958634	486	418	54	320	4741	4741	443	361	74
US 20130281399	380	262	7	246	861	0	408	857	98
WO 2012043791	321	254	11	213	1810	0	100	žádné reakce	
WO 2010146236	309	207	2	208	804	804	256	208	33
WO 2013174736	278	252	64	221	žádné reakce		193	378	47
US 20060089496	275	255	2	37	93	93	301	132	46
WO 2013092756	271	251	4	59	437	437	276	333	35
WO 2012025638	229	162	10	146	415	415	174	155	18
WO 2013184934	198	137	2	124	939	0	227	1067	54
WO 2014074778	196	118	24	107	280	0	145	130	34
WO 2003105845	186	136	0	112	328	0	218	151	38
WO 2013045495	176	71	6	65	825	825	92	462	50
US 20120115903	155	113	17	101	845	839	133	158	44
WO 2013 095304	138	68	0	79	80	80	76	60	7
WO 2008090140	125	99	2	76	426	426	108	92	23
WO 2014198478	105	82	101	82	250	0	žádné sloučeniny		
US 20110313167	85	57	2	35	31	30	89	32	25
US 6403604	81	30	28	31	36	36	40	29	2
WO 2015165970	73	59	2	59	818	0	Reaxys patent neobsahuje		
EP 2003120	73	44	19	38	357	311	91	37	34
WO 2008111096	63	57	10	18	52	52	62	40	13
EP 2738165	32	0	0	16	1	0	10	4	15
EP 735043	16	7	5	6	0	0	2	1	–
WO2014037391	12	4	0	2	1	1	žádné sloučeniny		
WO 201511010	9	2	0	8	3	0	7	2	4
WO 2008092665	12	7	1	6	8	8	11	7	2

jsou ... *Claimed and exemplified substances with supporting data*⁵. Je nutné si uvědomit, že termín *Claimed* znamená přesně výše uvedené případy, kdy jsou v patentových nárocích uvedeny sloučeniny s přesným názvem nebo strukturou, ale buď vůbec bez jakýchkoliv dalších údajů, nebo jen vágně přiřazené k obecné syntetické metodě. I když se nám to moc nelíbí, tak pokud je chemická sloučenina přesně pojmenována v patentových nárocích, jednoduše existuje.

V bázi dat SciFinder se v posledních letech objevil nový termín související s patentovými dokumenty, a to *Prophetic Substances in Patents*. Obrátíme-li se opět na informační materiály CAS, dozvíme se, že se jedná o sloučeniny ... *Specific but uncharacterized substances (e.g. reactants, isolated intermediates, products) identified by chemical name or structure, including structures displayed in a table*⁵. Zdá se, že této definici odpovídají právě výše zmíněné případy a že se jedná o snahu producenta umožnit rozlišení mezi sloučeninami plně charakterizovanými od sloučenin uvedených jen strukturou nebo názvem. Podrobnější zásady včetně příkladů jsou evidentně v manuálech excerptorů, které i v případě jejich dostupnosti by praktikující chemik asi stejně nestudoval. Použili jsme proto přístup, který v podobném, ale daleko závažnějším případě využil Ben Wagner⁶, a na souboru konkrétních patentů, byť omezeném, jsme se pokusili alespoň orientačně zjistit, jak se představy chemiků o rozsahu excerptce sloučenin z patentů shodují s realitou a ilustrovat tak označení *Prophetic Substances*.

Patenty byly vybrány v zásadě náhodně z oblasti farmaceutické chemie, nebo obecně chemie biocidů; jediným omezením byl požadavek na přítomnost generického (Markushova) vzorce a role *Prophetic Substances* v programu SciFinder. Dále byl výběr omezen na patenty evropské, americké a patenty WO, protože pro porovnání rozsahu excerptce v obou nejdůležitějších chemických bázích dat s bází Reaxys byl vzat ohled na to, že Reaxys jiné patenty nezpracovává. Určitý problém, v zásadě ale málo významný, se projevil při evidentně rozdílné praxi zpracovávání patentových rodin v případě báze SciFinder a Reaxys. V případě CAS je považováno za standardní praxi zpracovat první patent, který se dostane na stůl excerptorů, což je sice asi stejné i v případě báze Reaxys, ale komplexita patentů může vést k tomu, že takové první patentové dokumenty mohou být různé. Nejčastěji se takové situace vyskytují v případě patentů WO. V případě báze SciFinder bývají rozsahy excerptovaných sloučenin u patentů WO a druhého záznamu např. patentu EP prakticky stejné, v případě báze Reaxys je situace o něco méně přehledná. Setkáváme se i s případy, kdy jsou zpracovány dva záznamy patentů z jedné patentové rodiny jak se stejným rozsahem excerptce, tak i bez excerptce nebo s rozsahem výrazně odlišným. Přehled zpracovaných patentů a excerptovaných dat je sumarizován v tab. I.

3. Excerptce chemických sloučenin z patentů a jejich analýza

Máme-li otevřený bibliografický záznam patentového dokumentu v bázi dat, můžeme bezprostředně zobrazit soubor všech sloučenin, které byly z daného dokumentu excerptovány. V programu SciFinder k tomu slouží tlačítko *Get Substances*, v bázi Reaxys pak nabídka *Show All Substances* v záznamu patentu. Tato data by měla představovat všechny chemické sloučeniny, které byly při zpracování daného dokumentu převedeny do báze dat jako chemické sloučeniny. Ty jsou uvedeny v tab. I pod databází SciFinder a Reaxys, v obou případech ve sloupcích označených *All*. I když z náhodného výběru nelze dělat kategorické závěry, zdá se, že mimořádně velké počty excerptovaných sloučenin se vyskytují navzdory obecnosti generických vzorců spíše výjimečně. Vzhledem k tomu, že se jedná o patenty téměř výhradně z oblasti farmaceutické chemie, nejsou počty excerptovaných sloučenin v řádu set překvapující, i když jsou samozřejmě vyšší, než v typickém časopiseckém článku. Porovnání počtu excerptovaných sloučenin v bázích SciFinder a Reaxys, což by měly být bez ohledu na možnost odlišných excerptních zásad v zásadě srovnatelné hodnoty, nevykazuje kromě prvních případů mimořádně velkých počtů systematické odchylky, někdy naopak přináší i překvapivé shody.

Podstatně zajímavější jsou možnosti přesnější specifikace excerptovaných sloučenin, které nabízí především báze SciFinder. Pod nabídkou *Get Substances* při otevřeném bibliografickém záznamu patentu můžeme požadovat zobrazení buď všech excerptovaných sloučenin, tedy *All*, nebo jen sloučenin zařazených do některé z 13 kategorií (v terminologii bází CAS se jedná o tzv. *Roles*). Pro posouzení informací o chemických sloučeninách obsažených v daném patentu byly vybrány kategorie *Preparation*, *Reactant or Reagent* a hlavně zajímavá kategorie *Prophetics in Patents*. Do první kategorie by měly být zařazeny sloučeniny, pro které je v daném patentu popsána jejich syntéza, což bývá hlavní důvod pro přihlášení patentu. Počty syntetizovaných sloučenin jsou podle očekávání ve všech případech nižší, rozdíl by se měl objevit ve sloupci *Reactants or Reagents*. Nemůžeme ale očekávat, že součtem sloupců *Preparation* a sloupců *Reagents or Reactants* dostaneme celkový počet sloučenin. Jako reaktanty mohou být, a evidentně také jsou, uvedeny meziprodukty, které jsou současně označeny jako sloučeniny syntetizované.

Nejzajímavější jsou samozřejmě data ve sloupci *Prophetics in Patents*. Počty sloučenin takto označených jsou ve většině případů výrazně nižší, ale nechybí ani patenty, ve kterých se počet všech sloučenin, nebo přípravených sloučenin od počtu *Prophetic* liší jen málo. Oficiální charakteristika těchto sloučenin v návodech práce s programem SciFinder je uvedena výše s odkazem na materiál CAS⁵. Příklady z konkrétních patentů ukázaly, že se jedná typicky o sloučeniny, které jsou v patentu přesně označeny, ať již vzorcem, nebo systematickým názvem (zpravidla obojím způsobem), ale jejich příprava je odkazována na obecnou metodiku více či méně vzdálenou od

syntéz uvedených v příkladech, nebo není uvedena vůbec. Nejsou pro ně tedy uvedena žádná konkrétní data, body tání, hustoty, spektra, výtěžky reakcí apod. Přesto jsou takové sloučeniny zahrnuty do báze REGISTRY, protože jsou *exemplified* a *specific* a tedy *de iure* existují. Ve většině případů studovaného souboru patentů byly sloučeniny označeny jako *Prophetic* současně zařazeny do kategorie *Preparation*, čímž lze vysvětlit, že se počty sloučenin uvedených v těchto dvou kategoriích příliš neliší. V každém případě ale kategorie *Prophetic* dává možnost udělat si představu, pro kolik sloučenin jsou uvedeny explicitní preparativní postupy a kolik sloučenin je k takto detailně uvedeným syntézám více či volněji přiřazeno a poskytuje tedy v programu SciFinder obecně užitečnou informaci.

Další možností, jak si udělat představu o rozsahu sloučenin s uvedenými podmínkami přípravy, jsou informace o reakcích uvedených v záznamu daného patentu, což dostaneme využitím nabídky *Get Reactions* v programu *SciFinder*. Velký rozdíl mezi excerpovanými sloučeninami a záznamem reakcí v patentu č. 1 v tabulce zřetelně indikuje, že příprava naprosté většiny sloučenin je přiřazena jen k několika konkrétním popisům experimentů, což je v souladu i s uvedeným počtem sloučenin v kategorii *Prophetics*. Je jen nutné upřesnit, že počty reakcí v programu SciFinder jsou kombinací všech individuálních reakčních kroků vícestupňových syntéz, neboli jednotlivé syntetické kroky jsou zaznamenány formou reakčního schématu jednak jako jednostupňové, ale postupně dále jako dvou-, tří-, čtyř- i vícestupňové až po celkové syntetické schéma zvlášť. Takže např. pětistupňová syntéza může být v programu SciFinder zaznamenána celkem 15 záznamy reakcí, čímž se vysvětlují velké počty záznamů konkrétních reakcí v některých patentech. Program SciFinder všechny jedno i vícestupňové reakce zobrazuje samostatně, báze Reaxys pracuje v současné době s podobnou koncepcí, ale vícestupňové reakce zobrazuje v prostředí aplikace *SynthesisPlans*. Mimořádné počty v případě patentu č. 2 v tabulce ilustrují nestandardizované formy formulace patentů; v daném případě jsou experimentální podmínky uvedeny sice obecně, ale v rozsáhlých tabulkách jsou konkrétně specifikovány reakční podmínky, jako jsou rozpouštědla, teploty, doby reakcí apod. Velmi užitečnou nedávno zavedenou informací jsou detailní experimentální podmínky uvedené jako součást záznamu patentu (obecně jakéhokoliv záznamu dokumentu obsahující chemické reakce). Program SciFinder je uvádí pod označením *Experimental conditions*, báze Reaxys pod nabídkou *Has preparation*. Ale opět závisí na úsudku excerpatora, co přijme jako detailní popis syntézy. Shodný počet reakcí a k nim náležejících experimentálních podmínek v patentu č. 2 naznačuje, že výše uvedený způsob popisu syntéz je považován za detailní podmínky. V každém případě ale tento údaj může eliminovat potřebu studovat často obsáhlé popisy patentů. V případě báze Reaxys je u převzatých experimentálních detailů užitečná informace uvedením stránky, nebo jiné formy lokalizace daného experimentálního popisu v textu patentu.

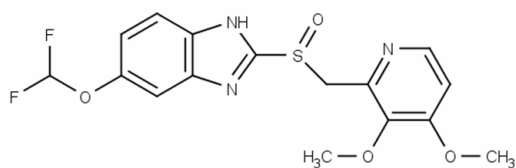
Báze Reaxys zatím kategorii *Prophetic* nemá; pokud je patent v bázi Reaxys zpracován jako záznam, jsou takové sloučeniny uvedeny v souboru *Substances*. Inspekce záznamů v bázi Reaxys ale ukázala, že jsou do této báze excerpovány i sloučeniny, pro které najdeme v patentech jen odkaz na obecnou metodu přípravy, nebo sloučeniny vyjmenované explicitně jako výčet patentových nároků, tedy v podstatě sloučeniny typu *Prophetics*. V zásadě se ukazuje, že překryv množin sloučenin excerpovaných z bázi SciFinder a Reaxys není, a z principu excerpční práce ani nemůže být stoprocentní, rozdíly se ale nezdají být zásadní.

4. Možnosti zadání strukturních rešerší v režimu *Markush*

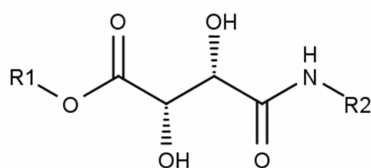
Generické neboli Markushovy vzorce jsou v každém případě záležitostí strukturního vyhledávání, a proto by nástroje označované jako *Markush Search*, nebo jen *Markush*, měly vycházet z jejich nakreslení ve strukturních editorech. To ale platí, navíc jen částečně, pouze pro strukturní editor programu SciFinder; tam sice najdeme nabídku na volbu kromě *Exact* a *Substructure* také *Markush Search*, ale pro zadání struktury můžeme využít jen nabídnuté alternativy substituentů, alternativních atomů, nebo velikosti kruhů. Báze Reaxys má stejné možnosti, ale nenabízí vyhledávací režim *Markush*. Jen poměrně „skrytě“ otevírá možnost zahrnout do výsledků rešerší i dokumenty s Markushovými vzorci⁷. Zatímco výsledek substrukturního vyhledávání je soubor struktur sloučenin vyhovujících zadání, výsledek hledání v režimu *Markush* je soubor bibliografických odkazů, z nichž některé obsahují očekávané Markushovy vzorce, ale setkáme se i s odkazy, kde v abstraktu uváděné vzorce mají spíše nepravděpodobnou podobu. Jak se má chemik v těchto výsledcích orientovat?

Především je nutné zdůraznit, že shoda mezi zadanou strukturou a strukturou obsaženou v patentu je založena na podobnosti generické nebo analogické, z nichž ani jeden termín nemá přesnou definici. Materiály CAS uvádějí, že vyhledávací režim *Markush* najde i sloučeniny podobné (*Similar*). Patenty nalezené na základě strukturní rešerše v režimu *Markush* by tedy měly obsahovat sloučeniny vyhovující těmto zásadám.

Pro vysvětlení nápadného nesouladu Markushových vzorců při prohlížení nalezených bibliografických záznamů je nutné dodat následující skutečnosti. Abstrakty bibliografických záznamů patentů typicky přebírají generické vzorce reprezentující sloučeniny v patentu chráněné pokud možno v jejich nejobecnější a z grafického hlediska také v nejrůznější podobě (viz např. obr. 2 až 7). V textu patentů se ale mohou vyskytovat další takové vzorce, které např. z důvodů přehlednosti reprezentují užší část celé množiny, anebo představují jiný důvod patentové ochrany, než jaký je uveden ve stručném textu abstraktu. Pokud se tyto skutečnosti ilustrovat na příkladu rešerše v režimu *Markush* pro známý farmaceutický preparát Pantoprazol (obr. 2). Zadáním tohoto vzorce do strukturního editoru

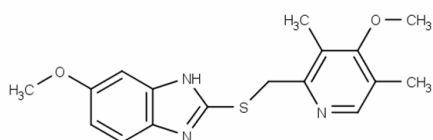


Obr. 2. Pantoprazol

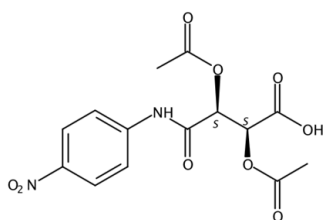


Obr. 3. Markušův vzorec v abstraktu patentu nalezeném na základě zadané struktury Pantoprazolu

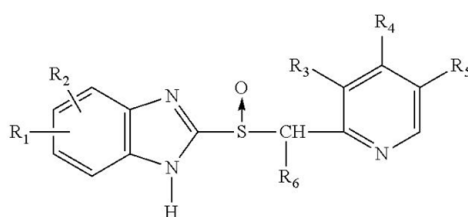
programu SciFinder a volby režimu *Markush Patents by Structure* (při volbě z rozbalené nabídky *More options* a *Explore by Structure* ve strukturním zobrazení) s alternativou *Substructure* dostaneme cca 176 odkazů na patenty, z nichž hned první čínský patent obsahuje jako součást abstraktu Markušův vzorec zcela se vymykající zadání (obr. 3). Zobrazíme-li si ale všechny sloučeniny, které byly excerpovány z tohoto patentu, nalezneme sloučeninu velmi blízkou zadané struktuře Pantoprazolu (obr. 4); vyhledávací algoritmus tedy funguje. Bližší inspekce pak ukáže, že Markušův vzorec, který je součástí



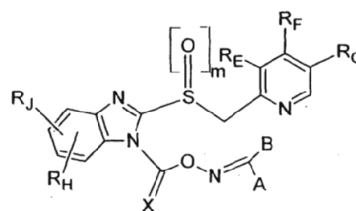
Obr. 4. Generická struktura sloučeniny nalezená jako generická varianta Pantoprazolu



Obr. 5. Struktura sloučeniny odpovídající Markušovu vzorci v patentu na obr. 3



Obr. 6. Typická podoba Markušova vzorce v abstraktu patentu nalezeného při strukturní rešerši analogu Pantoprazolu v režimu Markush



Obr. 7. Markušův vzorec v textu patentu s rozsáhlou variabilitou

abstraktu prvního čínského patentu (obr. 3), se vztahuje k jiným v patentu zmíněným sloučeninám, např. sloučenině na obr. 5. Zda sloučeniny odpovídající struktuře na obr. 3 nebo struktuře na obr. 5 jsou hlavní součástí patentových nároků, lze zjistit jen studiem textu patentu a nikoliv strukturní rešerší v režimu *Markush*. Strukturní rešerše v režimu *Markush* ale nalezne i odkazy na patenty, které obsahují Markušovy vzorce s velmi bohatou variabilitou substituentů nebo atomů (obr. 6 a 7).

Pro zachování autenticity nebyly vzorce 1 až 7 překreslovány, nýbrž byly vesměs zkopírovány z původních patentů.

Zpracování souboru bibliografických odkazů pro nalezení relevantních informací prostřednictvím Markušových vzorců je tedy možné buď individuální inspekcí souboru sloučenin excerpovaných v jednotlivých dokumentech, anebo lze zkusit selekci struktur pomocí nástroje *Get Substance* a následně *Refine* a *Chemical Structure*. Ovšem zde již máme k dispozici jen alternativu substrukturního vyhledávání, tedy omezení jen na modifikace zadané struktury.

Jak již bylo zmíněno, báze Reaxys nenabízí jiný vyhledávací režim než *Exact*, nebo *Substructure*. Ale na druhé straně pokud využijeme možnosti požádat o zařazení odkazů na patenty s Markušovým vzorcem⁷, najdeme v tomto souboru odkazů samostatné záznamy „sloučenin“, které jsou součástí takových vzorců. Takový záznam obsahuje strukturní vzorec tak, jak je uveden v patentech, doplněný výčtem substituentů a dalších strukturních složek ve

stejně podobě, tedy jako seznam R1, R2, ... Rx a další definice. Jinak řečeno, báze dat Reaxys registruje pro individuální Markushovy vzorce samostatná registrační čísla, samozřejmě jako *Reaxys Registry Numbers*, kterými doplňuje definice substituentů ve výše zmíněné podobě jejich seznamu. Takové registrační záznamy jsou označeny jako *Markush structure*. Jestliže tedy je výsledkem strukturální rešerše v bázi Reaxys jedna sloučenina v případě požadavku na *Exact* rešerši, nebo více sloučenin v případě režimu *Substructure*, pak pro každou takto nalezenou sloučeninu můžeme využít dalších možností práce s nalezeným záznamem (*Display further options and data*), kde najdeme alternativu *View related Markush*. Následně pak dostaneme seznam Markushových vzorců, které se vyskytují v souvislosti s danou strukturou s odkazem na bibliografickou citaci. Pokud pro danou strukturu se nabídka *View related Markush* nevyskytuje, znamená to, že nebyl nalezen žádný odkaz na patent obsahující relevantní generický vzorec.

5. Závěr

Lze konstatovat, že i přes nutně omezený soubor testovaných bibliografických záznamů patentů je možné formulovat určité poznatky týkající se rešeršní problematiky patentů obsahujících generické, tj. Markushovy vzorce. Především se v podstatě potvrzuje, že excerpce takových patentů do chemickýchází dat se řídí „rozumnými“ zásadami, totiž že jsou excerpovány sloučeniny, jejichž existence je nepochybná a potvrzená buď identifikačními daty, nebo i způsobem přípravy. Aspekt, který praktikujícím chemikům uniká, spočívá ve faktu, že sloučenina, v patentu jednoznačně identifikovaná buď vzorcem, nebo systematickým názvem, z právního hlediska jednoduše existuje bez ohledu na to, že můžeme mít pochybnosti o tom, že byla skutečně připravena a identifikována. Označení takových sloučenin termínem *Prophetic* použitým v programu SciFinder je nepochybně užitečné, i když jsou takové sloučeniny současně uvedeny jako připravené, což zpravidla vychází z odkazů na obecné metody syntéz. Využití různých kategorií pro chemické sloučeniny v programu SciFinder, včetně kategorie *Prophetic*, umožňuje udělat si alespoň přibližnou představu o charakteru daného patentu a v něm popsáných, resp. chráněných sloučenin. Některé takové možnosti vyplývají z dat v tabulce I. Naproti tomu strukturální rešerše v podobě zadání Markushových vzorců podle představ chemiků, jak by se zdálo z nabídky vyhledávacího režimu Markush, jsou poněkud zavádějící. Ve skutečnosti můžeme formulovat obecný vzorec jen jako vzorec pro substrukturální rešerši s využitím přednastavených substituentů a dalších strukturálních atributů a volba režimu Markush následně poskytne nikoliv jen sloučeniny substrukturálního charakteru, ale i sloučeniny, které je možné označit jako generické vzhledem k zadané struktuře. Je pochopitelné, že generický vztah závisí na

tom, jak je logika vyhledávání struktur naprogramována. Určitou nevýhodou je obtížná identifikace takových generických sloučenin, zvláště pokud celkový počet sloučenin excerpovaných z patentu je vysoký.

Rešeršní logika v bázi Reaxys je v tomto ohledu relativně bližší chápání syntetických chemiků, a to především v tom, že definice Markushova vzorce v patentu, resp. více takových upřesňujících vzorců v jednom patentu, je excerpována jako taková včetně definic substituentů Rx a dalších atributů a je zaznamenána jako samostatný záznam, neboli jako jedna „sloučenina“, tedy Markushova sloučenina. I když je následně zpracování souboru výsledků strukturálních rešerší se zájmem právě o generické sloučeniny opět náročnější, postup je přímočařejší a více odpovídá logice provádění strukturálních rešerší. Lze uzavřít, že nové nástroje v nejdůležitějších chemických bázích dat umožňující práci s Markushovými vzorci představují určitě užitečné rozšíření rešeršních možností chemických patentů, dosud vyhrazených spíše profesionálním patentovým rešeršérům.

Autoři by rádi poděkovali za finanční podporu Národnímu programu udržitelnosti I číslo LO-1215 (MŠMT – 34870/2013).

LITERATURA

1. Eugene A. Markush (1887) byl maďarský chemik, který přišel do Spojených Států před 1. světovou válkou, po válce založil firmu *Pharma Chemical Corporation*, která se stala později součástí koncernu Bayer.
2. Jedná se o US patent 1506316. Text originálního patentu jak v bázi dat Espacenet, tak v bázi US Patent and Trade Office žádný vzorec ani samostatné nákresy neobsahuje. Problematika generických vzorců ale vyplývá z popisu předmětu patentu: „*The process for manufacture of dyes which comprises coupling with a halogen-substituted pyralazone, a diazotized unsulphonated material selected from the group consisting of aniline, homologues of aniline, and halogen substitution products of aniline*” in August 1924.
3. http://www.stn-international.de/stn_content.html, staženo 8.9.2016.
4. <https://www.cas.org/content/markush>, staženo 8.9.2016.
5. <https://www.cas.org/content/prophetics>, staženo 8.9.2016.
6. Wagner A. Ben: *J. Chem. Inf. Model.* 46, 767 (2006).
7. Požadavek na zahrnutí odkazů na patenty s Markushovými vzorci můžeme v bázi Reaxys vyjádřit zaškrtnutím této možnosti po otevření nabídky *More options*.
8. Markush DARC User Manual, Derwent World Patents Index, Edition 1, 2008 The Thomson Corporation. ISBN: 978 1 905935 14 7
http://ip-science.thomsonreuters.com/m/pdfs/mgr/Markush_Darc_User_Manual.pdf, staženo 31.10.2016.

Pozn. redakce:

Existuje celá řada editorů chemických vzorců, které Markushovy vzorce nakreslí, ale pouze některé je správně interpretují. Mezi ty, které chápou tyto strukturní vzorce správně, patří m.j. programy, jako strukturní editor z databáze SciFinder (Chemical Abstracts Service), program ACD/ChemSketch (Advanced Chemical Development Inc.) či program MarvinSketch (ChemAxon Ltd.).

J. Šilhánek, E. Benešová, and P. Kačer
(Department of Organic Technology, University of Chemistry and Technology, Prague): **Empirical Testing of New Tools for Searching Generic (Markush) Structures in Main Chemistry Databases**

Generic or Markush structures in chemistry patents represent a very useful and convenient description of compounds claimed in patent documents. On the other hand, chemists face an uncertainty about a real existence of chemical compounds represented by a given Markush structure. Particularly for a synthetic chemist, generic or Markush structure brings uncertainty about which specific

compound corresponding to a generic formula was really synthesized and identified and which not (and is included for the maximum extent of patent claims). This problem should in first instance be solved by chemists creating bibliographic records for chemistry databases, primarily SciFinder and Reaxys. In case of a really synthesized compound, it is included in the existing list of known chemical compounds and receives the so called Registry number, most importantly the CAS RN. SciFinder database offers a chemist some help by setting up a new category of compounds with only partial description under the name *Prophetic in Patents*. Markush formulas represent a serious problem when we search for particular chemical compounds. Both chemical databases, SciFinder and Reaxys, recently offered new searching tools which make searching for generic compounds possible. The aim of this work is to test these searching tools on a random set of chemistry patents which contain both Markush formulas and compounds in the Prophetic category. Despite the limited set of patents tested, it was possible to formulate some conclusions about the applicability of these tools. Particularly the category *Prophetic in Patents* seems to be useful, as well as individual registration of Markush structures with relevant Reaxys Registry Number.