

Tak už známe v chemii 100 000 000 sloučenin!!!

V roce 1999 byl na tomto místě otištěn úvodník s mírně provokativním názvem „Kolik je vlastně chemických sloučenin?“ (Chem. Listy 93, 545 (1999)), který reflektoval v té době první zpřístupnění chemickýchází dat s logickou otázkou, kolik je v nich vlastně sloučenin. Tato otázka mohla být položena také díky koncepci registračních čísel CAS RN, což jsou de facto pořadová čísla chemických sloučenin přidělovaná novým sloučeninám chronologicky tak, jak na ně abstraktoři této instituce narazí v publikacích nebo patentech či jiných zdrojích. A jelikož tato čísla byla velmi rychle přidělena i všem sloučeninám zachyceným v rejstřících CA od r. 1907, mohli jsme se tak dozvědět, kolik těch sloučenin vlastně už známe, což v případě jejich seznamů v tištěných rejstřících nebylo zas tak snadné. Mimochodem, v letošním roce je to už 50 let od systematického přidělování těchto registračních čísel, což je nesporně významným přínosem CAS k rozvoji chemie jako vědní disciplíny.

Aby kolegyně a kolegové nemuseli hledat, kolik že těch sloučenin v r. 1999 (přesněji k datu 10.8.1999) tehdy bylo, zopakují tehdejší závěr:

Celkem je registrováno 20 433 827 chemických látek, z čehož je: 16 174 057 látek organických a anorganických a 4 259 770 tzv. „sequences“, neboli obecně biopolymerů.

Ve skupině „látek organických a anorganických“ jsou také polymery a řada dalších skupin, jejichž počty jsou ale vzhledem k počtu „klasických“ sloučenin málo významné.

Motivací pro zopakování tohoto úvodníku je skutečnost, že v červnu letošního roku dosáhl počet registrovaných chemických sloučenin čísla:

100 000 000, tedy rovných sto miliónů !!!

Je nutné zdůraznit, že se jedná o sto miliónů individuálních sloučenin v pravém slova smyslu, a nikoliv o registrační číslo 1000000-00-9 (poslední číslice je kontrolní), které sice asi bylo přiděleno, je ale odstraněno. Pokud by to někoho zajímalo, tak ona stomiliónová sloučenina má CAS RN 1786400-23-4 a jedná se o nijak mimořádně složitou strukturu ze skupiny benzodiazepinů publikovanou v patentu WO 2015081280 firmy Coferon, Inc. z USA pro aplikaci v případech akutní myeloidní leukémie. Důvod pro nesoulad mezi CAS RN a počtem sloučenin je způsobem upřesňováním vztahu mezi strukturou a CAS RN a tedy rušením čísel (Deleted CAS RN) s absolutním zákazem jejich opětovného použití.

Jestliže nyní srovnáme údaje o počtu sloučenin v roce 1999 se současným stavem, je zřejmé, že za uplynulých 16 let došlo k téměř neuvěřitelnému pětinasobnému nárůstu. Doporučuji podívat se na webovské stránky CAS, kde najdeme malý marketingový trik CAS, okno s číslem, jehož poslední číslice průběžně nepravidelně narástá. Každý nárůst znamená, že právě byla do báze REGISTRY přidána další nová sloučenina, neboli můžeme sledovat, jak nám sloučeniny přibývají v přímém přenosu a změřit si, kolik jich přibude za minutu, hodinu, den atd. Chemici, kteří ještě pamatují, jak se hledaly informace o existenci dané sloučeniny v tištěných rejstřících CA před pouhými cca 15 lety, si mohou zkusit představit, jak rozsáhlé by byly rejstříky názvů sloučenin dnes a jak bychom v nich hledali tu pravou ve sto miliónech jejich systematických názvů.

Na svých stránkách uvádí CAS při této příležitosti řadu dalších zajímavých čísel, např. že téměř 94 miliónů sloučenin obsahuje cyklické struktury, nebo že za celou dobu své existence je v bázích CA zpracována vědecká informace z oblasti chemie z cca 50 000 titulů vědeckých časopisů, tedy i zaniklých nebo vydávaných pod jinými názvy. Přesto, že excerpce chemických reakcí byla zahájena až v devadesátých letech v podobě báze CASREACT, díky zpětné registraci a dohodám o sdílení dat obsahuje tato báze dnes téměř 68 miliónů jedno a vícečetných reakčních schémat, plus cca 14 miliónů tzv. „half-reaction“ obsahujících jen pravou stranu s produkty. Neskrývejme ale skutečnost, že tato báze obsahuje jen reakce z oblasti organické a metaloorganické chemie a žádná reakce anorganických sloučenin.

Ti, kteří mají rádi prorocké předpovědi, se mohou při této příležitosti dočíst na stránkách CAS, že za dalších 50 let bude chemických sloučenin 650 miliónů !!! (trajektorie nárůstu viz Chem. Listy 101, 593 (2007)). A pokud si chemik nedovede představit, odkud se ty sloučeniny vezmou, pak doporučuji otevřít si program SciFinder a na nějakou velmi jednoduchou cyklickou strukturu náhodně umístit pár opět jednoduchých substituentů. A pokud ne hned napoprvé, pak určitě jen po několika pokusech zjistí, že taková velmi jednoduchá sloučenina v oněch sto miliónech jednoduše ještě není. Neboli chemikové mají do budoucna o práci spolehlivě postaráno.

Jaroslav Šilhánek