

## ODHADOVÁNÍ SYNTETICKÉ DOSTUPNOSTI PŘI POČÍTAČOVÉM NÁVRHU LÉČIV

MILAN VORŠILÁK<sup>a,b</sup> a DANIEL SVOZIL<sup>a,b</sup>

<sup>a</sup> CZ-OPENSREEN: Národní infrastruktura pro chemickou biologii, Ústav molekulární genetiky AV ČR v.v.i., Vídeňská 1083, 142 20 Praha 4, <sup>b</sup> CZ-OPENSREEN: Národní infrastruktura pro chemickou biologii, Laboratoř informatiky a chemie, Fakulta chemické technologie, Vysoká škola chemicko-technologická v Praze, Technická 5, 166 28 Praha 6  
Daniel.Svozil@vscht.cz

Došlo 8.8.17, přijato 6.10.17.

Klíčová slova: chemický prostor, syntetická dostupnost, retrosyntetická analýza, molekulární komplexita, virtuální screening, počítačový návrh léčiv

### Obsah

1. Úvod
2. Vnímání syntetické dostupnosti chemiky
3. Komplexita struktur
4. Retrosyntetická analýza
5. Dostupnost výchozích látek
6. Dostupný, nedostupný a obtížně dostupný chemický prostor
7. Závěr

### 1. Úvod

Nalezení nového léčiva je obecně složitý a dlouho trvající proces, provázený častěji neúspěchy než úspěchy. Jedním z možných scénářů při hledání nových léčiv je nalézt v chemické knihovně aktivní sloučeniny pomocí testování s vysokou propustností („high throughput screening“, HTS)<sup>1,2</sup>, ze kterého vzejde pro další zpracování seznam kandidátů seřazený podle afinity k testovanému molekulovému cíli. Podobně funguje i počítačový návrh léčiv, jen s tím rozdílem, že struktur může být řádově více a že se látky netestují experimentálně, ale v počítači pomocí technik zmíněných v souběžně publikovaných článcích<sup>3–5</sup>. Předpokládané vlastnosti v počítači vybraných struktur je ale vždy třeba experimentálně ověřit. Pro úspěšné experimentální testování je samozřejmě nutné navrženou strukturu syntetizovat, což nemusí být triviální. Proto je potřeba do prioritizace struktur také zahrnout odhad jejich syntetické dostupnosti. Je zřejmé, že snadnost přípravy je důležitým ukazatelem i u předem připravených nebo nakoupených chemických látek. Problém s dostupností u již připra-

vených látek se stává důležitým až při potřebě zvýšit produkci vybrané látky při důkladnějším testování v pozdějších fázích a dále při případné výrobě.

Syntetická dostupnost není exaktně měřitelná veličina, ale dá se spíše považovat za skrytou vlastnost sloučeniny a může korelovat s mnoha dalšími vlastnostmi. Koncept syntetické dostupnosti v sobě zahrnuje několik aspektů ovlivňujících přípravu organické sloučeniny. Sem patří např. dostupnost výchozích látek, složitost přípravy produktu odvíjející se od počtu reakčních kroků a jejich výtežnosti, či reaktivita a stabilita produktu. Navíc se syntetická dostupnost pro danou látku mění v čase, s tím, jak jsou v odborné literatuře publikovány nové a nové postupy organické syntézy. Látka, již nelze současnými postupy připravit v požadovaném množství či kvalitě, může být naopak v budoucnu syntetizována zcela rutinně. Nejednoznačné vnímání pojmu syntetická dostupnost dokresluje i fakt, že je popisována toliko na kvalitativní úrovni, kdy se struktury charakterizují jako např. hůře či snáze synteticky dostupné, realizovatelné, nebo „přívětivé“ k syntéze<sup>6–8</sup>.

### 2. Vnímání syntetické dostupnosti chemiky

Odhad syntetické dostupnosti jde v medicíně často ruku v ruce s odhadem „drug-likeness“ (tj. mohla by daná látka fungovat jako léčivo?) nebo „lead-likeness“ (tj. mohlo by být z dané látky, tzv. „leadu“, připraveno strukturálně optimalizované léčivo?)<sup>3</sup>, absence toxicity a metabolické stability, protože kandidáti bez těchto vlastností se také nehodí pro další výzkum<sup>9</sup>. Chemici, často se lišící ve svých odbornostech, se nezdívka neshodnou na strukturách, které se mají nebo naopak nemají vyřadit z důvodu jejich syntetické dostupnosti<sup>10</sup>. Navíc ani jednotlivec nemívá neměnný názor a často zařadí strukturu při novém hodnocení do jiné třídy syntetizovatelnosti<sup>11</sup>. Za účelem odstranění subjektivity hodnocení syntetizovatelnosti se proto počítá s konsenzuální hodnotou získanou od několika chemiků. Ovšem ani hodnocení více experty není dokonalé. Chemici se mohou při hodnocení řídit více instinktem než striktními pravidly. Navíc přeceňují vliv jednotlivých ukazatelů na své hodnocení a struktury tak hodnotí podle menšího počtu kritérií, než se domnívají<sup>12</sup>. Všechny tyto faktory mají vliv na výsledný odhad syntetizovatelnosti. Pokud se např. snadnost přípravy struktur vyjádří číselnou hodnotou, mají největší rozptyl odhadu ty, které mají průměrnou syntetizovatelnost<sup>13</sup>. Kromě praktického odhadu syntetické dostupnosti experty v organické či medicíně, který je časově a logisticky náročný, byla pro tento účel vyvinuta i celá řada počítačových metod. Počítačově se syntetická dostupnost struktury odhaduje pomocí tří následujících technik: první je výpočet kom-

plexity samotného produktu (tedy čím komplexnější struktura je, tím obtížnější je ji syntetizovat), druhou je pak retrosyntetická analýza (tedy čím větší počet syntetických kroků je k přípravě produktu zapotřebí, tím hůře je syntetizovatelný) a třetí je založena na hodnocení dostupnosti výchozích látek potřebných pro přípravu výsledného produktu.

### 3. Komplexita struktur

Molekulární komplexita je vlastnost, kterou je též obtížné definovat. Na rozdíl od syntetické dostupnosti je však pro člověka snadněji představitelná, lze ji posoudit přímo ze struktury a je v čase konstantní. Dá se očekávat, že složitější struktura bude i obtížněji syntetizovatelná. Toto pravidlo však není absolutní a některé reakce mohou z jednoduchých reaktantů vytvořit složité sloučeniny. Naopak některé zdánlivě jednoduché sloučeniny mohou být velmi obtížně připraveny. Molekulární komplexita však nehraje roli pouze v odhadu syntetické dostupnosti, ale je také spojována se selektivitou biologických cílů, s úspěšností klinických studií a s bezpečností látky samotné<sup>14</sup>.

Prvotní snahy o kvantitativní popsání strukturní komplexity vycházely z grafové reprezentace a topologie struktury<sup>6,15,16</sup>. Například Bertzův komplexitní index sestává ze dvou příspěvků: jedním z nich je topologie počítaná pomocí počtu podgrafů s dvěma hranami (počet „propanů“) a druhým je distribuce heteroatomů, při jejímž výpočtu se pracuje se Shannonovou entropií<sup>6</sup>. Dalším indexem, popisujícím složitost struktury pomocí matice konektivity, je tzv. „total walk count“ (TWC), který sčítá všechny možné průchody grafem s počtem hran až do počtu všech hran grafu. V TWC se odráží hlavně větvení struktury, ale nerozlišuje mezi typy atomů<sup>17</sup>.

Jiný přístup k hodnocení komplexity představuje Whitlockův index, který se nedrží striktně matematického popisu grafu, ale snaží se o intuitivní výpočet komplexity<sup>18</sup>. Ve Whitlockově indexu přispívají k celkové komplexitě počty kruhů, heteroatomů, nenasycených vazeb a chirálních center. Nedostatkem Whitlockova indexu je, že nestojně velké kruhy přispívají ke komplexitě stejnou měrou a uhlíkaté substituenty se nezapočítávají vůbec. Z tohoto důvodu vyvinul R. Barone index (Baroneho index), ve kterém každý atom přispívá podle své valence a každý kruh podle své velikosti<sup>19,20</sup>. Všechny výše uvedené komplexitní metriky mají tu nevýhodu, že významně korelují s velikostí struktury a tudíž i s molekulovou hmotností. Taková závislost je ovšem někdy nežádoucí, neboť např. substituce fluoru jódem reálnou komplexitu příliš nezvyšuje, ale vysoká atomová hmotnost jódu vede k velkému nárůstu některých komplexitních indexů. Toto omezení má překonat metrika SMCM (synthetic and molecular complexity)<sup>21</sup>, která hodnotí atomy a vazby mezi nimi podle jejich relativních elektronegativit vztažených k elektronegativitě atomu uhlíku. V SMCM indexu kruhy ke komplexitě přispívají a pokud jsou jiné než pěti- a šestičlenné, tak jejich příspěvek je dvojnásobný. SMCM kom-

plexitu navíc zvyšuje přítomnost spojených kruhů, spiro a můstkových atomů, zatímco předdefinované fragmenty komplexitu snižují. Dle našich zkušeností však ani SMCM index zcela neodstraní závislost na molekulové hmotnosti a pokud člověk potřebuje v analýzách tuto závislost eliminovat, jeví se jako nejvhodnější přístup rozdělit sloučeniny do košů (angl. bins) dle molekulové hmotnosti a analyzovat každý z košů nezávisle<sup>22</sup>.

Samotná grafová reprezentace a z ní vypočítaná komplexita však nemůže podchytit příspěvky všech možných fragmentů (např. řetězec 4 kyslíků). Proto P. Ertl z farmaceutické firmy Novartis navrhl SAScore<sup>13</sup>, které zohledňuje informaci o dostupnosti molekulových fragmentů v již syntetizovaných látkách a tak nepřímo i informaci o snadnosti syntézy. Při odvození SAScore postupoval autor tak, že analyzoval 1 milión náhodně vybraných struktur z veřejně dostupné databáze Pubchem<sup>23</sup>. Sloučeniny rozdělil na fragmenty a z jejich četnosti určil jejich příspěvky k výsledné syntetizovatelnosti. Hlavní myšlenka SAScore je ta, že čím častěji se fragment nachází v databázi, tím je snadněji syntetizovatelný, a naopak méně častý je hůře syntetizovatelný. Kromě fragmentů ještě do výsledného SAScore přispívají chirální, spiro a můstkové atomy, kruhy větší než devítičlenné a počet těžkých atomů. SAScore bylo později vylepšeno optimalizací numerických hodnot odrážejících důležitosti jednotlivých příspěvků<sup>24</sup>, které jsou původní Ertlové práci určené empiricky.

### 4. Retrosyntetická analýza

Průkopníkem v oblasti retrosyntetické analýzy a celkově návrhu syntéz je americký chemik Elias James Corey, za což byl odměněn Nobelovou cenou za chemii pro rok 1990. Základní principy retrosyntetické analýzy postuloval již téměř před 60 lety, kdy přišel s jednoduchou myšlenkou, že spíše než začínat návrh syntézy s předpokládanými výchozími látkami je vhodnější cílovou strukturu dekomponovat zpětně pomocí hypotetických transformací. Tyto transformace odpovídají zpětným syntetickým krokům, odtud tedy název retrosyntéza. Postupuje se od produktu k prekurzorům, na které se aplikují stejné metody, až se nakonec dojde k dostupným, většinou jednodušším, výchozím látkám. Pro použití zpětné reakce musí být ve struktuře nalezen retron, motiv spojený se syntetickou reakcí, na který může být tato aplikována. Například pro aldolovou transformaci musí struktura obsahovat podstrukturu HO-C-C-C=O. Těchto retronů může být ve struktuře několik, a proto nemusí retrosyntetický návrh postupovat jen jednou cestou, ale rozvětňuje se do stromu<sup>25</sup>.

Na konci 60. let začalo být využíváno možnosti počítačů při vytváření návrhů retrosyntézy a vznikl tak zcela nový obor „počítačový návrh organické syntézy“ (angl. computer-aided organic synthesis, CAOS). V té době vznikly programy OCSS<sup>26</sup> a následně LHASA<sup>27</sup> a SECS<sup>28</sup>. První systémy nenavrhovaly kompletní syntézu, ale inter-

aktivně nabízely chemikovi cesty, jakým směrem se vydat. Aby chemik mohl intuitivně zadávat chemickou strukturu, bylo tehdy nutné pro kreslení místo dnes běžné myši využít grafického tabletu s perem. Počítač nakreslenou strukturu převedl do strojově čitelné podoby – matic konektivit (spojovací tabulky) atomů a vazeb se stereochemií<sup>29</sup>. Matice mohla původně obsahovat jen/až 36 atomů bez započítání vodíků<sup>30</sup>.

Matice konektivit popisuje jen atomy a vazby mezi nimi, ale jejich analýzou se získávají znalosti o vyšších strukturních celcích, jako jsou funkční skupiny, kruhy, symetrie a aromaticita. Funkční skupiny významně ovlivňují návrh syntézy, a proto se klasifikují podle podobnosti v reaktivitě (např. nitro-, kyano- a karbonylovou skupinu spojuje kyselý vodík v alfa-poloze)<sup>26,27</sup>. Analytické nalezení kruhů patří také mezi důležité úkoly při počítačovém návrhu organické syntézy. Algoritmicky je snadné nalézt všechny kruhy, ale protože spojené a přemostěné kruhy mohou být detegovány vícekrát, je potřeba z nich vybrat ty synteticky významné a rozdělit je na reálné kruhy a pseudokruhy. Algoritmus SSSR (angl. smallest set of smallest rings) např. u bicyklo[2.2.1]heptanu (norbornan) nalezne jen pětičlenné kruhy, ale ignoruje potenciálně důležité kruhy šestičlenný, který je potřeba doplnit heuristicky<sup>30</sup>.

Jakmile jsou vztahy ve struktuře popsány, postupuje návrh do fáze strategického plánování, která se spoléhá na heuristická pravidla. Nejprve se odstraní reaktivní funkční skupiny a poté se aplikují složitější strategie, které musí buď explicitně zvolit uživatel, nebo jsou navrženy zakomponovanou heuristikou, jako je tomu např. u systému SYNCHEM<sup>31</sup>. Tyto strategie mohou cílit na potenciální výchozí látku, nebo vedou k výběru strategických kruhů a vazeb, které je možné snadno odštěpit<sup>32,33</sup>. Některé strategie také cílí na možnost aplikovat vzdálenější transformace, po nichž dojde k tak významnému zjednodušení struktury, že jejich aplikace se vyplatí i přes nutnost nejprve přidat do struktury chybějící retron<sup>34,35</sup>.

Celý retrosyntetický přístup vyžaduje databázi reakcí a výchozích látek. Pro zápis reakcí se využíval jazyk CHMTRN koncipovaný (s ohledem na snadnou použitelnost) chemiky, kteří původní databáze vytvářeli. Čím rozsáhlejší databáze transformací je, tím se zvyšuje pravděpodobnost vzniku extrémně hlubokých syntetických stromů. Aby k tomu nedocházelo, je nutné regulovat maximální hloubku stromů nebo opouštět větve s nevhodně navrženými strukturami. Například program RASA u hlubokých cest ukončuje hledání a přechází k výpočtu komplexity navržených koncových struktur<sup>36</sup>. Problematická je rovněž nutnost udržovat databázi transformací aktuální, proto některé přístupy místo transformací založených na skutečných reakcích používají transformace více abstraktní, které popisují výsledné strukturní změny, ovšem již neříkají, jak konkrétně se k těmto změnám synteticky dobrat. Při tomto formálním přístupu je ale kladen větší důraz na ukončování nevhodných větví, protože bez něj roste množství nalezených cest exponenciálně. Do této skupiny patří např. programy IGOR a EROS<sup>37,38</sup>.

Přestože bylo do počítačového návrhu organické syntézy investováno nemálo úsilí a vzniklo mnoho algoritmů a programů, nedošly tyto nikdy významnějšího uplatnění a rozšíření<sup>39</sup>. Naopak se v průběhu let rozšířilo využívání reakčních databází, jako jsou např. Reaxys či SciFinder<sup>40</sup>, do kterých jsou nejnovější chemické transformace přidávány velmi rychle, což je jeden z klíčových problémů CAOS. Proto nová generace programů jako ARChem nebo ICSYNTH přichází s napojením na reakční databáze, odkud se automaticky generují databáze transformací, které tak mohou být řádově větší než databáze vytvářené ručně<sup>41</sup>. S aplikovatelností retrosyntetického přístupu při virtuálním screeningu, v jehož průběhu se pracuje se stovkami tisíc či s miliony struktur, je spjat ovšem problém rychlosti retrosyntetických analýz. Ty lze totiž stěží označit za vysoce propustné, a tudíž jejich uplatnění při zhodnocení syntetické dostupnosti např. *de novo* navržených virtuálních struktur je velmi omezené.

## 5. Dostupnost výchozích látek

Z předchozí kapitoly vyplývá, že návrh kompletního syntetického postupu je velmi náročný. Pro úsporu výpočetního času a zdrojů je možné vypustit hledání návrhu syntetické cesty a zaměřit se jen na hledání podobnosti mezi navrhovanou strukturou a výchozími látkami shromážděnými ve vhodné databázi. K tomu se používá jak vyhledávání substrukturního (tj. nalézání přesných shod ve struktuře), tak i vyhledávání podobnostního<sup>42</sup>.

Hledání přesných shod se využívá zejména při použití abstraktní reprezentace struktury, kdy je např. zachována pouze kostra struktury bez rozlišení atomů a vazeb. Program SST<sup>43</sup> používá dvě hladiny abstrakce. Vyšší abstrakce popisuje jen skelet struktury a u nižší je navíc indikováno, kde jsou připojeny funkční skupiny. Testovaná struktura je nejdříve převedena do vyšší abstrakce, která je vyhledána v databázi stejně reprezentovaných výchozích látek. Pokud se povede nalézt více výchozích látek, je podle nižší abstrakce vybrána vhodnější látka. Na podobném principu funguje program SESAM s tím rozdílem, že se nezaměřuje na celou strukturu, ale tu rozdělí na menší části, které pak hledá v databázi výchozích látek<sup>44</sup>. Tyto aplikace ignorují stereochemii, která je ale u návrhu látek s chirálními centry klíčová.

Návrhem chirálních látek se zabývá program CHIRON<sup>45</sup>, který uplatňuje podobnostní vyhledávání. Při rozkladu navrhované struktury na fragmenty se snaží najít maximálně podobné látky s důrazem na uhlíkovou kostru, stereochemii a funkční skupiny. Funkční skupiny jsou posuzovány podle snadnosti své změny. Také kostra nemusí být rigidní, u kruhů s nenasyčenou vazbou se počítá s možností jejího rozštěpení. Podobnostní vyhledávání na rozdíl od exaktního zvyšuje množství nalezených výchozích látek, což je vykoupeno vyššími výpočetními nároky. Pro dosažení co nejlepších výsledků vyhledávání je důležité, podobně jako u retrosyntetického přístupu, udržovat databázi vstupních látek aktuální<sup>42</sup>.

Předchozí koncepty jsou často kombinovány. Nalezení potenciálního výchozího materiálu může být doprovázeno následným návrhem syntézy, jak získat testovanou strukturu, tedy opačně než v případě retrosyntetického přístupu. Takový systém je využíván např. programem WODCA<sup>39</sup>. Další kombinací přístupů nabízí program CAESA<sup>46</sup>, jenž rozděluje navrhovanou strukturu na části, které lze nalézt mezi výchozími látkami a které mezi nimi nalézt nelze. Nenalezené fragmenty přispívají k celkovému skóre svojí komplexitou. Program SYLVIA<sup>10</sup> přidává k podobnosti s výchozími látkami a ke komplexitě struktury též informaci o reakčních centrech, které jsou definovány v externí databázi. Tato reakční centra naznačují, kde dochází k syntéze, jedná se tedy o jakousi zjednodušenou obdobu retrosyntetického přístupu.

## 6. Dostupný, nedostupný a obtížně dostupný chemický prostor

Při počítačovém procházení chemického prostoru<sup>47</sup> se často pracuje se strukturami virtuálními, doposud nesyntetizovanými. Jeden ze způsobů návrhu virtuálních struktur je kombinování existujících molekulových fragmentů. To je výhodné i z hlediska syntetické dostupnosti navržené struktury, neboť volbou vhodných fragmentů je možno zvýšit šance na syntetizovatelnost výsledného produktu. Do hry ale mohou negativně vstoupit sterické efekty objemných rozvětvených fragmentů<sup>48</sup>. Při tvorbě úplně nových chemotypů při *de novo* návrhu se pak snadno dostaneme do situace, že některé struktury budou reálně nepřipravené. Pokud bychom byli schopni rozdělit chemický prostor na synteticky dostupný a nedostupný a uměli se pohybovat jen v tom dostupném, problém s virtuálním návrhem by se výrazně zjednodušil.

Část synteticky dostupného chemického prostoru je snadno přístupná z veřejných databází sloučenin<sup>49,23</sup>. Databáze komerčně dostupných sloučenin ZINC, kterou je možno považovat za největší databázi svého druhu, obsahuje (k červnu 2017) 360 milionů sloučenin. Nicméně synteticky dostupný chemický prostor se stále zvětšuje (např. ZINC se za 6 let zvětšil 10krát, cit.<sup>50</sup>), a proto není stále jasné, kde leží jeho hranice. Ty se snaží zmapovat počítačově generované databáze GDB, jež obsahují kompletní chemický prostor sestávající ze sloučenin omezených počtem těžkých atomů (tj. všech atomů těžších než vodík). Takto omezený prostor je i tak obrovský a např. databáze GDB-17 obsahuje 166 miliard všech možných struktur s maximálně 17 atomy uhlíku, dusíku, kyslíku nebo síry<sup>51,52</sup>. Databáze GDB nejsou však, bohužel, rozděleny na synteticky dostupný a nedostupný podprostor, jen jsou z nich odstraněny nestabilní kombinace atomů (enoly, některé vazby spojené heteroatomy, alleny).

Pro počítačovou klasifikaci struktur, např. pomocí metod strojového učení, na synteticky dostupné a nedostupné je potřeba mít k dispozici zástupce z obou kategorií. Synteticky dostupné struktury jsou snadno k dispozici ve veřejných databázích, zato pro nedostupné žádná taková

databáze neexistuje. Toto omezení se pokusila překonat skupina profesora Karypise, která navrhla<sup>53</sup> algoritmus pro hledání obtížně syntetizovatelných struktur. Vycházeli z toho, že známý chemický prostor není rovnoměrně osídlen a tudíž že sloučeniny z hustě okupovaného prostoru budou snadno synteticky dostupné, zatímco sloučeniny z řídkého prostoru obtížně. Hlavním nedostatkem tohoto přístupu je však skutečnost, že chemický prostor není invariantní vůči změně souřadného systému<sup>54,55</sup>. Jak bylo popsáno v jiném článku tohoto čísla Chemických listů<sup>47</sup>, chemický prostor je možno popsat v různých souřadných systémech (např. různé sady deskriptorů apod.) a vzdálenostní vztahy mezi strukturami, potřebné pro určení hustoty osídlení prostoru, se mezi jednotlivými souřadnými systémy mohou lišit i velmi výrazně. Tuto skutečnost obchází algoritmus Nonpher<sup>22</sup>, který využívá upraveného algoritmu molekulového morfování<sup>56</sup>. V molekulárním morfování se postupuje tak, že se výchozí struktura po etapách mění v cílovou. V každé etapě se z kandidátních struktur vygenerují nové a ty, které jsou nejpodobnější cíli, se vyberou jako kandidáti v další etapě. Generování nové struktury probíhá jednoduchou transformací jako přidej/uber atom/vazbu. V Nonpheru byl algoritmus molekulárního morfování upraven tak, že probíhá bez cílení na konkrétní strukturu a je tedy úplně náhodný. Kritériem pro ukončení morfování struktury je zvýšení její komplexity nad maximální hodnoty, které jsou odvozeny od podobně velkých syntetizovatelných struktur v databázi ZINC. Pomocí algoritmu Nonpher lze tedy vygenerovat libovolně velkou množinu obtížně syntetizovatelných struktur, které je možno využít v následných výpočetních studiích.

## 7. Závěr

Pro prioritizaci tisíců sloučenin se běžně používají počítačové metody, jež se snaží více či méně úspěšně předpovědět fyzikálně-chemické a biologické vlastnosti ligandů. Mezi takové vlastnosti patří např. drug-, lead- či NP-likeness (NP je označení pro přírodní látku, angl. natural product)<sup>57</sup>, afinita k molekulovému cíli, působení na jiné cíle (angl. off-target effects) nebo právě syntetická dostupnost. Syntetická dostupnost však není objektivně měřitelná veličina, ale její posouzení je čistě subjektivní záležitostí. Ruční posouzení syntetické dostupnosti je časově a logisticky náročný proces, neboť v jeho průběhu je třeba shromáždit, za účelem vytvoření konsenzuálního závěru, názory několika (mnoha) odborníků z oblasti organické či medicínské chemie. Tito experti běžně disponují různými znalostmi a zkušenostmi z oboru organické syntézy a jejich závěry ohledně syntetické dostupnosti té které sloučeniny bývají mnohdy dosti rozdílné a nejednotné. Metody počítačového odhadu syntetické dostupnosti, popsané v tomto článku, mají výhodu v tom, že jsou mnohem konzistentnější než manuální přístup, na druhou stranu je jejich největším nedostatkem skutečnost, že jejich predikce není možno srovnat s žádným pevně daným etalonem synteticky těžko dostupných látek. Existující metody počíta-

čového odhadu syntetické dostupnosti lze rozdělit do tří hlavních kategorií, které posuzují složitost syntézy produktu na základě následujících hledisek: složitost (komplexita) produktu, délka retrosyntetické cesty vedoucí k syntéze produktu a podobnost produktu s výchozími sloučeninami. Přes všechny své nevýhody se však počítačové metody odhadu syntetizovatelnosti organických sloučenin začínají čím dál tím více prosazovat v procesu počítačového návrhu léčiv, především díky vývoji metod a algoritmů pro generování virtuálních sloučenin pocházejících z dosud neprobádaných oblastí chemického prostoru.

*Tento článek vznikl za podpory MŠMT v rámci Národního programu udržitelnosti I projekt LO1220 (CZ-OPENSREEN).*

#### Seznam zkratek

HTS	testování s vysokou propustností
TWC	total walk count
SMCM	synthetic and molecular complexity
SSSR	smallest set of smallest rings
CAOS	computer-aided organic synthesis
NP	natural product

#### LITERATURA

- Popr M., Sedlák D., Bartůněk P.: Chem. Listy 111, 772 (2017).
- Muller T., Sedlák D., Bartůněk P.: Chem. Listy 111, 766 (2017).
- Svozil D.: Chem. Listy 111, 738 (2017).
- Škuta C., Svozil D.: Chem. Listy 111, 747 (2017).
- Šícho M., Svozil D.: Chem. Listy 111, 754 (2017).
- Bertz S. H.: J. Am. Chem. Soc. 103, 3599 (1981).
- Warr W. A.: Mol. Inf. 33, 469 (2014).
- Daina A., Michielin O., Zoete V.: Sci. Rep. 7, 1 (2017).
- Ajay, Walters W. P., Murcko M. A.: J. Med. Chem. 41, 3314 (1998).
- Bonnet P.: Eur. J. Med. Chem. 54, 679 (2012).
- Lajiness M. S., Maggiora G. M., Shanmugasundaram V.: J. Med. Chem. 47, 4891 (2004).
- Kutchukian P. S., Vasilyeva N. Y., Xu J., Lindvall M. K., Dillon M. P., Glick M., Coley J. D., Brooijmans N.: PLoS One 7, e48476 (2012).
- Ertl P., Schuffenhauer A.: J. Cheminf. 1, 1 (2009).
- Méndez-Lucio O., Medina-Franco J. L.: Drug Discovery Today 22, 120 (2017).
- Bonchev D., Mekenyan O., Trinajstić N.: Int. J. Quantum Chem. 17, 845 (1980).
- Randic M.: J. Am. Chem. Soc. 97, 6609 (1975).
- Ruecker G., Ruecker C.: J. Chem. Inf. Comput. Sci. 33, 683 (1993).
- Whitlock H. W.: J. Org. Chem. 63, 7982 (1998).
- Barone R., Chanon M.: J. Chem. Inf. Comput. Sci. 41, 269 (2001).
- Huang Q., Li L.-L., Yang S.-Y.: J. Mol. Graphics Modell. 28, 775 (2010).
- Allu T. K., Oprea T. I.: J. Chem. Inf. Model. 45, 1237 (2005).
- Voršilák M., Svozil D.: J. Cheminf. 9, 1 (2017).
- Kim S., Thiessen P. A., Bolton E. E., Chen J., Fu G., Gindulyte A., Han L., He J., He S., Shoemaker B. A., Wang J., Yu B., Zhang J., Bryant S. H.: Nucleic Acids Res. 44, D1202 (2015).
- Fukunishi Y., Kurosawa T., Mikami Y., Nakamura H.: J. Chem. Inf. Model. 54, 3259 (2014).
- Corey E. J.: Angew. Chem. Int. Ed. Engl. 30, 455 (1991).
- Corey E. J., Wipke W. T.: Science 166, 178 (1969).
- Pensak D. A., Corey E. J., v knize: *Computer-Assisted Organic Synthesis*, American Chemical Society, Washington D.C. 1977.
- Wipke W. T., Ouchi G. I., Krishnan S.: Artif. Intell. 11, 173 (1978).
- Jirát J., Svozil D.: Chem. Listy 111, 710 (2017).
- Todd M. H.: Chem. Soc. Rev. 34, 247 (2005).
- Gelernter H. L., Sanders A. F., Larsen D. L., Agarwal K. K., Boivie R. H., Spritzer G. A., Searleman J. E.: Science 197, 1041 (1977).
- Corey E. J., Perersson G. A.: J. Am. Chem. Soc. 94, 460 (1972).
- Corey E. J., Howe W. J., Orf H. W., Pensak D. A., Petersson G.: J. Am. Chem. Soc. 97, 6116 (1975).
- Cook A., Johnson A. P., Law J., Mirzazadeh M., Ravitz O., Simon A.: Wiley Interdiscip. Rev.: Comput. Mol. Sci. 2, 79 (2012).
- Corey E. J., Howe W. J., Pensak D. A.: J. Am. Chem. Soc. 96, 7724 (1974).
- Huang Q., Li L.-L., Yang S.-Y.: J. Chem. Inf. Model. 51, 2768 (2011).
- Ugi I., Bauer J., Bley K., Dengler A., Dietz A., Fontain E., Gruber B., Herges R., Knauer M., Reitsam K., Stein N.: Angew. Chem. Int. Ed. Engl. 32, 201 (1993).
- Gasteiger J., Jochum C., v knize: *Organic Compounds: Syntheses/ Stereochemistry/Reactivity*, Springer-Verlag, Berlin 1978.
- Ihlenfeldt W.-D., Gasteiger J.: Angew. Chem. Int. Ed. Engl. 34, 2613 (1996).
- Jindřich J.: Chem. Listy 111, 731 (2017).
- Ravitz O.: Drug Discovery Today: Technol. 10, e443 (2013).
- Baber J. C., Feher M.: Mini-Rev. Med. Chem. 4, 681 (2004).
- Wipke W. T., Rogers D.: J. Chem. Inf. Comput. Sci. 24, 71 (1984).
- Mehta G., Barone R., Chanon M.: Eur. J. Org. Chem. 1998, 1409 (1998).
- Hanessian S., Franco J., Larouche B.: Pure Appl. Chem. 62, 1887 (1990).
- Gillet V. J., Myatt G., Zsoldos Z., Johnson A. P.: Perspect. Drug Discovery Des. 3, 34 (1995).
- Čmelo I., Svozil D.: Chem. Listy 111, 724 (2017).

48. Douguet D., Munier-Lehmann H., Labesse G., Pochet S.: *J. Med. Chem.* **48**, 2457 (2005).
49. Sterling T., Irwin J. J.: *J. Chem. Inf. Model.* **55**, 2324 (2015).
50. Irwin J. J., Sterling T., Mysinger M. M., Bolstad E. S., Coleman R. G.: *J. Chem. Inf. Model.* **52**, 1757 (2012).
51. Fink T., Bruggesser H., Reymond J.-L.: *Angew. Chem. Int. Ed. Engl.* **44**, 1504 (2005).
52. Ruddigkeit L., van Deursen R., Blum L. C., Reymond J.-L.: *J. Chem. Inf. Model.* **52**, 2864 (2012).
53. Podolyan Y., Walters M. A., Karypis G.: *J. Chem. Inf. Model.* **50**, 979 (2010).
54. Shanmugasundaram V., Maggiora G. M., Lajiness M. S.: *J. Med. Chem.* **48**, 240 (2005).
55. Sheridan R. P., Kearsley S. K.: *Drug Discovery Today* **7**, 903 (2002).
56. Hoksza D., Skoda P., Vorsilak M., Svozil D.: *J. Cheminf.* **6**, 1 (2014).
57. Ertl P., Roggo S., Schuffenhauer A.: *J. Chem. Inf. Model.* **48**, 68 (2008).

**M. Voršilák<sup>a,b</sup> and D. Svozil<sup>a,b</sup>** (<sup>a</sup> CZ-OPENSREEN: National Infrastructure for Chemical Biology, Institute of Molecular Genetics of the Academy of Sciences of the Czech Republic, Prague, <sup>b</sup> CZ-OPENSREEN: National Infrastructure for Chemical Biology, Laboratory of Informatics and Chemistry, University of Chemistry and Technology, Prague): **Estimation of Synthetic Accessibility during Computational Drug Design**

Virtual screening makes it possible to design and then test up to millions of structures. On the other hand, not every virtual structure can readily be prepared and tested in reality. Therefore, it is necessary to exclude unsuitable candidates, ideally at the beginning of virtual screening campaign. Three main approaches are used to filter out synthetically unfeasible structures: similarity to starting material, retrosynthetic route and complexity of the structure. These techniques vary in computational demands; they can therefore be used in various stages of screening, and/or combined to improve estimation of synthetic feasibility.