

## MOLEKULOVÉ DOKOVÁNÍ JAKO NÁSTROJ PRO VIRTUÁLNÍ NÁVRH LÉČIV

MARTIN ŠÍCHO<sup>a,b</sup>, 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  
Martin.Sicho@vscht.cz, Daniel.Svozil@vscht.cz

Došlo 24.7.17, přijato 10.10.17.

Klíčová slova: molekulové dokování, počítačový návrh léčiv, strukturní metody návrhu léčiv, virtuální screening

## Obsah

1. Úvod
2. Základní principy molekulového dokování
  - 2.1. Konformační vyhledávání
  - 2.2. Skórování
    - 2.2.1. Fyzikální skórovací funkce
    - 2.2.2. Empirické skórovací funkce
3. Aktuální vývoj a témata výzkumu v molekulovém dokování
4. Závěr

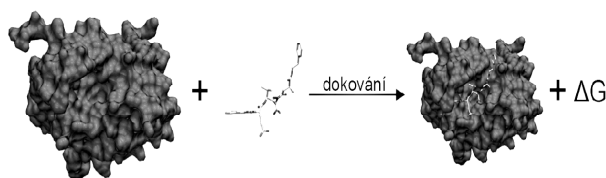
## 1. Úvod

V posledních několika dekádách došlo díky významným pokrokům v medicíně, genomice a chemické a strukturní biologii k nahromadění velkého objemu dat, která zachycují vlastnosti nejrůznějších chemických látek v širokém biologickém kontextu. Například díky rozvoji vysoce výkonných experimentálních metod, jako je kombinatoriální chemie a HTS (high-throughput screening), je možné rychle syntetizovat i tisíce látek a stanovit jejich farmakodynamické vlastnosti ve velice krátkém čase<sup>1</sup>. Navíc lze díky NMR<sup>2</sup> a rentgenové krystalografii<sup>3</sup> získat trojrozměrné struktury proteinů a nukleových kyselin a udělat si tak představu o jejich dynamickém chování<sup>4</sup>. Klíčovou volně dostupnou databází, shromažďující strukturní data biomakromolekul, je Protein Data Bank<sup>5</sup> (PDB), která v době vytvoření tohoto článku (červen 2017) obsahovala přes 130 000 vyřešených struktur nukleových kyselin, proteinů a jejich komplexů. Data týkající se interakce malých ligandů s biomakromolekulami jsou bohatým

zdrojem informací použitelných ve strukturních metodách návrhu léčiv (structure-based drug design, SBDD)<sup>6–8</sup>.

Hlavní metodou strukturního návrhu léčiv je molekulové dokování, které by se dalo definovat jako počítačové modelování interakce (v drtivé většině případů nevazebné) mezi dvěma chemickými látkami (nejčastěji mezi nízkomolekulárním ligandem a proteinem) za účelem predikce konformace rovnovážného stavu a do jisté míry i změny volné vazebné energie (obr. 1). Schopnost přesně a rychle vypočítat změnu volné vazebné energie při vzniku komplexu ligand-protein by znamenala revoluci, neboť by bylo možné virtuálně testovat (viz jiný článek v tomto čísle Chemických listů) milióny sloučenin ještě před finančně a časově náročnými laboratorními experimenty. Tento cíl je však i za použití stávajících výpočetních prostředků velmi vzdálen, neboť modelování fyzikálně-chemických dějů při tvorbě komplexu ligand-protein<sup>9,10</sup> je velice obtížné. I přesto se molekulové dokování od svého vzniku na přelomu 70. a 80. let minulého století vyvinulo od poměrně jednoduchých počítačových simulací v komplexní disciplíně, která i přes mnohé nedostatky pomáhá zefektivnit vývoj léčiv a dalších látek<sup>11</sup>.

Tento článek si klade za cíl přiblížit čtenáři molekulové dokování, ale nejedná se o kompletní rešerši, neboť tato oblast je velice široká a není možné ji v tomto formátu dostatečně podrobně pokrýt. V naší diskusi se tedy omezíme pouze na hlavní témata týkající se predikce vzniku komplexu biomolekula-ligand a na některé současné směry ve vývoji molekulového dokování. Pro detailnější náhled do aktuálního stavu v této oblasti lze vřele doporučit pravidelně vycházející články Yurievové a spol.<sup>12–14</sup> nebo některou z dalších nedávných rešerší<sup>15–22</sup>. Pro podrobnou diskusi dostupného software lze doporučit nedávnou práci Pagadala a spol.<sup>23</sup>. My zde pouze zmíníme, že mnoho programů je možné pro akademické nebo nekomerční účely získat zcela zdarma – např. AutoDock Vina<sup>24</sup> (open sour-



Obr. 1. Schéma ilustrující dva problémy nejčastěji řešené při molekulovém dokování, tj. predikce struktury komplexu mezi nízkomolekulárním ligandem a proteinem a predikce změny volné vazebné energie tohoto procesu ( $\Delta G$ ). Zdroj: [https://en.wikipedia.org/wiki/File:Docking\\_representation\\_2.png](https://en.wikipedia.org/wiki/File:Docking_representation_2.png)

ce), UCSF DOCK<sup>25</sup> (neplacená akademická licence) nebo HADDOCK<sup>26</sup> (zdarma pro nekomerční použití).

## 2. Základní principy molekulového dokování

Před dokováním samotným je často třeba vyřešit řadu problémů, jako je vhodná reprezentace proteinu a ligandu včetně tautomerů, stereoisomerů a protonačního stavu při fyziologickém pH (cit.<sup>27,28</sup>). Především volba vhodné reprezentace proteinu představuje nelehký úkol, neboť pro jeden protein je často k dispozici i více struktur různé kvality a je tedy třeba se rozhodnout, které struktury budou pro modelování nejlepší. Krom toho je někdy též třeba explicitně zahrnout některé molekuly vody, které mohou při vytvoření komplexu hrát významnou roli<sup>29</sup>. Po výběru vhodné reprezentace již zpravidla následuje samotné dokování, které by se dalo konceptuálně rozdělit na dvě fáze: (1) konformační vyhledávání (conformational sampling) a (2) skórování (scoring).

### 2.1. Konformační vyhledávání

V této fázi se algoritmy snaží o co nejkompletnější pokrytí možných konformací obou partnerů ve struktuře komplexu. Ve většině případů to znamená zavedení neurůznějších pravidel a heuristických omezení, neboť vzhledem k počtu možných řešení nelze postihnout všechny možné vzájemné konfigurace biomolekuly a ligandu, tzv. pózy. Tato omezení samozřejmě mohou vést k neschopnosti algoritmu nalézt konfiguraci, která skutečně odpovídá minimálnímu energetickému stavu, ale i tak je většina dnešních algoritmů v tomto relativně úspěšná<sup>13</sup>.

Pro vytvoření množiny možných póz existuje mnoho přístupů, které lze obecně rozdělit do dvou hlavních skupin<sup>30,31</sup>: (1) systematické a (2) stochastické. Systematické metody většinou postupným prozkoumáváním jednotlivých stupňů volnosti systematicky generují jednotlivá dílčí řešení, která se postupně podrobují detailnější optimalizaci směrem ke stavu s nejnižší energií. Oproti tomu při stochastickém vyhledávání se zpočátku náhodně vygeneruje několik konfigurací, ze kterých se pak na základě různých pravděpodobnostních pravidel postupně vybírají energeticky nejvýhodnější pózy. Jako příklad systematického vyhledávání lze uvést program Glide<sup>32</sup> nebo FlexX<sup>33</sup>. Stochastické metody jsou pak implementovány např. v programech GOLD<sup>34</sup> a PLANTS<sup>35,36</sup>.

### 2.2. Skórování

V molekulovém dokování hrají významnou roli tzv. skórovací funkce (scoring functions), které jsou zodpovědné za identifikaci pózy, která je z termodynamického hlediska nejpravděpodobnější. Toto je však též Achillovou patou molekulového dokování. Přesné a rychlé vyhodnocení vazebné energie jednotlivých konfigurací je totiž díky komplexním termodynamickým dějům<sup>9,10</sup> probíhajících při

interakci proteinu a ligandu velice složité. Níže uvádíme nejdůležitější vztahy, které popisují vznik komplexu, a jejich stručný popis.

Změnu Gibbsovy volné energie při přechodu mezi vázanou a nevázanou formou proteinu a ligandu je možné vyjádřit vztahem:

$$\Delta G = -RT \ln \frac{[PL]}{[P][L]} \quad (1)$$

kde  $\Delta G$  je změna Gibbsovy energie,  $R$  je univerzální plynová konstanta,  $T$  je teplota systému,  $[PL]$  je koncentrace komplexu proteinu-ligand a  $[P]$  a  $[L]$  potom koncentrace partnerů mimo komplex. Gibbsova volná vazebná energie vyjadřuje celkový termodynamický potenciál systému pro vytvoření komplexu protein-ligand a lze ji stanovit experimentálně. Tvorba komplexu probíhá spontánně, pokud je  $\Delta G$  záporné (systém se posune do stavu s nižší absolutní energií).

Změnu volné vazebné energie lze též vyjádřit jako kombinaci příspěvků podle termodynamického vztahu:

$$\Delta G = \Delta H - T\Delta S \quad (2)$$

kde  $\Delta G$  je změna Gibbsovy energie,  $\Delta H$  je změna entalpie,  $T$  je opět teplota systému a  $\Delta S$  je změna entropie. Entalpie je globální vlastnost celého systému a mimo změny energie způsobených nevazebnými interakcemi mezi ligandem a proteinem (např. vodíkové vazby nebo van der Waalsovy interakce) zahrnuje i změny, které jsou vyvolané interakcí s rozpouštědlem. Při navázání ligandu totiž dochází k přerušení vodíkových vazeb mezi proteinem a molekulami vody, které se ve vazebném místě nachází. Přerušení těchto vazeb je většinou při vytvoření komplexu nevýhodné a musí být vykompenzováno jinými interakcemi, jako je např. vytvoření obdobných nevazebných interakcí s molekulou ligandu. Entropie představuje úroveň neuspořádanosti systému, která vznikne přesouváním tepla (podle druhého zákona termodynamiky) z míst s vyšší teplotou do míst s nižší teplotou. K entropii systému přispívá několik složek, které zahrnují mimo jiné např. ztrátu rotačních a translačních stupňů volnosti systému při vytvoření komplexu nebo solvatační entropii, která zahrnuje změnu vyvolanou zanořením části povrchu ligandu do vazebného místa (snížení plochy na povrchu ligandu, která je v nevázaném stavu v kontaktu s rozpouštědlem).

Odhadnout důležitost a velikost entalpických a entropických příspěvků je velmi obtížné a často i specifické pro konkrétní systém<sup>10</sup>. V současné době neexistuje žádná univerzální skórovací funkce, která by dokázala tento problém vyřešit. V literatuře již bylo publikováno více než 100 skórovacích funkcí, které lze podle jejich charakteru zhruba rozdělit na čtyři základní skupiny<sup>20</sup>: (1) fyzikální, (2) empirické, (3) deskriptorové a (4) znalostního potenciálu. V následujících sekcích se budeme podrobněji věnovat pouze fyzikálním a empirickým funkcím, neboť tyto dvě skupiny reprezentují dva hlavní směry skórovacích přístupů. Podrobnější rozbor a příklady deskriptorových

funkcí a funkcí znalostního potenciálu lze nalézt v literatuře<sup>37–39</sup>.

### 2.1.1. Fyzikální skórovací funkce

Fyzikální funkce se v literatuře dlouho označovaly jako funkce na principu silových polí (force-field based functions), neboť většina z nich stanovuje energii konkrétní pózy podle funkčního předpisu, který vychází z klasické mechaniky. Například pro jedno z nejpoužívanějších silových polí, AMBER<sup>40</sup>, má tento předpis tvar:

$$E = \Delta E_{\text{vazby}} + \Delta E_{\text{úhly}} + \Delta E_{\text{dihed}} + \Delta E_{\text{vdW}} + \Delta E_{\text{estat}} \quad (3)$$

kde  $E$  je celková energie pózy a  $\Delta E_{\text{vazby}}$ ,  $\Delta E_{\text{úhly}}$ ,  $\Delta E_{\text{dihed}}$ ,  $\Delta E_{\text{vdW}}$  a  $\Delta E_{\text{estat}}$  jsou sumy energetických příspěvků z délek jednotlivých vazeb, velikostí vazebných úhlů, rotace kolem dihedrálních úhlů, van der Waalsových interakcí mezi jednotlivými páry atomů a elektrostatických interakcí, opět mezi jednotlivými atomovými páry. Například pod  $\Delta E_{\text{vazby}}$  se v tomto předpisu skrývá model vazby jako harmonického oscilátoru definovaného předpisem  $\sum_{\text{vazby}} K_r (r - r_{\text{eq}})^2$ , kde  $K_r$  je tuhost pružiny (pevnost vazby),  $r$  je skutečná vzdálenost atomů v dané póze a  $r_{\text{eq}}$  je tzv. rovnovážná vzdálenost.  $K_r$  a  $r_{\text{eq}}$  jsou parametrizované pro každý typ vazby a vazebných partnerů na základě experimentálních dat<sup>40</sup>.

V dnešní době existuje mnoho fyzikálních skórovacích funkcí s různými předpisy, které mohou např. obsahovat speciální úpravy pro vodíkové vazby nebo další energetické příspěvky, především k přesnějšímu zahrnutí vlivu rozpouštědla a konformační flexibility ligandu a proteinu. Velice přesnou metodou pro odhad volné vazebné energie je tzv. perturbace volné energie (free energy perturbation, FEP), kde se silové pole používá k popisu energetiky ve spojení s podrobným prozkoumáním flexibility obou partnerů pomocí metody Monte Carlo nebo molekulové dynamiky<sup>41</sup>. FEP sice umožňuje stanovení volné vazebné energie s chybou menší než 1 kcal mol<sup>-1</sup> (cit.<sup>42</sup>), avšak tato metoda je výpočetně velice náročná a její využití v SBVS je výrazně omezené. Obdobou FEP je i metoda tzv. „lineární interakční energie“ (linear interaction energy, LIE), která opět využívá metod Monte Carlo nebo simulací molekulové dynamiky, ale k popisu energetiky slouží lineární vztah s empirickým členem<sup>43</sup>. Mezi další často používané metody využívající popisu energetiky pomocí silových polí patří MM-PBSA (molekulová mechanika s implicitním modelem solvatace podle Poissona a Boltzmannova, Poisson-Boltzmann Surface Area) a MM-GBSA (molekulová mechanika s implicitním modelem solvatace podle Bornova generalizovaného modelu, Generalized Born Surface Area)<sup>44</sup>, které modelují rozpouštědlo jako kontinuální dielektrikum a zavádějí oproti FEP a LIE do výpočtu další aproximace<sup>42</sup>. Metody LIE a MM-PBSA/MM-GBSA sice nabízejí o něco větší rychlost než FEP, ale možnost jejich využití pro plnohodnotný SBVS je stále výrazně limitovaná<sup>21</sup> a většinou se používají pouze jako nástroj pro validaci a podrobnější interpretaci výsledků získaných dokováními<sup>44,13</sup>. S těmito postupy též souvisí rapidní pokrok v oblasti hardware, díky kterému se v poslední době staly mnohem dostupnějšími<sup>45</sup>.

Díky svému teoretickému ukotvení mohou fyzikální skórovací funkce a metody stanovení volné vazebné energie dosahovat vysoké přesnosti. Především metody kombinující molekulovou mechaniku a semiempirické kvantové výpočty<sup>46,47</sup> by mohly v budoucnu pomoci výrazně zlepšit popis jednotlivých interakcí. Experimenty s těmito komplexnějšími fyzikálními přístupy však zatím nevedly k jednoznačně lepším výsledkům, než je použití FEP<sup>46,47</sup>, a jejich výpočetní náročnost je činí méně praktické pro SBVS ve velkém měřítku.

### 2.1.2. Empirické skórovací funkce

Empirické skórovací funkce představují další stupeň aproximace změny Gibbsovy volné energie ( $\Delta G$ ), která je vyjádřena vztahem:

$$\Delta G = \sum_i W_i \cdot \Delta G_i \quad (4)$$

kde  $\Delta G_i$  jsou jednotlivé příspěvky ke změně volné energie a  $W_i$  jsou jejich váhy, které se nastavují pomocí vícenásobné lineární regrese na základě experimentálních dat získaných zkoumáním afinity již známých ligandů k nejrůznějším molekulovým cílům. Typickým příkladem empirické skórovací funkce je např. často používaná ChemScore<sup>48</sup>.

Největší nevýhodou empirických funkcí je jejich závislost na experimentálních datech, která jsou často nevyvážená vzhledem k různým proteinovým rodinám a obsahují výsledky pro ligandy, které pokrývají jen omezenou část chemického prostoru. Vytvořit tedy empirickou skórovací funkci, která by byla dostatečně obecná a fungovala pro všechny systémy, je velice obtížné. Na druhou stranu široké možnosti parametrizace empirických funkcí umožňují odvodit jednoduché vztahy, které jsou sice specifické pro konkrétní systém, ale v daném kontextu mohou fungovat velice dobře a rychle.

## 3. Aktuální vývoj a témata výzkumu v molekulovém dokování

V této sekci se zaměříme na některé nejintenzivněji prozkoumávané oblasti dokování a na nové výzvy, které tuto oblast čekají v nadcházejících letech. Těmi jsou především správné zahrnutí flexibility proteinu<sup>49,50</sup>, modelování molekul vody, dostatečná validace nových i stávajících přístupů<sup>51</sup>, celkové zvyšování škálovatelnosti a kvality SBVS<sup>13</sup>, větší integrace dokování s experimentálními metodami<sup>16</sup>, alternativní molekulové cíle (např. nukleové kyseliny nebo proteinová rozhraní)<sup>19,52</sup>, reverzní<sup>22</sup> a kovalentní dokování<sup>18</sup> a dokování molekulových fragmentů<sup>13,21</sup>. Všechna tato témata zde bohužel nejde kompletně pokrýt, ale pokusíme se zde shrnout alespoň ta nejdůležitější z nich.

Asi největšími dvěma problémy současných dokovacích metod je jak přistupovat k flexibilitě proteinu a k modelování vlivu vody ve vazebném místě. Kvůli úspoře času se často při modelování flexibility berou

v potaz pouze rezidua ve vazebném místě a různé programy k němu přistupují různě (např. konceptuálně poměrně rozdílná jsou řešení v programech AutoDock<sup>53</sup> a MedusaDock<sup>54</sup>). Kompletní flexibilitu proteinu lze do dokovacího procesu zahrnout dokováním do skupiny konformací proteinu (ensemble docking<sup>50</sup>), které jsou buď zjištěné experimentálně, nebo vycházejí ze simulace molekulové dynamiky. Právě skloubení simulací molekulové dynamiky a dokování<sup>55</sup> je v dnešní době velice atraktivní a díky zvyšujícímu se výkonu moderního hardware též dostupnější.

Možná ještě složitější problém než flexibilita je modelování molekul vody, které mají především vliv na změnu solvatační entropie, ale mohou se i přímo podílet na vazbě mezi ligandem a proteinem<sup>9</sup>. Existuje velké množství různých metodologií, které se soustředí na různé aspekty tohoto problému<sup>29,13</sup>. Patří sem jak implicitní modely vody, které na ni nahlíží jako na kontinuální dielektrikum<sup>44</sup>, tak i metody, které modelují vodu explicitně<sup>13</sup>. Jako příklad lze uvést přístup, který je součástí programu FlexX<sup>33,56</sup>, nebo o něco komplexnější WScore<sup>57</sup>. Vzhledem k tomu, že se voda v různých systémech na vazbě podílí v různé míře, je velice těžké přijít s opravdu obecným řešením<sup>58</sup>. I tato oblast tedy stále zůstává otevřenou kapitolou a bude i v následujících letech vyžadovat zvláštní pozornost.

Je třeba též zmínit, že molekulovým dokováním lze řešit i jiné problémy než interakci proteinu a nízkomolekulárního ligandu. Jako příklad lze uvést dokování fragmentů, které souvisí s fragmentovými metodami návrhu léčiv (fragment-based drug discovery, FBDD<sup>59</sup>), při kterém se studuje afinita molekul menších než běžné ligandy (zpravidla látek menších než 250 Da, cit.<sup>21</sup>). Cílem je pak spojit tyto fragmenty ve větší celky, které mají mnohem větší biologický účinek. K dokování fragmentů lze sice přistupovat stejně jako k dokování větších celků, je ovšem třeba vzít v potaz také jistá specifika<sup>60</sup>. Zejména se jedná o nižší úspěšnost identifikace správné pózy, neboť rozdíly ve vazebných energiích různých vazebných módů fragmentů jsou mnohem menší, a proto znamenají pro klasické skórovací funkce, které byly původně vyvinuté a otestované pro dokování kompletních struktur, větší výzvu. Shrnutí některých nedávných vylepšení fragmentového dokování lze nalézt v článku Yurievové a spol.<sup>13</sup>

Dalším zajímavým cílem je též rozhraní protein-protein<sup>19</sup>. Protože většina signálů mezi buňkami se přenáší pomocí interakcí mezi proteiny, je tento přístup k návrhu nových léčiv velice atraktivní. Při dokování molekul do tohoto rozhraní je však třeba navrhnout strukturu komplexu protein-protein, která většinou není z experimentálních dat dostupná. Vzhledem k velikosti obou partnerů je nemožné vyzkoušet a energeticky vyhodnotit všechny možné vzájemné orientace nebo vzít do úvahy, třeba i jen částečně, flexibilitu obou partnerů. I přesto je počítačových nástrojů pro podporu tohoto typu dokování poměrně mnoho a jedním z kompletních řešení je např. již dříve zmíněný program HADDOCK<sup>26</sup>.

Mimo proteiny lze též cílit i na nukleové kyseliny včetně RNA<sup>52</sup>. Například proces translace v prokaryotních

buňkách je regulován pomocí úseků nekódující RNA, které lze použít jako cíl nejrůznějších antibiotik<sup>61</sup>. Jelikož je RNA velice plastická, je zahrnutí flexibility při dokování velice důležité. Je též třeba dobře zvolit i skórovací funkci, neboť záporně nabitě fosfáty vyžadují zahrnutí dalekosáhlých elektrostatických interakcí a díky své větší chemické diverzitě formují i řadu velice specifických interakcí<sup>62</sup>. Dokovat nukleové kyseliny lze např. ve volně dostupném programu rDOCK<sup>63</sup>.

V poslední době se značná pozornost věnuje též modelování kovalentní vazby, neboť kovalentní inhibitory se vyznačují vysokou účinností a dlouhotrvajícím účinkem<sup>64</sup>. Jedním z příkladů implementace kovalentního dokování a momentálně asi nejkompletnějším a nejdostupnějším řešením je program CovalentDock<sup>65,66</sup>, který příspěvek kovalentní vazby vyjadřuje pomocí Morseho potenciálu. Kovalentní dokování je též součástí komerčního software Schrödinger ve formě modulu CovDock, který oproti standardnímu dokování nabízí několik úprav jak pro identifikaci reziduí, které lze spojit kovalentní vazbou, tak pro výpočet skóre<sup>67,68</sup>.

Zajímavou aplikací je též reverzní dokování (reverse/inverse docking), tedy dokování jedné látky do spektra různých cílů. Tato metodologie má především význam při vytváření toxického profilu látek a nalezení nového terapeutického využití již známých léčiv<sup>22,69,70</sup>. Největší komplikací při reverzním dokování jsou nekonzistentní výsledky dokovacích programů pro různé proteiny a proteinové rodiny<sup>51</sup>. Je to především kvůli různé míře zahrnutí flexibility a skórovacím funkcím, které nejsou dostatečně obecné a dosahují různé kvality pro různé systémy.

Na závěr této kapitoly je též třeba zmínit význam porovnávání a validace nových a stávajících přístupů pro identifikaci silných a slabých stránek současných nástrojů v molekulovém dokování. K tomuto účelu slouží validační datové množiny sestávající jak ze známých ligandů s krystalograficky určenou konfigurací ve vazebném místě a jejich naměřenou afinitou k danému biomolekulárnímu cíli (actives), tak z látek, které mají podobné fyzikálně-chemické vlastnosti jako známé aktivní ligandy, ale nevytváří s cílem terapeutického zásahu komplex (decoys). Populárními volně dostupnými validačními množinami jsou např. DUD<sup>58</sup>, DUD-E<sup>71</sup>, Astex Diverse Set<sup>72</sup> nebo DEKOIS<sup>73</sup>. Pokud jde o porovnání dokovacích programů, tak to často probíhá formou nejrůznějších soutěží a výzev (např. CSAR<sup>74</sup>, D3R Grand Challenge<sup>75</sup> nebo CAPRI<sup>76</sup>).

#### 4. Závěr

V této práci jsme se pokusili čtenáři stručně přiblížit disciplínu molekulového dokování. Byly představeny základní pojmy a vysvětleny některé principy konformačního vyhledávání a skórování dokovacích póz. Byla též diskutována problematika současných metod a byly představeny některé nové směry v této oblasti. V budoucnu se dá čekat na poli molekulového dokování další vývoj a to především směrem k zahrnutí kompletní flexibility proteinu do modelovacího procesu a odvození lepších skórovacích

funkcí, které budou schopné dosahovat konzistentních výsledků mezi různými systémy a přesnější predikce volné vazebné energie. Klíčovými faktory v tomto vývoji zřejmě též budou nové metody více integrované se simulacemi molekulové dynamiky a kvantově-chemickými výpočty. Nemalou roli, a to především ve zpřesňování a zvyšování propustnosti, bude mít též zlepšování hardware a vývoj metod, které povedou k jeho efektivnějšímu využití. Dá se čekat i další nárůst v počtu experimentálních dat, což by mohlo přispět nejen k vývoji přesnějších a obecnějších skórovacích funkcí, ale i k vytvoření diverzních datových množin pro testování nových přístupů. Je třeba též zmínit význam soutěží a vytváření dostatečně obsáhlých validačních množin, které se nepřímo podílejí na vývoji obecných a účinných metod s vysokou propustností.

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

#### Seznam symbolů

HTS	high-throughput screening
SBDD	structure-based drug design
SBVS	structure-based virtual screening
FBDD	fragment-based drug discovery
FEP	free energy perturbation
LIE	linear interaction energy
MM	molekulová mechanika
PBSA	Poisson-Boltzmann Surface Area
GBSA	Generalized Born Surface Area

#### LITERATURA

- Liu R., Li X., Lam K. S.: *Curr. Opin. Chem. Biol.* **38**, 117 (2017).
- Ziarek J. J., Peterson F. C., Lytle B. L., Volkman B. F.: *Methods Enzymol.* **493**, 241 (2011).
- Shi Y.: *Cell* **159**, 995 (2014).
- van den Bedem H., Fraser J. S.: *Nat. Methods* **12**, 307 (2015).
- Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H., Shindyalov I. N., Bourne P. E.: *Nucleic Acids Res.* **28**, 235 (2000).
- Taboureau O., Baell J. B., Fernández-Recio J., Villoutreix B. O.: *Chem. Biol.* **19**, 29 (2012).
- Jorgensen W. L.: *Acc. Chem. Res.* **42**, 724 (2009).
- Jorgensen W. L.: *Science* **303**, 1813 (2004).
- Bissantz C., Kuhn B., Stahl M.: *J. Med. Chem.* **53**, 5061 (2010).
- Du X., Li Y., Xia Y.-L., Ai S.-M., Liang J., Sang P., Ji X.-L., Liu S.-Q.: *Int. J. Mol. Sci.* **17**, 144 (2016).
- Brown N., v knize: *In Silico Medicinal Chemistry*. Royal Society of Chemistry, 2015.
- Yuriev E., Agostino M., Ramsland P. A.: *J. Mol. Recognit.* **24**, 149 (2011).
- Yuriev E., Holien J., Ramsland P. A.: *J. Mol. Recognit.* **28**, 581 (2015).
- Yuriev E., Ramsland P. A.: *J. Mol. Recognit.* **26**, 215 (2013).
- Wong C. F.: *Expert Opin. Drug Discovery* **10**, 1189 (2015).
- Sturlese M., Bellanda M., Moro S.: *Mol. Inf.* **34**, 513 (2015).
- London N., Raveh B., Schueler-Furman O.: *Curr. Opin. Struct. Biol.* **23**, 894 (2013).
- Kumalo H. M., Bhakat S., Soliman M. E. S.: *Molecules* **20**, 1984 (2015).
- Sable R., Jois S.: *Molecules* **20**, 11569 (2015).
- Liu J., Wang R.: *J. Chem. Inf. Model.* **55**, 475 (2015).
- Irwin J. J., Shoichet B. K.: *J. Med. Chem.* **59**, 4103 (2016).
- Rognan D.: *Drug Discovery Today: Technol.* **10**, e403 (2013).
- Pagadala N. S., Syed K., Tuszynski J.: *Biophys. Rev.* **9**, 91 (2017).
- Trott O., Olson A. J.: *J. Comput. Chem.* **31**, 455 (2010).
- Allen W. J., Balias T. E., Mukherjee S., Brozell S. R., Moustakas D. T., Lang P. T., Case D. A., Kuntz I. D., Rizzo R. C.: *J. Comput. Chem.* **36**, 1132 (2015).
- Dominguez C., Boelens R., Bonvin A. M. J. J.: *J. Am. Chem. Soc.* **125**, 1731 (2003).
- Klebe G.: *Drug Discovery Today* **11**, 580 (2006).
- Madhavi Sastry G., Adzhigirey M., Day T., Annabhimouju R., Sherman W.: *J. Comput. Aided Mol. Des.* **27**, 221 (2013).
- Wong S. E., Lightstone F. C.: *Expert Opin. Drug Discovery* **6**, 65 (2011).
- Guedes I. A., Magalhães C. S. d., Dardenne L. E.: *Biophys. Rev.* **6**, 75 (2014).
- Ferreira L. G., dos Santos R. N., Oliva G., Andricopulo A. D.: *Molecules* **20**, 13384 (2015).
- Friesner R. A., Banks J. L., Murphy R. B., Halgren T. A., Klicic J. J., Mainz D. T., Repasky M. P., Knoll E. H., Shelley M., Perry J. K., Shaw D. E., Francis P., Shenkin P. S.: *J. Med. Chem.* **47**, 1739 (2004).
- Rarey M., Kramer B., Lengauer T., Klebe G.: *J. Mol. Biol.* **261**, 470 (1996).
- Jones G., Willett P., Glen R. C.: *J. Mol. Biol.* **245**, 43 (1995).
- Korb O., Stützle T., Exner T. E.: *PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design*; International Workshop on Ant Colony Optimization and Swarm Intelligence, 2006.
- Korb O., Stützle T., Exner T. E.: *Swarm Intelligence* **1**, 115 (2007).
- Gabel J., Desaphy J., Rognan D.: *J. Chem. Inf. Model.* **54**, 2807 (2014).
- Huang S.-Y., Zou X., v knize: *Annual Reports in Computational Chemistry* (Wheeler R. A., ed.), Elsevier 2010.
- Pason L. P., Sotriffer C. A.: *Mol. Inf.* **35**, 541 (2016).
- Cornell W. D., Cieplak P., Bayly C. I., Gould I. R.,

- Merz K. M., Ferguson D. M., Spellmeyer D. C., Fox T., Caldwell J. W., Kollman P. A.: *J. Am. Chem. Soc.* **117**, 5179 (1995).
41. Michel J., Foloppe N., Essex J. W.: *Mol. Inf.* **29**, 570 (2010).
  42. Parenti M. D., Rastelli G.: *Biotechnol. Adv.* **30**, 244 (2012).
  43. Åqvist J., Luzhkov V. B., Brandsdal B. O.: *Acc. Chem. Res.* **35**, 358 (2002).
  44. Genheden S., Ryde U.: *Expert Opin. Drug Discovery* **10**, 449 (2015).
  45. Perez A., Morrone J. A., Simmerling C., Dill K. A.: *Curr. Opin. Struct. Biol.* **36**, 25 (2016).
  46. Yilmazer N. D., Korth M.: *Comput. Struct. Biotechnol. J.* **13**, 169 (2015).
  47. Ryde U., Söderhjelm P.: *Chem. Rev.* **116**, 5520 (2016).
  48. Verdonk M. L., Cole J. C., Hartshorn M. J., Murray C. W., Taylor R. D.: *Proteins: Struct., Funct., Bioinf.* **52**, 609 (2003).
  49. Lexa K. W., Carlson H. A.: *Q. Rev. Biophys.* **45**, 301 (2012).
  50. Antunes D. A., Devaurs D., Kaviraki L. E.: *Expert Opin. Drug Discovery* **10**, 1301 (2015).
  51. Wang Z., Sun H., Yao X., Li D., Xu L., Li Y., Tian S., Hou T.: *Phys. Chem. Chem. Phys.* **18**, 12964 (2016).
  52. Stefaniak F., Chudyk E. I., Bodkin M., Dawson W. K., Bujnicki J. M.: *WIREs Comput. Mol. Sci.* **5**, 425 (2015).
  53. Morris G. M., Huey R., Lindstrom W., Sanner M. F., Belew R. K., Goodsell D. S., Olson A. J.: *J. Comput. Chem.* **30**, 2785 (2009).
  54. Ding F., Yin S., Dokholyan N. V.: *J. Chem. Inf. Model.* **50**, 1623 (2010).
  55. Alonso H., Bliznyuk A. A., Gready J. E.: *Med. Res. Rev.* **26**, 531 (2006).
  56. Rarey M., Kramer B., Lengauer T.: *Proteins: Struct., Funct., Bioinf.* **34**, 17 (1999).
  57. Murphy R. B., Repasky M. P., Greenwood J. R., Tubert-Brohman I., Jerome S., Annabhimoju R., Boyles N. A., Schmitz C. D., Abel R., Farid R., Friesner R. A.: *J. Med. Chem.* **59**, 4364 (2016).
  58. Huang N., Shoichet B. K., Irwin J. J.: *J. Med. Chem.* **49**, 6789 (2006).
  59. Erlanson D. A., Fesik S. W., Hubbard R. E., Jahnke W., Jhoti H.: *Nat. Rev. Drug Discovery* **15**, 605 (2016).
  60. Verdonk M. L., Giangreco I., Hall R. J., Korb O., Mortenson P. N., Murray C. W.: *J. Med. Chem.* **54**, 5422 (2011).
  61. Colameco S., Elliot M. A.: *Biochemical Pharmacology* **133**, 29 (2017).
  62. Šponer J., Leszczynski J., Hobza P.: *Biopolymers* **61**, 3 (2001).
  63. Ruiz-Carmona S., Alvarez-Garcia D., Foloppe N., Garmendia-Doval A. B., Juhos S., Schmidtke P., Barril X., Hubbard R. E., Morley S. D.: *PLoS Comput. Biol.* **10**, e1003571 (2014).
  64. Singh J., Petter R. C., Baillie T. A., Whitty A.: *Nat. Rev. Drug Discovery* **10**, 307 (2011).
  65. Ouyang X., Zhou S., Su C. T. T., Ge Z., Li R., Kwok C. K.: *J. Comput. Chem.* **34**, 326 (2013).
  66. Ouyang X., Zhou S., Ge Z., Li R., Kwok C. K.: *Nucleic. Acids. Res.* **41**, W329 (2013).
  67. Toledo Warshaviak D., Golan G., Borrelli K. W., Zhu K., Kalid O.: *J. Chem. Inf. Model.* **54**, 1941 (2014).
  68. Zhu K., Borrelli K. W., Greenwood J. R., Day T., Abel R., Farid R. S., Harder E.: *J. Chem. Inf. Model.* **54**, 1932 (2014).
  69. Xie L., Xie L., Bourne P. E.: *Curr. Opin. Struct. Biol.* **21**, 189 (2011).
  70. Westermaier Y., Barril X., Scapozza L.: *Methods* **71**, 44 (2015).
  71. Mysinger M. M., Carchia M., Irwin J. J., Shoichet B. K.: *J. Med. Chem.* **55**, 6582 (2012).
  72. Hartshorn M. J., Verdonk M. L., Chessari G., Brewerton S. C., Mooij W. T. M., Mortenson P. N., Murray C. W.: *J. Med. Chem.* **50**, 726 (2007).
  73. Bauer M. R., Ibrahim T. M., Vogel S. M., Boeckler F. M.: *J. Chem. Inf. Model.* **53**, 1447 (2013).
  74. Carlson H. A., Smith R. D., Damm-Ganamet K. L., Stuckey J. A., Ahmed A., Convery M. A., Somers D. O., Kranz M., Elkins P. A., Cui G., Peishoff C. E., Lambert M. H., Dunbar J. B.: *J. Chem. Inf. Model.* **56**, 1063 (2016).
  75. *D3R | D3R Grand Challenge*. <https://drugdesigndata.org/about/grand-challenge>, staženo 24.9. 2017.
  76. *CAPRI Challenge Home*. <http://www.ebi.ac.uk/msd-srv/capri/>, staženo 24.9. 2017.
- M. Šicho<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): **Molecular Docking as a Tool to Virtually Develop Drugs**
- Bringing a new drug to the market is a lengthy, risky and expensive endeavor. Money spent on developing new drugs keeps going up each year, which is disproportional to the number of drugs brought to the market. Therefore, it is important to find ways to reduce costs and improve the effectiveness of drug discovery and development. Underlined by fast-paced developments in algorithms and processing power of modern hardware, computational methods have shown great potential in achieving this goal and molecular docking is an important tool in this toolbox. In this work, we briefly introduce the very basic principles of molecular docking and review some important contemporary challenges and developments in this field.