

## **Računarski modeli za predviđanje rastvorljivosti lekova**

**Erić Slavica\*, Kalinić Marko, Popović Aleksandar**

Institut za farmaceutsku hemiju, Farmaceutski fakultet, Univerzitet u  
Beogradu, Vojvode Stepe 450, 11000 Beograd, Srbija

---

### **Kratak sadržaj**

Rastvorljivost leka u vodi je faktor koji može značajno da utiče na bioraspoloživost peroralno primenjenog leka, kao i na njegovu raspodelu u organizmu. Razmatranjem rastvorljivosti u ranim fazama otkrića i razvoja leka smanjuje se mogućnost neuspeha u daljem razvoju leka. Računarske metode za predviđanje rastvorljivosti lekova omogućavaju analizu kombinatornih baza podataka, identifikaciju potencijalno problematičnih jedinjenja i isključivanje onih čija je rastvorljivost neadekvatna. Pored predviđanja rastvorljivosti na osnovu hemijske strukture, analizom ovih modela moguće je detaljnije razjasniti odnose hemijske strukture i rastvorljivosti ispitivanih lekova i optimizovati strukture u cilju poboljšanja rastvorljivosti, pri čemu bi njihova aktivnost ostala nepromenjena. Razvoj ovakvih modela je kompleksan proces koji zahteva razmatranje velikog broja faktora koji mogu uticati na uspešnost predviđanja konačnog modela. U ovom radu su prikazani različiti pristupi koji se koriste u razvoju računarskih modela za predviđanje rastvorljivosti. I pored intenzivnog rada na razvoju ovih modela tokom protekle decenije, pouzdanost predviđanja rastvorljivosti lekova različitih struktura još uvek ostaje veliki izazov. Kvalitet dostupnih eksperimentalnih podataka koji se koriste u modelovanju rastvorljivosti se u sve većoj meri prepoznaje kao jedan od glavnih uzroka ograničene pouzdanosti većine do sada predloženih modela. Iskorišćenje punog potencijala razvijenih pristupa modelovanja rastvorljivosti uslovljeno je širom dostupnošću pouzdanih podataka za rastvorljivost određenih pod identičnim eksperimentalnim uslovima.

***Ključne reči:*** predviđanje rastvorljivosti, rastvorljivost lekova, računarski modeli

---

\*Autor za korespondenciju: seric@pharmacy.bg.ac.rs

## 1. Uvod

Rastvorljivost, po najjednostavnijoj definiciji, predstavlja maksimalnu količinu supstance koju je, pri konstantnoj temperaturi i pritisku, moguće rastvoriti u određenoj zapremini rastvarača. Rastvorljivost i brzina rastvaranja lekovite supstance u vodi predstavljaju dva važna parametra koji mogu, posebno kod peroralne primene, biti ograničavajući faktori za obim i brzinu resorpcije, kao i raspodelu leka na ciljnom mestu dejstva [1-4]. Veoma slabo rastvorljivi lekovi, čak i ukoliko pokazuju dobru permeabilnost kroz biološke membrane i veliku potentnost, često nemaju adekvatnu biološku raspoloživost [5]. S druge strane, vrlo lako rastvorljivi lekovi su uglavnom polarne i slabo lipofilne strukture čija permeabilnost može biti ograničena. Iz tih razloga je neophodno uspostavljanje odgovarajućeg kompromisa između hidrofilnosti potrebne za rastvaranje u telesnim tečnostima i lipofilnosti koja uslovljava prolazak leka kroz biološke membrane. Pravovremeno razmatranje fizičko-hemijskih karakteristika potencijalnih lekova ima ključno mesto u ranim fazama otkrivanja i razvoja novih lekova [6-9].

Računarsko predviđanje rastvorljivosti leka na osnovu hemijske strukture, pre sinteze, veoma je značajno iz sledećih razloga: a) smanjuju se troškovi za sintezu i razvoj leka čija bi rastvorljivost bila neadekvatna; b) na osnovu analize dobijenih modela, postiže se bolji uvid u strukturne karakteristike koje utiču na rastvorljivost ispitivanih jedinjenja i c) omogućava se modifikacija hemijske strukture potencijalnog leka u cilju poboljšanja rastvorljivosti, pri čemu aktivnost leka ostaje nepromenjena. Pre nego što se pristupi optimizaciji rastvorljivosti potencijalnog leka, potrebno je izvršiti analizu odnosa strukture i dejstva potencijalnog leka i identifikovati funkcionalne grupe ili delove molekulske strukture koji su od značaja za biološku aktivnost. Nakon modifikacije strukture u cilju poboljšanja rastvorljivosti, potrebno je izvršiti ponovna biološka ispitivanja aktivnosti i neželjenih dejstava leka, jer dodatne funkcionalne grupe modifikovanog leka mogu da stupe u interakciju sa drugim ciljnim mestima u organizmu ili interferiraju u metaboličkim reakcijama. U nekim slučajevima modifikovanje rastvorljivosti potencijalnog leka nije moguće, jer strukturne karakteristike koje je potrebno modifikovati u cilju optimizacije rastvorljivosti istovremeno su neophodne za dejstvo leka, što je veoma značajno utvrditi u ranim fazama otkrivanja i razvoja novog leka.

Rad na razvijanju matematičkih modela za predviđanje rastvorljivosti pre sinteze potencijalnog leka je u poslednjoj deceniji doveo do značajnih rezultata, ali je pouzdanost većine razvijenih modela i dalje ograničena određenim faktorima [10]. U ovom radu su prikazani različiti pristupi modelovanja rastvorljivosti sa ciljem identifikacije ograničavajućih faktora za njihovu

pouzdaniju primenu, kao i načina da se poboljša uspešnost predviđanja postojećih modela i razviju novi modeli veće pouzdanosti.

## **2. Izbor seta za uspostavljanje modela i kvalitet eksperimentalnih podataka**

Modelovanje rastvorljivosti započinje izborom baza podataka koje će biti korišćene za uspostavljanje modela. Postoje dva osnovna aspekta razmatranja u izboru seta: izbor komponenata i kvalitet eksperimentalnih podataka.

### **2.1. Izbor komponenata, lokalni i globalni modeli**

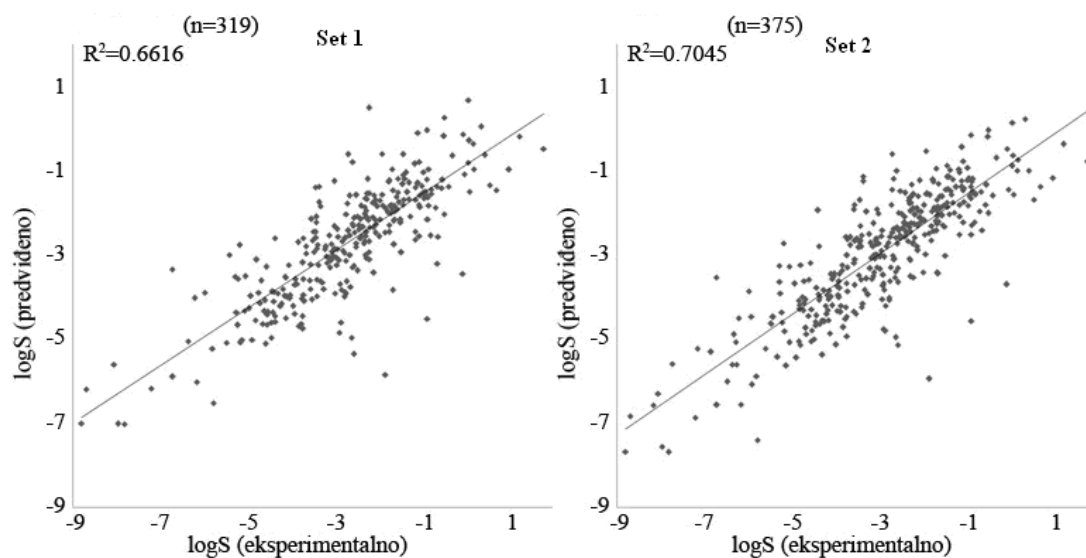
Raznovrsnost molekulskih struktura u ispitivanoj bazi podataka utiče na širinu hemijskog prostora u okviru kojeg je razvijen model, sa odgovarajućom tačnošću, primenljiv. Pokušaji da se korišćenjem modela predvidi rastvorljivost strukture koja se nalazi van definisanog hemijskog prostora najčešće rezultuju neupotrebljivim vrednostima, zato je važno da se za svaki model definiše i domen njegove primenljivosti [11]. Priroda fenomena rastvorljivosti je takva da brojne strukturne karakteristike, sa različitim doprinosom, utiču na rastvorljivost jedinjenja. U okviru kongeneričke grupe struktura (grupe strukturno sličnih jedinjenja), razlike u rastvorljivosti je relativno lako dovesti u vezu sa određenim strukturnim karakteristikama koje za tu grupu jedinjenja u najvećoj meri određuju rastvorljivost. Model razvijen korišćenjem kongeneričkog seta predstavlja lokalni model čija je primenljivost ograničena na mali deo hemijskog prostora ali čija su predviđanja, u okviru tog domena, često veoma pouzdana. Za razvijanje globalnog modela potreban je set podataka za uspostavljanje modela (engl. *training set*) koji obuhvata veliki deo odgovarajućeg hemijskog prostora. Sa farmaceutskog aspekta, hemijski značajan prostor čini tek mali deo ukupnog hemijskog prostora [12]. Izbor komponenata mora biti takav da odražava osobine lekova, što znači da u setu za uspostavljanje modela treba da budu uglavnom zastupljene strukture lekova koje su dovoljno heterogene. Od podjednakog značaja su i interval vrednosti rastvorljivosti obuhvaćen setom za uspostavljanje modela i homogena raspodela komponenata u značajnom hemijskom prostoru. Izbor odgovarajućeg seta za uspostavljanje modela omogućava da rezultujući model nije previše složen da bi bio upotrebljiv ali i da istovremeno daje pouzdana predviđanja za strukture u okviru njegovog domena primenljivosti. Globalni modeli se mogu koristiti za analizu raznovrsnih struktura, što nije slučaj sa lokalnim modelima, ali je tačnost njihovih predviđanja nešto manja. Predviđanje rastvorljivosti u okviru male, homogene grupe, uspešnije je korišćenjem lokalnih modela. Izbor pristupa, konačno, zavisi od namene.

## 2.2. Kvalitet eksperimentalnih podataka

Kvalitet eksperimentalnih podataka je faktor koji značajno utiče na uspešnost modela za predviđanje rastvorljivosti [11]. Ponekad se u istom modelu koriste podaci za rastvorljivost koji su dobijeni različitim metodama, ili se pod eksperimentalnim podatkom za rastvorljivost podrazumevaju potpuno različiti termini. Zbog toga je, pre svega, potrebno definisati vrednosti za rastvorljivost koje se koriste u uspostavljanju modela. Termodinamička rastvorljivost odgovara koncentraciji rastvorene supstance u zasićenom rastvoru kada se on nalazi u ravnoteži sa najstabilnijim kristalnim oblikom supstance u višku [13]. Termodinamička rastvorljivost neutralnog oblika supstance predstavlja osnovnu ili prirodnu rastvorljivost (engl. *intrinsic solubility*), čije se vrednosti koriste za modelovanje rastvorljivosti. Rastvorljivost je, za dati rastvarač i pri datoj temperaturi, funkcija molekulske strukture i može se izmeniti samo uvođenjem strukturnih promena u molekulu. Konceptualno gledano, eksperimentalno određivanje rastvorljivosti trebalo bi da bude jednostavno. Međutim, određivanje rastvorljivosti je mnogo komplikovanije jer se ravnoteža rastvora u praksi retko zaista i postiže, te na izmerenu rastvorljivost može uticati više faktora, kao što je brzina prelaska metastabilnih u stabilan kristalni oblik supstance, kao i niz drugih parametara koji utiču na ravnotežu rastvora, usled čega se eksperimentalno određena rastvorljivost značajno može razlikovati od stvarne [1]. Ova odstupanja ukazuju na to da eksperimentalno određena vrednost više nije samo funkcija strukture. Efektivno, na vrednosti eksperimentalno određene rastvorljivosti lekovite supstance utiču: čistoća upotrebljene supstance, kristalni oblik, dužina inkubacije, pH i temperatura, sastojci pufera i hemijska stabilnost supstance. S obzirom da su za razvoj globalnih modela neophodne vrednosti rastvorljivosti za stotine jedinjenja, prilikom formiranja seta za uspostavljanje modela najčešće se koriste slobodne ili komercijalno dostupne baze podataka o rastvorljivosti. Ovi podaci najčešće potiču iz velikog broja izvora i dobijeni su primenom različitih eksperimentalnih protokola, što doprinosi ukupnoj eksperimentalnoj grešci koja se unosi u model. Pokazano je da se, između različitih laboratorija, vrednosti rastvorljivosti za isto jedinjenje razlikuju u proseku za 0,6 logS [14] dok se, u pojedinim slučajevima, nailazi i na dvostruke do trostruke razlike u izmerenim vrednostima [15]. Kako model ne može prevazići tačnost eksperimentalnih podataka korišćenih za njegovo uspostavljanje [16], greška predviđanja modela koji je uspostavljen korišćenjem eksperimentalnih podataka za rastvorljivost iz različitih laboratorija i pod različitim eksperimentalnim uslovima verovatno neće biti manja od 0,6 logS.

Kvalitet eksperimentalnih podataka može presudno da utiče na kvalitet modela. Definisane jasne kriterijume prilikom prikupljanja podataka

(korišćenje samo onih vrednosti koje odgovaraju osnovnoj rastvorljivosti, dobijenih pod istim eksperimentalnim uslovima) može redukovati eksperimentalnu grešku koja se unosi u model [17]. Korišćenjem uniformno dobijenih eksperimentalnih podataka, značajno se može poboljšati uspešnost u predviđanju rezultujućih modela. U našem prethodnom radu, ispitivan je uticaj različitih eksperimentalnih podataka za uspostavljanje modela na preciznost predviđanja rastvorljivosti [18]. Prikazani grafici (**Slika 1**) pokazuju u kojoj meri je upotreba eksperimentalnih podataka dobijenih upotrebom identičnih eksperimentalnih modela od značaja za preciznost predviđanja rastvorljivosti. Dodatkom malog broja komponenata čije su vrednosti za rastvorljivost određene pod istim eksperimentalnim uslovima u set za uspostavljanje modela, tačnost predviđanja prikazanih preliminarnih modela značajno se poboljšava. Može se zaključiti da dostupnost većih količina kvalitetnih eksperimentalnih podataka predstavlja preduslov za poboljšanje uspešnosti predviđanja postojećih i razvoj novih modela za predviđanje rastvorljivosti lekova [11, 19-20].



**Slika 1. Korelacije između predviđenih i eksperimentalno određenih vrednosti osnovne rastvorljivosti ( $\log S$ ), pre (Set 1) i nakon (Set 2) dodatka uniformno određenih eksperimentalnih podataka u set za uspostavljanje modela [18].**

**Figure 1. Correlation between predicted and experimental intrinsic solubility values ( $\log S$ ), before (Set 1) and after (Set 2) the inclusion of uniformly determined experimental data into the training set [18].**

### 3. Pristupi u modelovanju rastvorljivosti

Tokom protekle decenije prikazan je veliki broj različitih modela za predviđanje rastvorljivosti. Po nekim autorima, veliko interesovanje za ovu oblast svedoči o njenoj važnosti, ali i o činjenici da do sada nije predložen nijedan u potpunosti zadovoljavajući pristup [21].

Mada ne postoji univerzalno prihvaćena klasifikacija, prvu grupu čine modeli koji sadrže deskriptore koji su eksperimentalno određeni. Više autora je razvilo relativno jednostavne modele za procenu rastvorljivosti korišćenjem eksperimentalno određenih vrednosti tačke topljenja i/ili particionog koeficijenta [22-26]. Korišćenjem tačke topljenja, model uzima u obzir interakcije koje se odigravaju u čvrstoj fazi, dok particioni koeficijent (logP) predstavlja dobar pokazatelj sposobnosti supstance da stupa u interakcije sa molekulama vode. Iako se logP danas može, sa velikom tačnošću, predvideti primenom računarskih metoda, to nije slučaj sa tačkom topljenja, čije je izračunavanje jedan od najvećih izazova u domenu *in silico* predviđanja fizičko-hemijskih karakteristika supstance [27]. Da bi se predvidela rastvorljivost potencijalnih lekova upotrebom modela koji je uspostavljen korišćenjem eksperimentalno određenih deskriptora, potrebno je sintetisati ispitivanu supstancu i za nju takođe eksperimentalno odrediti deskriptore, na osnovu kojih se vrši predviđanje rastvorljivosti. Zbog toga su ovi modeli danas većinom samo od istorijskog značaja, jer nisu primenljivi za ranu analizu i dizajn kombinatornih biblioteka [20]. Nedavno je objavljena metoda za predviđanje rastvorljivosti uz upotrebu izračunatih vrednosti za tačku topljenja i particioni koeficijent [28], ali pouzdanost ovog pristupa još nije u potpunosti izvesna. Korelacija između rastvorljivosti i izračunatog particionog koeficijenta je, međutim, primenjena i u većem broju novijih modela [29, 30].

Drugu grupu modela čine tzv. fragmentni modeli. Ovi modeli se zasnivaju na pretpostavci da svaki deo molekule (atom ili funkcionalna grupa) daje konstantan i kumulativan doprinos ukupnoj rastvorljivosti jedinjenja. Primenom algoritama koji na osnovu dvodimenzionalne strukture (broja atoma i njihove topologije) raščlanjavaju molekulu, sa ili bez preklapanja, na predefinisane fragmente, dobija se veliki broj fragmentnih deskriptora za svako od jedinjenja u setu. Korelisanjem vrednosti fragmentnih deskriptora sa eksperimentalno određenim vrednostima rastvorljivosti, svakom od deskriptora pridružuje se odgovarajući korekcionni faktor, što predstavlja osnovu modela. U literaturi se može pronaći veći broj ovakvih modela [31-35]. Nedostatak ovog pristupa je u tome što na rastvorljivost utiču i drugi faktori, kao što su intramolekulske vodonične veze i elektronski efekti supstituenata, koji se ne mogu modelovati primenom isključivo fragmentnog pristupa [21]. Takođe, razvoj pouzdanog modela zahteva definisanje velikog broja fragmenata, jer

primena modela na jedinjenje koje sadrži fragmente koji nisu bili obuhvaćeni setom za uspostavljanje modela kao rezultat daje predviđanje male tačnosti. Korišćenje deskriptora vezanih za elektro-topološko stanje fragmenta ne mora nužno dati bolje rezultate od klasičnog pristupa zasnovanog na doprinosu pojedinih atoma [16]. I pored navedenih nedostataka, budući da ne zahtevaju računarski zahtevnu optimizaciju struktura, ovi modeli imaju veoma veliku primenu u ranim ispitivanjima rastvorljivosti potencijalnih lekova.

Treću grupu modela čine modeli zasnovani na različitim molekulskim deskriptorima izračunatim na osnovu dvodimenzionalne ili trodimenzionalne strukture jedinjenja. U ovim modelima često se koristi i particioni koeficijent kao molekulski deskriptor. Iz dvodimenzionalnih (2D) struktura mogu se izračunati samo jednostavni konstitucionni, topološki i drugi deskriptori koji ne zahtevaju precizno poznavanje trodimenzionalne (3D) strukture. Dvodimenzionalni deskriptori poslužili su za razvoj značajnog broja modela [36-40]. Razvojem takvih modela izbegava se problem geometrijske optimizacije struktura, štede se kompjuterski resursi, ali se ne koriste geometrijski, većina elektronskih i kvantno-hemijskih deskriptora koji potencijalno mogu da opisuju karakteristike od značaja za rastvorljivost.

Geometrijska optimizacija podrazumeva pronalaženje konformacije molekule za koju postoji velika verovatnoća da predstavlja konformaciju najniže moguće energije [41]. Za geometrijsku optimizaciju dostupan je veliki broj metoda zasnovanih na različitim teorijskim modelima molekule kao sistema ili multiparametarskih funkcija. To mogu biti molekularno-mehaničke (MM) metode, kao i kvantno-mehaničke (QM) metode, od kojih su najširu primenu našle semi-empirijske optimizacije, *ab initio* optimizacije i metode zasnovane na teoriji funkcionalne gustine (engl. *Density Functional Theory*, DFT). Od navedenih pristupa, u optimizaciji struktura za potrebe razvijanja modela za predviđanje rastvorljivosti najčešće su korišćene MM i semi-empirijske metode. MM optimizacija je brza ali ne uzima u obzir elektronsku strukturu molekule, kao što je to slučaj sa semi-empirijskom, koja zahteva značajnije kompjuterske resurse. Razmatranje pristupa optimizacije struktura veoma je značajno. S jedne strane, metoda optimizacije određuje propusnu moć finalnog modela ali se izbor metode optimizacije najčešće podređuje zahtevima računanja željenih deskriptora u okviru kojih se traže oni koji su najznačajniji za rastvorljivost. Izabranom metodom optimizacije uniformno se određuju konformeri najmanje energije za svako jedinjenje u setu jer se, u suprotnom, vrednosti deskriptora mogu značajno razlikovati, što dovodi do greške pri uspostavljenju odnosa brojnih vrednosti deskriptora sa rastvorljivošću. U pojedinačnim slučajevima, autori su uspeli da pokažu značajnu neosetljivost dobijenih modela na razlike u konformaciji korišćenoj za izračunavanja

molekulskih deskriptora [42], ali uspešnost takvog pristupa predstavlja retkost u literaturi.

Većina današnjih softverskih paketa omogućava računanje velikog broja deskriptora (hiljadu i više), usled čega je izbor najznačajnijih deskriptora za rastvorljivost ispitivanog seta podataka veoma značajna [43]. Pored statističke značajnosti, odabrani deskriptori bi, istovremeno, trebalo da budu i objašnjivi u kontekstu fenomena koji se modeluje. Heuristička metoda (HM), kao relativno jednostavna metoda, pokazala se kao pouzdan pristup za selekciju najznačajnijih deskriptora na osnovu statističke značajnosti [44, 45]. Ova metoda omogućava eliminaciju međusobno visoko korelisanih deskriptora, kao i onih čija je vrednost nedovoljno varijabilna u okviru seta, te omogućava pronalaženje najboljih regresionih modela za unapred definisani broj izabranih deskriptora. U našem prethodnom radu, primenom HM pokazano je u kojoj meri upotreba eksperimentalnih podataka određenih upotrebom identične metode u setu za uspostavljanje modela ima uticaj na izbor najznačajnijih deskriptora za predviđanje rastvorljivosti [18]. HM se može upotrebiti i za uspostavljanje preliminarnih modela, čije performanse mogu biti veoma bliske, iako slabije, od modela dobijenih značajno složenijim pristupima, pri čemu primena HM zahteva 2-5 puta manje vremena i nema ograničenja u pogledu veličine seta [46]. Pored linearnih metoda, genetski algoritmi (GA) su takođe primenjeni u izboru najznačajnijih deskriptora za rastvorljivost, pri čemu je tumačenje pojedinačnog udela deskriptora u ukupnom modelu komplikovanije. Primenom GA kombinuju se deskriptori pri čemu se kvalitet kombinacija procenjuje nekom regresionom metodom (multilinearna regresija, metoda najmanjih kvadrata), vrši se izbor najboljih rešenja uz stohastičko uključanje manjeg broja lošijih (radi održavanja varijabilnosti i sprečavanja prerane konvergencije), a zatim kombinuju izabrane kombinacije pri čemu se ovaj proces ponavlja kroz predefinisani broj generacija ili do postizanja željenih performansi. Baveći se problemom izbora deskriptora za predviđanje rastvorljivosti i lipofilnosti, uz primenu inovativnog pristupa automatskoj proceni značajnosti deskriptora, pojedini autori su pokazali da GA može biti brz i pouzdan pristup izboru deskriptora čak i za velike setove podataka [47].

Bez obzira na odabrani pristup, kako je pokazano da sa porastom broja dostupnih deskriptora raste i mogućnost slučajnih korelacija [48], statističku značajnost deskriptora ne bi trebalo posmatrati odvojeno od mogućnosti njihovog tumačenja u okviru fenomena rastvorljivosti.

Modeli zasnovani na deskriptorima izvedenim iz optimizovanih 3D struktura ili kombinaciji 2D i 3D deskriptora česti su u literaturi [42, 49-54]. Za razliku od fragmentnih, u slučaju ovih modela ne može se javiti problem „nedostajućeg fragmenta”, jer je moguće izračunati sve molekulske deskriptore

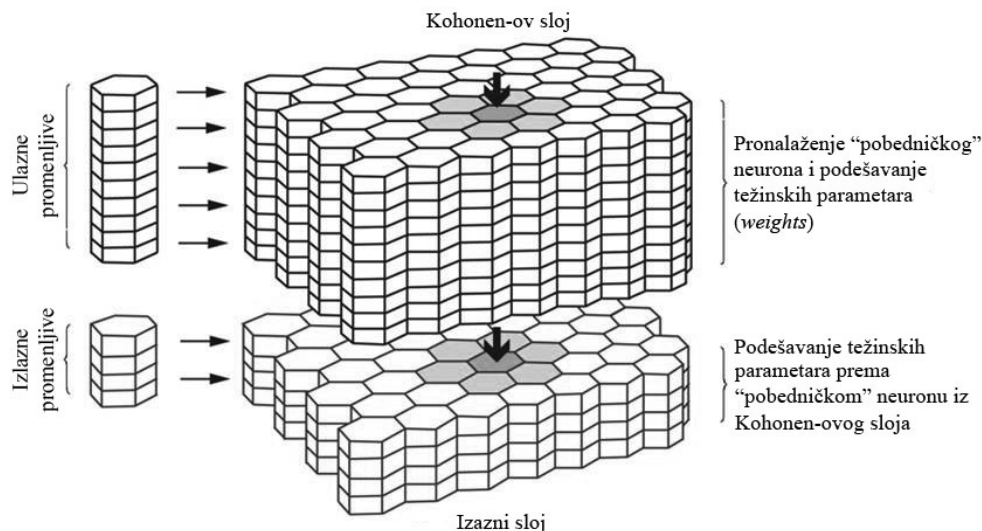


za svaku strukturu. Modeli se često zasnivaju na deskriptorima kao što su površina dostupna rastvaraču (*Solvent Accessible Surface Area*, SASA), polarna molekulska površina (*Polar Surface Area*, PSA), izračunati particioni koeficijent (ClogP/SlogP), hidrofobnost, broj donora/akceptora vodoničnih veza i/ili površina koju oni zauzimaju, te je njihova dodatna prednost i jednostavno tumačenje u okviru osobine koju opisuju.

Konačno, uprkos velikoj različitosti navedenih pristupa, moć predviđanja prikazanih modela između ovih grupa ne pokazuje značajne razlike. Primera radi, Wang i saradnici razvili su, korišćenjem istog seta za uspostavljanje modela, dva modela, jedan zasnovan na fragmentnim deskriptorima a drugi na površini dostupnoj rastvaraču (SASA) kao osnovnim deskriptorom. Pokazano je da je uspešnost predviđanja ova dva modela slična, i da model koji uključuje SASA pokazuje tek neznatne prednosti [55]. Bergström i saradnici pokazali su da globalni model razvijen samo pomoću 2D deskriptora ima veću uspešnost predviđanja od modela razvijenog korišćenjem samo 3D deskriptora, ali da model koji kombinuje 2D i 3D deskriptore vrši bolja predviđanja od oba prethodna [15]. U radovima nekih autora pokazano je da nije postignuto značajno poboljšanje performansi uključenjem 3D deskriptora [27, 56]. Ovakvi rezultati i ilustrativno govore o činjenici da nijedan od navedenih pristupa ponaosob, za sada, nije dokazano superiorniji od ostalih.

#### **4. Linearne i nelinearne metode za predviđanje rastvorljivosti**

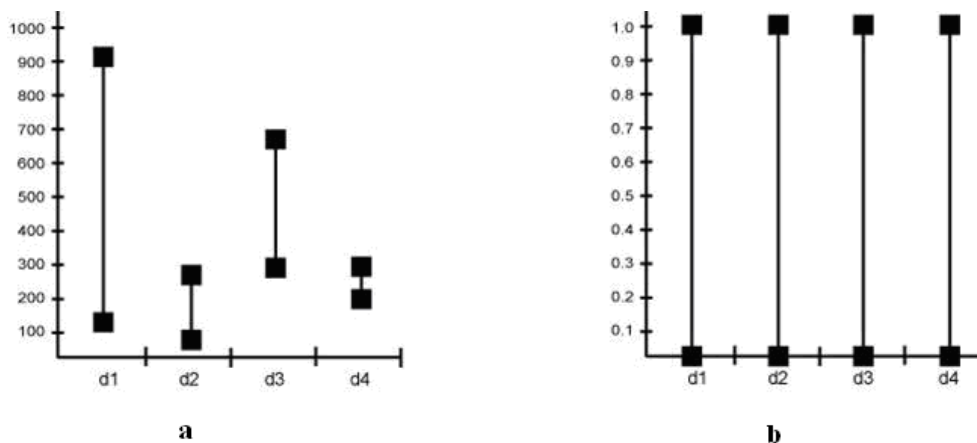
Uspostavljanje korelacije između rastvorljivosti i odabranih deskriptora može se vršiti primenom linearnih i nelinearnih matematičkih metoda. Multilinearna regresija (engl. *multilinear regression*, MLR) i metoda najmanjih kvadrata (engl. *partial least squares method*, PLS) dugo su predstavljali osnovu regresione analize u domenu predviđanja rastvorljivosti. Danas je za ovu namenu prilagođen i primenjen čitav spektar nelinearnih metoda kao što su veštačke neuronske mreže (engl. *artificial neural networks method*, (ANN)), genetski algoritmi (GA), *random forest* regresija (RF), *support vector machine* (SVM) i druge metode mašinskog učenja. Osnovna prednost neuronskih mreža (**Slika 2**) i nelinearnih metoda generalno, jeste u sposobnosti modelovanja složenih, nelinearnih odnosa bez pretpostavki o izgledu modela [57].



**Slika 2. Arhitektura kontrapropagirajućih veštačkih neuronskih mreža (CPANN).**

**Figure 2. Illustration of the architecture of counter-propagation artificial neural network (CPANN).**

Pod pretpostavkom da odnos između strukture i rastvorljivosti nije linearan, što bi bilo očekivano s obzirom na složenost samog fenomena rastvorljivosti, modelovanje primenom nelinearnih metoda trebalo bi da osigura veću uspešnost predviđanja rezultujućih modela. Međutim, modeli dobijeni primenom neuronskih mreža i sličnim pristupima često predstavljaju „*black box*” modele, jer je njihovo tumačenje veoma kompleksno. Dok primena linearne regresione analize obezbeđuje uvid u relativni značaj svake promenljive u modelu poređenjem odgovarajućih regresionih koeficijenata, modeli dobijeni primenom neuronskih mreža ne omogućavaju direktnu analizu relativne značajnosti pojedinih deskriptora (**Slika 3**). Takođe, sa povećanjem broja iteracija kroz koje se ANN model trenira, raste i mogućnost da rezultujući model bude „pretreniran” (engl. *overtraining*). Kao posledica „pretreniranja”, tačnost predviđanja rastvorljivosti za komponente seta korišćenog za uspostavljanje modela može da bude veća od tačnosti samih eksperimentalnih podataka, dok se preciznost predviđanja rastvorljivosti uglavnom drastično smanjuje kada se model primeni na eksterni set koji sadži komponente koje nisu bile uključene u set za uspostavljanje modela.



**Slika 3. Grafički prikaz udela pojedinačnih deskriptora u modelu za predviđanje rastvorljivosti koji je uspostavljen primenom linearnih metoda (a) i veštačkih neuronskih mreža (b).**

**Figure 3. Graphical representation of the contribution of single descriptors in models for solubility prediction established by linear methods (a) and artificial neural networks methods (b).**

U pogledu poređenja uspešnosti linearnih i nelinearnih metoda u literaturi ne postoji konsenzus. U nekim slučajevima predviđanje rastvorljivosti je bolje primenom linearnih [27, 39], a u drugim slučajevima primenom nelinearnih metoda [31, 32, 56, 58]. Ovim razlikama verovatno doprinose i različitosti u deskriptorima iskorišćenim za uspostavljanje odnosa između strukture i rastvorljivosti. Konačno, linearne i nelinearne metode nisu uzajamno isključive i često se koriste u međusobnoj sprezi. Kako pojedini autori zaključuju, nelinearne metode mogu imati značajno mesto u odabiru deskriptora kao i u poboljšanju uspešnosti modela, onda kada linearne metode daju lošije rezultate [59].

## 5. Validacija i uspešnost predviđanja modela

Validacija modela se može izvesti na različite načine: korišćenjem internog ili eksternog test seta, unakrsnom validacijom, validacijom izostavljanjem po jedne komponente (*leave-one-out*) itd. Međutim, čini se da rezultati dobijeni validacijom modela često ne odgovaraju uspešnosti predviđanja modela u realnom okruženju, kada se on primeni na raznovrsnije strukture lekova. Pokazano je da većina modela pokazuje najmanju uspešnost predviđanja u intervalu niske rastvorljivosti za koje je, zapravo, najvažnije i

postići precizna predviđanja [60]. Ovaj problem uglavnom nije jasno izražen kod modela objavljenih u literaturi, jer se validacija najčešće vrši u daleko širem intervalu rastvorljivosti, čime se stiče optimističniji utisak o sposobnosti predviđanja modela. Delaney-ev koncept „prikladnosti nameni” (engl. *fit-for-purpose*) drugi autori su proširili i predložili procenu uspešnosti predviđanja literaturnih modela za interval rastvorljivosti od 0,1 µg/mL do 250 µg/mL, koji najbolje predstavlja interval od značaja za analizu jedinjenja u fazi otkrića leka [10]. Svrshodniji dizajn test setova, doprineo bi, dakle, sticanju boljeg uvida u stvarnu primenljivost modela. Taskinen i Norinder su procenili da bi se primenom većine postojećih, komercijalnih i literaturnih modela, primenjenih na test set koji je okvirno reprezentativan za strukture lekova, dobile predviđene vrednosti sa RMSE (kvadrat prosečne standardne greške) u intervalu od 1 do 1,5 logS [11].

## 6. Perspektive predviđanja rastvorljivosti

Sve intenzivnije interesovanje za uticaj kvaliteta eksperimentalnih podataka na uspešnost predviđanja razvijenih modela možda je i najbolje ilustrovano nedavno objavljenom publikacijom, u kojoj su autori prikazali vrednosti rastvorljivosti za 100 struktura lekova, određene pod identičnim eksperimentalnim uslovima, i pozvali istraživače da na osnovu ponuđenih 100 pouzdanih merenja razviju model za predviđanje rastvorljivosti 32 druga jedinjenja [61]. Publikovani rezultati istraživača pružili su interesantan uvid u trenutna dostignuća u razvoju pouzdanih modela za predviđanje rastvorljivosti [62] i otvorili prostor za ponovno razmatranje dostupnih pristupa, njihove uspešnosti i preporuka u razvoju novih modela [63]. Veća dostupnost kvalitetnih eksperimentalnih vrednosti za rastvorljivost, određenih pod identičnim uslovima, trebalo bi da pruži veće mogućnosti za razvoj modela velike pouzdanosti i da omogući potpuno iskorišćenje potencijala razvijenih metoda za modelovanje rastvorljivosti.

Takođe, modelovanja interakcija koje se odigravaju u čvrstom stanju i određivanje faktora koji određuju kristalnu strukturu supstance je još jedan pristup koji bi mogao značajno da doprinese poboljšanju modela za predviđanje rastvorljivosti. Većina dostupnih modela uzima u obzir samo faktore koji utiču na solvataciju, zanemarujući time važan početni korak u procesu rastvaranja supstance, koji podrazumeva razgradnju strukture kristalne rešetke [19]. Računarsko modelovanje parametara koji određuju gustinu i energiju kristalne rešetke i uspostavljanje veze ovih parametara sa rastvorljivošću doprinelo bi poboljšanju uspešnosti predviđanja modela. Razvoj pouzdanih pristupa za računarsko modelovanje interakcija u čvrstoj fazi omogućio bi, dakle, razvoj novih modela za predviđanje rastvorljivosti koji bi celovitije oslikavali

celokupan fenomen rastvorljivosti, što bi značajno moglo da se odrazi na tačnost i pouzdanost njegovog predviđanja.

## 7. Zaključak

Modeli za računarsko predviđanje rastvorljivosti mogu značajno doprineti racionalizaciji procesa dizajniranja i razvoja novog leka. I pored značajnih napredaka u razvijanju novih pristupa modelovanju, rastvorljivost ostaje parametar koji je teško predvideti, a većina razvijenih modela nedovoljno primenljiva u realnom okruženju. Za dalji napredak u ovoj oblasti potreban je veći broj kvalitetnih eksperimentalnih podataka za rastvorljivost struktura lekova, poseban pristup dizajniranju reprezentativnih setova za uspostavljanje modela i kritička procena domena primenljivosti i pouzdanosti razvijenih modela. Permanentna optimizacija modela unošenjem eksperimentalnih podataka za rastvorljivost određenih pod identičnim uslovima, kao i povećanje domena primenljivosti, može predstavljati jedan od načina poboljšanja pouzdanosti modela. Računarsko modelovanje interakcija u čvrstoj fazi takođe ima potencijal da značajno unapredi kvalitet i tumačenje modela za predviđanje rastvorljivosti.

## Literatura

1. Bhattachar SN, Deschenes LA, Wesley JA. Solubility: it's not just for physical chemists. *Drug Discov Today* 2006; 11(21/22):1012-1017.
2. Hörter D, Dressman JB. Influence of physicochemical properties on dissolution of drugs in the gastrointestinal tract. *Adv Drug Deliv Rev* 1997; 25:3-14.
3. Zhao YH, Abraham MH, Le J, Hersey A, Luscombe CN, Beck G et al. Rate-limited steps of human oral absorption and QSAR studies. *Pharm Res* 2002; 19(10):1446-1457.
4. Amidon GL, Lennernäs H, Shah VP, Crison JR. A theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability. *Pharm Res* 1995; 12(3):413-420.
5. Curatolo W. Physical chemical properties of oral drug candidates in the discovery and exploratory development settings. *Pharm Sci Technol Today* 1998; 1(9):387-393.
6. Van de Waterbeemd H, Gifford E. ADMET in silico modelling: towards prediction paradise?. *Nat Rev Drug Discov* 2003; 2(3):192-203.

7. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 1997; 23:3-26.
8. Di L, Kerns EH. Profiling drug-like properties in discovery research. *Curr Opin Chem Biol* 2003; 7:402-408.
9. Stegemann S, Leveiller F, Franchi D, de Jong H, Lindén H. When poor solubility becomes an issue: from early stage to proof of concept. *Eur J Pharm Sci* 2007; 31:249-261.
10. Johnson SR, Zheng W. Recent progress in the computational prediction of aqueous solubility and absorption. *AAPS J* 2006; 8(1):E27-E40.
11. Taskinen J, Norinder U. In silico predictions of solubility. In: Taylor JB, Triggle DJ. eds. *Comprehensive Medicinal Chemistry II* (vol 5). Amsterdam: Elsevier, 2007: 627-48.
12. Lipinski CA. Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods* 2000; 44(1):235-249.
13. Avdeef A. Physicochemical Profiling (Solubility, Permeability and Charge State). *Curr Top Med Chem* 2001; 1:277-351
14. Katritzky AR, Wang Y, Sild S, Tamm T, Karelson M. QSPR studies on vapor pressure, aqueous solubility, and the prediction of water-air partition coefficients. *J Chem Inf Comput Sci* 1998; 38(4):720-725.
15. Bergström CAS, Wassvik CM, Norinder U, Luthman K, Artursson P. Global and local computational models for aqueous solubility prediction of drug-like molecules. *J Chem Inf Comput Sci* 2004; 44(4):1477-1488.
16. Balakin KV, Savchuk NP, Tetko IV. In silico approaches to prediction of aqueous and DMSO solubility of drug-like compounds: trends, problems and solutions. *Curr Med Chem* 2006; 13(2):223-241.
17. Rytting E, Lentz KA, Chen XQ, Qian F, Venkatesh S. Aqueous and cosolvent solubility data for drug-like organic compounds. *AAPS J* 2005; 7(1):E78-E102.
18. Erić S, Kalinić M, Popović A, Makić H, Civić E, Bektašević M. The importance of accuracy of experimental data for the prediction of solubility. *J Serb Chem Soc* 2010; 75(4):483-495.
19. Berström CAS. In silico predictions of drug solubility and permeability: two rate-limiting barriers to oral drug absorption. *Basic Clin Pharmacol Toxicol* 2005; 96:156-161.
20. Jorgensen WL, Duffy EM. Prediction of drug solubility from structure. *Adv Drug Deliv Rev* 2002; 54:355-366.
21. Faller B, Ertl P. Computational approaches to determine drug solubility. *Adv Drug Deliv Rev* 2007; 59:533-545.
22. Irmann F. Eine einfache Korrelation zwischen Wasserlöslichkeit und Struktur von Kohlenwasserstoffen und Halogenkohlenwasserstoffen. *Chem Ing Tech* 1965; 37:789-798.

23. Hansch C, Quinlan JE, Lawrence GL. Linear free-energy relationship between partition coefficients and the aqueous solubility of organic liquids. *J Org Chem* 1968; 33(1):347-350.
24. Yalkowsky SH, Valvani SC. Solubility and partitioning I: solubility of nonelectrolytes in water. *J Pharm Sci* 1980; 69(8):912-922.
25. Jain N, Yalkowsky SH. Estimation of the aqueous solubility I: application to organic nonelectrolytes. *J Pharm Sci* 2001; 90(2):234-252.
26. Myrdal P, Ward GH, Dannenfelser RM, Mishra D, Yalkowsky SH. AQUAFAC I: aqueous functional group activity coefficients; application to hydrocarbons. *Chemosphere* 1992; 24(8):1047-1061.
27. Hughes LD, Palmer DS, Nigsch F, Mitchell JB. Why are some properties more difficult to predict than others? A study of QSPR models of solubility, melting point and logP. *J Chem Inf Model* 2008; 48(1):220-232.
28. Jain P, Yalkowsky SH. Prediction of aqueous solubility from SCRATCH. *Int J Pharm* 2010; 385(1):1-5.
29. Pinsuwan S, Yalkowsky SH. Correlation of octanol/water solubility ratios and partition coefficients. *J Chem Eng Data* 1995; 40(3):623-626.
30. Box KJ, Comer JEA. Using measured pKa, logP and solubility to investigate supersaturation and predict BCS class. *Curr Drug Metab* 2008; 9(9):869-878.
31. Huuskonen J. Estimation of aqueous solubility for a diverse set of organic compounds based on molecular topology. *J Chem Inf Comput Sci* 2000; 40(3):773-777.
32. Tetko IV, Tanchuk VY, Kasheva TN, Villa AEP. Estimation of aqueous solubility of chemical compounds using E-state indices. *J Chem Inf Comput Sci* 2001; 41(6):1488-1493.
33. Klopman G, Zhu H. Estimation of the aqueous solubility of organic molecules by the group contribution approach. *J Chem Inf Comput Sci* 2001; 41(2):439-445.
34. Hou TJ, Xia K, Zhang W, Xu XJ. ADME evaluation in drug discovery. 4. Prediction of aqueous solubility based on atom contribution approach. *J Chem Inf Comput Sci* 2004; 44(1):266-275.
35. Du-Cuny Lei, Huwyler J, Wiese M, Kansy M. Computational aqueous solubility prediction for drug-like compounds in congeneric series. *Eur J Med Chem* 2008; 43:501-512.
36. Gao H, Shanmugasundaram V, Lee P. Estimation of aqueous solubility of organic compounds with QSPR approach. *Pharm Res* 2002; 19(4):497-503.
37. Cheng A, Merz KM. Prediction of aqueous solubility of a diverse set of compounds using quantitative structure-property relationships. *J Med Chem* 2003; 46(17):3572-3580.
38. Delaney JS. ESOL: Estimating aqueous solubility directly from molecular structure. *J Chem Inf Comput Sci* 2004; 44(3):1000-1005.

39. Catana C, Gao H, Orrenius C, Stouten PFW. Linear and nonlinear methods in modeling the aqueous solubility of organic compounds. *J Chem Inf Model* 2005; 45(1):170-176.
40. Huuskonen J, Livingstone DJ, Manallack DT. Prediction of drug solubility from molecular structure using a drug-like training set. *SAR QSAR Environ Res* 2008; 19(3/4):191-212.
41. Cramer CJ. *Essentials of Computational Chemistry: Theories and Models*. 2nd ed. Chichester: John Wiley & Sons Ltd, 2004: 42.
42. Wang J, Krudy G, Hou T, Zhang W, Holland G, Xu X. Development of reliable aqueous solubility models and their application in druglike analysis. *J Chem Inf Model* 2007; 47(4):1395-1404.
43. Mager DE. Quantitative structure-pharmacokinetic/pharmacodynamic relationships. *Adv Drug Deliv Rev* 2006; 58:1326-1356.
44. Erić S, Šolmajer T, Nović M, Oblak M, Agbaba D. Quantitative structure-activity relationships of  $\alpha_1$  adrenergic antagonists. *J Mol Model* 2004; 10(2):139-150.
45. Erić S, Šolmajer T, Župan J, Nović M, Oblak M, Agbaba D. Prediction of selectivity of  $\alpha_1$ -adrenergic antagonists by counterpropagation neural network (CP-ANN). *Il Farmaco* 2004; 59(5):389-395.
46. Ma W, Zhang X, Luan F, Zhang H, Zhang R, Liu M et al. Support vector machine and the heuristic method to predict the solubility of hydrocarbons in electrolyte. *J Phys Chem A* 2005; 109(15):3485-3492.
47. Wegner JK, Zell A. Prediction of aqueous solubility and partition coefficient optimized by a genetic algorithm based descriptor selection method. *J Chem Inf Comput Sci* 2003; 43(3):1077-1084.
48. Topliss JG, Edwards RP. Chance factors in studies of quantitative structure-activity relationships. *J Med Chem* 1979; 22:1238-1244.
49. Jorgensen WL, Duffy EM. Prediction of drug solubility from Monte Carlo simulations. *Bioorg Med Chem Lett* 2000; 10:1155-1158.
50. Bruneau P. Search for predictive generic model of aqueous solubility using Bayesian neural nets. *J Chem Inf Comput Sci* 2001; 41(6):1605-1616.
51. Berström CAS, Norinder U, Luthman K, Artursson P. Experimental and computational screening models for prediction of aqueous drug solubility. *Pharm Res* 2002; 19(2):182-188.
52. Chen XQ, Cho SJ, Li Y, Venkatesh S. Prediction of aqueous solubility of organic compounds using a quantitative structure-property relationship. *J Pharm Sci* 2002; 91(8):1838-1852.
53. Schwaighofer A, Schroeter T, Mika S, Laub J, ter Laak A, Sülzle D et al. Accurate solubility prediction with error bars for electrolytes: a machine learning approach. *J Chem Inf Model* 2007; 47(2):407-424.
54. Duchowicz PR, Talevi A, Bruno-Blanch LE, Castro EA. New QSPR study for the prediction of aqueous solubility of drug-like compounds. *Bioorg Med Chem* 2008; 16:7944-7955.



55. Wang J, Hou T, Xu X. Aqueous solubility prediction based on weighted atom type counts and solvent accessible surface areas. *J Chem Inf Model* 2009; 49(3):571-581.
56. Palmer DS, O'Boyle NM, Glen RC, Mitchell JBO. Random forest models to predict aqueous solubility. *J Chem Inf Model* 2007; 47(1):150-158.
57. Taskinen J, Yliruusi J. Prediction of physicochemical properties based on neural network modeling. *Adv Drug Deliv Rev* 2003; 55:1163-1183.
58. Gharagheizi F. QSPR studies for solubility parameter by means of genetic algorithm-based multivariate linear regression and generalized regression neural network. *QSAR Comb Sci* 2008; 27(2):165-170.
59. Yan A, Gasteiger J, Krug M, Anzali S. Linear and nonlinear functions on modeling of aqueous solubility of organic compounds by two structure representation methods. *J Comput Aided Mol Des* 2004; 18:75-87.
60. Delaney JS. Predicting aqueous solubility from structure. *Drug Discov Today* 2005; 10(4):289-295.
61. Llinàs A, Glen RC, Goodman JM. Solubility challenge: can you predict solubilities of 32 molecules using a database of 100 reliable measurements?. *J Chem Inf Model* 2008; 48(7):1289-1303.
62. Hopfinger AJ, Esposito EX, Llinàs A, Glen RC, Goodman JM. Findings of The challenge to predict solubility. *J Chem Inf Model* 2009; 49(1):1-5.
63. Hewitt M, Cronin MTD, Enoch SJ, Madden JC, Roberts DW, Dearden JC. In silico prediction of aqueous solubility: The solubility challenge. *J Chem Inf Model* 2009; 49(11): 2572-2587.

# Computational models for the prediction of drug solubility

Erić Slavica\*, Kalinić Marko, Popović Aleksandar

Department of Pharmaceutical Chemistry, Faculty of Pharmacy,  
University of Belgrade, Vojvode Stepe 450, 11000 Belgrade, Serbia

---

## Summary

Aqueous solubility of a drug is a factor which can significantly influence its oral bioavailability, and can also affect the drug distribution in the body. Consideration of aqueous solubility in early stages of drug discovery and development is vital in reducing the incidence of late-stage drug development failures. The application of computational models for solubility prediction could provide the screening of combinatorial libraries, helping single-out potentially problematic and eliminate compounds with inadequate solubility. In addition to the prediction of solubility from chemical structure, the interpretation of such models can give an insight into structure-solubility relationships and can guide the optimization of structures in order to provide better solubility whilst retaining the activity of the investigated drugs. Development of such models is a complex process that requires consideration of numerous factors which can impact the final model's performance. Different solubility modeling approaches are discussed in this article. Despite intensive research on model development, prediction of the solubility of diverse drugs remains a challenging task. The quality of available experimental data used for modeling of solubility is increasingly recognized as one of the main causes for the limited reliability of many of the proposed models. Therefore, the full potential of the developed modeling methods will only be achieved by greater availability of reliable data obtained by same experimental methodology.

**Key words:** solubility prediction, drug solubility, computational models

---

\*corresponding author: seric@pharmacy.bg.ac.rs