

Originalni naučni rad

UDK 575.087.1:575.18

DOI 10.7251/SVR1510029M

RAZVOJ MODELA ZA PREDVIĐANJE STEPENA SLADIVOSTI KORIŠTENJEM METODA GENETSKOG PROGRAMIRANJA

**Prof. dr Halid Makić¹, mr sc Aida Korčić, dr sc Bahrudin Hrnjica, mr sc.
Samira Dedić, mr sc Samira Hotić**

Biotehnički fakultet, Univerzitet u Bihaću

Mr sc Džemila Agić

Centar za ekologiju i energiju Tuzla

Apstrakt: Sladila su prirodne ili sintetičke tvari koje ostavljaju slatki osjećaj i posjeduju nikakvu ili zanemarivu prehrambenu vrijednost u odnosu na opseg slatkoće. U svijetu postoji znatan interes za razvoj i sintezu novih sladila, a razvijene zemlje uspostavljaju trend za smanjenjem kalorija, pa se pribjegava i korištenju vještačkih zaslađivača. Zbog zdravstvenog rizika upotrebe vještačkih zaslađivača, pribjegava se istraživanju zaslađivača sa visokim stepenom slatkoće, a smanjuje količinski unos.

U ovom radu, razvijeni su kvantitativni modeli za predviđanje logSw (logaritam relativne sladivosti) od 132 spoja s molekulskom masom 152,146-804,872 i slatkoće od 0 do 300000.

Kako bi našli bolju vezu između sladivosti i strukture zaslađivača, koristit će se velik broj prirodnih i umjetnih sladila različitih struktura. Stupanj slatkoće je koreliran s kvantnohemijskim i drugim molekulskim deskriptorima korištenjem evolucijskog modeliranja, zatim izvršena analiza dobivenih modela. Cijeli set podataka nasumično je podijeljen u trening set uključujući 26 spojeva i testni set uključujući 8 spojeva, koje zastupa 11 odabranih molekularnih deskriptora. Vrijednosti logSw su predviđene pomoću GPdotNET softvera.

Za testni set, koeficijenti korelacije 0,999326.

Odabir deskriptora za razvoj modela je odabran na način da može tumačiti i podržati AH/B sistem modela po Shallenbergeru i Acreeu.

Ključne riječi: *zaslađivač, relativna sladivost, neutonske mreže, evolucijsko modeliranje, model, predviđanje, deskriptori.*

UVOD

Zaslađivači su sintetizirane ili prirodne supstance koje se koriste za zamjenu šećera, znatno su veće slatkoće, zavisno od vrste sladila od 10 do 3000 puta. Slatkoća predstavlja bezdimenzijsku veličinu i služi za prikaz relativne slatkoće neke supstance u usporedbi sa saharozom, gdje saharoza nosi vrijednost 1. U svijetu postoji znatan interes za razvoj i sintezu novih

¹ Halid Makić, halid_btf@yahoo.com, 037 228059

sladila, a razvijene zemlje uspostavljaju trend za smanjenjem kalorija, pa se pribjegava i korištenju vještačkih zaslađivača. Također osjetan je i porast rasprava u vezi sigurnosti zaslađivača koji su u zadnje vrijeme dominantni u upotrebi (saharin i ciklambat). Potraga za novim sladilima još uvijek je komplicirana zbog činjenice da odnos između hemijskih tvari i percepcije zaslađivača nije najbolje riješen. U biti sigurnost pogodnih tvari koje se koriste u ishrani mora biti određena. Također i neki drugi kriteriji moraju biti ispunjeni kao na primjer spoj mora biti adekvatno topiv i stabilan u širokoj granici pH i temperature, mora imati čist slatki okus bez drugih ili post okusnih djelovanja, te mora osigurati slatki osjećaj kao što je to glukoza.

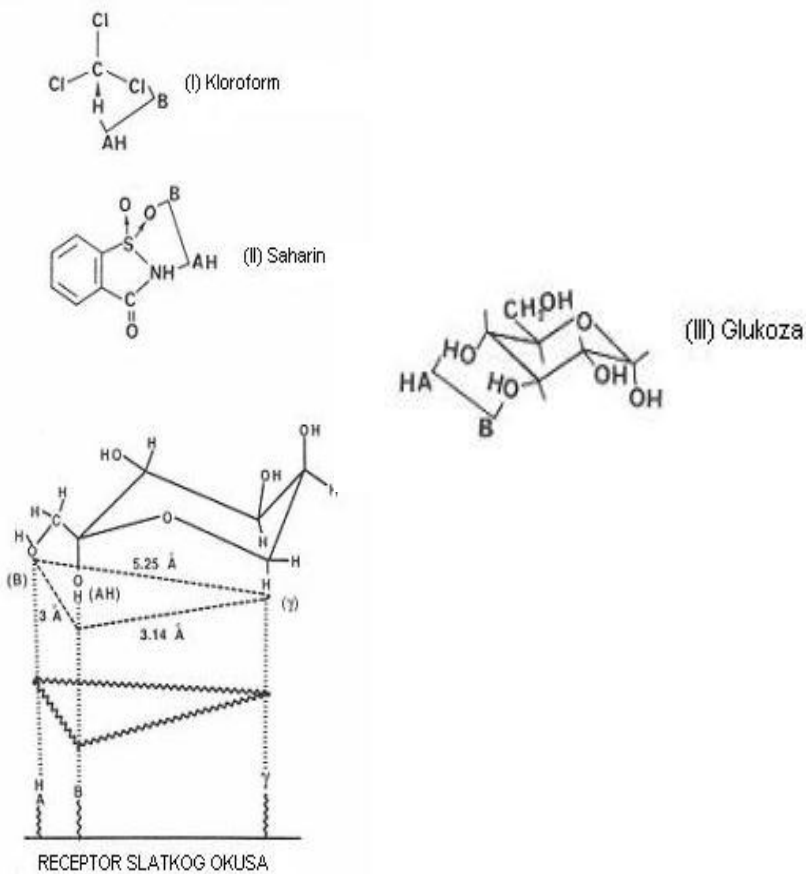
Molekularno modeliranje je univerzalni izraz koji se koristi za teorijske metode (kvantno-mehaničke i empirijske) i kompjuterske tehnike, ali i izgled i ponašanje molekula. Molekularno modeliranje koristi se za tumačenje i predviđanje eksperimentalnih rezultata u hemiji i fizici za koje ne postoje relevantni eksperimentalni podaci, promjene strukture u različitim uvjetima i čini osnovu za razvoj novih materijala i dizajn bioloških aktivnih tvari. Kompjuterska hemija se može definisati kao primjena matematičkih i teorijskih principa za rješavanje hemijskih problema. Možemo reći da je molekularno modeliranje, podgrupa kompjuterske hemije, koja se bavi predviđanjem ponašanja individualnih molekula u nekom hemijskom sistemu.

Istraživanja na sintezama vještačkih zaslađivača su započeta još u 19. vijeku i do danas je razvijeno više različitih klasa sladila. Trend u razvoju zaslađivača ide u pravcu povećanja relativne slatkoće, odnosno povećanje efekta slatkoće sa što manjom koncentracijom zaslađivača. Većina istraživanja u pogledu razvoja modela za predviđanje relativne slatkoće i veze između strukture zaslađivača i relativne slatkoće idu u pravcu QSAR istraživanja i provedena su na određenom skupu srodnih molekula. Do 1963. godine, kada su Robert Shallenberger i Terry Acree predložili AH-B teoriju slatkoće (*Shallenberger i Acree, 1963*). Jednostavno rečeno, oni su predložili teoriju koja glasi: da bi spoj imao okus slatkog, spoj mora sadržavati donatora vodikove veze (AH) i Lewis bazu (B) udaljene oko 0,3 nanometara. Prema toj teoriji AH-B jedinica sladilo veže s odgovarajućom AH-B jedinicom na biološku slatkoću preko receptora za proizvodnju osjećaja slatkoće.

Pored te teorije, Lemont Kier predlaže BX teoriju 1972. godine (*Lemont Kier, 1972.*). Dok su prethodni istraživači napomenuli da se među nekim skupinama spojeva razlikuju okusi slatkoće, ova teorija formalizira ova zapažanja predlažući da spoj mora imati trećinu obavezujućeg mjesta (označen X) koji bi mogao komunicirati s hidrofobnim receptorom na receptoru slatkoće. Kasnije statistički, istraživači su analizirali udaljenosti između pretpostavljenih AH, B i X mjesta u nekoliko vrsta slatkih tvari za procjenu udaljenosti između tih interakcija receptora na slatkoću.

MPA teorija multipoint privitka teoriju slatkoće predlažu Jean-Marie Tinti i Claude Nofre 1991. godine, koja je do sada najviše razrađena teorija (*Tinti i Nofre, 1991*). Ova teorija uključuje ukupno osam interakcija između mjesta zaslađivača i receptora slatkoće, iako sva sladila ne reaguju sa svih osam mjesta. Ovaj model uspješno je riješio napore usmjerene na pronalaženje vrlo jakih sladila, uključujući i najmoćnije vrste sladila poznatih do danas, guanidine sladila. Najmoćniji od tih, lugduname, je oko 225 puta slađi od saharoze.

Prije modernih teorija slatkoće, popularno je bilo povezivati slatkoću sa hidroksil (OH) grupama jer su šećeri zasićeni njima. Međutim, ovo je uskoro postalo meta kritika jer se polihidroksi grupe značajno razlikuju u slatkoći, i mnoge amino kiseline, neke metalne soli i nevezane materije, kao što je kloroform (CHCl_3) i saharin su također slatke. Ipak, bilo je očito da postoje neke zajedničke karakteristike među slatkim supstancama, i proteklih 75 godina teorija vezana za molekularne strukture i sladak okus je napredovala tako da zadovoljavajuće objašnjava zašto određene materije odaju sladak okus. Shallenberger i Acree (*Shallenberger i Acree, 1967*) su prvi predložili AH/B teoriju za tvar punu okusa zajedničku za sve komponente koje imaju sladak okus. Ta je tvar prvo posmatrana kao kombinacija kovalentno vezanog H protona i elektronegativne orbitale smještene na razdaljini od 3 \AA od protona. Tako, susjedni elektronegativni atomi molekule su ključni za slatkoću. Dalje, jedan od atoma mora imati proton vezan hidrogen vezom. Atomi kisika, azota i klora često popunjavaju ove uloge u slatkim molekulama i atomi kisika hidroksil – grupe mogu služiti za AH ili B funkciju u molekuli.



Slika 1: Shematski prikaz veze između AH/B strane β -D-fruktopiranoze i receptora za sladak okus (*Shallenberger i Acree, 1967*)

Stereohemijski zahtjevi su također obavezni i za AH/B grupu komponenti slatkog okusa tako da bi oni bili prikladni za stranu receptora. Za interakciju između aktivne grupe slatkih molekula i receptora okusa se trenutno smatra da se odvija kroz H-vezu AH/B komponenti do sličnih struktura u receptoru okusa. Treće je svojstvo također dodano teoriji da bi ona proširila svoju valjanost na intenzivno slatke supstance. Ovaj dodatak inkorporira prikladno stereohemijski raspoređene regione sklone lipidima na slatkim molekulama, obično označeni sa γ , koji su privučeni sličnim regionima sklonim lipidima na receptorima okusa. Ti regioni na slatkim molekulama su obično metilen ($-\text{CH}_2-$), metil ($-\text{CH}_3$) ili fenil ($-\text{C}_6\text{H}_5$) grupe. Kompletna molekula koja daje sladak okus je geometrijski smještena tako da triangulacijski kontakt svih aktivnih jedinica (AH, B i γ) sa receptorskim molekulama se dešava za intenzivno slatke supstance i ovakav raspored pravi trostranu teoriju slatkoće.

Tako je γ strana izrazito bitna osobina intenzivno slatkih supstanci, ali igra manju ulogu u slatkoći šećera (*Lemont Kier, 1972*). Ona, kako se čini, funkcioniše kroz pružanje pristupa određenim molekulama stranih receptora, i kao takva, utiče na intenzitet slatkoće. Kako su šećeri većinom hidrofobični, ovo svojstvo dolazi do izražaja u ograničenoj mjeri za neke šećere, kao što je fruktoza. Ova komponenta jedinice slatkog okusa je glavni činilac za varijacije u kvaliteti slatkoće koja se posmatra za različite slatke supstance. Ne samo da je bitno za vremenski intenzitet percepcije slatkoće, već je također i vezano za neke interakcije između slatkog i gorkog okusa koji odaju neke komponente.

Spillane (*William J. Spillane, 1996*) je naglasio da AH/B teorija naizgled funkcioniše jako dobro, iako prostorni, hidrofobični/hidrofilični i elektronički efekti također imaju značaja. Shallenberger objašnjava da se slatkoća inicira zbog koncentriranih intermolekularnih, neparalelnih interakcija vodikovih veza između glikofora (grč. *glyks* – slatko, *phoros* – nositi) i dipola receptora. Teškoću u objašnjavanju slatkoće molekula sa različitim hemijskom strukturom je također objasnio Shallenberger i to kako je to dovelo do formiranja drugačijih teorija slatkoće. Primjena teorije slatkoće je pokazala da ima jako velik uticaj u industriji hrane.

Veći eksperimenti sa velikim brojem šećera Birch-a i Lee-a (*Birch i Lee, 1971*) podupiru Shallenbergerovu teoriju slatkoće i naznačuju da je četvrta hidroksil grupa glukopiranosida od jedinstvene važnosti u određivanju slatkoće, vjerovatno donirajući proton AH grupi. Izgleda da je primarna alkohol grupa malog značaja za slatkoću. Zamjena acetil ili azid grupa daje veliku gorčinu šećeru, dok zamjena benzoil grupa uzrokuje bezukusnost.

Kompjuterska hemija se može definisati kao primjena matematičkih i teorijskih principa za rješavanje hemijskih problema. Možemo reći da je molekularno modeliranje, podgrupa kompjuterske hemije, koja se bavi predviđanjem ponašanja individualnih molekula u nekom hemijskom sistemu (*Leach 2001*). Danas je u upotrebi mnogo različitih MM modela i polja sila, od kojih Gaussian uključuje: *Amber, Dreiding* i *UFF*. MM metoda koristi se

uglavnom za račune s vrlo velikim molekulama za koje su tačniji postupci praktično neizvodljivi. Tehnike molekularne mehanike su često empirijski metodi zasnovani na principima klasične fizike, a takvi su računski najbrži.

Kompjuterski molekularni modeli su rezultat matematičkih jednačina koje opisuju položaj i ponašanje elektrona i jezgara. Razvijeno je nekoliko modela koji se koriste zavisno od simulacije. U osnovi postoje dva konceptijski različita postupka koji u većoj ili manjoj mjeri mogu poslužiti u predviđanju molekulskih struktura. Matematički modeli su podijeljeni na klasično-mehaničke i kvantno-mehaničke principe (*Carson i Cobelli, 2001*):

Klasično – mehanički pristup gleda na molekule kao na skup atoma i veza koji se tretiraju kao lopte i rastezljiva nit. Informacije kao što su atomski radijus i promjena dužine veze se koriste da bi se pronašao najbolji položaj atoma. To je brz, i u većini slučajeva relativno tačan metod za lociranje optimalne geometrije molekula. Evo nekoliko metoda molekularne mehanike; MM+ metoda koja koristi opšte polje sila i metoda AMBER, BIO+, OPLIS, koje koriste specijalizirano polje sila.

Kvantno-mehaničke metode rješavaju Schrödingerovu jednadžbu na dva načina; semiempirijski i ab initio (što znači od početka). Semiempirijske metode koriste eksperimentalne podatke da pojednostave Schrödingerovu jednadžbu, da bi mogla brže da se riješi. Za ovo pojednostavljenje su razvijene mnoge metode uključujući i Huckel, Extended Huckel, INDO-S, MINDO. Svaka od njih ima skup parametara koji su zasnovani na eksperimentalnim mjerenjima za mnoga jedinjenja. Neke od semiempirijskih metoda su: Expended Huckel, INDO, CNDO, MINDO/3, MNDO, AM1, PM3 (uključuje prelazne metale), ZINDO/1, ZINDO/S.

Ab initio metode, s druge strane koriste samo matematičke aproksimacije. Ove metode su jako teoretski "čiste", ali s druge strane znatno kompjuterski komplikovanije i oslanjaju se na kompjutere velikih brzina. Neke od STO-G1 metoda su napravljene da poboljšaju tačnost ab initio metoda koristeći manji broj podataka.

Pitanje koje se postavlja je, kako je moguće da se predvidi svojstvo spojeva koji su slabo istraženi ili koji su tretirani nekom od ranijih metoda. Kod ispitivanja neophodno je da se uzme veći broj spojeva koji se tretiraju u klasičnom pristupu koji obuhvata sintezu i izolaciju spoja i na kraju potrebna mjerenja. Na ovakav način se vrši predviđanje određenih svojstava spoja.

U ovom radu su korištene Shallenberger i Acree teorija za predviđanje slatkoće molekula. Kao osnovna veličina korištena je relativna slatkoća odnosno njena logaritamska vrijednost. Relativna slatkoća je veličina koja prikazuje slatkoću neke supstance u poređenju sa saharozom. Hemijski molekularni deskriptor je rezultat logičkog i matematičkog postupka koji transformiše hemijske informacije u numeričke vrijednosti. Svaka molekula nosi tačno određenu vrstu informacija a koju posmatramo na različitim nivoima, te na osnovu tog dolazimo do željenih rezultata. Rad se zasniva na različitim strukturama molekula koje imaju određena fizičko-hemijska i

biološka svojstva. Kod razvoja modela za predviđanje slatkoće koristi se set od 132 molekule koje imaju odgovarajuću vrijednost deskriptora i relativne slatkoće. Kod svake molekule korišteno je 11 različitih deskriptora. U razvoju modela svaka je tretirana sa 13 deskriptora i 132 relativne slatkoće što u zbiru iznosi 1716 podataka. Karakteristike molekula mogu se posmatrati na različitim nivoima pri čemu svaka nosi tačno određenu vrstu informacija.

METODE RADA

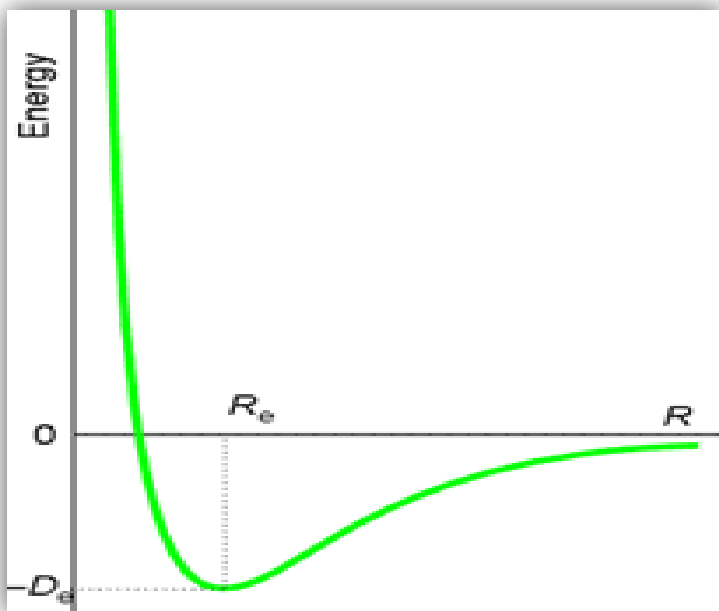
U ovom radu su razvijeni modeli za predviđanje relativne slatkoće na temelju struktura molekula zaslađivača. U tu su svrhu je korišteno genetičko (evolucijsko) modeliranje. Kod evolucijskog modeliranja dobili smo set od 34 molekule koje su filtrirane. Postupak rada za razvoj modela uključuje 4 koraka koji uključuju pripremu skupa podataka koji sadrži spojeve s eksperimentalnom slatkoćom. Potom smo radili izračun molekularnih deskriptora za opis i odabir odgovarajućih deskriptora za predviđanje slatkoće i podjelu cijelog seta podataka na trenig set i testni set nasumično. Na kraju smo radili pravljenje modela pomoću metode genetičkog modeliranja. U potrazi za odgovor koji je zadan ovim radom nepohodni su programi visoke kompleksnosti, njihova primjena nalazi se u rješavanju matematičkih i logističkih problema te modeliranju molekularnih sistema.

U tu svrhu korišteni su sljedeći programi:

1. Gaussian03, GaussView 4.1 (*Gaussian 03, 2004*)
2. GPdotNET (GPdotNET, 2003)

Gaussian (*Gaussian 03, 2004.*) je programski paket za proračun molekularnih svojstava. Paket se sastoji od 80 zasebnih programa integriranih u jedinstvenu cjelinu sa automatskim pozivanjem iz glavnog programa. Sadrži oko 800.000 naredbi. Popis njegovih autora sadrži imena od 82 istraživača. Po broju i raznovrsnosti integriranih postupaka Gaussian je najbogatiji paket za molekularno modeliranje. Zbog svoje univerzalnosti, dostupnosti izvornog koda i povoljne cijene za akademske ustanove, programski paket Gaussian predstavlja svjetski standard za modeliranje ab initio i drugim metodama. Ove metode mogu predvidjeti energiju, molekularne strukture i vibracijsku frekvenciju molekularnih sistema, optička i magnetska svojstva molekula i tvari zajedno s brojnim molekularnim svojstvima koji proizlaze iz ovih osnovnih vrsta računanja.

Izračunavanja se mogu izvoditi na sistemima u fazi gasa ili otopine, i u njihovim osnovnom stanju ili probuđenom stanju. Gaussian može poslužiti kao moćan alat za istraživanje područja hemijskih interesa kao zamjenski efekt, mehanizama reakcija, potencijalnih energetske površina i podražajni energija. Gaussian je u mogućnosti serijski ili paralelno da obrađiva 8 procesnih izračunavanja. Može izračunati energiju i druga povezana svojstva za nekoliko teških atoma u samo nekoliko minuta. Može podnijeti poslove i od nekoliko stotina atoma i može predvidjeti strukturu molekula koje imaju i do nekoliko stotina atoma na istoj veličini kompjuterskog sistema. Odgovarajući veći sistemi mogu se izvoditi na superkompjuterima, na osnovu njihovih posebnih CPU izvedenih karakteristika.



Slika 2. Potencijalna energija kao funkcija udaljenosti unutar jezgre za dvoatomne molekule (Campbel i sar., 2006)

GPdotNET (B. I. Hrnjica 2014) je alat umjetne inteligencije za primjenu genetičkog programiranja i genetičkih algoritama u modeliranju i optimizaciji raznih problema inženjerske prirode. Aplikacija je rađena u NET (mono) načinu u C# programskom jeziku i može se pokrenuti i u Windows i u Linux operativnom sistemu. Razvoj programa je počeo 2006. godine uz postdiplomski rad za modeliranje i optimizaciju evolucijskih algoritama. Kao projekt otvorenog izvora, GPdotNET je prvi put objavljen 5. novembra 2009. godine na stranici www.codeplex.com. GPdotNET je veoma lak za korištenje. Osoba bez znanja o genetičkom programiranju i genetičkom algoritmu, može primjeniti te metode u programu za pronalaženje rješenja problema. Aplikacija se može koristiti za modeliranje bilo kakvog procesa u inženjerstvu, koji se može predstaviti egzaktnim podacima, kao i u obrazovanju za podučavanje studenata o evolucijskim metodama, najviše genetičkom programiranju i genetičkom modeliranju.

GPdotNET podržava slijedeće tipove optimizacije i modeliranja:

1. **Model za konkretne podatke** – modeliranje sa ili bez predviđanja konkretnih podataka koristeći simboličku regresiju sa genetičkim programiranjem
2. **Modeliranje i optimizaciju za konkretne podatke** – modeliranje sa ili bez predviđanja konkretnih podataka koristeći simboličku regresiju sa genetičkim programiranjem i optimiziranje dobivenog GPdotNET modela koristeći genetički algoritam

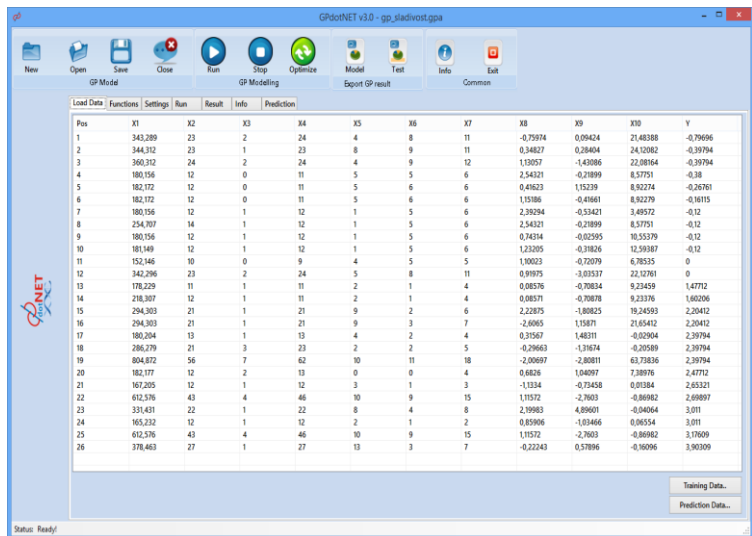
3. **Model vremenskog intervala** – modeliranje vremenskog intervala i predviđanje podataka koristeći simboličku regresiju sa genetičkim programiranjem

4. **Optimizacija analitičke funkcije** – optimizacija analitički definirane funkcije koristeći genetički algoritam.



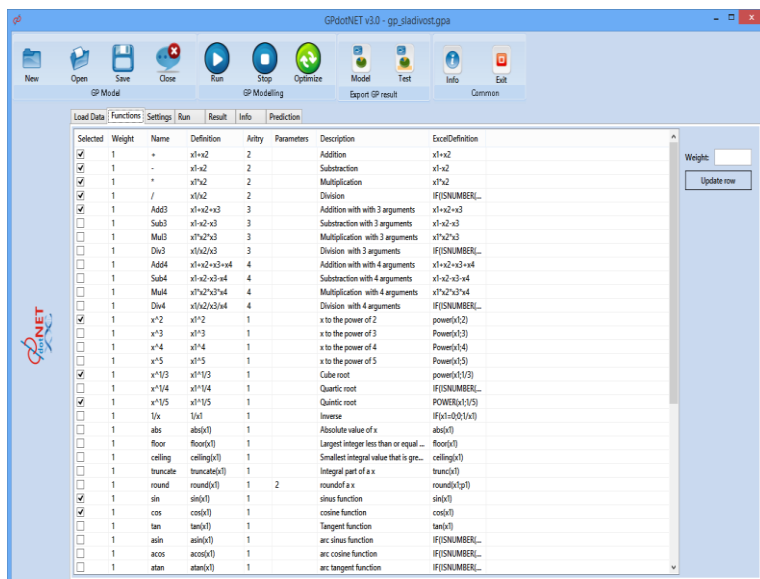
Slika 3: Korisničko sučelje programa GPdotNET (Hrnjica B.I. 2014)

Matematičko modeliranje metodom genetičkog programiranja pomoću GPdotNET alata za modeliranje, optimizaciju i predikciju pomoću evolucijskih algoritama. Učitavanje i prikaz podataka za treniranje GP modela



Slika 4: Učitavanje i prikaz podataka (Hrnjica B.I. 2014)

Definisanje skupa funkcija koje učestvuju u simulaciji i traženju modela.



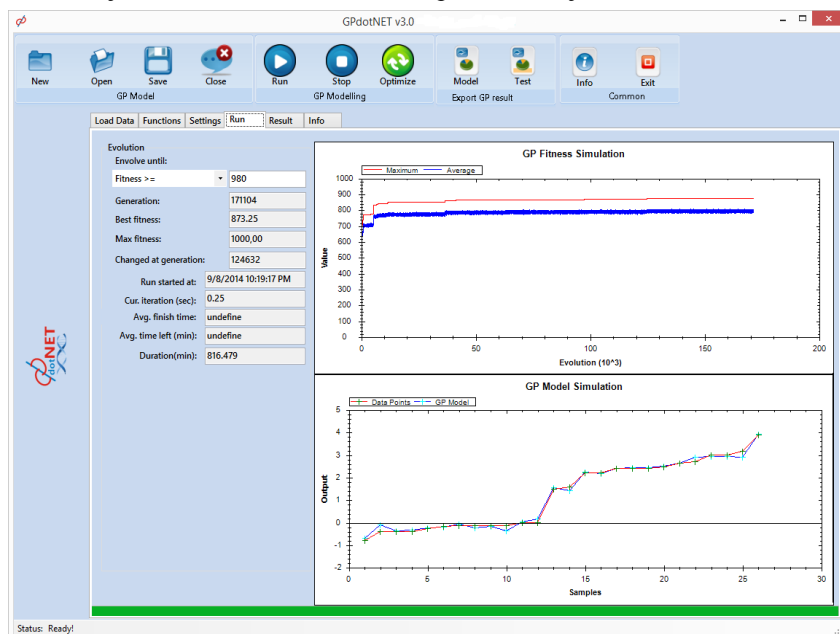
Slika 5: Definisanje skupa funkcija (Hrnjica B.I. 2014.)

REZULTATI I DISKUSIJA

Jedan od prvih pokušaja da se sistemski nađu korelacije između strukture molekule i njegove slatkoće je izradio njemački kemičar, Georg Cohn, 1914. i do danas su urađene mnoge studije i korišteni različiti pristupi u predviđanju, te iznalaženju kvantitativne veze između strukture zaslađivača i relativne slatkoće. Mnogi istraživači su analizirali odnos između slatkoće različite vrste sladila i njihove strukture primjenom kvantitativnih (QSAR) modela odnosa strukture i aktivnosti/svojstva (Walters, 1995). Među prvima su godine 1966, Deutsch i Hansch proveli QSAR istraživanja na nizu od 9 slatkih spojeva supstituenata nitroanilina, i pokazali da koeficijent distribucije oktanol/voda i hidrofobnost supstituenata igraju važnu ulogu u ispoljavanju sladivosti tih spojeva (Deutsch & Hansch, 1966). Od 1980. do 1981, Iwamura istražuje slatkoću na skupini spojeva perillartine i derivata anilina, te nekih analoga L -aspartil dipeptida korištenjem QSAR modela, i pokazao je visoku korelaciju sladivosti i strukture susptituenata (Iwamura, 1980, Iwamura, 1981). Spillane smatra da je slatkoća također bitno u odnosu na hidrofobnost spoja (William J. Spillane, Ryder i Walsh, 1996). To je pokazao na skupini derivata amino-sukcininata. Tarko je uvideo da na jačinu sladivosti spoja povoljno utječe veličina molekule. Analizirao je set za razvoj modela od 97 molekula ($R^2 = 0,83$) i testni set koji je uključivao 24 molekule ($R^2 = 0,86$) (Tarko & Lupescu, 2006).

U novije vrijeme istraživanja Yanga i saradnika, bazirana na QSAR i QSPR korištenjem raznolikog skupa podataka, i testiranjem modela pomoću drugih različitih skupova podataka. Ovakve studije imaju za cilj izgradnju prikladnih modela za predviđanje slatkoće skupa od većeg broja spojeva, koji

je veći nego skup ranije objavljenih studija (Yang i sar., 2011). Potreba pribjegavanju razvoju modela na različitom skupu molekula leži u činjenici da je raspon slatkoće vrlo širok, širok raspon molekulske mase, strukturna i konformacijska raznolikost molekula. Na temelju takvog koncepta baziran je i ovaj rad, s tim da se ovdje koristi primjena vještačke inteligencije i generičkog modeliranja na različitom setu molekula prirodnih i vještačkih zaslađivača.



Slika 6: Simulacija traženja rješenja (Hrnjica B.I. 2014)

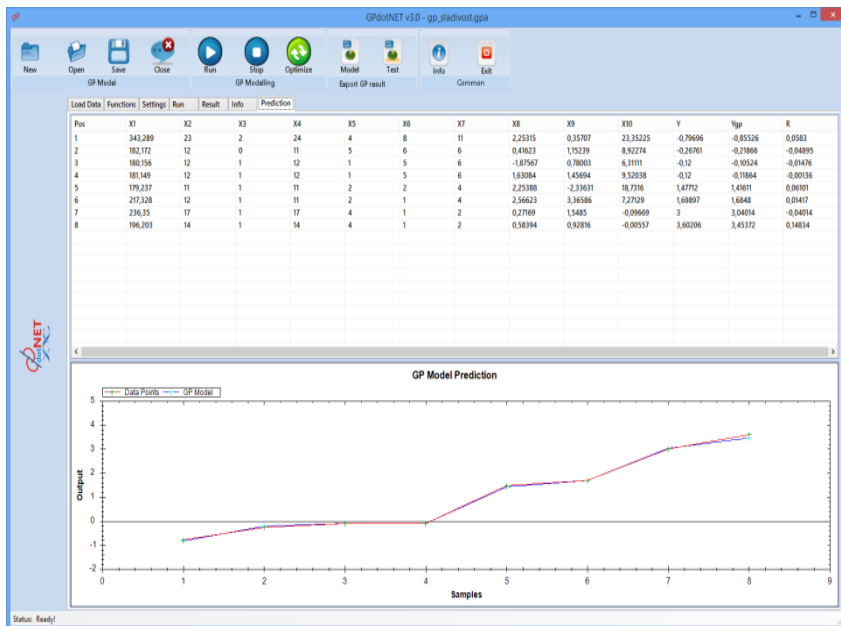
Kako nam je cilj ovog istraživanja naći vezu između struktura molekule zaslađivača i relativne slatkoće, razviti prihvatljiv model za predviđanje sa što manjim odstupanjima izračunatih i eksperimentalnih vrijednosti, provedena je potrebna procedura i korišteno matematičko modeliranje pomoću evolucijskih algoritama.

Model je razvijen koristeći program *GPdotNET V3.0- artificial intelligence tool* (B.I.Hrnjica <http://gpdotnet.codeplex.com>), a analiza izvršena koristeći 12 deskriptora kao ulazne varijable. Prosječno odstupanje između izračunatih i eksperimentalnih vrijednosti iznosi 0,048, koeficijent korelacije 0,999. Iz rezultata je vidljivo da je najbolje rješenje pronađeno u 124.632 generaciji, sa fitness vrijednosti od 873,25. Matematički model koji je dobijen dat je u obliku „drveta izraza“ a prikazan je na slici 6.

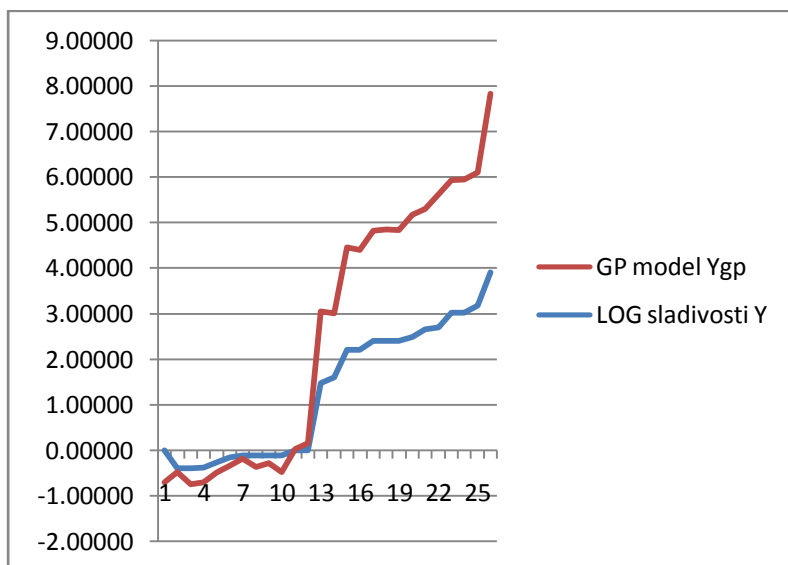
Dobiveni rezultati su znatno bolji u odnosu na QSAR predikcije, što je vidljivo iz reference objavljene u ranije objavljenim radovima (Yang i sar., 2011). Obrazloženje leži u činjenici da smo u radu koristili veći broj dekriptora, a time i varijabli, što je pogodovalo korištenju genetskog algoritma. Pored toga genetski algoritam ima niz prednosti:

- optimizacija sa kontinualnim ili diskretnim varijablama
- ne zahtjeva deriviranje funkcije cilja

- simultano traženje iz širokog intervala funkcije cilja
- korištenje velikog broja varijabli
- prilagođenost za paralelne računare
- optimizacija varijabli sa ekstremno složenim funkcijama cilja
- obezbjeđuje skup optimalnih varijabli, a ne samo jedno rješenje

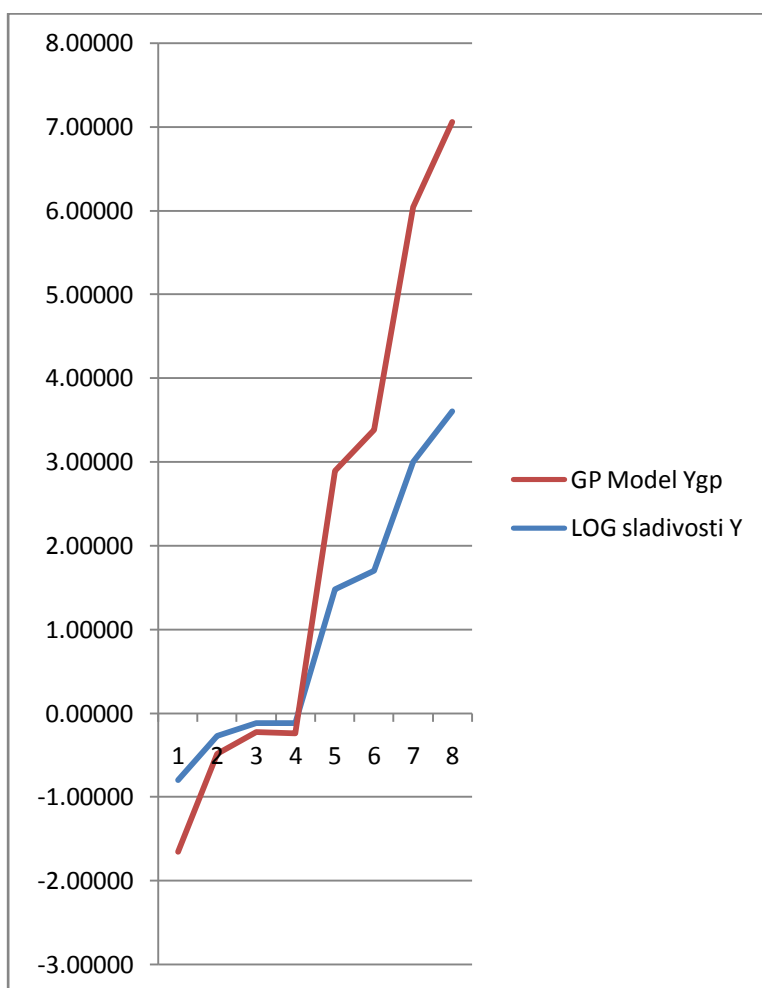


Slika 7: Simulacija testnih podataka (Hrnjica B.I. 2014)



Slika 8: Grafički prikaz pogreške trening seta

Ove prednosti dolaze do izražaja kada klasičnim metodama ne možemo izvršiti zadovoljavajuću optimizaciju. Naravno, genetički algoritam nije najbolje rješenje za rješavanje svakog problema. Npr. klasične metode rješavanja su znatno bolje i brže izračunavaju optimum kada se tiče analitičkih funkcija koje imaju nekoliko varijabli (2,3, i sl). U našem smo radu pronašli da je kombinacija deskriptora korištena za predviđanje relativne sladivosti komponente usko vezana za $\log Sw$. S obzirom da nema ekstremnih odstupanja eksperimentalnih i izračunatih vrijednosti (tabela 4.), zaključili smo da je sladivost visoko vezana sa svojstvima atoma, brojem donorskih i akceptorskih veza, molekulske mase, ali i strukturnih svojstava koje su obuhvatili dati deskriptori. To indicira da odabrani molekularni deskriptori mogu dobro interpretirati AH/B sistem.



Slika 9: Grafički prikaz pogreške testnog seta

Tabela 4: Testni set molekula

Iz razvijenog matematičkog modela dat u jednačini, moguće je dobiti vrijednosti relativne sladivosti korištenjem molekularnih deskriptora za bilo koju novo sintetiziranu molekulu. To potvrđuje i činjenica da je model razvijen na trening setu, a testiran na test setu nasumice odabranih molekula. Korelacija za trening set je 0,99611, a prosječno odstupanje je 0,098346.

Redni br.		1	2	3	4	5	6	7	8
Molekularna masa	X1	343,289	182,172	180,156	181,149	180,156	217,328	236,35	196,203
Broj atoma	X2	23	12	12	12	11	12	17	14
Broj prstenova	X3	2	0	1	1	1	1	1	1
Broj veza	X4	24	11	12	12	11	11	17	14
Broj rotirajućih veza	X5	4	5	1	1	2	2	4	4
Donori H-veza	X6	8	6	5	5	2	1	1	1
Akceptori H-veza	X6	11	6	6	6	4	4	2	2
Dipol 1	X7	2,25315	0,41623	-1,87567	1,63084	2,25388	2,56623	0,27169	0,58394
Dipol 2	X8	0,35707	1,15239	0,78003	1,45694	-2,33631	3,36586	1,54850	0,92816
HF	X9	23,35225	8,92274	6,31111	9,52038	18,73160	7,27129	-0,09669	-0,00557
LOG slatkoće	X10	-0,79696	-0,26761	-0,12000	-0,12000	1,47712	1,69897	3,00000	3,60206
GP Model	Ygp	-0,85525878	-0,21865886	-0,10523482	-0,1186441	1,41609755	1,684805	3,04014097	3,45372271
RESIDUAL	R	0,06	-0,05	-0,01	0,00	0,06	0,01	-0,04	0,15

Korelacija za testni set je 0,999326, a prosječno odstupanje je 0,0475

ZAKLJUČAK

U ovom radu razvijeno je i testirano matematičko modeliranje metodom genetičkog programiranja pomoću GPdotNET alata za modeliranje, optimizaciju i predikciju pomoću evolucijskih algoritama. Pokazalo se da odabrani molekularni deskriptori mogu dobro predvidjeti relativnu slatkoću i mogu također interpretirati teoriju slatkog okusa sladila, kao što je AH/B teorija koju je osnovao Shalenberger i Acree i koja je dosta istražena.

U području istraživanja zaslađivača, glavni trend za sintezu visoke slatkoće sladila za zamjenu postojećih sladila proučavanjem mehanizma za slatkoću. Iako puno faktora utječe na slatkoću spoja, ovaj rad razmatra samo važnost molekularne strukture sladila bez razmatranja komplicirane interakcije između zaslađivača i njegova receptora. Modeli izgrađeni u radu će biti korisni za dizajn novih intenzivnijih sladila.

Matematičko modeliranje metodom genetskog programiranja pomoću GPdotNET alata za modeliranje, optimizaciju i predikciju relativne slatkoće pomoću evolucijskih algoritama daje dobra slaganja između eksperimentalnih i izračunatih vrijednosti. Prosječno odstupanje između izračunate i eksperimentalne vrijednosti iznosi 0,045 a koeficijent korelacije 0,999.

**DEVELOPMENT OF A MODEL OF PREDICTING THE SWEETNESS LEVEL
BY USING METHODS OF GENETIC PROGRAMMING**

**Professor Halid Makić PhD, Aida Korčić MA, Bahrudin Hrnjica PhD,
Samira Dedić MA, Samira Hotić MA, Džemila Agić MA**

Abstract: Sweeteners are natural or synthetic substances that leave a sweet feeling and have no or negligible nutritional value when compared to the scope of sweetness. In the world, there is a considerable interest in the development and synthesis of new sweeteners, and developed countries are establishing a trend for calories reduction, so it resorts to the use of artificial sweeteners. Due to the health risks of artificial sweetener use, it is being resorted to study sweeteners with a high degree of sweetness, and reduced quantity input.

In this paper, the development of quantitative models for predicting logSw (logarithm of the relative sweetness) from 132 compounds with a molecular weight of 152.146 to 804.872 and sweetness from 0 to 300000.

In order to find a better link between sweetness and the structure of sweeteners, a great number of natural and artificial sweeteners of different structures will be used. The degree of sweetness is correlated with quantum chemical and other molecular descriptors using evolutionary modeling, followed by an analysis of the obtained models. The entire dataset is randomly divided into a training set of 26 compounds and test set including 8 compounds, represented by 11 selected molecular descriptors. Values of logSw are provided by GPdotNET software.

For the test set, the coefficients of correlation are 0.999326.

The choice of descriptors for model development is selected in a way that could be interpreted to support the AH/B model system according to Shallenberger and Acree.

Keywords: *sweetener, relative sweetness, neural networks, evolutionary modeling, model, predict, descriptors.*

LITERATURA

1. Birch & Lee, (1971): The chemical basis of sweetness in model sugars, London.
2. Campbell, Neil A.; Brad Williamson; Robin J. Heyden (2006): *Biology: Exploring Life*. Boston, Massachusetts: Pearson Prentice Hall.
3. Carson E. i Cobelli C. (Eds.) (2001): *Modelling Methodology for Physiology and Medicine*, Academic Press, San Diego, CA..
4. Gaussian 03, Revision C.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Wallingford CT, 2004.

5. Hrnjica, B.I. (Bahrudin) (2014): *GPdotNET V3.0- artificial intelligence tool* [kompjuterski program], <http://gpdotnet.codeplex.com>, zadnja posjeta 19/06/.
6. Iwamura, H. (1980): Structure-taste relationship of perillartine and nitro-andcyanoaniline deriavtives. *Journal of Medicinal Chemistry*, 23, 308–312.
7. Iwamura, H. (1981): Structure-sweetness relationship of L-aspartyl dipeptide analogues. A receptor site topology. *Journal of Medicinal Chemistry*, 24, 572–583.
8. Lemont Kier (1972): Molecular theory of sweet taste, *Journal of Pharmaceutical Sciences*, Volume 61, Issue 9, pages 1394-1397.
9. MATLAB and Statistics Toolbox Release (2012b): The MathWorks, Inc., Natick, Massachusetts, United States.
10. McCaleb, Rob (1997): "Controversial Products in the Natural Foods Market". Herb Research Foundation.
11. Leach, R., (2001): *Molecular Modelling: Principles and Applications*.
12. Schallenberger, R.S. i Acre, T.E. (1967): Molecular theory of sweet taste. *Nature*, 216, 480-482.
13. Tarko, L., Lupescu, I. (2006): QSAR Studies on amino-succinamic acid derivatives sweeteners. *Arkivoc* 13: 22–40.
14. Tinti, J.M., Nofre, C. (1991): Design of sweeteners, Walters DE Washington.
15. Yang, X., Chong, Y., Yan, A., Chen, J. (2011): In-silico prediction of sweetness of sugars and sweeteners. *Food Chem* 128(3):653–8.
16. William, J. Spillane, Caroline, A. Ryder, M. Rachel Walsh, Patrick, J. Curran, Danny G. Concagh, Siobhan N. Wall, (1996): Sulfamate sweeteners.
17. Walters Tinti J.M., Nofre, C. (1991): Design of sweeteners, Walters DE Washington.