

Različne predstavitve kemijskih struktur

V kemiji je opis kemijskih struktur bistvenega pomena za večino obdelav eksperimentalnih podatkov, ki se tičejo lastnosti in opisov kemijskih elementov, spojin, mešanic in procesov s katerimi se obdelujejo ter končnih izdelkov teh procesov.

V prvem obdobju kemijskih raziskav so za opise predmetov kemijskega raziskovanja zadostovala splošna imena, ki so bila uveljavljena od prej (voda, zrak, kislina, itd...), kmalu so si posamezniki pričeli izmišljati imena glede na najbolj očitne lastnosti spojin s katerimi so se ukvarjali in jih preiskovali. Kasneje se je pokazalo, da za uspešno izmenjavo različnih kemijskih znanj potrebujejo poenoteno imenovanje spojin ali celo standardizacijo. Zato je skrbel poseben oddelek mednarodne organizacije za čisto in uporabno kemijo IUPAC (International Union of Pure and Applied Chemistry). Oddelek za nomenklaturo kemijskih spojin se je ukvarjal z uvajanjem enotnih kemijskih imen po vsem svetu, s sprejemanjem pravili za poimenovanje spojin, z usklajevanjem različnih idej in novih pogledov na tem področju, s standardizacijo itd.

Do pričetkov 50. in 60. let preteklega stoletja, ko so se pričele uveljavljati računalniške in avtomatizirane obdelave lastnosti kemijskih spojin, je IUPAC-ova notacija kemijskih spojin povsem zadostovala. Z naraščanjem števila spojin in hkrati z rastočo zmogljivostjo računalnikov pa so rasle tudi zahteve po najrazličnejših obdelavah. S tem so, žal, začeli naraščati tudi problemi v zvezi z zapisovanjem kemijskih struktur. Glavni problem je v tem, da bi morala biti iz imena avtomatično razpoznavna vsaj 2-d če že ne 3-d struktura kemijske spojine. Poudarek je na besedi avtomatično, kar pomeni s pomočjo računalnika. Problem se je izkazal za precej težjega, kot so v začetku predvidevali in napovedovali računalniški strokovnjaki. Drugi del istega problema je avtomatična generacija imen za nove spojine iz njihove 2-dimenzionalne slike. Število vseh spojin, ki so jih poznali bodisi v raziskovalnih laboratorijih ali v industriji je bilo še v leta 1975 na vsem svetu okrog 3 milijone, leta 2009 pa govorimo že o 46 milijonih spojinah (leta 2006 28 milijonov, 2003 23 milijonov). Začasna rešitev obeh problemov je bila v uvedbi t.i. CAS registrske številke (Chemical Abstract Registry Number) ali "CAS Reg. No.". Slabost CAS številke je v tem, da na prvi pogled kemiku ne pove ničesar o kemijski strukturi ali lastnostih spojine. Struktura in lastnosti so dosegljive samo s pomočjo računalnika in to šele v primeru, ko imamo hkrati s CAS številko tudi dostop do podatkovnih baz, ki vsebujejo povezave med CAS številkami, imeni spojin, njihovimi strukturami in lastnostmi.

Naslednja težava CAS številke je v tem, da z njo, razen identifikacije spojin, lastnosti in strukture, ki je dosegljiva, ne moremo početi nič uporabnega, npr., primerjave med dvema spojinama. Oziroma, kar je najvažnejše, s CAS številko ne moremo matematično obdelovati skupine spojin oziroma njihovih struktur.

What is CAS REGISTRY?	<p>CAS REGISTRY is the most authoritative collection of disclosed chemical substance information, containing more than 46 million organic and inorganic substances and 60 million sequences. (view current numbers)</p> <p>CAS REGISTRY covers substances identified from the scientific literature from 1957 to the present, with additional substances going back to the early 1900s.</p>
What is a CAS Registry Number (CASRN)?	<p>Each CAS Registry Number (often referred to as a CAS Number):</p> <ul style="list-style-type: none">• Is a unique numeric identifier• Designates only one substance• Has no chemical significance• Is a link to a wealth of information about a specific chemical substance
What does a CAS Registry Number look like?	<p>A CAS Registry Number is a numeric identifier that can contain up to 10 digits, divided by hyphens into three parts.</p> <p>The right-most digit is a check digit used to verify the validity and uniqueness of the entire number. For example, 58-08-2 is the CAS Registry Number for caffeine.</p>
How does CAS assign Registry Numbers?	<p>A CAS Registry Number is assigned to a substance when it enters the CAS REGISTRY database. Numbers are assigned in sequential order to unique, new substances identified by CAS scientists for inclusion in the database.</p>

<p>What kinds of compounds does the CAS REGISTRY contain?</p>	<p>CAS REGISTRY contains a wide variety of substances, including the world's largest collection of:</p> <ul style="list-style-type: none"> •Organic compounds •Inorganic compounds •Metals •Alloys •Minerals •Coordination compounds •Organometallics •Elements •Isotopes •Nuclear particles •Proteins and nucleic acids •Polymers •Nonstructurable materials (UVCBs)
<p>How frequently is the CAS REGISTRY database updated?</p>	<p>Approximately 12,000 new substances are added each day. See the most up-to-date CAS REGISTRY figures.</p>

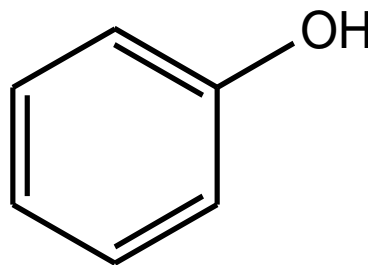
The Latest CAS Registry Number® and Substance Count		
Date	05/11/2009 08:42:17 EST	
Count	46,726,072	organic and inorganic substances
	60,936,919	sequences
CAS RN	1145355-08-3	is the most recent CAS Registry Number

<p>Why have CAS Registry Numbers become the world standard?</p>	<p>Since CAS Registry Numbers are not dependent upon any system of chemical nomenclature, they can:</p> <ul style="list-style-type: none">• Provide a reliable common link between the various nomenclature terms used to describe substances• Serve as an international resource for chemical substance identifiers used by scientists, industry, and regulatory bodies
<p>Why do regulatory agencies rely on CAS Registry Numbers?</p>	<p>Chemical compounds can be described in many different ways:</p> <ul style="list-style-type: none">• Molecular formula• Structure diagram• Systematic names• Generic names• Proprietary or trade names• Trivial names <p>A CAS Registry Number, however, is unique and specific to only one substance regardless of how many other ways the substance can be described.</p> <p>Governmental agencies have found CAS Registry Numbers ideal for keeping track of substances:</p> <ul style="list-style-type: none">• They are unique• They can be validated quickly and reliably• They are internationally recognized

Where can I find CAS REGISTRY Numbers?	<p>Find CAS Registry Numbers in:</p> <ul style="list-style-type: none">• SciFinder• STN databases• CD-ROM products• CAS printed publications• Other publications<ul style="list-style-type: none">○ Governmental regulatory agency commercial chemical inventories○ Selected Elsevier databases○ Handbooks, guides, technical reports○ Other printed reference works○ Chemical catalogs <p>In a hurry? Contact CAS Registry Number Lookup Service and ask for 24-hour turnaround (Mon-Fri).</p>
How can I obtain or request assignment of a CAS Registry Number?	<p>CAS Client Services offers you several options for confirming and assigning CAS Registry Numbers and obtaining CA Index Names. Note that there is a fee associated with this service.</p>
Why is it important to come to CAS for Registry Numbers?	<p>While printed publications may contain outdated information, the CAS REGISTRY database is updated daily providing your best source for the latest CAS Registry Number information.</p> <p>CAS is the source and final authority for CAS Registry Numbers. Other resources may have associated an incorrect CAS Registry Number with a compound, and use of that incorrect CAS Registry Number in searches could lead you to irrelevant or inappropriate information.</p>

CAS - številka fenola

108-95-2



ChemFinder.Com
Database & Internet Searching

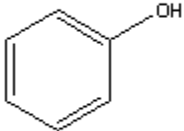


Search

Enter a Chemical Name, CAS Number, Molecular Formula or Weight.
Use * for partial names (e.g. ben*).
Search here for free. For professional searching, use [ChemINDEX](#).

<http://chembiofinder.cambridgesoft.com/chembiofinder/SimpleSearch.aspx>

ChemFinder – result for phenol

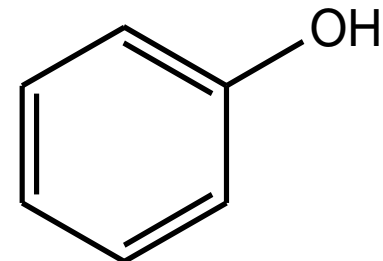
108-95-2	Tools	OpenChem
	BUY AT CHEMACX.COM	VIEW LINKS
	VIEW CHEMDRAW STRUCT	ADD COMPOUND
	VIEW CHEM3D MODEL	ADD/CHANGE PROPERTY
		ADD LINK
	CAS RN Lookup	
	THE MERCK INDEX	
	NCI DATABASE	

ChemFinder – result for phenol

Formula	C_6H_6O	Molecular Weight	94.1128
CAS RN	108-95-2	Melting Point (°C)	40.5
ACX Number	X1001304-8	Boiling Point (°C)	181.7
Density	1.07	Vapor Density	3.24
Refractive Index	1.5418	Vapor Pressure	0.35
Evaporation Rate		Water Solubility	Slightly sol. 8.28 g/100 mL
Flash Point (°C)	79		

CAS - številka fenola

108-95-2



<http://pubchem.ncbi.nlm.nih.gov/>



PubChem Compound

GO

PubChem Compound

GO

1: CID: 996



phenol, carboic acid ...

IUPAC: phenol

MW: 94.111 | MF: C6H6O

2: CID: 8750



Sodium phenate, Phenol sodium ...

IUPAC: sodium phenoxide

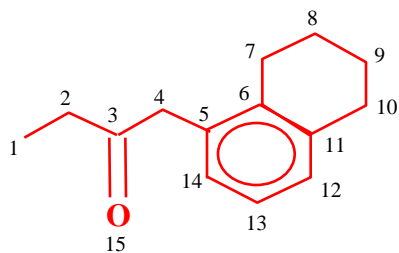
MW: 116.093 | MF: C6H5NaO

Matrika stičnosti in povezovalna tabela

Računalniška obdelava kemijskih struktur je bila v prvem obdobju omejena na dvodimenzionalne predstavitve. Zaradi omejenih računalniških zmogljivosti, so bile možne le osnovne obdelave, n.pr. tiskanje 2-dimenzionalne slike z linijskimi tiskalniki, ki so imeli na voljo le 128 alfanumeričnih znakov, razstavljanje kemijskih struktur na fragmente, iskanje struktur z danimi substituentami ali skeleti ipd.

Glavna izhodna oblika za obdelavo kemijskih struktur z računalnikom, ki je v uporabi še danes (seveda ima danes celo vrsto izboljšav in dodatkov), je povezovalna tabela. Matematično gledano je povezovalna tabela izšla iz kvadratne matrike stičnosti, ki pove kateri atomi v molekuli so medseboj povezani in kateri ne.

Struktura



Matrika stičnosti:- 1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
3	0	1	1	1	0	0	0	0	0	0	0	0	0	0	1
4	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
5	0	0	0	1	1	1	0	0	0	0	0	0	0	0	1
6	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0
7	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0
8	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
9	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
10	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0
11	0	0	0	0	0	1	0	0	0	1	1	1	0	0	0
12	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0
13	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0
14	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0
15	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1

Matrika stičnosti:- 2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	C	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	C	1	0	0	0	0	0	0	0	0	0	0	0	0
3	0	1	C	1	0	0	0	0	0	0	0	0	0	0	2
4	0	0	1	C	1	0	0	0	0	0	0	0	0	0	0
5	0	0	0	1	C	a	0	0	0	0	0	0	0	a	0
6	0	0	0	0	a	C	1	0	0	0	a	0	0	0	0
7	0	0	0	0	0	1	C	1	0	0	0	0	0	0	0
8	0	0	0	0	0	0	1	C	1	0	0	0	0	0	0
9	0	0	0	0	0	0	0	1	C	1	0	0	0	0	0
10	0	0	0	0	0	0	0	0	1	C	1	0	0	0	0
11	0	0	0	0	0	a	0	0	0	1	C	a	0	0	0
12	0	0	0	0	0	0	0	0	0	0	a	C	a	0	0
13	0	0	0	0	0	0	0	0	0	0	0	a	C	a	0
14	0	0	0	0	a	0	0	0	0	0	0	0	a	C	0
15	0	0	2	0	0	0	0	0	0	0	0	0	0	0	O

Povezovalna tabela

	1	Atom	Vez	Atom	Vez
1	C	2	1		
2	C	3	1		
3	C	4	1	15	2
4	C	5	1		
5	C	6	a	14	a
6	C	7	1	11	a
7	C	8	1		
8	C	9	1		
9	C	10	1		
10	C	11	1		
11	C	12	a		
12	C	13	a		
13	C	14	a		
14	C				
15	O				

Povezovalna tabela je skrajšana oblika matrike stičnosti. Njena prednost je v tem, da z naraščanjem velikosti molekule (števila atomov v molekuli) narašča linearno (samo v dolžino) in ne kvadratnično v obe smeri kot matrika stičnosti. Poleg tega lahko v vsako vrstico, ki opisuje posamezen ne-vodikov atom, dodajamo lastnosti ali opise tega atoma. V današnjih oblikah povezovalnih tabel so najpogostejše informacije v vsaki vrstici (x,y,z) koordinate in naboj atoma. Lahko so dodani še Van der Waalsovi radii, oddaljenost do atoma števila 1, ali kaka druga informacija, ki je vezana na ta atom. Nekatere povezovalne tabele imajo na koncu dodano še informacije o posameznih obročih, t.j.: število atomov v posameznih obročih in seznam števil atomov v vsakem od njih.

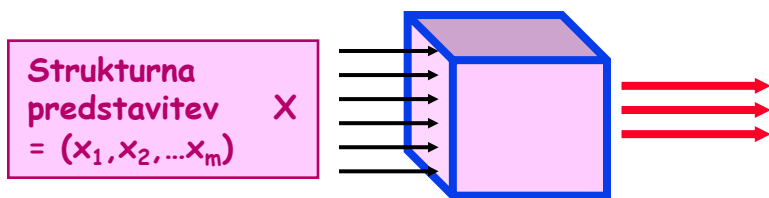
Uniformne predstavitve kemijskih struktur

Največja slabost povezovalnih tabel je v tem, da jih ne moremo uporabljati pri izračunu modelov. V nadaljevanju bomo obravnavali le take načine predstavitve kemijskih struktur, ki jih je možno uporabiti pri skupinskih obdelavah struktur (grupiranje, modeliranje, klasifikacija itd.). Zahteve, ki jih mora strukturna predstavitev izpolniti, da jo lahko uporabimo za računalniško obdelavo večjega števila struktur hkrati, so:

1. enoličnost,
2. uniformnost,
3. reverzibilnost in
4. translacijska in rotacijska invarianca.

ad 1. Enoličnost pomeni, da ima vsaka kemijska struktura eno samo notacijo in obratno, vsaki notaciji pripada ena sama kemijska strukturo.

ad 2. Uniformnost je ena najvažnejših zahtev za računalniško obdelavo. Z njo zagotovimo uresničitev pogoja, da je vsaka struktura predstavljena z istim številom (m) enakih "deskriptorjev".



ad 3. Reverzibilnost ali obrnljivost omogoča, da kemijsko strukturo rekonstruiramo iz predstavitve. Ta lastnost je tesno povezana z zahtevo po uniformnosti, če uniformnost ni izpolnjena, je obrnljivost sorazmeroma lahko dosegljiva in obratno, če je predstavitev uniformna, je reverzibilnost izvedljiva zelo težko.

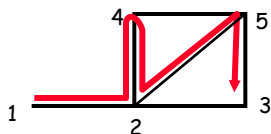
ad 4. Translacijska in rotacijska invarianca je v določeni meri zajeta že v prvi zahtevi, žal pa ni temu vedno tako. Nekatere predstavitve so lahko enolične, vendar pa rotacijsko niso invariantne, kar pomeni, da je treba strukture pred izračunom njihove notacije orientirati v skupnem internem koordinatnem sistemu. Taka orientacija oziroma določitev internega koordinatnega sistema ima dobre in slabe lastnosti. Slabost je v tem, da je treba tak koordinatni sistem sploh uvesti in drugič, da je treba vsako spojino prilagajati temu koordinatnemu sistemu, kar ni enostavno.

Uniformna predstavitve kemijskih struktur

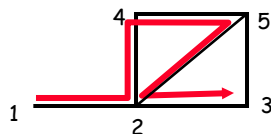
Do danes ne poznamo še nobene strukturne predstavitve, ki bi ustrezala vsem štirim zahtevam, omenjenih na prejšnji strani. Najpogosteje se odpovemo reverzibilnosti, to je temu, da bi kemijsko strukturo lahko izvedeli, če je podana samo predstavitev. Tipični predstavniki takih predstavitev so kemijski deskriptorji. Kemijski deskriptorji so vrednosti, ki jih izračunamo na podlagi različnih informacij o strukturah spojin. Če želimo predstaviti strukturo ali skupino kemijskih struktur, jih opišemo z m deskriptorji: $X = (d_1, d_2, \dots, d_m)$.

Deskriptorji kemijskih lastnosti so zelo različni. Najčešče uporabljeni deskriptorji so:

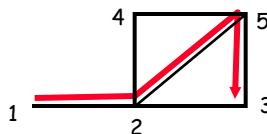
1. Konstitucijski deskriptorji, ki odražajo samo kemijsko sestavo spojine, ne da bi se pri tem ozirali na geometrijo ali strukturo molekule. Primeri zanje so:
 - skupno število vseh in/ali posameznih atomov (C, H, O, S, N, F, Cl, Br, I, P itd.) v molekuli,
 - relativna števila atomov glede na število vseh atomov v molekuli,
 - število vseh vezi in absolutna ter relativna števila enojnih, dvojnih, trojnih in aromatskih vezi v molekuli,
 - število vseh obročev, število benzenskih obročev, št. benzenskih obročev/št. atomov v molekuli,
 - molekularna teža itd.
2. Topološki deskriptorji, odražajo topologijo (razporeditev atomov) v molekuli. Znanih je nekaj desetih različnih topoloških indeksov, ki opisujejo različne topološke lastnosti molekule. Glavne topološke lastnosti "sprehod", "veriga", "pot" in "cikel" so prikazane na spodnjem primeru na isti molekuli (na istem grafu):



Sprehod je poljubna neprekinjena povezava med dvema točkama (atomoma) v grafu (molekuli), ki dovoljuje večkratni obisk iste točke in celo prehod iste vezi.



Veriga ima ostrejšje zahteve kot sprehod: ne dovoljuje dvakratne uporabe iste vezi, lahko pa obiščemo dvakrat isto točko.



Pot ima najostrejšje zahteve: ne dovoljuje dvakratne uporabe iste vezi niti dvakratnega obiska iste točke.

Če se pot začne in konča v isti točki se imenuje cikel

Topološki deskriptor je običajno vsota nekaj specifičnih topoloških lastnosti vseh atomov v molekuli. Npr.: Wienerjev indeks:

$$W = \sum_i \sum_j d_{ij}$$

je polovična vsota vseh najkrajših poti d_{ij} med vsemi možnimi pari parov atomov i in j v molekuli.

3. Geometrijski deskriptorji zahtevajo poznavanje (x,y,z) koordinat vseh atomov. Primeri geometrijskih indeksov so:
- vztrajnostni momenti okrog osi x, y in z. Npr.: $I_x = \sum_i m_i r_{ix}^2$ m_i je masa i-tega atoma, r_i pa oddaljenost od osi x,
 - molekularni volumen,
 - površina molekule (na Van der Waalsovih radiih, ali glede na kakšne druge enote - prostorske bloke),
 - gravitacijski indeks $G = \sum_i m_i m_j / r_{ij}^2$
4. Elektrostatski deskriptorji odražajo lastnosti povezane z naboji na atomih, ki jih dobimo z različnimi kvantno-mehanskimi izračuni. Primer takega indeksa je:
- topološki elektronski indeks $T = \sum_i q_i q_j / r_{ij}^2$ pri katerem lahko produkte nabojev q_i in q_j jemljemo z vseh ali samo iz povezanih atomov.
5. Kvantno-kemijski deskriptorji so tisti, ki jih računamo na podlagi energijskih nivojev dobljenih s kvantno-kemijskimi izračuni. Primeri za take indekse so:
- vrednosti, ki so asociirane s koeficienti najvišje in najnižje zasedenih molekularnih orbital (HOMO/LUMO highest/ lowest occupied molecular orbital),
 - delni naboji izračunani z Millikanovo populacijsko analizo in izpeljane vrednosti,
 - skupna energija, energijske razlike med HOMO in LUMO, različni potenciali itd,
6. Termodinamski deskriptorji, kot npr.:
- termodinamska toplota formiranja molekule pri 300 °K,
 - vibracijska entalpija in vibracijska entropija molekule pri 300 °K,
 - translacijska entalpija in translacijska entropija molekule pri 300 °K,
 - rotacijska entropija molekule pri 300 °K itd.

Avtokorelacijski deskriptor A

S pomočja vsakega od naštetih deskriptorjev se da narediti še posebne avtokorelacijske deskriptorje, ki so v nekem smislu "celostni" opisi molekul. Splošna oblika za komponento a(i) avtokorelacijskega deskriptorja $A = [(a(0), a(1), a(2), \dots, a(d), \dots, a(n))]$ se da napisati kot avtokorelacijska funkcija:

$a(d) = \sum_j L(j) * L(j+d)$ kjer je $L(j+d)$ lastnost L na atomu, ki je od j-tega atoma oddaljen d vezi. Vsota teče po vseh atomih molekule.

"Spektru-podobna" predstavitev kemijskih struktur

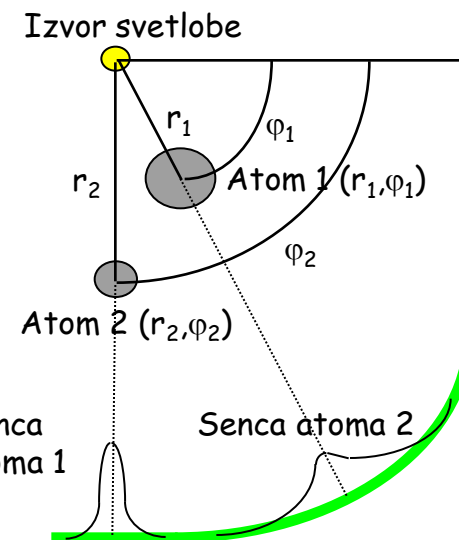
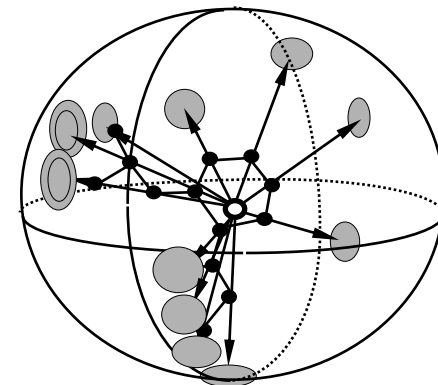
"Spektru-podobna" predstavitev je enolična, uniformna in reverzibilna predstavitev kemijskih struktur, ki pa žal ni rotacijsko invariantna in zateva predhodno ureditev vseh struktur v isti koordinatni sistem. Zamisel "spektru-podobne" predstavitve izhaja iz spektroskopije. Molekula je postavljena v navidezno kroglo v sredini katere je izvor svetlobe, ki na obod krogle riše sence atomov. Položaj, velikost in ostrina senc so odvisni od relativnega položaja vsakega atoma glede na svetilo. Če projiciramo atome samo na tri pravokotne ekvatorje omenjene krogle, dobimo tri "spektre" katerih ločljivost določa natančnost s katero opišemo molekulsko strukturo. Obliko sence $I(\varphi_i)$ j -tega atoma, katerega položaj je v 2-dimenzionalnem prostoru določen s polarnima koordinatama ρ_j in, v odvisnosti od kota φ_i , ki teče od 0 do 360°, opišemo z Lorentzovo krivuljo $I(\varphi_i)$:

$$I_j(\varphi_i) = \frac{\rho_j}{(\varphi_i - \varphi_j)^2 + \sigma_j^2}$$

$$s(\varphi_i) = \sum_{j=1}^N \frac{\rho_j}{(\varphi_i - \varphi_j)^2 + \sigma_j^2}$$

$$s(\varphi_i) = \sum_{j=1}^N \frac{\sqrt{x_j^2 + y_j^2}}{[\varphi_i - \arccos(\frac{x_j}{\sqrt{x_j^2 + y_j^2}})]^2 + \sigma_j^2}$$

Lorentzova krivulja ima zvonasto obliko, maksimum pri kotu φ_j in širino na polovični višini maksimuma σ . Celoten "spekter" molekule $s(\varphi_i)$ je sestavljen iz vsote posameznih Lorentzovih senc vseh N atomov. "Spektralna" predstavitev molekule je torej aditivna in lahko v primerih ko obravnavamo molekule z istim skeletom vse sence atomov skeleta enostavno odštejemo od celotnih spektrov. Če polarne koordinate pretvorimo v kartezične, dobimo malo bolj zapleteno obliko, ki pa jo lahko razstavimo na projekcije v treh ravninah (x,y) , (x,z) in (y,z) , lahko pa uporabljamo tudi samo en del (če je molekula planarna).



"Spekter" z ločljivostjo 1 kotne stopinje (razlika med dvema kotoma $\varphi_{m+1} - \varphi_m = 1^\circ$) v projekciji na tri ekvatorialne kroge lahko v celoti prikažemo kot vektor s 1080 komponentami s_i , $i=1, \dots, 1080$.

$$\mathbf{S} = (s_1, s_2, s_3, \dots, s_{357}, s_{358}, s_{359}, s_{360}, s_{361}, s_{362}, s_{363}, \dots, s_{718}, s_{719}, s_{720}, s_{721}, s_{722}, s_{723}, \dots, s_{1078}, s_{1079}, s_{1080})$$

$$s_i = \sum_{j=1}^N \frac{\sqrt{x_j^2 + y_j^2}}{[\varphi_i - \arccos(\frac{x_j}{\sqrt{x_j^2 + y_j^2}})]^2 + \sigma_j^2}$$

(x,y)

$$s_i = \sum_{j=1}^N \frac{\sqrt{x_j^2 + z_j^2}}{[\varphi_i - \arccos(\frac{x_j}{\sqrt{x_j^2 + z_j^2}})]^2 + \sigma_j^2}$$

(x,z)

$$s_i = \sum_{j=1}^N \frac{\sqrt{y_j^2 + z_j^2}}{[\varphi_i - \arccos(\frac{y_j}{\sqrt{y_j^2 + z_j^2}})]^2 + \sigma_j^2}$$

(z,y)

