

UNIVERSITATEA "BABEȘ-BOLYAI" CLUJ-NAPOCA
FACULTATEA DE CHIMIE ȘI INGINERIE CHIMICĂ

TEZĂ DE DOCTORAT

REZUMAT

LAVINIA LORENA PRUTEANU

CONDUCĂTOR DE DOCTORAT
PROF. DR. LORENTZ JÄNTSCHI

2018

UNIVERSITATEA "BABEȘ-BOLYAI" CLUJ-NAPOCA
FACULTATEA DE CHIMIE ȘI INGINERIE CHIMICĂ

**SIMILARITATEA MOLECULARĂ ÎN INVESTIGAREA PROPRIETĂȚILOR CHIMICE
ȘI ACTIVITĂȚILOR BIOLOGICE**

REZUMAT

CONDUCĂTOR DE DOCTORAT
PROF. DR. LORENTZ JÄNTSCHI

2018

CUPRINS

INTRODUCERE	1
STADIUL ACTUAL AL CUNOAȘTERII	4
CAPITOLUL 1. STRUCTURA MOLECULARĂ	5
1.1. Optimizarea geometriei moleculare	7
1.2. Simetria și proprietățile moleculare	8
1.3. Grafuri de structură chimică	10
CAPITOLUL 2. ASOCIEREA STRUCTURII MOLECULARE CU PROPRIETĂȚILE ȘI ACTIVITĂȚILE BIOLOGICE CARACTERISTICE	17
2.1. Clustering - Gruparea moleculelor	17
2.2. Măsuri de calcul ale similarității	18
2.3. Matrici de calcul a similarității	24
2.4. Descriptorii moleculari și amprente numerice	31
2.5. Relații cantitative structură chimică - proprietate/activitate	38
2.6. Machine learning/ Învățare mecanică	40
CAPITOLUL 3. COMPORTAMENTUL MOLECULAR EXPRIMAT LA NIVEL TRANSCRIPTOMIC	43
3.1. Metode de evaluare a exprimării genice - Microarray vs RNA-Seq	44
3.2. Analiza moleculară	48
3.2.1. Gene exprimate diferențial	49
3.2.2. <i>Signature matching</i>	51
3.2.3. Rețele de interacțiune de tip proteină-proteină	52
3.2.4. Rețele de co-expresie genică	53
3.3. Aplicabilitatea noțiunilor de similaritate în contextul exprimării genice	54

CONTRIBUȚII PERSONALE	56
CAPITOLUL 4. CADRUL CERCETĂRII	57
4.1. Aranjamentele moleculare și proprietățile acestora	58
4.2. Similaritatea în clase de compuși	59
4.3. Relații cantitative structură-proprietate	61
4.4. Comportamentul arsenatilor <i>in vitro</i>	62
4.5. Efectele terapeutice ale lactobacililor urmărite <i>in vitro</i>	64
CAPITOLUL 5. MODELE DE HIDRATARE A IONILOR MONOVALENȚI	66
5.1. Materiale și metode	66
5.1.1. Optimizarea geometriei	66
5.1.2. Gruparea moleculelor ion-apă	67
5.1.3. Construirea cuștilor de congeneri	67
5.1.4. Simetria și stabilitatea structurală	69
5.2. Rezultate și discuții	70
5.2.1. Moleculele stabile ion-apă	71
5.2.2. Simetria moleculară	73
5.2.3. Electronegativitatea	74
5.3. Concluzii	80
CAPITOLUL 6. SIMILARITATEA MOLECULARĂ ÎN CLASE DE ANTI-INFLAMATOARE	81
6.1. Materiale și metode	81
6.1.1. Colectarea datelor	81
6.1.2. Similaritatea utilizând scheletul molecular	84
6.1.3. Analiza componentelor principale și maparea datelor	89
6.2. Rezultate și discuții	89
6.2.1. Generarea legăturii de filogenie dintre specii	90
6.2.2. Clusterii de similaritate	99
6.2.3. Maparea similarității utilizând scaffold-urile moleculare	104

6.2.4. Analiza componentelor principale	106
6.3. Concluzii	109

CAPITOLUL 7. MODEL DE RELAȚIE CANTITATIVĂ STRUCTURĂ-PROPRIETATE ÎN CLASE DE STEROIDE 110

7.1. Materiale și metode	110
7.1.1. Optimizarea geometriei	110
7.1.2. Construcția hipermoleculei	111
7.1.3. Descriptorii topologici	111
7.1.4. Pozițiile semnificative	113
7.2. Rezultate și discuții	114
7.2.1. Model de regresie obținut pe baza pozițiilor semnificative	114
7.2.2. Validarea modelului prin procedura leave-one-out	116
7.2.3. Validarea modelului prin procedura training vs. Test	116
7.3. Concluzii	118

CAPITOLUL 8. EFECTUL TERAPEUTIC AL ARSENAȚILOR ASUPRA LINIILOR CELULARE DE CANCER DE SÂN 119

8.1. Materiale și metode	119
8.1.1. Cultura de celule și tratamentul	119
8.1.2. Analiza microarray	120
8.1.3. Analiza bioinformatică a datelor de microarray	121
8.2. Rezultate și discuții	124
8.2.1. Evaluarea efectului arsenăților în cultura de matrigel	124
8.2.2. Evaluarea efectului arsenăților asupra autofagiei și apoptozei	125
8.2.3. Vizualizarea datelor în funcție de răspunsul celulelor în urma tratamentului	128
8.2.4. Analiza componentelor principale	130
8.2.5. Analiza setului semnificativ de gene exprimate diferențial	131
8.3. Concluzii	134

CAPITOLUL 9. EFECTUL TERAPEUTIC AL LACTOBACILILOR ASUPRA CELULELOR INTESTINALE EPITELIALE PORCINE LA NIVEL DE EXPRESIE GENICĂ	136
9.1. Materiale și metode	136
9.1.1. Cultura de celule	136
9.1.2. Prepararea mixului de lactobacili (LB)	136
9.1.3. Tratamentul bacterian	137
9.1.4. Extracția ARN-ului total pentru analiza microarray	137
9.1.5. Analiza microarray	138
9.1.6. Analiza statistică a datelor de microarray	138
9.1.7. Validarea datelor de exprimare genică	139
9.2. Rezultate și discuții	140
9.2.1. Rezultatele microarray	140
9.2.2. Clasificarea funcțională a genelor exprimate diferențial	141
9.2.3. Validarea prin Real-Time PCR	146
9.2.4. Analiza căilor celulare	146
9.3. Concluzii	148
CONCLUZII ȘI PERSPECTIVE VIITOARE	150
ANEXE	154
REFERINȚE	167
LISTA PUBLICAȚIILOR	

Rezumat scurt

În lucrarea prezentată este creat un context interdisciplinar utilizând noțiuni de chimie și de biologie-biochimie investigând comportamentul molecular al structurilor chimice cât și efectele lor la nivelul sistemelor biologice. Scopul studiului a fost de a descrie și înțelege legătura dintre structura moleculară a compușilor anorganici și organici, gruparea compușilor pe baza similarității și efectele terapeutice ale unor compuși anorganici și organici la nivel celular. Compușii testați *in vitro* includ arsenati și amestecuri de probiotice, pornind în prealabil de la înțelegerea comportamentului molecular prin prisma similarității a compușilor anorganici (reprezențați prin ioni metalici) și organici (reprezențați prin clase de compuși antiinflamatori, derivați steroidici) în apă și respectiv *in silico*. Stadiul actual al cunoașterii structurat în trei capitole (Capitolul 1. Structura moleculară; Capitolul 2. Asocierea structurii moleculare cu proprietățile și activitățile biologice caracteristice; Capitolul 3. Comportamentul molecular exprimat la nivel transcriptomic) conține o parte din informația privind legătura dintre structura moleculară în raport cu geometria și topologia moleculară și a grupării compușilor pe baza similarității precum și comportamentul molecular la nivel transcriptomic în scop terapeutic. În jurul conceptelor prezentate în stadiul actual, au derivat o serie de abordări prezentate în partea de „Contribuții personale” structurate în cinci capitole (Figura 0). Înțelegerea comportamentului molecular al compușilor anorganici în apă (Capitolul 5), al compușilor organici *in silico* (Capitolul 6) precum și a relațiilor dintre proprietățile structurale și activitatea biologică (Capitolul 7) a continuat cu caracterizarea efectului compușilor la nivel celular în scop terapeutic (desfășurată în Capitolele 8 și 9).

Cuvinte cheie

Similaritate moleculară, proprietăți chimice, indici topologici, descriptori și amprente numerice, anti-inflamatoare, activități biologice, clase de compuși, arsenati, probiotice, exprimare genică, transcriptomică

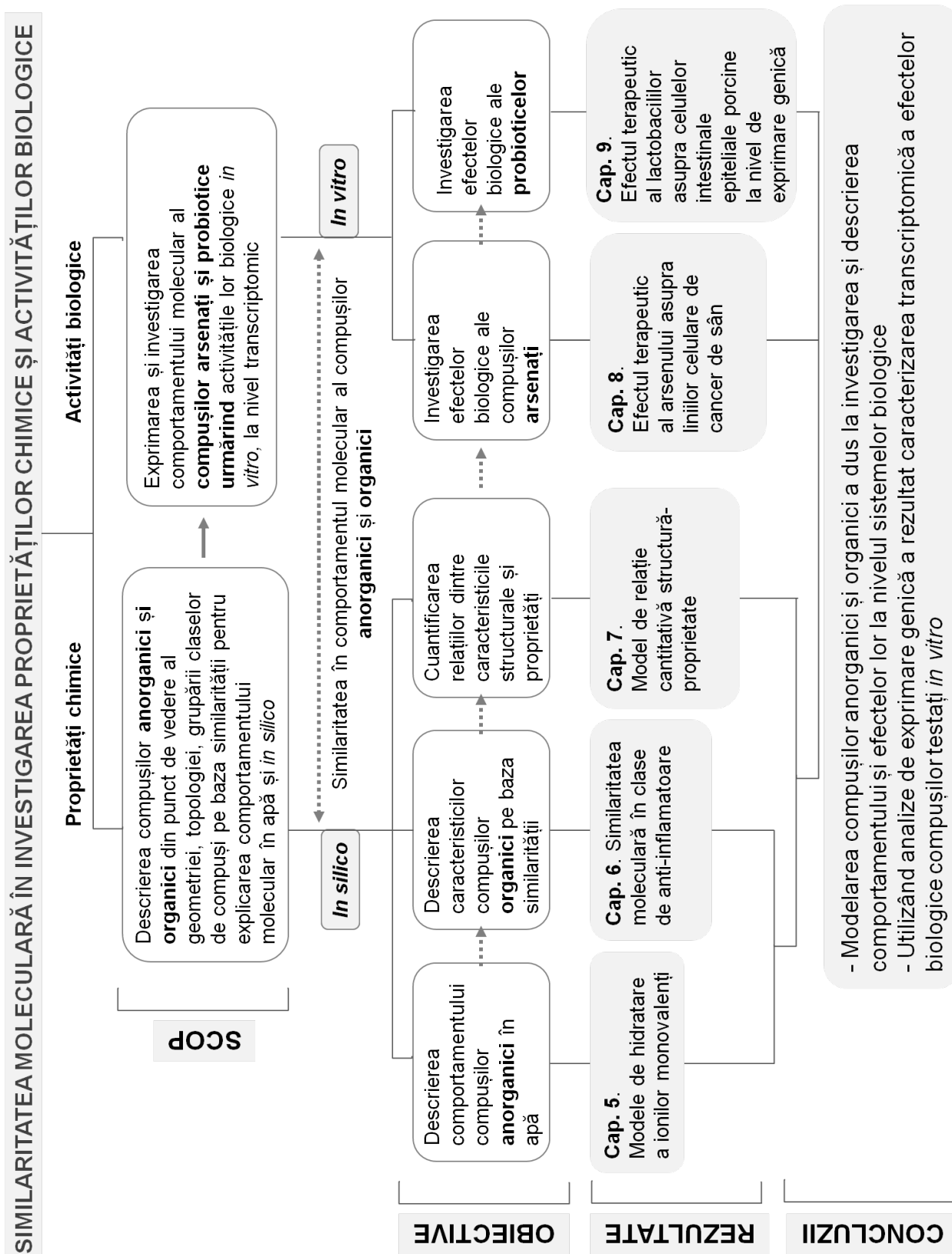


Figura 0. Reprezentarea cadrului cercetării

CAPITOLUL 1. STRUCTURA MOLECULARĂ

Proprietățile moleculare și relațiile dintre structura chimică și proprietăți/bioactivități au servit în construcția modelelor de interacțiune și a efectelor compușilor asupra sistemelor biologice, inclusiv la nivel de exprimare genică. Interdisciplinaritatea este considerată un factor fundamental în activitatea de cercetare desfășurată.

1.3. Grafuri de structură chimică

Un graf (Figura 1.3.a) este definit ca fiind o pereche de două seturi, un set cu un număr finit de puncte (vârfuri) și un set de perechi neordonate de puncte distincte, în timp ce un diagraf (Figura 1.3.b) conține un set cu număr finit de puncte împreună cu un set de perechi ordonate de puncte distincte. Într-un multigraf (Figura 1.3.c) se consideră două puncte ce pot fi unite prin mai mult de o linie (Diudea et al., 2001).

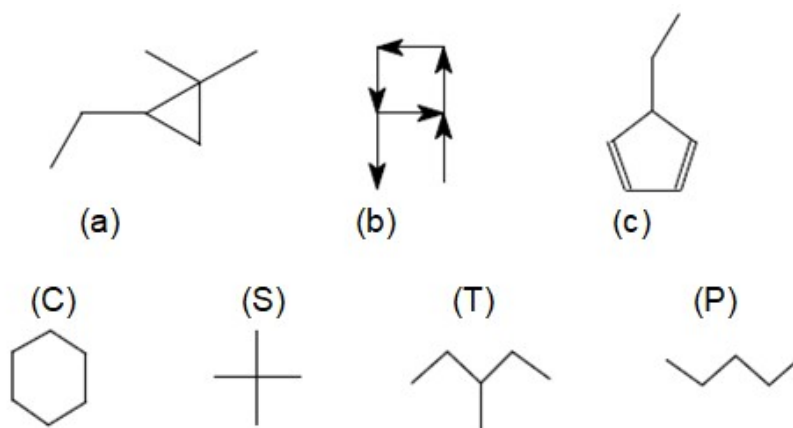
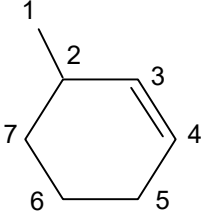


Figura 1.3. Reprezentări grafuri. Graf (a), diagraf (b), multigraf (c); C (cycle - ciclu): lanț care pleacă și revine în unul și același vârf; S (star - stea): set de vârfuri unite cu un vârf comun (Sv' , cu $v' = v-1$); T (tree - arbore): structură ramificată; P (path) cale: lanț (catenă) neramificat(ă) - adaptat după (Diudea et al., 2001).

Reprezentările numerice corespunzătoare grafurilor (Figura 1.3) sunt integrate în algoritmi de calcul. Reprezentările pot fi de tip funcțional pentru formule moleculare empirice, greutatea moleculară (Harsa et al., 2014), diagrame constituționale (2D), coordonate atomice (Kochev et al., 2003), indici topologici, matrici de adiacență,

distanță (Diudea et al., 2007, 2001; Harsa et al., 2014), etc., câteva astfel de exemple fiind redată în Tabelul 1.1:

Tabel 1.1. Exemple de reprezentări funcționale pentru structuri moleculare

Denumiri	Reprezentări
Graf numerotat	
Tabel de conectivitate	1,1,2 2,1,1 2,2,1 3,2,4 4,3,2
Notăție liniară	SMILES: <chem>CC(C(=O)C1=CC=CC=C1)O</chem> ROSDAL: 1-2-3-4=5-6-7-2
Matrice de adiacență	$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$
Matrice de distanță	$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 0 & 1 & 2 & 3 \\ 2 & 1 & 0 & 1 & 2 \\ 3 & 2 & 1 & 0 & 1 \\ 4 & 3 & 2 & 1 & 0 \end{pmatrix}$

Similaritatea dintre două entități (atomi, molecule, fragmente moleculare, chiar și grupuri de molecule) este definită ca fiind o funcție a proprietăților lor cumulate, bazată fiind pe relațiile uneia cu altele ce aparțin aceluiași grup, această similaritate poate fi de asemenea estimată pe baza topologiei și a geometriei moleculare.

CAPITOLUL 2. ASOCIEREA STRUCTURII MOLECULARE CU PROPRIETĂȚILE ȘI ACTIVITĂȚILE BIOLOGICE CARACTERISTICE

Pentru analiza similarității dintre molecule (similaritatea inter-moleculară), una dintre soluții este de a grupa moleculele în funcție de cât de multe asemănări sunt găsite între acestea, iar analiza acestor grupuri poate fi realizată urmărind similaritatea la nivelul grupurilor (clusterilor).

2.1. Clustering - Gruparea moleculelor

Clustering sau gruparea este o modalitate prin care diferite seturi de obiecte/molecule sunt atribuite anumitor clase în funcție de gradul lor de similaritate redat prin intermediul caracteristicilor/proprietăților spațiale sau temporale. Pentru a identifica caracteristicile specifice se utilizează calcule empirice (Çöltekin et al., 2010; Dodge, 2011).

2.2. Măsuri de calcul ale similarității

Definind similaritatea moleculară ca fiind o măsură a gradului de suprapunere a câte două molecule într-un anumit spațiu, amintim în cele ce urmează o serie de măsuri cu ajutorul cărora se poate caracteriza similaritatea dintre molecule.

Tabel 2.1. Măsuri de calcul ale similarității frecvent utilizate reprezentate prin ecuațiile binare (0;1) pentru a câte două molecule (A;B)

Măsuri	Denumiri	Formule de calcul
Distanțe	Euclidiană	$D_{A,B} = [a + b - 2x]^{1/2}$
	Hamming (Manhattan)	$D_{A,B} = a + b - x$
	Soergel	$D_{A,B} = \frac{1 - x}{a + b - x}$
Coeficienți de similaritate	Tanimoto (Jaccard)	$S_{A,B} = \frac{x}{a + b - x}$
	Dice (Czekanowski)	$S_{A,B} = \frac{2x}{a + b}$

Măsur	Denumiri	Formule de calcul
Coeficienți de corelație	Cosine (Ochiai)	$S_{A,B} = \frac{x}{\sqrt{ab}}$
	Pearson	$S_{A,B} = \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$
Notății	A, B - două molecule distincte a - caracteristică molecula A b - caracteristică molecula B c - caracteristică distinctă (altă decât a) molecula A d - caracteristică distinctă (altă decât b) molecula B $x = a \cap b$ $D_{A,B}$ = distanța dintre două molecule $S_{A,B}$ = similaritatea dintre două molecule	

Pentru expresia similarității a câte două molecule, cele mai comune măsuri considerate sunt distanțele Euclidiană, Hamming (Manhattan), Soergel (Albrecht et al., 2004; Allen et al., 2001; Bero et al., 2017), coeficienții de similaritate Tanimoto (Rogers și Tanimoto, 1960), Indicele Dice (Czekanowski) (Bero et al., 2017; Willett et al., 1998) și coeficienți de corelație Cosine (Ochiai) (Willett et al., 1998), Pearson (Pearson, 1895) (Tabel 2.1).

Tanimoto sau coeficientul Jaccard (Reynolds et al., 1992) este un coeficient standard, dependent de absența/prezența unei caracteristici, reprezentând astfel o măsură 2D (Ma et al., 2011; Willett et al., 1998). În tabelul de mai jos (Tabelul 2.2) este ilustrat principiul coeficientului Tanimoto.

Tabel 2.2. Ilustrarea principiului de calcul al coeficientului Tanimoto

a	b	c	d	a	a	d	c	b	A = 3
a	b	b	c	d	a	c	a	a	B = 4
1	1	0	0	0	1	0	0	0	$ A \cap B = 3$
$\text{Tanimoto } (A_a, B_a) = \frac{ A \cap B }{ A + B - A \cap B } = 0.75$									
a, b, c, d - caracteristici ale moleculelor A, respectiv B									

Astfel luând două molecule (A,B) și având caracteristicile reprezentative (a, b, c, d) fiecărei molecule, se consideră 0 absența unei anumite caracteristici (spre exemplu

absența a unui atom într-o anumită poziție în cazul unei molecule, față de cealaltă) iar cu 1 prezența respectivei caracteristici (Tabelul 2.2)..

Rezultatele obținute în urma măsurilor de similaritate (Figura 2.1) pot fi utilizate în stabilirea de topologii, în clasificarea și gruparea componentelor ce aparțin unui set de molecule de interes (Diudea et al., 2001).

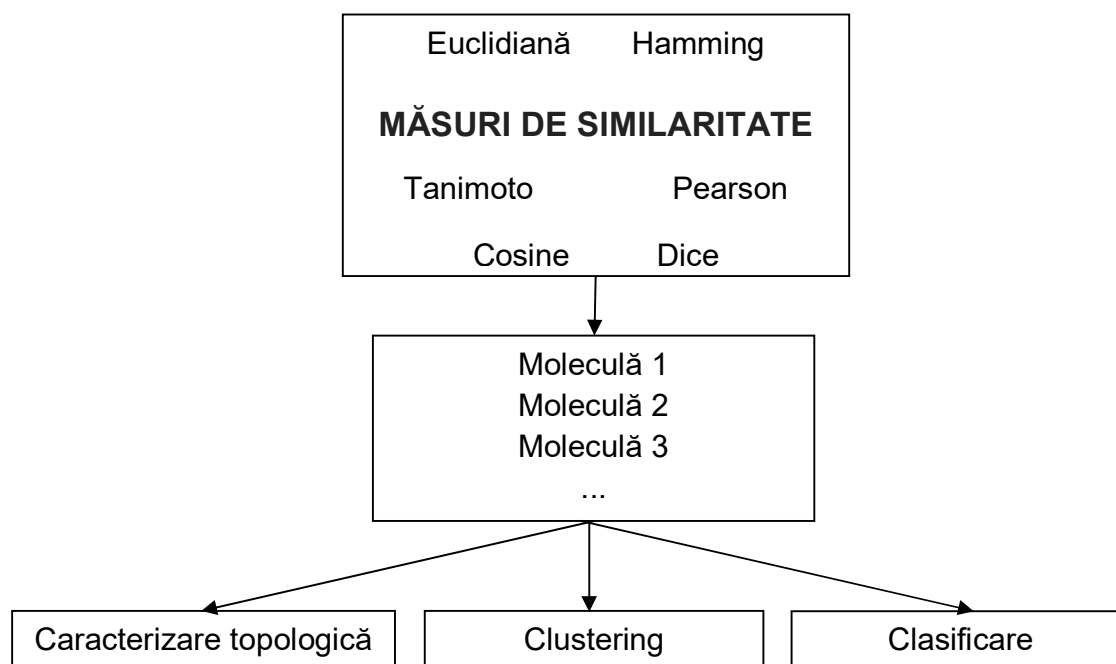


Figura 2.1. Măsurile de similaritate pot conduce la caracterizarea moleculelor din punct de vedere topologic, la gruparea pe seturi de molecule (clustering) în funcție de similaritate și chiar la clasificarea acestora pe baza anumitor proprietăți/caracteristici.

2.3. Matrici de calcul a similarității

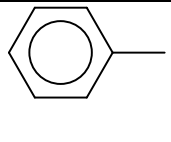
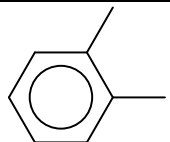
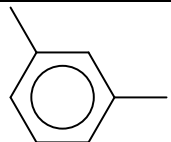
Matricile pot reprezenta o măsură a distanței dintre două puncte/atomi (A, B) (matricea de adiacență) în cadrul unei molecule aflate într-un spațiu, ce se exprimă printr-o funcție $d(A, B)$, unde $d(A, B) = 0$ dacă $A = B$ și $d(A, B) = 1$ în caz contrar. Rezultatul unei asocieri de valori numerice corespunzătoare distanțelor atomice în cadrul moleculelor are în vedere 3 axiome și anume:

- 1) $d(A, B) \geq 0$ și $d(A, B) = 0, A=B$ (reflexivitate)
- 2) $d(A, B) = d(B, A)$ (simetrie sau comutativitate)

3) $d(A, B) + d(B, C) \geq d(A, C)$ (subaditivitatea)

Măsurile de similaritate în spațiul 2D permit reprezentarea topologică a structurilor chimice (Tabel 2.3):

Tabel 2.3. Reprezentări moleculare distincte (molecula 1 – m1, molecula 2 – m2, molecula 3 – m3)

m1	m2	m3
		
m1, m2, m3 = molecula 1, 2, 3		

Se pot reprezenta astfel distanțele între atomi (Jäntschi, 2000) - oricare dintre matricile exemplificate mai jos (Tabel 2.4, Tabel 2.5, Tabel 2.6, Tabel 2.7) e aplicabilă similar. Se construiesc matricile de adiacență (Tabel 2.4) corespunzătoare structurilor chimice (m1, m2, m3) (Tabel 2.3):

Tabel 2.4. Reprezentare matrice de adiacență pentru moleculele m1, m2, m3

[Ad]	1	2	3	4	5	6	7
1	0	1	1	0	0	0	0
2	1	0	0	1	0	0	0
3	1	0	0	0	0	1	0
4	0	1	0	0	1	0	0
5	0	0	0	1	0	1	0
6	0	0	1	0	1	0	1
7	0	0	0	0	0	1	0
m1							

[Ad]	1	2	3	4	5	6	7	8
1	0	1	1	0	0	0	0	0
2	1	0	0	1	0	0	0	0
3	1	0	0	0	0	1	0	1
4	0	1	0	0	1	0	0	0
5	0	0	0	1	0	1	0	0
6	0	0	1	0	1	0	1	0
7	0	0	0	0	0	1	0	0
8	0	0	1	0	0	0	0	0
m2								

[Ad]	1	2	3	4	5	6	7	8
1	0	1	1	0	0	0	0	1
2	1	0	0	1	0	0	0	0
3	1	0	0	0	0	1	0	0
4	0	1	0	0	1	0	0	0
5	0	0	0	1	0	1	0	0
6	0	0	1	0	1	0	1	0
7	0	0	0	0	0	1	0	0
8	1	0	0	0	0	0	0	0
m3								

Se știe că numerotarea atomilor este arbitrară, deci folosirea acestor matrici în această formă nu e posibilă. Suma elementelor în m2 și m3 este aceeași (8 în [Ad]) (Tabel 2.4), 61 în [Di] (Tabel 2.5)) deci suma elementelor nu discriminează cele două structuri. Problema numerotării se poate elimina dacă se consideră indici topologici (Diudea et al., 2001).

Matricea de distanță (Tabel 2.5) poartă o informație suplimentară față de adiacență - în locul valorilor de "0" (exceptând diagonala principală) există acum distanțele

topologice (numărul de legături) între atomi (Jäntschi, 2000; Minkin, 1999).

Tabel 2.5. Reprezentare matrice de distanță pentru moleculele m1, m2, m3

[Di]	1	2	3	4	5	6	7
1	0	1	1	2	3	2	3
2	1	0	2	1	2	3	4
3	1	2	0	3	2	1	2
4	2	1	3	0	1	2	3
5	3	2	2	1	0	1	2
6	2	3	1	2	1	0	1
7	3	4	2	3	2	1	0
m1							

[Di]	1	2	3	4	5	6	7	8
1	0	1	1	2	3	2	3	2
2	1	0	2	1	2	3	4	3
3	1	2	0	3	2	1	2	1
4	2	1	3	0	1	2	3	4
5	3	2	2	1	0	1	2	3
6	2	3	1	2	1	0	1	2
7	3	4	2	3	2	1	0	3
8	2	3	1	4	3	2	3	0
m2								

[Di]	1	2	3	4	5	6	7	8
1	0	1	1	2	3	2	3	1
2	1	0	2	1	2	3	4	2
3	1	2	0	3	2	1	2	2
4	2	1	3	0	1	2	3	3
5	3	2	2	1	0	1	2	4
6	2	3	1	2	1	0	1	3
7	3	4	2	3	2	1	0	4
8	1	2	2	3	4	3	4	0
m3								

În Tabelul 2.6 și Tabelul 2.7 au fost sunt considerate secvențe de numere și descriptori topologici (aplicabile atât la linii cât și la coloane, rezultând același lucru) care permit analiza de similaritate intramoleculară (în interiorul moleculei) (Tabelul 2.6) și respectiv intermoleculară (între molecule) (Tabelul 2.7):

Tabel 2.6. Reprezentare matrice de distanță pentru moleculele m1, m2, m3 – similaritate intramoleculară

[Di]	1	2	3	4	5	6	7	0.1.2.3.4	Σ
1	0	1	1	2	3	2	3	1.2.2.2.0	12
2	1	0	2	1	2	3	4	1.2.2.1.1	13
3	1	2	0	3	2	1	2	1.2.3.1.0	11
4	2	1	3	0	1	2	3	1.2.2.2.0	12
5	3	2	2	1	0	1	2	1.2.3.1.0	11
6	2	3	1	2	1	0	1	1.3.2.1.0	10
7	3	4	2	3	2	1	0	1.1.2.2.1	15
m1									

[Di]	1	2	3	4	5	6	7	8	0.1.2.3.4	Σ
1	0	1	1	2	3	2	3	2	1.2.3.2.0	14
2	1	0	2	1	2	3	4	3	1.2.2.2.1	16
3	1	2	0	3	2	1	2	1	1.3.3.1.0	12
4	2	1	3	0	1	2	3	4	1.2.2.2.1	16
5	3	2	2	1	0	1	2	3	1.2.3.2.0	14
6	2	3	1	2	1	0	1	2	1.3.3.1.0	12
7	3	4	2	3	2	1	0	3	1.1.2.3.1	18
8	2	3	1	4	3	2	3	0	1.1.2.3.1	18
m2										

[Di]	1	2	3	4	5	6	7	8	0.1.2.3.4	Σ
1	0	1	1	2	3	2	3	1	1.3.2.2.0	13
2	1	0	2	1	2	3	4	2	1.2.3.1.1	15
3	1	2	0	3	2	1	2	2	1.2.4.1.0	13
4	2	1	3	0	1	2	3	3	1.2.2.3.0	15
5	3	2	2	1	0	1	2	4	1.2.3.1.1	15
6	2	3	1	2	1	0	1	3	1.3.2.2.0	13
7	3	4	2	3	2	1	0	4	1.1.2.2.2	19
8	1	2	2	3	4	3	4	0	1.1.2.2.2	19
m3										

Se poate observa că secvențele de numere "1.2.3.1.0", "1.2.2.2.0", "1.3.2.1.0", "1.2.2.1.1", "1.1.2.2.1" identifică corect similaritățile între atomi - similaritate intramoleculară.

Nu același lucru se poate spune despre sumă deși în acest caz este și ea un identificator al similarității. Concluzia este că pentru analiza de similaritate între atomi secvențele de tipul "1.2.3.1.0", "1.2.2.2.0", "1.3.2.1.0", "1.2.2.1.1" trebuie folosite (Jäntschi, 2000). Pentru analiza de similaritate intermoleculară (Tabel 2.7) o modalitate de extindere este de a însuma, iar alta este de a păstra distinctibilitatea.

Tabel 2.7. Reprezentare matrice de distanță pentru moleculele m1, m2, m3 – similaritate intermoleculară

[Di]	1	2	3	4	5	6	7	0.1.2.3.4	Σ
1	0	1	1	2	3	2	3	1.2.2.2.0	12
2	1	0	2	1	2	3	4	1.2.2.1.1	13
3	1	2	0	3	2	1	2	1.2.3.1.0	11
4	2	1	3	0	1	2	3	1.2.2.2.0	12
5	3	2	2	1	0	1	2	1.2.3.1.0	11
6	2	3	1	2	1	0	1	1.3.2.1.0	10
7	3	4	2	3	2	1	0	1.1.2.2.1	15
0.1.2.3.4								0.7.0.0.0.. 0.1.5.1.0.. 0.0.5.2.0.. 5.2.0.0.0	
								7.14.14.10.2	

m1

[Di]	1	2	3	4	5	6	7	8	0.1.2.3.4	Σ
1	0	1	1	2	3	2	3	2	1.2.3.2.0	14
2	1	0	2	1	2	3	4	3	1.2.2.2.1	16
3	1	2	0	3	2	1	2	1	1.3.3.1.0	12
4	2	1	3	0	1	2	3	4	1.2.2.2.1	16
5	3	2	2	1	0	1	2	3	1.2.3.2.0	14
6	2	3	1	2	1	0	1	2	1.3.3.1.0	12
7	3	4	2	3	2	1	0	3	1.1.2.3.1	18
8	2	3	1	4	3	2	3	0	1.1.2.3.1	18
									0.7.0.0.0.. 0.2.4.2.0.. 0.0.4.4.0.. 4.4.0.0	
									8.16.20.16.4	

m2

[Di]	1	2	3	4	5	6	7	8	0.1.2.3.4	Σ
1	0	1	1	2	3	2	3	1	1.3.2.2.0	13
2	1	0	2	1	2	3	4	2	1.2.3.1.1	15
3	1	2	0	3	2	1	2	2	1.2.4.1.0	13
4	2	1	3	0	1	2	3	3	1.2.2.3.0	15
5	3	2	2	1	0	1	2	4	1.2.3.1.1	15
6	2	3	1	2	1	0	1	3	1.3.2.2.0	13
7	3	4	2	3	2	1	0	4	1.1.2.2.2	19
8	1	2	2	3	4	3	4	0	1.1.2.2.2	19
									8.16.20.14.6	

m3

Astfel, considerând matricea (Tabel 2.7), și dacă se calculează distanța Minkowski (Minkowski, 1953) între doi atomi/vârfuri (o formă generalizată a distanței Manhattan - vezi Tabelul 2.1), $d = \sum |p_i - q_i|$, unde p_i = număr de vârfuri la distanță i pentru molecula A, q_i = număr de vârfuri la distanță i pentru molecula B, d = distanța, obținem (Tabelul 2.7):

$$d(m1,m2) = |7-8| + |14-16| + |14-20| + |10-16| + |2-4| = 1 + 2 + 6 + 6 + 2 = 17$$

$$d(m1,m3) = |7-8| + |14-16| + |14-20| + |10-14| + |2-6| = 1 + 2 + 6 + 4 + 4 = 17$$

$$d(m2,m3) = |8-8| + |16-16| + |20-20| + |16-14| + |4-6| = 4.$$

2.4. Descriptorii moleculari și amprente numerice

Luarea în considerare a mai multor descriptori moleculari (vezi Figura 2.3) este foarte frecvent regăsită în analizele de similaritate (Bender et al., 2009), astfel încât prin combinații diferite ale acestora rezultă noi descriptori iar mai apoi aplicabilitatea lor urmează metode de corelație sau analize de tip componente principale.

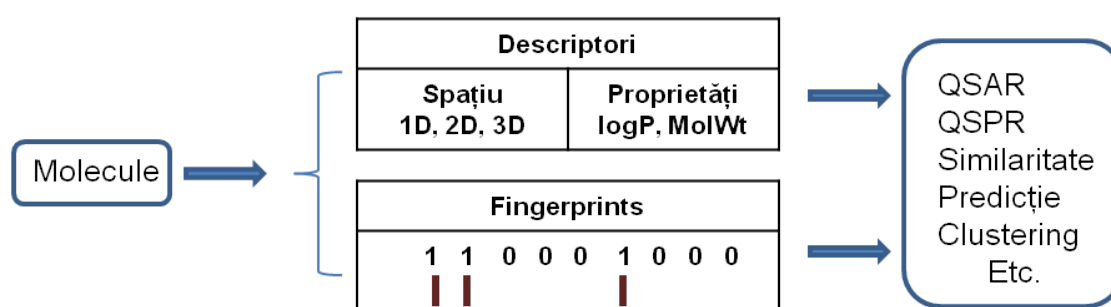


Figura 2.3. Utilizarea descriptorilor și amprentelor moleculare în chimie. Calcularea similarității dintre două molecule consideră descriptori moleculari în funcție de spațiul (1D, 2D, 3D) și proprietățile moleculelor (lipofilicitatea - logP, greutatea moleculară - MolWt) iar calculele numerice se realizează cu ajutorul amprentelor numerice (fingerprints) de tip 1100 ... 00. Rezultatele permit inițierea analizelor precum QSAR, QSPR, similaritatea moleculară, de predicție, de grupare (clustering), etc. - figură adaptată după (Dong et al., 2015).

Ampretele moleculare (fingerprints) (Doucet și Weber, 1996; Rarey și Dixon, 1998) nu reprezintă informații referitoare la coordonatele unei molecule, însă acestea codifică o structură moleculară într-o serie de cifre binare (ex. 01001) (vezi Figura 2.3).

Trei tipuri de indici topologici sunt descriși în literatură ca fiind cel mai adesea utilizați, și anume:

- Indicele Wiener: reprezentări numerice întregi, obținute din grafurile corelate cu unele proprietăți (Singh et al., 2008);
- Indici de conectivitate moleculară: de această dată, aceștia sunt reprezentați prin numere reale și sunt obținute din grafuri cu numere întregi corelate cu unele proprietăți (Diudea et al., 2001);

- Indicii topologici reprezentați de numere reale derivate din grafuri cu numere reale corelate cu anumite proprietăți (Bender, 2005).

Scaffold-urile moleculare ("scheletul" unei molecule) reprezintă o metodă des întâlnită în căutarea similarității (Egieyeh et al., 2016; Velkoborsky și Hoksza, 2016). Utilizând ca model o moleculă sub formă de scaffold, dintr-un set de molecule, se creează un model "query" pe baza căruia se urmărește similaritatea (Willett, 2011).

2.5. Relații cantitative structură chimică - proprietate/activitate

Prin aplicarea metodelor de tip QSAR (Quantitative Structure-Activity Relationship) sau QSPR (Quantitative Structure-Property Relationship) este posibilă descrierea relației structură chimică - activitate biologică, relației proprietate - activitate biologică sau chiar relației structură chimică - proprietate fizică/chimică, în urma cărora se pot analiza și colecta date atât din punct de vedere cantitativ cât și din punct de vedere calitativ, ulterior fiind posibil să se interpreteze și gradul de similaritate dintre anumiți compuși.

2.6. Machine learning/ Învățare mecanică

Metodele de *machine learning* (învățare mecanică) permit aplicarea algoritmilor de calcul (ex. Matrici, modele liniare) în mediul informatic (prin limbaje de programare și medii de calcul specifice – R în RStudio) (RStudio, 2014) pe seturi reale de date experimentale. Dintre alternativele ce pot fi utilizate sunt:

- Se pot genera combinații de descriptori chimici structurali și secvențe ale țintelor cu ajutorul analizelor încorporate în platforme de tip SVMs (support vector machines) (Bleakley și Yamanishi, 2009; Jiang et al., 2007; Mishra et al., 2010);
- Măsurarea similarității (Ding et al., 2013): are loc generarea matricilor de similaritate pentru perechi de compuși, corelate cu structurile lor chimice. Printre metodele ce le includ, amintim regresia Kernel (Chen et al., 2015), BLM (bipartite local method) (Bleakley și Yamanishi, 2009), PKM (pairwise kernel method) (Ding et al., 2013), metoda celor mai mici pătrate (Legendre, 1805), Laplacian (Terfloth, 2003; Thomson et al., 2003), interacțiunile Gaussiene (Doucet și Weber, 1996; Jäntschi et al., 2015), matricea Bayesian (Nidhi et al., 2006).

CAPITOLUL 3. COMPORTAMENTUL MOLECULAR EXPRIMAT LA NIVEL TRANSCRIPTOMIC

Transcriptomul este definit prin reprezentarea ansamblului tuturor moleculelor de ARN dintr-o celulă sau o populație de celule (Trapnell et al., 2011). Dintre acestea amintim tipurile de ARN codificatoare (ARNc) și necodificatoare (ARNnc), ARN lung non-codificator (ARNlnc), ARN nuclear mic (ARN-sn), ARN ribozomal (ARN-r).

Cel mai frecvent tip de ARN studiat este ARN-ul mesager datorită funcției sale de a transporta informația genetică de la ADN, necesară pentru sinteza de proteine (Wang et al., 2009). Această varietate a permis dezvoltarea de diferite tehnologii pentru a înțelege și prezice exprimarea genică la nivelul sistemelor biologice.

Efectele compușilor sunt urmărite la nivelul sistemelor biologice (Bose, 2013) prin intermediul răspunsurilor exprimate la nivel de genom, predicția mecanismelor de acțiune fiind posibilă. Relația dintre răspunsurile biologice și modul de acțiune a unui compus descrisă la nivel genic, poate conduce la identificarea de biomarkeri de prognostic sau diagnostic.

3.1. Metode de evaluare a exprimării genice - Microarray vs RNA-Seq

Tehnicile microarray permit măsurarea modificărilor în expresia a mii de gene în urma expunerii acestora la acțiunea unui anumit compus medicamentos (Yu et al., 2006). Principiul de funcționare se bazează pe atașarea moleculelor de ADN pe o suprafață solidă, numită adesea *chip* (Figura 3.2).

În comparație cu analiza microarray, noile tehnici cum ar fi RNA-Seq, pe lângă scurtarea perioadei de extragere și secvențiere a ADN-ului, permit analiza fragmentelor de ADN foarte vechi și totodată analiza modificărilor continue la nivelul transcriptomului, permițând cuantificarea directă a exprimării genice.

Principiul RNA-Seq are la bază izolarea ARN-ului, conversia în ADN-ul complementar (ADNc) (Secvențe 1, Secvențe 2), controlul calității, prepararea librăriei de secvențe și secvențierea acestora pe o platformă de *next generation sequencing* (NGS).

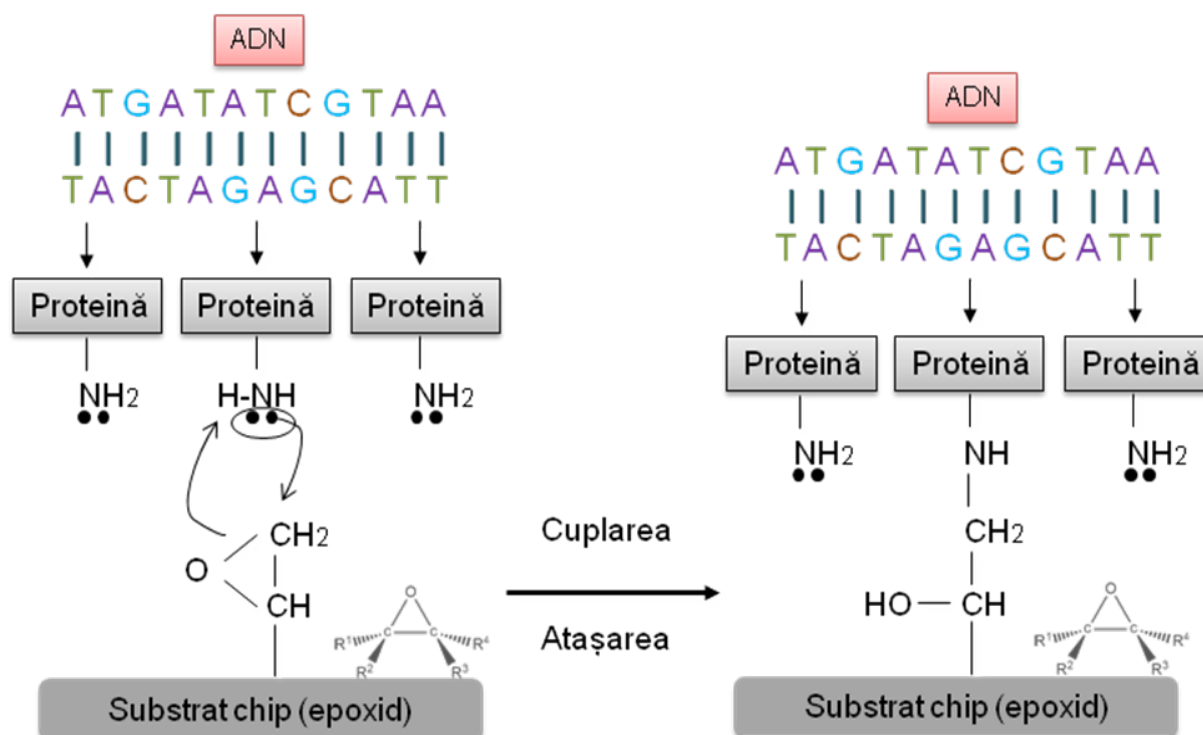


Figura 3.2. Principiul microarray de ataşare a moleculelor de ADN pe suprafaţa solidă a chip-ului prin legături covalente prin grupările amino-alifaticе (NH₂) - figură realizată cu ajutorul instrumentului de lucru – ChemBioDraw Ultra 12.0 (Milne, 2010) şi adaptată după (Stears et al., 2003).

Tehnica RNA-Seq (RNA sequencing) în comparaţie cu microarray, permite o analiză mai detaliată din punctul de vedere al:

- măsurării exprimării genice;
- înţelegerii *splicing*-ului (combinarea) alternativ al genelor;
- modificărilor post-transcripţie.

Datele generate se pot analiza pentru identificarea unor noi factori de transcripţie, a unor noi combinaţii de gene alternative şi înţelegerea funcţionalităţii transcriptomului (ARNm, ARNnc, ARNinc) (Kukurba şi Montgomery, 2016) (Figura 3.3).

Dacă microarray se bazează pe potenţialul de hibridizare a probelor ce sunt marcate cu secvenţele de ADNc ţintă, RNA-Seq utilizează tehnici avansate de secvenţiere cum ar fi *Next-Generation Sequencing* (NGS) pentru secvenţarea ARN prin secvenţiere directă a catenelor de ADNc (Figura 3.3). În ambele cazuri (Bourdon-

Lacombe et al., 2015), în urma obținerii genelor exprimate diferențial se poate urmări în continuare funcționalitatea celulară și expresia la nivelul căilor cu ajutorul metodelor statistice și pachetelor bioinformaticice ce vor fi amintite în cele ce urmează.

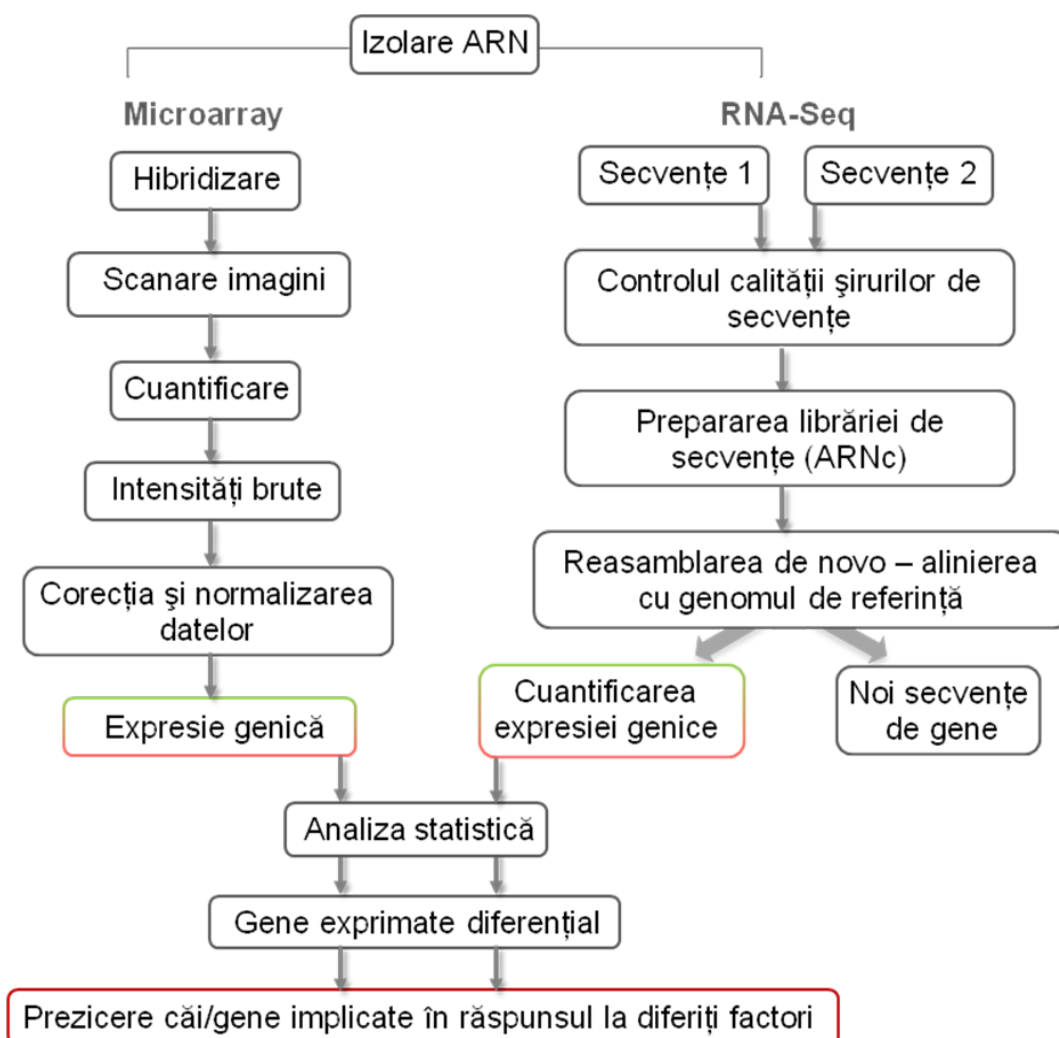


Figura 3.3. Principiul tehnicii Microarray vs RNA-Seq – adaptat după (Corney et al., 2013).

3.2. Analiza moleculară

Pentru analiza exprimării genice se pot aplica o serie de metode de screening pentru identificarea căilor și genelor exprimate ca răspuns la diferiți factori: determinarea genelor exprimate diferențial (DEGs); potrivirea secvențelor - semnelor (signature matching); rețele (network) de interacțiuni proteină-proteină (I-P-P) (PPI network);

rețele de coexprimare genică (coexpression network) (Alexander-Dann et al., 2018) (Figura 3.4).

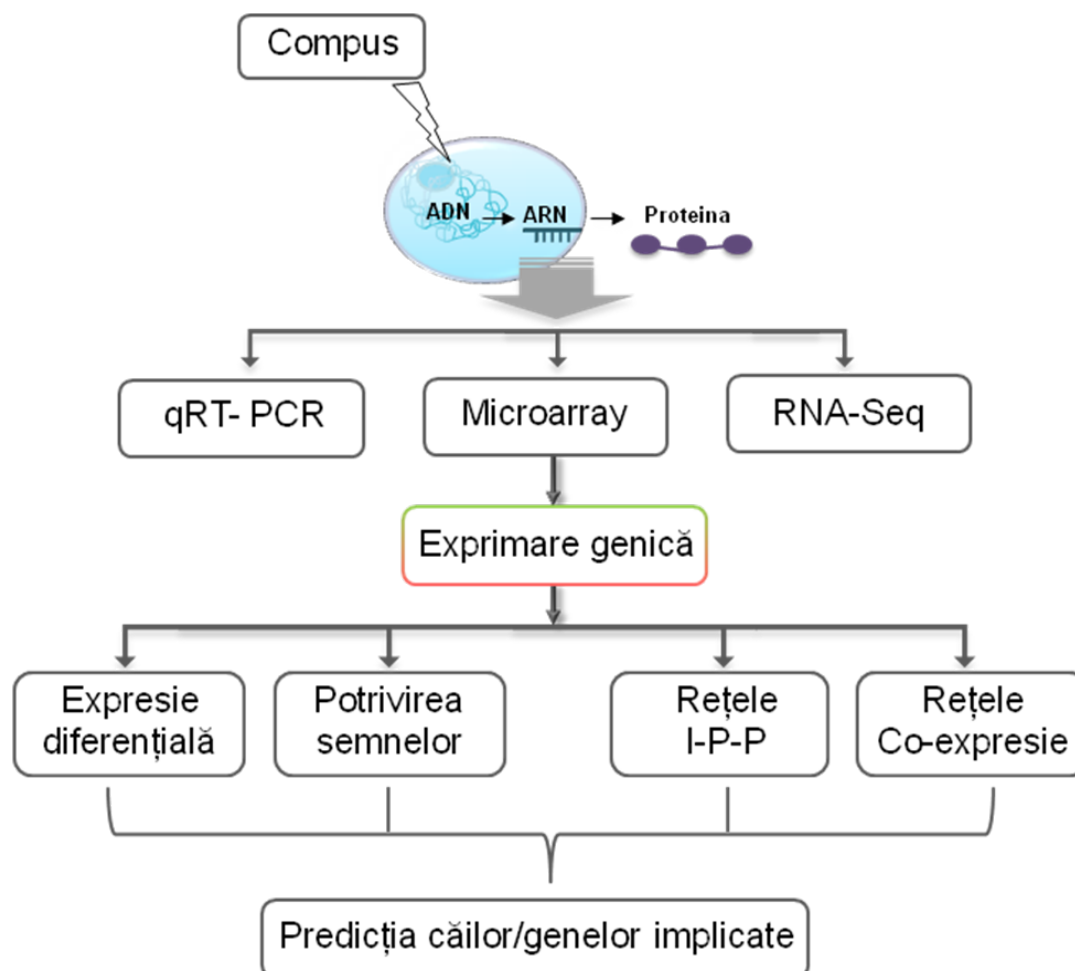


Figura 3.4. Reprezentarea tipurilor de analiză a exprimării genice în urma expunerii celulare la un anumit compus - adaptat după (Alexander-Dann et al., 2018).

3.2.1. Gene exprimate diferențial

Genele exprimate diferențial sunt utile în procesul de identificare a biomarkerilor specifici cum ar fi biomarkerii tumorali, de toxicitate, de prognostic sau diagnostic (Shi et al., 2008).

Un marker ideal este detectat înainte de simptomele patologice tradiționale și este caracterizat prin specificitate și o sensibilitate ridicată precum și de relațiile mecanice de relevanță biologică (Rininger et al., 2000). O genă este considerată a fi exprimată

diferențial dacă diferența observată dintre două condiții diferite este semnificativă statistic (Alexander-Dann et al., 2018; Anjum et al., 2016) Tabelul 3.1.

Tabelul 3.1. Măsurile analitice de determinare a DEGs

Metodă	Principiu	Formulă de calcul	Utilitate
Fold change	Se calculează raportul în exprimarea genică dintre probă și control. Este exprimat în logaritmul în baza 2, unde este egal cu 0, dacă nu există nicio diferență. În funcție de limita aleasă (cut-off), sunt determinate DEGs cu valori cuprinse între 0.5 și 2.	$F_g = \log_2(E_{ga}) - \log_2(E_{gb})$ F - fold change g - genă E - expresie a - condiție a (ex. tratat) b- condiție b (ex. netratat)	- ușor de interpretat - probe mici - apar variații
Rank product	Compararea rândurilor de gene în funcție de expresie prin aplicarea unui test nonparametric.	$R_g = \frac{E_{ga}}{E_{gb}}$ R - rația g - genă E - expresie a - condiție a (ex. tratat) b- condiție b (ex. netratat)	- comparare de rezultate de pe platforme diferite - utilizată în meta-analize
t-test, ANOVA	Compararea valorii medii a expresiei per genă în probe. Se bazează pe ipoteza nulă unde mediile sunt egale. În timp ce t-test este utilizat pentru compararea a două probe, ANOVA permite compararea pentru probe multiple.	$t = \frac{\text{media}(d)}{\frac{\sigma(d)}{\sqrt{n}}}$ d = $x_a - y_b$ x_a - media numărului de gene atribuite condiției a y_b - media numărului de gene atribuite condiției b $\sigma(d)$ - variabilitatea grupurilor de gene, σ = deviația standard n - numărul de perechi testate	- semnificație statistică ridicată - valorile sunt normal distribuite și au aceeași variație
Bayesian	Utilizează datele pentru predicția probabilităților expresiei diferențiate și deviația standard.	$P(A B) = \frac{P(B A) * P(A)}{P(B)}$ $P(A B)$ - probabilitatea condiției A dacă condiția B are loc $P(B A)$ - probabilitatea	- timp mai îndelungat - rezultate mai semnificative decât t-test

Metodă	Principiu	Formulă de calcul	Utilitate																
		condiției B dacă condiția A are loc P(A) - probabilitatea priorității condiției A P(B) - probabilitatea priorității condiției B																	
Test exact	Pentru compararea exprimării genice, utilizează numărarea reală a expresiilor.	$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!n!}$ unde, <table border="1"> <thead> <tr> <th></th><th>C_A</th><th>C_B</th><th>Total</th></tr> </thead> <tbody> <tr> <td>Gena X</td><td>a</td><td>b</td><td>a+b</td></tr> <tr> <td>Rest gene</td><td>c</td><td>d</td><td>c+d</td></tr> <tr> <td>Total</td><td>a+c</td><td>b+d</td><td>n</td></tr> </tbody> </table> C_A - condiție A C_B - condiție B		C _A	C _B	Total	Gena X	a	b	a+b	Rest gene	c	d	c+d	Total	a+c	b+d	n	- necesită un număr exact de copii de ARNm
	C _A	C _B	Total																
Gena X	a	b	a+b																
Rest gene	c	d	c+d																
Total	a+c	b+d	n																

Pentru măsurarea acestora se consideră o serie de metode statistice precum: fold change (nivel de exprimare genică) (Love et al., 2014; Tarca et al., 2006), Rank product (Breitling et al., 2004; Hong et al., 2006), metode liniare (ANOVA, t-test) (Trapnell et al., 2013), metode Bayesiene (Hardcastle și Kelly, 2010), testul exact (Auer și Doerge, 2010; Love et al., 2014), încorporate în diferite pachete de calcul precum *limma* (Tarca et al., 2006), weighted average difference method (Kadota et al., 2008), RankProd (Hong et al., 2006), Cuffdiff 2 (Trapnell et al., 2013), baySeq (Hardcastle și Kelly, 2010), DESeq2 (Love et al., 2014), edgeR (Robinson et al., 2009). Utilitatea și principiul metodelor este redată în Tabelul 3.1.

3.2.2. Signature matching

În aplicarea metodelor de *Signature matching* (potrivirea semnelor), DEGs sunt evaluate în comparație cu o librărie ce conține profile transcriptomice cu scopul de a prezice efectele potențiale (Hettne et al., 2013).

Comparând modificările în expresia genelor rezultate în urma expunerii la anumiți compuși, se pot corela efectele acestora asupra unui sistem biologic cu proprietățile structurale ale acestora, separat de ceea ce înseamnă o analiză de similaritate

moleculară. Utilizând de asemenea librăriile cu gene de referință, se poate măsura similaritatea dintre compuși în spațiul exprimării genice (Lamb et al., 2006).

3.2.3. Rețele de interacțiune de tip proteină-proteină

Rețelele de interacțiune proteină-proteină (Protein-Protein interaction network - PPI network) permit vizualizarea și caracterizarea răspunsului biologic după expunerea unui anumit compus (efect terapeutic sau efect toxic – vezi Figura 3.5) la nivel celular.

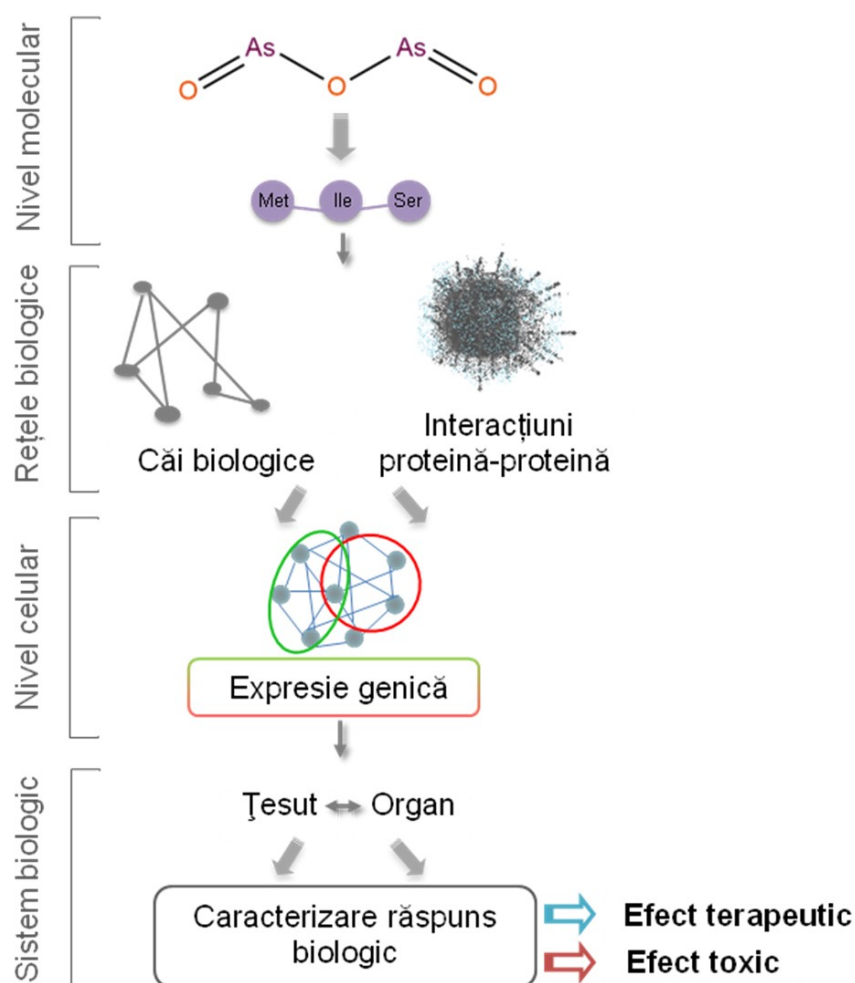


Figura 3.5. Reprezentarea exemplificată (compus: trioxid de arsen) a integrării datelor experimentale și computaționale la nivelul sistemelor biologice pentru caracterizarea răspunsurilor biologice prin intermediul rețelelor de asociere (căi biologice și interacțiunile dintre proteine) - figură adaptată după (Orozcoa et al., 2016).

În Figura 3.5, se poate observa cum la nivel molecular, trioxidul de arsen acționează asupra unor ținte proteice (targets). Căile biologice implicate și interacțiunile dintre proteine pot fi analizate prin prisma diferențierii expresiei genelor rezultate. Caracterizarea răspunsului biologic survine în funcție de predicția căilor implicate.

3.2.4. Rețele de co-expresie genică

Rețelele de co-expresie (Co-expression network) utilizează totalitatea măsurilor legate de transcriptom, ceea ce creează o puternică corelație între gene și efectele biologice. Metode aplicate până la ora actuală sunt regăsite în exemple precum: Context Likelihood of Relatedness (Taylor et al., 2008), Weighted Gene Co-expression Network Analysis' – WGCNA (Zhang și Horvath, 2005) .

Metodele permit corelația și compararea rețelelor de exprimare genică cu un puternic impact în predicția efectelor compușilor testați. Similaritățile și diferențele fiecărui răspuns în parte, pentru fiecare compus sunt astfel evidențiate. Dependența metodelor de co-expresie de determinarea corelațiilor între expresiile genice este asociată cu utilizarea unui număr minim necesare de replicate.

3.3. Aplicabilitatea noțiunilor de similaritate în contextul exprimării genice

Transferul conceptului de similaritate la nivelul exprimării genice a fost tot mai mult înaintat în ultimii ani. Au fost create întregi librării cu informații descriptive, astfel încât similaritatea compușilor se poate măsura în spațiul de expresie a genelor. Una dintre abordările curente referă maparea conectivității (Connectivity mapping) cu ajutorul căreia se poate descrie acțiunea unor compuși la nivelul genelor (Sirici et al., 2017).

Prin reprezentările numerice se pot analiza relațiile dintre compuși și expresia genelor. O serie de măsuri de calcul a similarității exprimării genice au fost dezvoltate până la ora actuală, utilizând grafuri ca și reprezentări ale genelor, unde nodurile se consideră a fi genele, iar muchiile legătura dintre gene, dacă există o relație între acestea (Yona et al., 2006).

CONTRIBUȚII PERSONALE

Similaritatea moleculară a fost ilustrată pornind de la descrierea comportamentului compușilor anorganici în apă (Capitolul 5) continuând cu descrierea caracteristicilor compușilor organici pe baza similarității (Capitolul 6) și cuantificarea relațiilor dintre caracteristicile structurale și proprietăți (Capitolul 7), ajungând la modelarea și caracterizarea la nivel transcriptomic a răspunsului compușilor anorganici (arsenați) (Capitolul 8), a probioticelor (Capitolul 9) în urma tratamentului liniilor celulare de cancer de sân, respectiv epiteliale intestinale porcine, *in vitro* (Fig. 4.1).

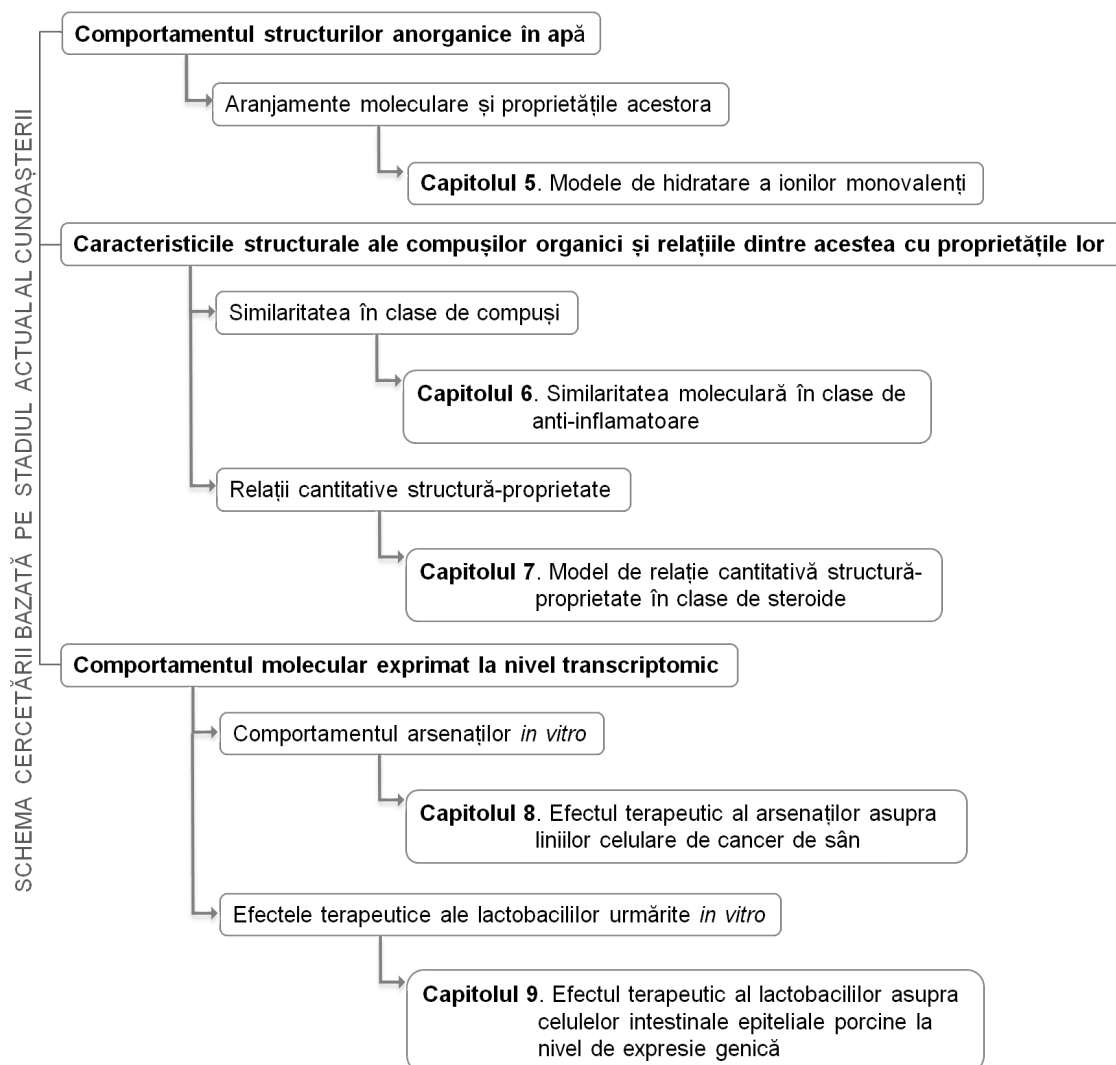


Figura 4.1. Schema cercetării bazată pe stadiul actual al cunoașterii

CAPITOLUL 4. CADRUL CERCETĂRII

4.1. Aranjamentele moleculare și proprietățile acestora

În vederea investigării și descrierii comportamentului structurilor anorganice în apă, au fost urmărite caracteristicile structurale și proprietățile acestora în cadrul aranjamentelor moleculare rezultate. În acest context, în capitolul 5 sunt prezentate modelele de hidratare obținute pe baza modelării interacțiunii a șase ioni monovalenți (NH_4^+ , F^- , Cl^- , Li^+ , Na^+ și K^+) cu molecule de apă (Pruteanu et al., 2016).

4.2. Similaritatea în clase de compuși

O colecție proprie de specii de plante pentru care se cunosc extrase de compuși cu potențial anti-inflamator a fost considerată. Similaritatea a fost calculată între compușii acestor plante și un set de compuși medicamentoși utilizați în tratamentul afecțiunilor inflamatorii. Astfel în urma studiului s-a dorit evidențierea altor surse naturale de compuși cu proprietăți similare.

4.3. Relații cantitative structură-proprietate

Pentru a valida un model de relație cantitativă structură-proprietate a fost considerat un set de 40 de molecule similare (Pruteanu et al., 2016) din baza de date PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) aparținând unei clase de steroide și proprietatea lor de lipofilicitate exprimată logaritmic prin coeficientul de partiție octanol-apă ($\log P$, de obicei < 5) făcând referire la raportul de concentrație dintre speciile neionizate ale unui compus.

4.4. Comportamentul arsenatilor *in vitro*

Comportamentul molecular a fost urmărit și la nivel celular, astfel în urma tratamentului a trei linii celulare de cancer de sân cu arsenati și o linie celulară normală (HUMEC) pentru control, s-au investigat modificările la nivel de exprimare genică. A fost propusă astfel evaluarea efectului arsenatilor la nivel transcriptomic și investigarea mecanismelor de acțiune (activarea apoptozei, autofagia, reducerea proliferării celulare) pentru a înțelege comportamentul arsenatilor *in vitro*.

4.5. Efectele terapeutice ale lactobacililor urmărite *in vitro*

În privința efectelor terapeutice ale compușilor organici, s-a considerat o soluție de probiotice reprezentative a trei specii de *Lactobacillus* (Taranu et al., 2018). În urma tratării celulelor epiteliale intestinale porcine (IPEC-1) cu amestecul de lactobacili, s-a urmărit comportamentul acestor compuși la nivel de expresie genică.

CAPITOLUL 5. MODELE DE HIDRATARE A IONILOR MONOVALENȚI

5.1. Materiale și metode

În Tabelul 5.1 sunt reprezentate valorile distanțelor calculate prin metode diferite (Experimental, MP4/6-31G*, MP3/6-31G*, MP2/6-31G*, M06-2X/6-31G*, HF/6-31G*, HF/3-21G*, HF/STO-3G) iar valorile cele mai apropiate între ele sunt date de calculul efectuat prin metodele: M06-2X/6-31G*, HF/3-21G* and MP3/6-31G* (Russell, 2006).

Tabel 5.1. Lungimea d(O—H) în apă: experimental vs. calculat (Russell, 2006)

Metoda	d(O—H) pm
Experimental	95.78
MP4/6-31G*	97.03
MP3/6-31G*	96.68
MP2/6-31G*	96.89
M06-2X/6-31G*	96.56
HF/6-31G*	94.73
HF/3-21G*	96.65
HF/STO-3G	98.92

5.1.2. Gruparea moleculelor ion-apă

Algoritmul de lucru care descrie această procedură este detaliată în pașii după cum urmează:

- Pas 1: molecule de apă au fost plasate în vecinătatea ionilor de interes

- Pas 2*: a fost efectuată optimizarea geometriei
- Pas 3*: mai multe molecule de apă au fost adăugate în spațiile neocupate din vecinătatea ionilor investigați

*Pașii 2 și 3 au fost repetați până nu a mai fost observată nicio schimbare în aranjamentele grupărilor formate din primul substrat.

Pașii descriși anteriori au fost aplicați pentru fiecare ion investigat în parte.

5.1.3. Construirea cuștilor de congeneri

Cuștile de dodecaedru au fost construite conform metodelor descrise în literatura de specialitate (Burnham et al., 2006; Grayson et al., 2009) pentru fiecare cation investigat (NH_4^+ , Li^+ , Na^+ și K^+) iar structurile obținute sunt prezentate în Figura 5.1.

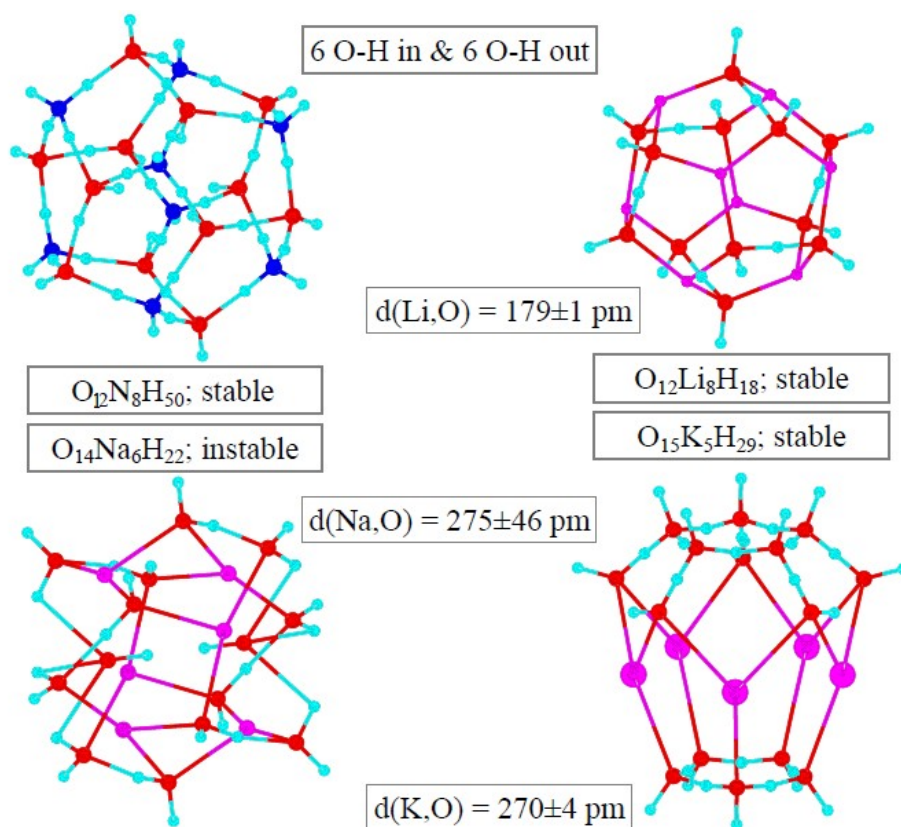


Figura 5.1. Grupările dodecaedru pentru cationii NH_4^+ , Li^+ , Na^+ și K^+ (O - roșu; H - albastru; N - albastru închis; Li, Na, respectiv K - roz)

Mediile lungimilor legăturilor au fost comparate cu ajutorul testului statistic ANOVA cu un nivel de semnificație 0.33% ($\alpha = 0.05$ ajustat de numărul de comparații considerate, în cazul de față, acest număr fiind reprezentat de numărul de ioni investigați; $\alpha^* = 0.05/[6*(6-1)/2]$).

Luând în considerare și testul Bonferroni (Biella et al., 2008) în urma rezultatelor au fost observate diferențe semnificative. Pentru a testa diferențele dintre unghiurile formate între grupările ion-apă a fost aplicat testul Friedman ANOVA (Pruteanu et al., 2016), analiza statistică fiind realizată utilizând software-ul Statistica (V. 8.) (<http://software.dell.com/products/statistica/>).

5.1.4. Simetria și stabilitatea structurală

În cazul amoniacului ($O_{12}N_8H_{50}$) șase legături O-H sunt stabilizate în interiorul grupării formate și șase legături O-H sunt stabilizate în afara structurii (după cum se poate observa în Figura 5.1).

Formarea și/sau stabilitatea grupărilor dodecaedru ce integrează moleculele de apă, au fost predominant investigate la formarea acestora cu amoniacul.

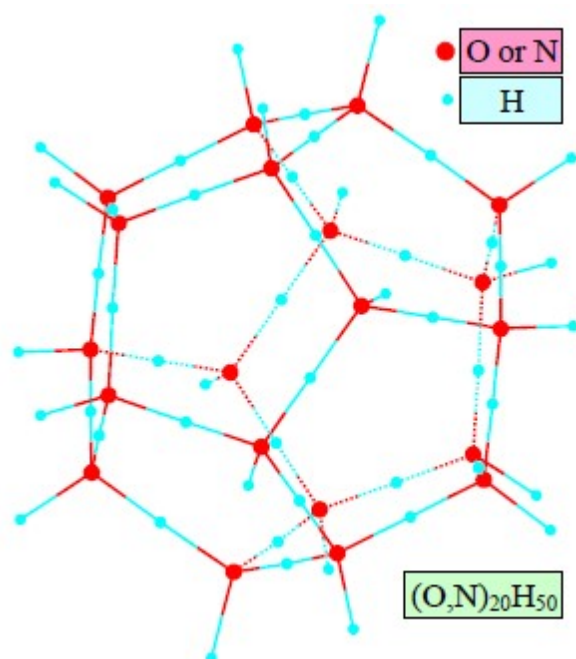


Figura 5.2. Forma generală pentru $O_xN_{20-x}H_{50}$ (O sau N - roșu; H - albastru)

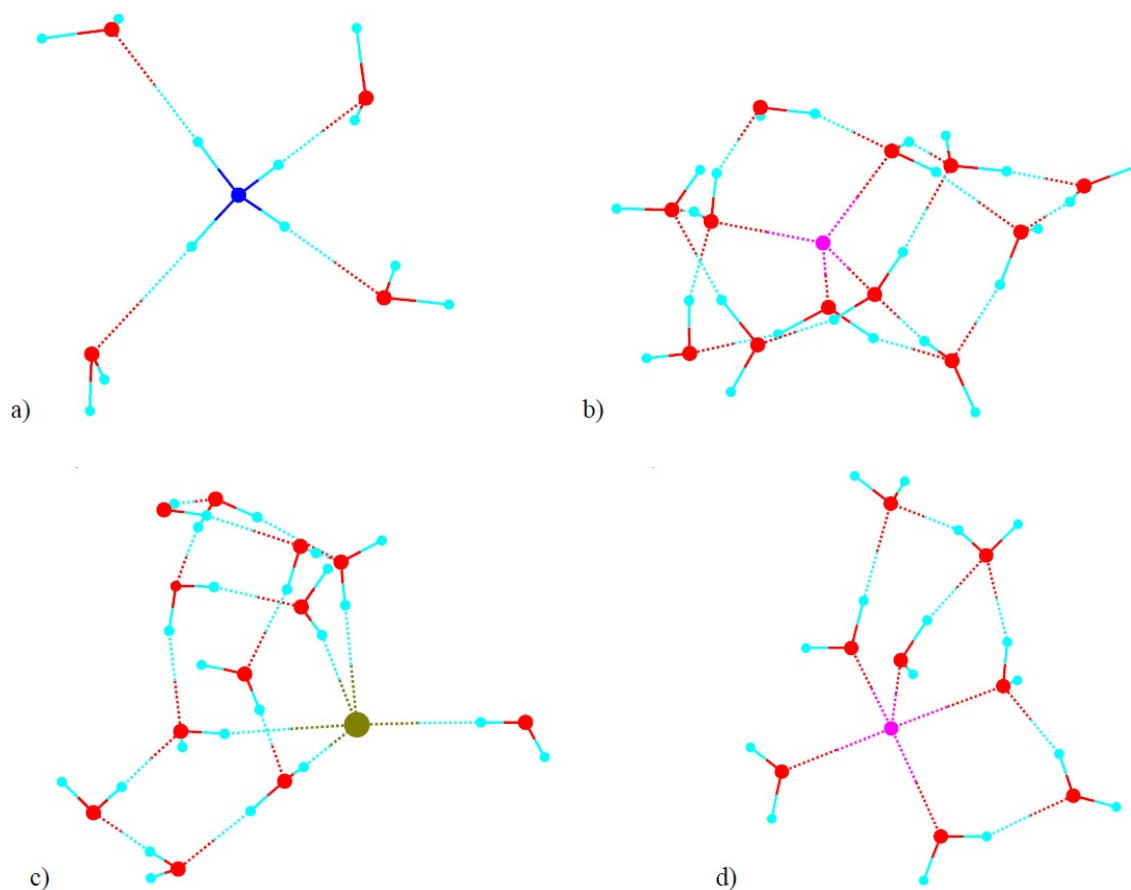
5.2. Rezultate și discuții

5.2.1. Moleculele stabile ion-apă

Grupările stabile ion-apă identificate în cazul de față, s-au dovedit a fi cele care s-au format cu:

- patru ($\text{NH}_4^+ \cdot 4\text{H}_2\text{O}$ și $\text{Li}^+ \cdot 4\text{H}_2\text{O}$);
- cinci ($\text{Cl}^- \cdot 5\text{H}_2\text{O}$ și $\text{Na}^+ \cdot 5\text{H}_2\text{O}$);
- șase ($\text{F}^- \cdot 6\text{H}_2\text{O}$ și $\text{K}^+ \cdot 6\text{H}_2\text{O}$) molecule de apă.

Grupările stabile menționate sunt reprezentate în figura de mai jos (Figura 5.3) cu mențiunea că grupările ion-apă sunt natural formate și fără constrângeri, dat fiind faptul că modelul *in silico* a fost efectuat cu moleculele de apă.



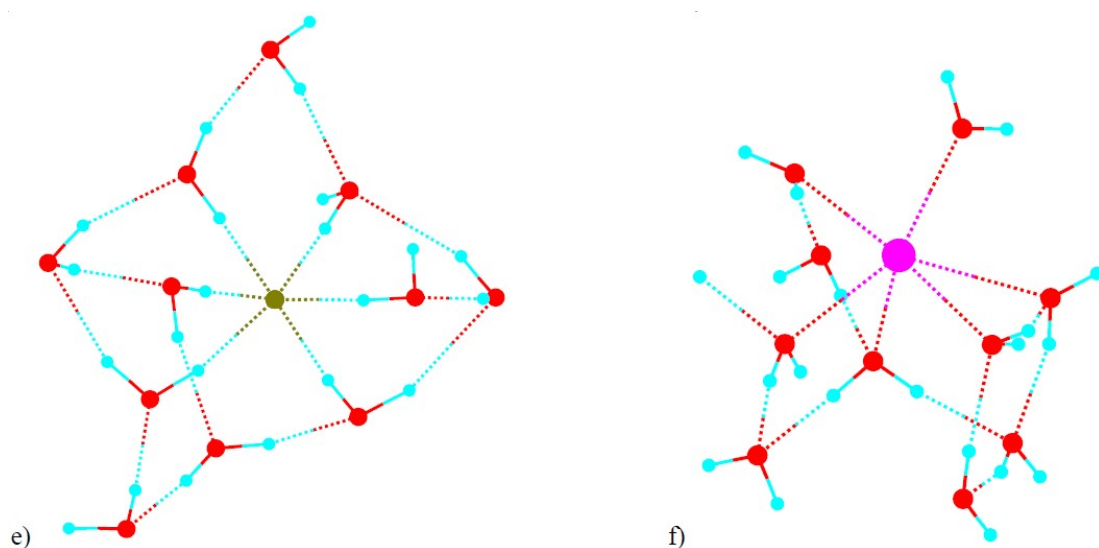


Figura 5.3. a) Gruparea $\text{NH}_4^+ \cdot 4\text{H}_2\text{O}$, b) Gruparea $\text{Li}^+ \cdot 4\text{H}_2\text{O}$, c) Gruparea $\text{Cl}^- \cdot 5\text{H}_2\text{O}$, d) Gruparea $\text{Na}^+ \cdot 5\text{H}_2\text{O}$, e) Gruparea $\text{F}^- \cdot 6\text{H}_2\text{O}$, f) Gruparea $\text{K}^+ \cdot 6\text{H}_2\text{O}$.

Primul ion analizat (NH_4^+) a fost selectat ca referință pentru reproductibilitatea calculului și pentru validarea metodei de analiză iar rezultatele obținute sunt în acord cu date cunoscute (Clegg și Brimblecombe, 1989; Galashev, 2013; Guerra et al., 2014; Janeiro-Barral și Mella, 2006).

Legăturile și unghiurile calculate reies ca fiind foarte aproape de valorile acestora când se află în stare naturală în mediul apos. Pentru $\text{Na}^+(\text{H}_2\text{O})_{20}$, $\text{Cl}^-(\text{H}_2\text{O})_{17}$, și $\text{Na}^+(\text{H}_2\text{O})_{100}$ utilizând algoritmul PBHaT (Burnham et al., 2006) a fost identificat minimum global. Acest algoritm este un hibrid capabil să eșantioneze eficient funcția de partiție de la minimumul global la starea lichidă.

5.2.2. Simetria moleculară

Analizând grupările ionilor cu apa se arată că aranjamentul moleculei de apă în jurul ionilor investigați nu este simetric chiar dacă acesta pare a fi simetric la prima vedere (Ex. Figura 5.3a). Adăugând ulterior mai multe molecule de apă, după ce legătura inițială a fost creată se poate observa că asimetria este tot mai evidentă.

Se poate spune că este de așteptat ca aranjamentul molecular să își piardă simetria în momentul în care intervin în apropiere orbitali cu energii superioare. De fapt, este

greu de crezut că la dizolvarea unui ion în apă, în soluții diluate, dispunerea moleculelor de apă va urma simetria așteptată a stării fundamentale a ionului *in vitro*.

Așa cum era de așteptat, dat fiind faptul că au fost investigate grupări diferite ion-apă, lungimile legăturilor și unghiurile dintre legături diferă de la o grupare la alta. Rezultatele sunt redată în tabelele (Tabelul 5.2 și Tabelul 5.3):

Tabel 5.2. Lungimea legăturilor și unghiurilor dintre legături în gruparea $\text{Cl}^- \cdot 5\text{H}_2\text{O}$ și gruparea $\text{F}^- \cdot 6\text{H}_2\text{O}$

Distanța		Unghiurile			
d(Atom...Atom)	pm	(Atom...Atom) d(Atom...Atom)	(°)	(Atom...Atom) d(Atom...Atom)	(°)
Gruparea $\text{Cl}^- \cdot 5\text{H}_2\text{O}$					
d(H...Cl)	265	(H...Cl) ₂₈₁ –(H...Cl) ₃₀₇	67	(H...Cl) ₂₇₄ –(H...Cl) ₂₈₁	96
d(H...Cl)	274	(H...Cl) ₂₇₄ –(H...Cl) ₃₀₇	76	(H...Cl) ₂₆₅ –(H...Cl) ₂₇₄	101
d(H...Cl)	281	(H...Cl) ₂₈₁ –(H...Cl) ₂₈₄	82	(H...Cl) ₂₆₅ –(H...Cl) ₂₈₄	129
d(H...Cl)	284	(H...Cl) ₂₇₄ –(H...Cl) ₂₈₄	82	(H...Cl) ₂₈₄ –(H...Cl) ₃₀₇	140
d(H...Cl)	307	(H...Cl) ₂₆₅ –(H...Cl) ₃₀₇	89	(H...Cl) ₂₆₅ –(H...Cl) ₂₈₁	146
Gruparea $\text{F}^- \cdot 6\text{H}_2\text{O}$					
d(H...F)	179	(H...F) ₁₇₉ –(H...F) ₁₈₄	77	(H...F) ₁₈₄ –(H...F) ₁₈₅	89
d(H...F)	182	(H...F) ₁₈₄ –(H...F) ₁₈₇	85	(H...F) ₁₇₉ –(H...F) ₁₈₅	90
d(H...F)	183	(H...F) ₁₈₂ –(H...F) ₁₈₃	85	(H...F) ₁₈₃ –(H...F) ₁₈₇	93
d(H...F)	184	(H...F) ₁₇₉ –(H...F) ₁₈₃	86	(H...F) ₁₈₂ –(H...F) ₁₈₇	109
d(H...F)	185	(H...F) ₁₈₅ –(H...F) ₁₈₇	87	(H...F) ₁₇₉ –(H...F) ₁₈₇	162
d(H...F)	187	(H...F) ₁₈₃ –(H...F) ₁₈₄	88	(H...F) ₁₈₂ –(H...F) ₁₈₄	165
		(H...F) ₁₇₉ –(H...F) ₁₈₂	89	(H...F) ₁₈₃ –(H...F) ₁₈₅	177

În cazul anionilor Cl^- și F^- , legătura dintre moleculele de apă și ioni este creată cu contribuția ionilor de hidrogen ai apei, regiunea din jurul anionilor fiind bogată în electroni. Așa se explică faptul că distanțele și unghiurile aranjamentelor moleculare rezultate, sunt date în raport cu acești atomi de hidrogen (vezi Tabel 5.2 pentru Cl^- și F^-).

Distanțele H...Cl⁻ în grupurile $\text{Cl}^- \cdot 5\text{H}_2\text{O}$ sunt apropiate de acele valori ale distanței dintre H...Cl⁻ în dimensiunea normală de clatrat, unde sunt posibile salturi de orientare cu unghiuri mari (Laage și Hynes, 2007).

După cum se poate observa pentru ionul clorură, nu există simetrie datorită efectului suplimentar dat de prezența energiei libere corespunzătoare orbitalilor de tip d (a se

vedea Tabel 5.2). O altă constatare este vizibilă la nivelul numărului de molecule de apă care înconjoară anionul.

5.2.3. Electronegativitatea

Diferența de electronegativitate dintre ionii fluorură și ionii clorură, este evidențiată, ionul fluorură fiind capabil să atragă în primul strat, 6 atomi de hidrogen ai moleculelor de apă, pe când, ionul clorură, având electronegativitatea mai scăzută, atrage doar 5 atomi de hidrogen ai moleculelor de apă.

Calculând deviația standard dintre valorile obținute ale unghiurilor și valorile așteptate, se observă că deviația standard devine de două ori în comparație cu cea obținută în gruparea ion fluorură-apă. Pentru aceste calcule a fost luată în considerare structura platonică cu cinci vârfuri și șase fețe ($6 \times 90^\circ$, $3 \times 120^\circ$, $1 \times 180^\circ$).

Luând în considerare unghiurile, pentru ionul fluorură dizolvat în apă, aranjamentul moleculelor de apă din primul strat ce înconjoară ionul F^- , este apropiat de ceea ce înseamnă un aranjament de tip pătrat bi-piramidal (Tabel 5.2).

În cazul de față, din cauza diferenței de electronegativitate dintre oxigen și ionul fluorură, se poate observa că aranjamentul este unul asimetric de tip pătrat bi-piramidal format de atomii de hidrogen ce înconjoară ionul fluorură.

Pentru cationi, lucrurile sunt exact invers, încât legătura dintre moleculele de apă și cationi este formată prin intermediul atomilor de oxigen. Datorită acestui fapt, și situația în care numărul de coordonare este în scădere cu creșterea numărului atomic observat în cazul anionilor (6 pentru F^- , 5 pentru Cl^-), se inversează în cazul cationilor (a se vedea Tabel 5.3).

Tabel 5.3. Lungimea legăturilor și unghiurilor dintre legături în grupările cationi-apă:
 $NH_4^+ \cdot 4H_2O$, $Li^+ \cdot 4H_2O$, $Na^+ \cdot 5H_2O$, $K^+ \cdot 6H_2O$

Distanța		Unghiurile			
d(Atom...Atom)	pm	(Atom...Atom) d(Atom...Atom)	(°)	(Atom...Atom) d(Atom...Atom)	(°)
Gruparea $NH_4^+ \cdot 4H_2O$					
d(O...H)	95	(H...O) ₉₅ —(O...H) ₉₅	105		

Distanța		Unghiurile			
d(N...H)	101	(H...N) ₁₀₁ –(N...H) ₁₀₁	109		
d(O...H)	208	(H...Cl) ₉₅ –(O...H) ₂₀₈	113		
Gruparea Li⁺·4H₂O					
d(Li...O)	193	(Li...O) ₁₉₃ –(Li...O) ₁₉₆	99	(Li...O) ₁₉₃ –(Li...O) ₂₀₀	111
d(Li...O)	194	(Li...O) ₁₉₄ –(Li...O) ₂₀₀	104	(Li...O) ₁₉₄ –(Li...O) ₁₉₆	116
d(Li...O)	196	(Li...O) ₁₉₆ –(Li...O) ₂₀₀	110	(Li...O) ₁₉₃ –(Li...O) ₁₉₄	117
d(Li...O)	200				
Gruparea Na⁺·5H₂O					
d(Na...O)	231	(Na...O) ₂₃₄ –(Na...O) ₂₃₉	84	(Na...O) ₂₃₄ –(Na...O) ₂₃₅	97
d(Na...O)	234	(Na...O) ₂₃₆ –(Na...O) ₂₃₉	85	(Na...O) ₂₃₁ –(Na...O) ₂₃₄	100
d(Na...O)	235	(Na...O) ₂₃₁ –(Na...O) ₂₃₉	88	(Na...O) ₂₃₄ –(Na...O) ₂₃₆	123
d(Na...O)	236	(Na...O) ₂₃₁ –(Na...O) ₂₃₅	93	(Na...O) ₂₃₁ –(Na...O) ₂₃₆	135
d(Na...O)	239	(Na...O) ₂₃₅ –(Na...O) ₂₃₆	94	(Na...O) ₂₃₁ –(Na...O) ₂₃₅	178
Gruparea K⁺·6H₂O					
d(O...K)	278	(O...K) ₂₈₅ –(O...K) ₂₉₃	58.1	(O...K) ₂₇₈ –(O...K) ₂₈₄	94.4
d(O...K)	279	(O...K) ₂₈₁ –(O...K) ₂₈₅	79.0	(O...K) ₂₈₁ –(O...K) ₂₉₃	100.5
d(O...K)	281	(O...K) ₂₈₄ –(O...K) ₂₉₃	79.1	(O...K) ₂₇₉ –(O...K) ₂₈₄	121.6
d(O...K)	284	(O...K) ₂₇₈ –(O...K) ₂₈₁	80.8	(O...K) ₂₈₄ –(O...K) ₂₈₅	127.8
d(O...K)	285	(O...K) ₂₈₁ –(O...K) ₂₈₄	81.3	(O...K) ₂₇₈ –(O...K) ₂₈₅	128.8
d(O...K)	293	(O...K) ₂₇₉ –(O...K) ₂₉₃	86.2	(O...K) ₂₇₉ –(O...K) ₂₈₁	157.1
		(O...K) ₂₇₉ –(O...K) ₂₈₅	86.3	(O...K) ₂₇₈ –(O...K) ₂₉₃	173.0
		(O...K) ₂₇₈ –(O...K) ₂₇₉	92.5		

Media lungimilor de legătură între grupările ion-apă investigate s-a dovedit a fi semnificativ diferită (testul ANOVA cu o valoare $p = 3 \cdot 10^{-10}$). Testul post-hoc Bonferroni a identificat diferențe semnificative în raport cu lungimile de legătură pentru următoarele perechi de grupări (diferențele au fost considerate semnificative conform nivelului de semnificație ajustat de 0.3333%):

- Media lungimii de legătură pentru gruparea $\text{NH}_4^+ \cdot 4\text{H}_2\text{O}$ s-a dovedit a fi puțin semnificativă în comparație cu cea observată la gruparea $\text{Cl}^- \cdot 5\text{H}_2\text{O}$ (unde $p=10^{-8}$), gruparea $\text{Na}^+ \cdot 5\text{H}_2\text{O}$ ($p=8.7 \cdot 10^{-6}$), și respectiv la gruparea $\text{K}^+ \cdot 6\text{H}_2\text{O}$ ($p=5 \cdot 10^{-9}$).
- Media lungimii de legătură pentru gruparea $\text{Li}^+ \cdot 4\text{H}_2\text{O}$ s-a dovedit a fi puțin semnificativă în comparație cu cea observată la gruparea $\text{Cl}^- \cdot 5\text{H}_2\text{O}$ ($p=2.3 \cdot 10^{-5}$), și gruparea $\text{K}^+ \cdot 6\text{H}_2\text{O}$ ($1.6 \cdot 10^{-5}$).

- Media lungimii de legătură pentru gruparea $\text{Cl}^- \cdot 5\text{H}_2\text{O}$ s-a dovedit a fi semnificativă în comparație cu cea observată la gruparea $\text{F}^- \cdot 6\text{H}_2\text{O}$ ($p=5 \cdot 10^{-7}$).
- Media lungimii de legătură pentru gruparea $\text{F}^- \cdot 6\text{H}_2\text{O}$ s-a dovedit a fi semnificativ scăzută în comparație cu cea observată la gruparea $\text{K}^+ \cdot 6\text{H}_2\text{O}$ ($p=2 \cdot 10^{-7}$).
- Analiza unghiurilor prezentate în Tabelul 5.2 și 5.3 a condus la următoarele:
- După cum era de așteptat, cel mai mic unghi dintre legături a fost observat la o grupare cu 6 molecule de apă, respectiv la gruparea $\text{K}^+ \cdot 6\text{H}_2\text{O}$ (unghi=58.1°).
- Nu au fost observate diferențe semnificative când unghiurile dintre legături au fost investigate din punct de vedere statistic (Statistica Friedman ANOVA=4.27 $p=0.5119$).

Rezultatele obținute în cazul de față sunt asociate cu conceptul de "diluție infinită", la care, de exemplu se poate spune că nu există alți ioni în vecinătate (Tabel 5.4).

Tabelul 5.4 conține:

- Rația dintre apă și amoniac pentru fiecare grupare de acest tip (H_2O și NH_3);
- Numărul ioni hidroniu (coloana H_3O^+) și ioni hidroxid (coloana HO^-) eliberați în urma formării grupărilor;
- Rata sarcinii eliberate per numărul total de molecule de apă implicate (coloana $[\text{H}_3\text{O}^+]/[\text{H}_2\text{O}]$);
- Reacția care conduce la formarea grupurilor (coloana "Reacția de formare" care reprezintă de asemenea cheia de verificare pentru calculele anterioare);
- Rata dintre atomii de azot și oxigen care corespund întregului aranjament în amestecul total al apei și amoniacului.

Tabel 5.4. Aranjamente pentru $\text{O}_x\text{N}_{20-x}\text{H}_{50}$

Nr.	Grupare	H_2O	NH_3	H_3O^+	HO^-	$[\text{H}_3\text{O}^+]/[\text{H}_2\text{O}]$	Reacția de formare	$\text{N}/(\text{N}+\text{O})\%$
0	O_0N_{20}	0	20	10	0	$+(10)/(0+10)$	$20\text{NH}_3 + 10\text{H}_2\text{O} \rightarrow \text{O}_0\text{N}_{20}\text{H}_{50} + 10\text{H}_3\text{O}^+$	$18/28=100$

Nr.	Grupare	H ₂ O	NH ₃	H ₃ O ⁺	HO ⁻	[⁺ / ⁻]/H ₂ O	Reacția de formare	N/(N+O)%
1	O ₂ N ₁₈	2	18	8	0	+(8)/(2+8)	18NH ₃ + 10H ₂ O → O ₂ N ₁₈ H ₅₀ + 8H ₃ O ⁺	18/28=64.3
2	O ₄ N ₁₆	4	16	6	0	+(6)/(6+4)	16NH ₃ + 10H ₂ O → O ₄ N ₁₆ H ₅₀ + 6H ₃ O ⁺	16/26=61.5
3	O ₅ N ₁₅	5	15	5	0	+(5)/(5+5)	15NH ₃ + 10H ₂ O → O ₅ N ₁₅ H ₅₀ + 5H ₃ O ⁺	15/25=60.0
4	O ₆ N ₁₄	6	14	4	0	+(4)/(4+6)	14NH ₃ + 10H ₂ O → O ₆ N ₁₄ H ₅₀ + 4H ₃ O ⁺	14/24=58.3
5	O ₈ N ₁₂	8	12	2	0	+(2)/(2+8)	12NH ₃ + 10H ₂ O → O ₈ N ₁₂ H ₅₀ + 2H ₃ O ⁺	12/22=54.5
6	O ₁₀ N ₁₀	10	10	0	0	+(0)/(0+10)	10NH ₃ + 10H ₂ O → O ₁₀ N ₁₀ H ₅₀	10/20=50.0
7	O ₁₂ N ₈	12	8	0	2	-(2)/(2+12)	8NH ₃ + 14H ₂ O → O ₁₂ N ₈ H ₅₀ + 2HO ⁻	8/22=36.4
8	O ₁₄ N ₆	14	6	0	4	-(4)/(4+14)	6NH ₃ + 18H ₂ O → O ₁₄ N ₆ H ₅₀ + 4HO ⁻	6/24=25.0
9	O ₁₅ N ₅	15	5	0	5	+(5)/(5+15)	5NH ₃ + 20H ₂ O → O ₁₅ N ₅ H ₅₀ + 5HO ⁻	5/25=20.0
10	O ₁₆ N ₄	16	4	0	6	+(6)/(6+16)	4NH ₃ + 22H ₂ O → O ₁₆ N ₄ H ₅₀ + 6HO ⁻	4/26=15.4
11	O ₁₈ N ₂	18	2	0	8	+(8)/(8+18)	2NH ₃ + 26H ₂ O → O ₁₈ N ₂ H ₅₀ + 8HO ⁻	2/28=07.1
12	O ₂₀ N ₀	20	0	0	10	+(10)/(10+20)	0NH ₃ + 30H ₂ O → O ₂₀ N ₀ H ₅₀ + 10HO ⁻	0/30=0.00

Unghiurile dintre legături în grupările ioni-apă pentru ionii investigați, au fost obținute cu succes iar grupurile formate cu un număr considerabil de molecule de apă pot explica dizolvarea ionilor investigați în apă.

”Numărul considerabil” de molecule sau denumit în acest caz ”Numărul magic de grupări”, la care acestea au fost investigate pentru:

- Li_nNa_{8-n}, Na_nK_{8-n}, și K_nLi_{8-n} (Fournier, 2008);
- (C₅H₅N)_n (H₂O)_m (n=1~2, m=1~4) (DeBlase et al., 2015);
- Metil terț-butil eter (MTBE) - grupări de apă (Di Palma și Bende, 2013);
- H⁺(NH₃)(piridină)(H₂O)_n, H⁺(NH₃)(piridină)₂(H₂O)_n (n = 18, 20 și 27) (Ryding et al., 2012);
- H⁺(NH₃)₅(H₂O)₂₀ (amoniu tetraedric încapsulat într-o structură de dodecaedru (H₂O)₂₀, găsită în clatrat) (Hvelplund et al., 2010).

Pentru $\text{Na}^+(\text{H}_2\text{O})_{20}$, $\text{Cl}^-(\text{H}_2\text{O})_{17}$, și $\text{Na}^+(\text{H}_2\text{O})_{100}$ utilizând algoritmul PBHaT (Burnham et al., 2006) a fost identificat minimum global. Acest algoritm este un hibrid capabil să eșantioneze eficient funcția de partiție de la minimul global la starea lichidă.

5.4. Concluzii

- Diferență semnificativă printre congeneri în tendința aranjamentelor în echilibru ca rezultat al interacțiunilor dintre moleculele de apă (în acest caz fiind utilizat modelul SM8);
- În absența altor ioni (la o diluție infinită) aranjamentul este în general cu simetrie alterată;
- În prezența altor ioni (în soluții concentrate) grupările de dodecaedru, conținând 8 atomi de litiu și 4 atomi de potasiu, sunt simetrice și stabile în timp ce grupările de dodecaedru conținând 6 atomi de sodiu sunt instabili în timp ce simetria este alterată.

CAPITOLUL 6. SIMILARITATEA MOLECULARĂ ÎN CLASE DE ANTI-INFLAMATOARE

6.1. Materiale și metode

- Colecție de date compuși naturali cu activitate antiinflamatoare, din literatura de specialitate;
- Aplicare măsuri de similaritate cu ajutorul programului DataWarrior.

6.1.1. Colectarea datelor

Tabel 6.1. Informația de interes colectată pentru specia *Dipteracanthus prostratus*

NCBI_ID Taxonomic	Compuși	SMILES	Uniprot_ID – Ținte biologice (Targets)
1052855	Acid protocatecuic	<chem>OC(=O)c1ccc(c(c1)O)O</chem>	O75496, P22748, P23280,
	Acid galic	<chem>OC(=O)c1cc(O)c(c(c1)O)O</chem>	P10145, P22748, P23280,
	Acid galic metil ester	<chem>COC(=O)c1cc(O)c(c(c1)O)O</chem>	P83916, P55789, B2RXH2,

De exemplu, din datele experimentale se arată că 15 specii de plante, aparținând familiei *Acanthaceae*, au fost studiate pentru efectul lor anti-inflamator și în conformitate cu sistemul ITIS sunt cunoscute 39 de genuri aparținând familiei *Acanthaceae* (Tabel 6.2).

Tabel 6.2. Speciile de plante (36) ce aparțin celor 8 familii, luate în considerare

Familie	Gen/specie	NCBI_ID_Taxonomic
Acanthaceae 42%	<i>Acanthus ilicifolius</i>	328098
	<i>Adhatoda vasica</i>	141317
	<i>Andrographis paniculata</i>	175694
	<i>Asteracantha longifolia</i>	883475
	<i>Asystasia gangetica</i>	141292
	<i>Barleria prionitis</i>	4189
	<i>Barleria lupulina</i>	101743
	<i>Dipteracanthus prostratus</i>	1052855
	<i>Elytraria acaulis</i>	640489
	<i>Nelsonia campestris</i>	4193
	<i>Phlogocanthus thyrsoiflorus</i>	526790
	<i>Pseuderanthemum palatiferum</i>	1685563
	<i>Rhinacanthus nasutus</i>	537489
	<i>Ruellia tuberosa</i>	441035
	<i>Thunbergia laurifolia</i>	504053
Achariaceae 8%	<i>Carpotroche brasiliensis</i>	1633205
	<i>Flacourtia indica</i>	210376
	<i>Gynocardia odorata</i>	124848
Agaricaceae 5%	<i>Agaricus blazei</i>	79798
	<i>Agaricus bisporus</i>	5341
Aizoaceae 6%	<i>Glinus oppositifolius</i>	764175
	<i>Trianthema portulacastrum</i>	3548
Alariaceae 3%	<i>Undaria pinnatifida</i>	74381
Alismataceae 3%	<i>Alisma plantago-aquatica</i>	262913
	subsp. <i>Orientalis</i>	
Altingiaceae 3%	<i>Liquidambar styraciflua</i>	4400
Amaranthaceae 30%	<i>Aerva javanica</i>	240009
	<i>Alternanthera philoxeroides</i>	381410
	<i>Alternanthera sessilis</i>	221762
	<i>Amaranthus spinosus</i>	124765
	<i>Celosia cristata</i>	124768
	<i>Salicornia herbacea</i>	259302
	<i>Suaeda maritima</i>	126913

Familie	Gen/specie	NCBI_ID_Taxonomic
	<i>Cyathula prostrata</i>	221766
	<i>Pupalia lappacea</i>	240105
	<i>Spinacia oleracea</i>	3562

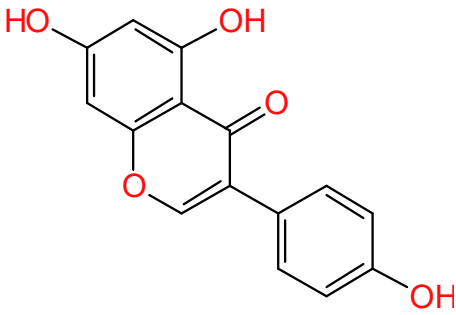
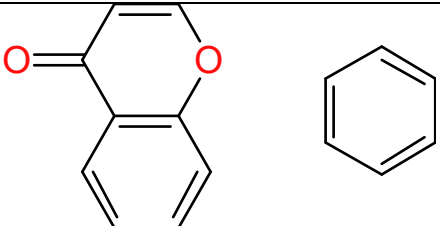
Anti-inflamatoarele ChEBI (în număr de 225 de compuși) au fost comparate cu un număr de 283 de compuși naturali extrași anterior din baza de date colectată (Tabel 6.4 din secțiunea Rezultate și discuții) din 35 specii de plante ce aparțin a 8 familii (Tabel 6.2).

6.1.2. Similaritatea utilizând scheletul molecular

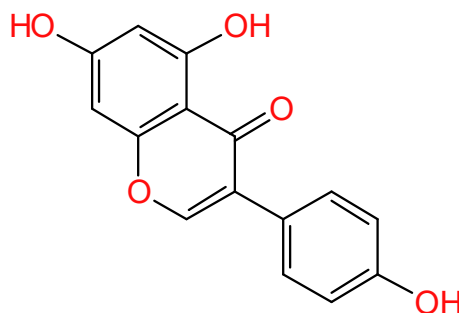
Din setul de molecule, se poate evidenția fragmentul central comun majorității moleculelor astfel încât substituenții sunt cei care diferă și conferă diferite conformații structurale.

Spre exemplu, în Tabelul 6.3, pentru molecula de genisteină (reprezentare DataWarrior), se pot observa tipurile de reprezentări moleculare după care s-au efectuat calculele în scopul caracterizării setului de compuși naturali colectat.

Tabel 6.3. Reprezentări de tip scaffold molecular luând ca și exemplu molecula de genisteină din setul de compuși naturali colectați - reprezentare din DataWarrior (Sander et al., 2015)

Genisteina 	
Tip scaffold	Reprezentare
(a) Sistemele ciclice simple	

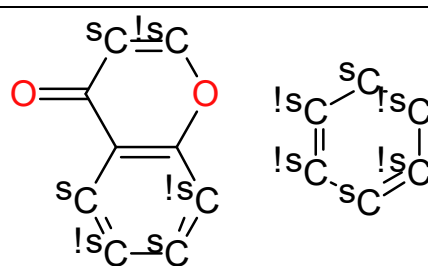
Genisteina



(b) Sisteme ciclice cu modele de substituție

s - substituenți

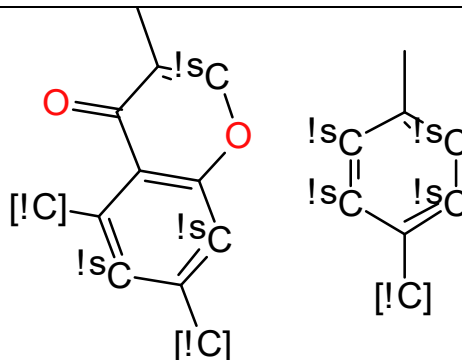
!s - substituent exo-ciclic, în afară de hidrogen



(c) Sisteme ciclice cu model de substituție carbon/hetero

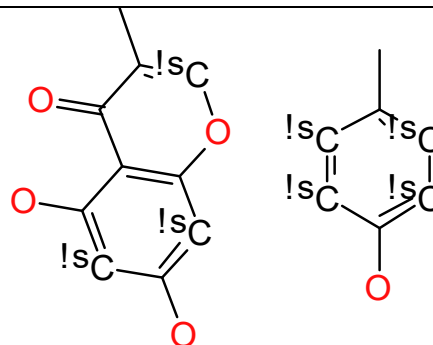
!s - substituent exo-ciclic, în afară de hidrogen

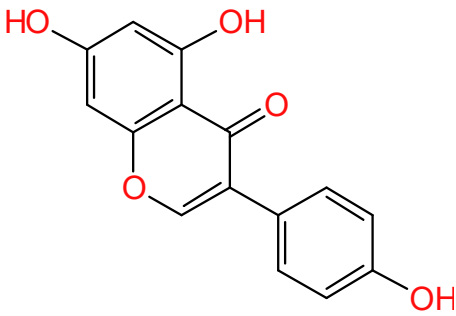
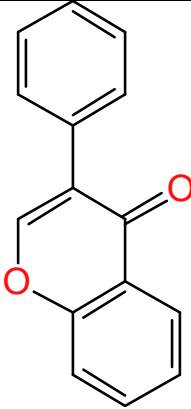
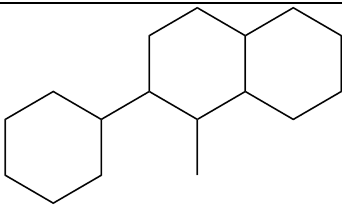
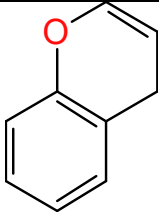
[!C] - orice atom substituent în afară de C sau H



(d) Sisteme ciclice fără modele de substituție atomică

!s - substituent exo-ciclic, în afară de hidrogen



Genisteina	
(e) Scaffold/Schelet Murcko	
(f) Graful scheletului Murcko	
(g) Sistemul ciclic central	

Cu ajutorul programului DataWarrior s-a calculat similaritatea prin diferite modalități în funcție de scopul propus și în funcție de metoda pe care dorim să o aplicăm. Pornind de la metoda de calcul simplificată a similarității moleculare la calculul similarității în spațiul 3D ce ține cont atât de geometria cât și de modul de legare a atomilor.

Rezultatul a constat în compararea moleculelor prin intermediul acestor descriptori specifici fiecărei molecule din set. Valorile rezultate (scorul de similaritate) bazate pe calcule euclidiene SOM (Self Organizing Map) (Sander et al., 2015) și măsura de

calcul Tanimoto (Willett, 2011), redau cât de multe au în comun compușii comparați, respectiv cât de similare sunt cele două molecule (secțiunea Rezultate și discuții).

6.1.3. Analiza componentelor principale și maparea datelor

Pe baza scaffoldurilor s-au construit grafice de vizualizare a similarității prin prisma și PCA (Principal component Analysis) în DataWarrior. De asemenea s-au generat moleculele din vecinătatea unei molecule principale din punct de vedere al similarității.

Pentru calcularea similarității a fost considerat coeficientul Tanimoto, iar prin descriptorii considerați, conform (Sander et al., 2015):

- S-au utilizat reprezentări binare ale fragmentelor substructurilor (FragFp);
- S-a considerat stereochemia, s-au numărat fragmentele în duplicat și s-au codificat heteroatomii (SkelSpheres).

Algoritmul de lucru a fost considerat în următorii pași:

- Poziționarea setului de molecule în spațiul 2D;
- Calcularea matricei similarității între toate moleculele;
- Localizarea a cele mai similare molecule vecine;
- Între a câte două molecule vecine, evidențierea forțelor de atracție, care cresc odată cu scorul de similaritate și distanța dintre acestea;
- Vizualizarea similarității și a moleculelor similare.

6.2. Rezultate și discuții

6.2.2. Clusterii de similaritate

Vizualizarea relațiilor de similaritate este evidențiată pe baza clusterilor formați prin corelația cu mai similari componenți din vecinătate. Datele au fost generate cu ajutorul programului DataWarrior în urma calculelor de similaritate comparând cele două seturi de molecule adnotate ANTIINFLAMMATORY (albastru) și CHEBI (roșu) (Figura 6.2).

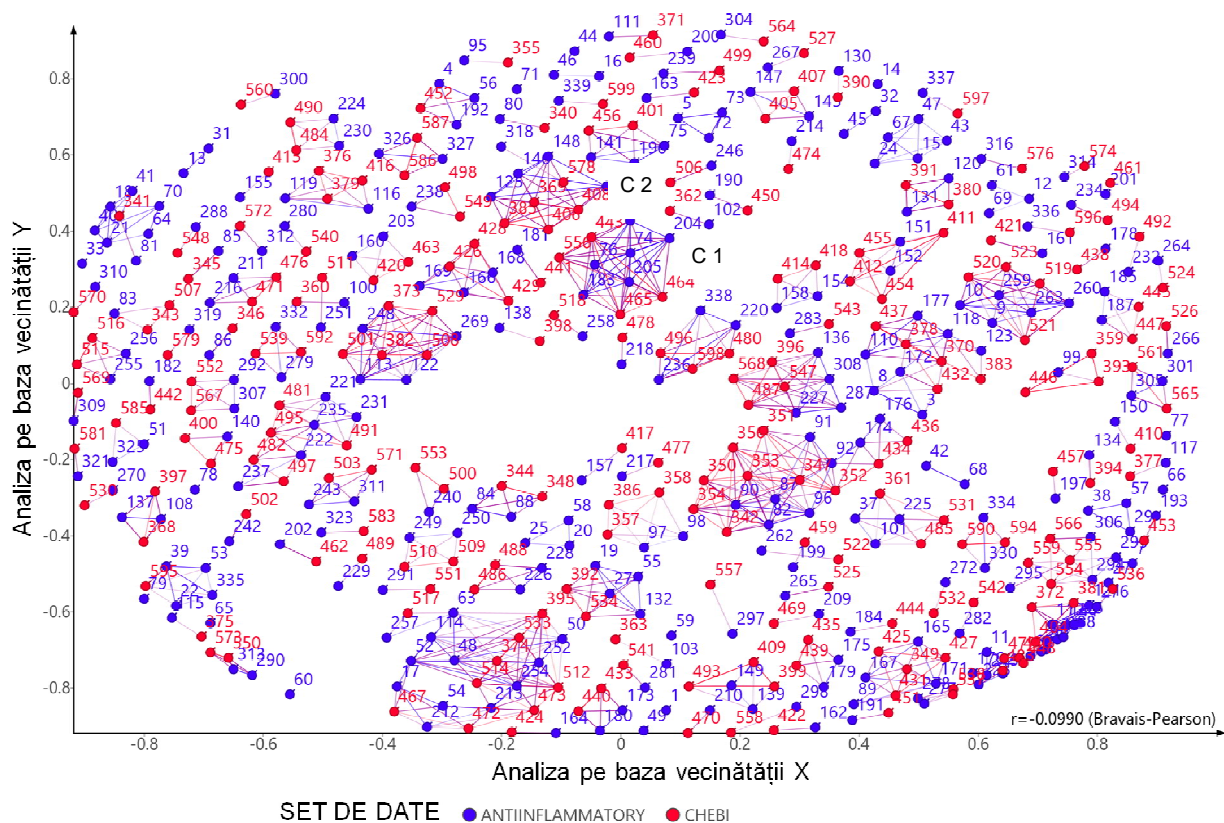


Figura 6.2. Clusterii de similaritate rezultați în urma calculării similarității pe baza vecinătății (vizualizare în DataWarrior) - exemplu cluster 1 (C1), cluster 2 (C2).

Această corelație a similarității dintre seturile de date a fost efectuată pe baza calculului coeficientului Bravais-Pearson, rezultând un coeficient de corelație foarte aproape de valoarea ideală 1, $r = 0.0990$ (Figura 6.2).

Clusterii formați ce pot fi observați în Figura 6.2 (exemplu cluster 1 format din moleculele corespunzătoare ID-urilor: 443, 556, 441, 74, 205, 76, 183, 518, 258, 465, 464; exemplu cluster 2 format din moleculele corespunzătoare ID-urilor: 146, 125, 549, 428, 38, 36, 578, 408, 400, 181) reprezintă subgrupuri de molecule, conectate între ele pe baza vecinătății ce aparțin ambelor seturi de date analizate, indicând o bună corelație între acestea în funcție de similaritate.

Pe baza descriptorilor structurali (SMILES) și SkelSpheres, în același timp considerând și scheletele structurale (scaffold-uri) de tip sistemele ciclice simple

corespunzătoare componentelor, a fost urmărită similaritatea în cadrul clusterilor formați (Figura 6.3 și Figura 6.4).

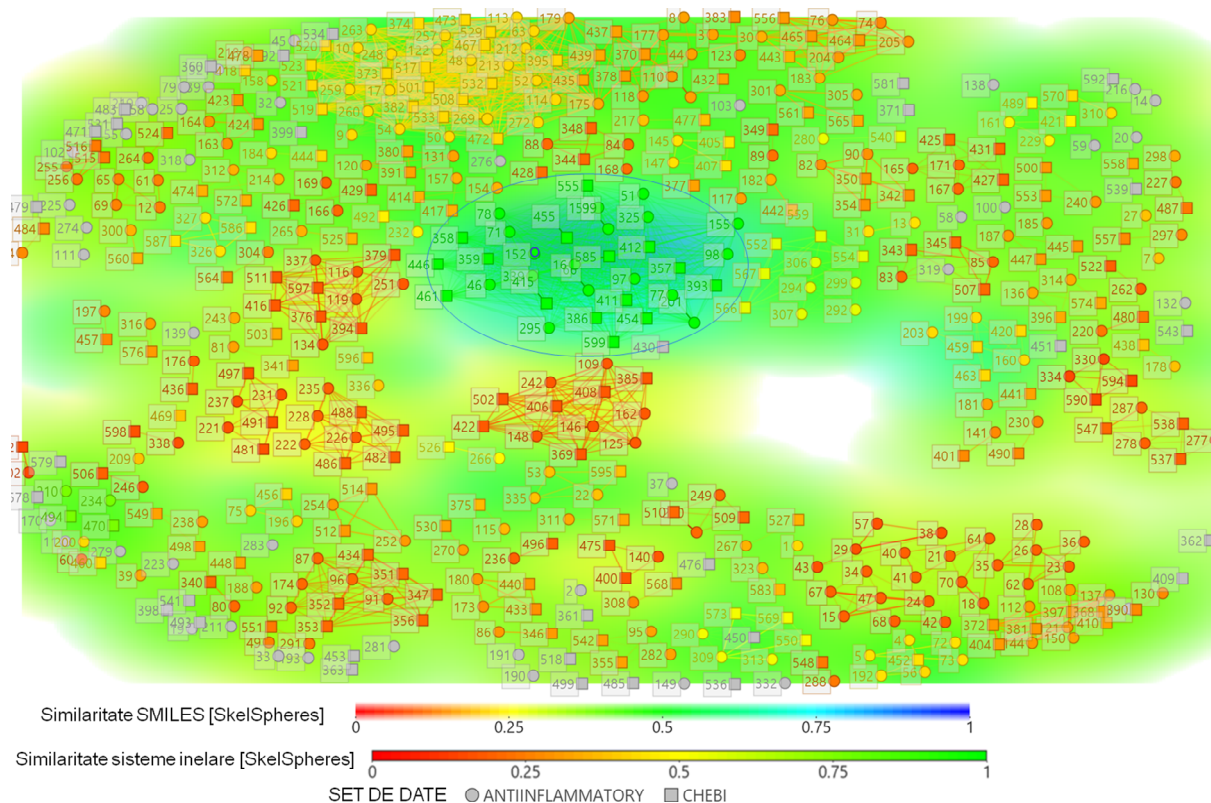


Figura 6.3. Similaritatea pe baza scaffold-urilor de tipul sistemelor ciclice simple (Ring Systems), SMILES și descriptorii SkelSpheres, formează clusterii ai căror componente se găsesc a fi foarte similari din punct de vedere al caracteristicilor structurale. Similaritatea raportată la componentul principal (lider) ID 152 (vezi Tabel Suplimentar 6.1) din setul de compuși naturali.

Un grad înalt de similaritate este evidențiat având ca și lideri (componente principale) de comparație, câte o moleculă din fiecare cluster. În figura 6.3 componentul principal este reprezentat de ID-ul 152 corespunzător acidului vanilic din setul de compuși naturali (ANTIINFLAMMATORY).

În figura 6.4 componentul principal este reprezentat de ID-ul 109 corespunzător acidului oleanolic din setul de compuși naturali. Relațiile de similaritate ale componentelor celor doi clusteri menționați sunt redată în Tabelul 6.5.

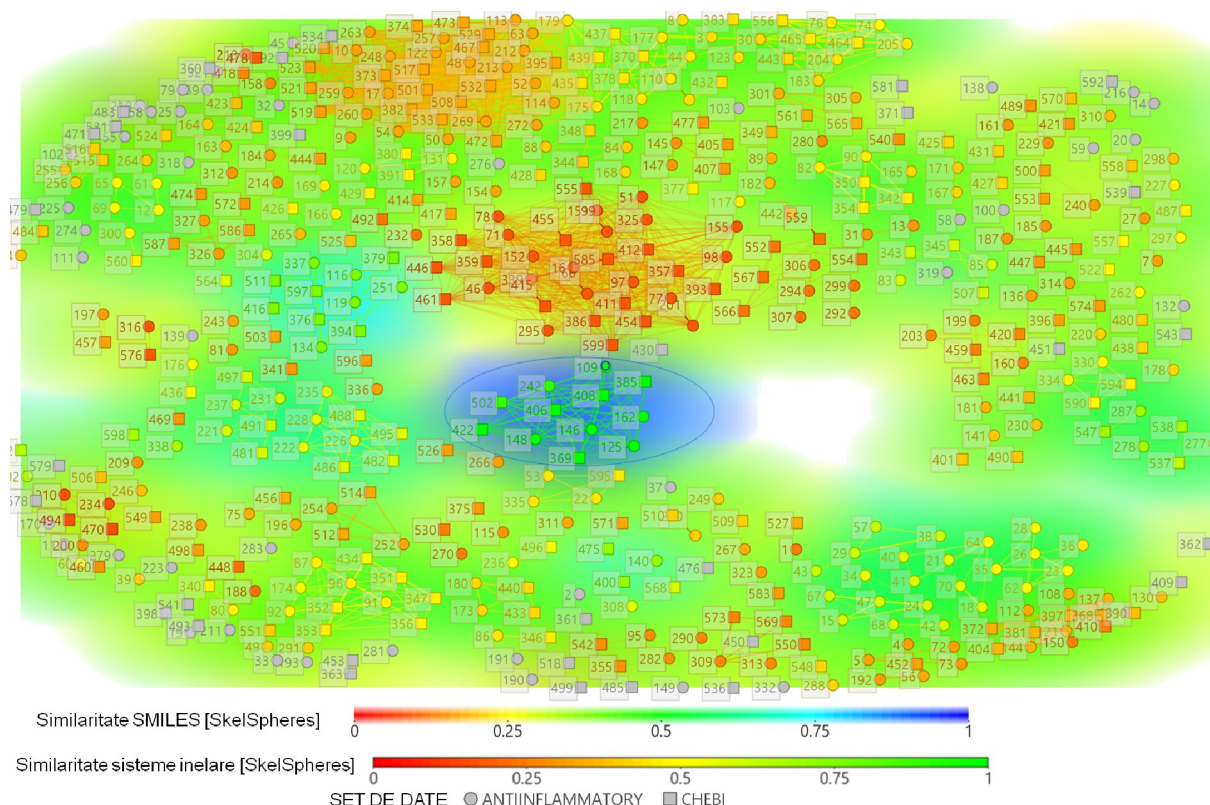
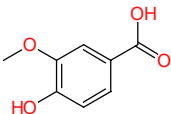
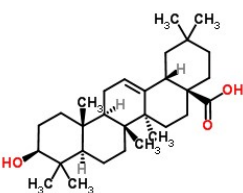
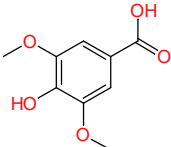
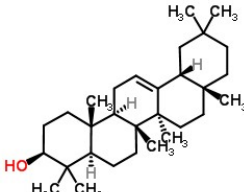
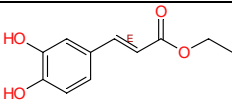
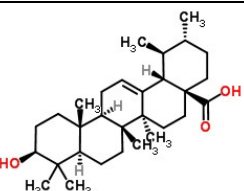
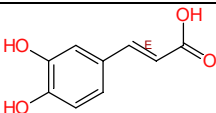
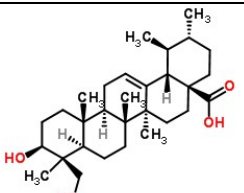
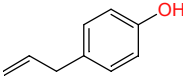
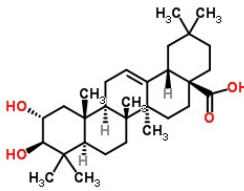
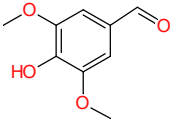
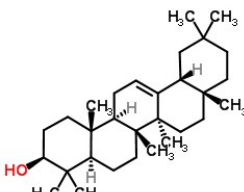
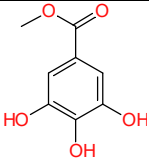
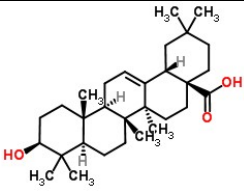
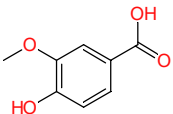
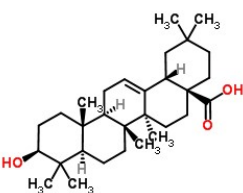
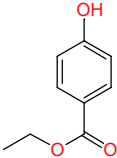
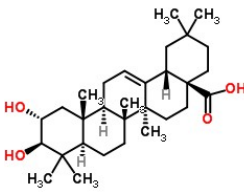
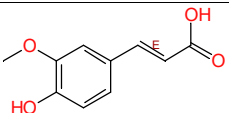
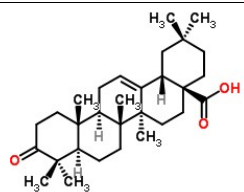
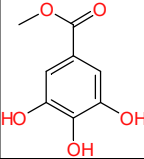
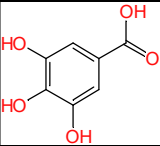
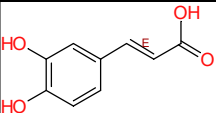
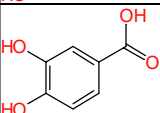
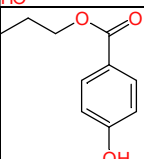


Figura 6.4. Similaritatea pe baza scaffold-urilor de tipul sistemelor ciclice simple (Ring Systems), SMILES și descriptorii SkelSpheres, formează clusterii ai căror componenți se găsesc a fi foarte similari din punct de vedere al caracteristicilor structurale. Similaritatea raportată la componetul principal (lider) ID 109 (vezi Tabel Suplimentar 6.1) din setul de compuși naturali.

Tabel 6.5. Relațiile de similaritate ale componenților celor doi clusteri de similaritate (cu liderii corespunzători considerați) reprezentanți ai figurilor 6.3 și 6.4.

Lider cluster 1 (ANTIINFLAMMATORY:A)				Lider cluster 2 (ANTIINFLAMMATORY:A)			
 152				 109			
ID	Structură 2D	Scor	Set	ID	Structură 2D	Scor	Set
455		1	CHEBI	242		1	A

Lider cluster 1 (ANTIINFLAMMATORY:A)				Lider cluster 2 (ANTIINFLAMMATORY:A)			
							
152				109			
454		1	CHEBI	162		1	A
599		1	CHEBI	406		0.934	CHEBI
585		1	CHEBI	385		0.933	CHEBI
555		1	CHEBI	408		0.927	CHEBI
461		1	CHEBI	422		0.910	CHEBI
446		1	CHEBI	369		0.900	CHEBI

Lider cluster 1 (ANTIINFLAMMATORY:A)				Lider cluster 2 (ANTIINFLAMMATORY:A)			
							
152				109			
46		0.916	A	148		0.877	A
66		0.887	A	502		0.877	CHEBI
99		0.858	A				
98		0.849	A				
325		0.845	A				
155		0.839	A				
71		0.829	A				

Luând în considerare câte un component lider din fiecare cluster de similaritate distinct format, se observă cum similaritatea globală calculată din punct de vedere al vecinătății se schimbă în acord cu aceștia (152 și 109). Axându-ne pe cei doi clusteri cu un scor de similaritate aproape egal cu 1, se observă legătura de similaritate dintre

compușii din cele două seturi (CHEBI și compușii naturali antiinflamatori). Astfel cei mai similari compuși din cei doi clusteri pot fi urmăriți în Tabelul 6.5.

6.2.3. Maparea similarității utilizând scaffold-urile moleculare

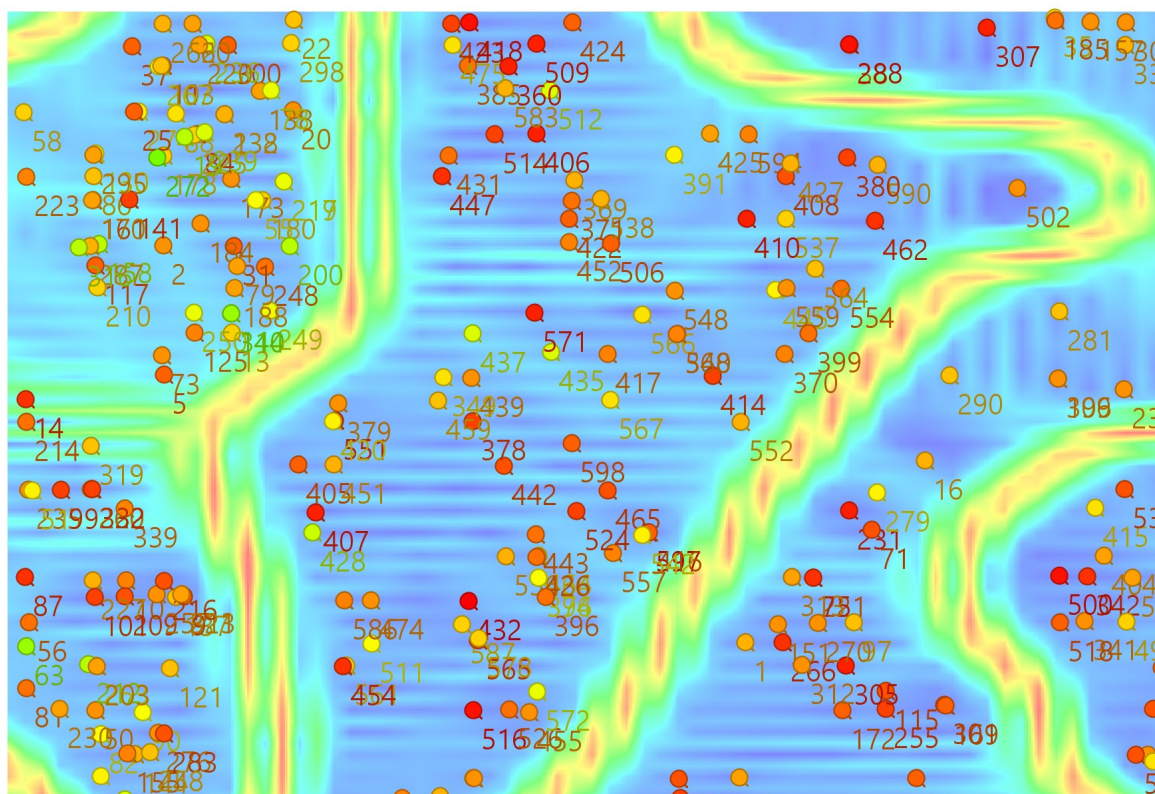


Figura 6.5. Self Organising Map pentru aranjarea moleculelor în funcție de asemănarea scaffold-urilor corespunzătoare compușilor și vizualizarea clusterilor de similaritate pe baza vecinătății și descriptorilor SkelSpheres (vizualizare în DataWarrior).

Prin SOM s-au luat în calcul și pot fi vizualizate valorile de similaritate exacte (Figura 6.5), iar în cazul PCA, scaffold-urile sunt separate în funcție de două sau trei componente principale, după cum se observă în cele ce urmează (Figura 6.6, Figura 6.7, Figura 6.8).

6.2.4. Analiza componentelor principale

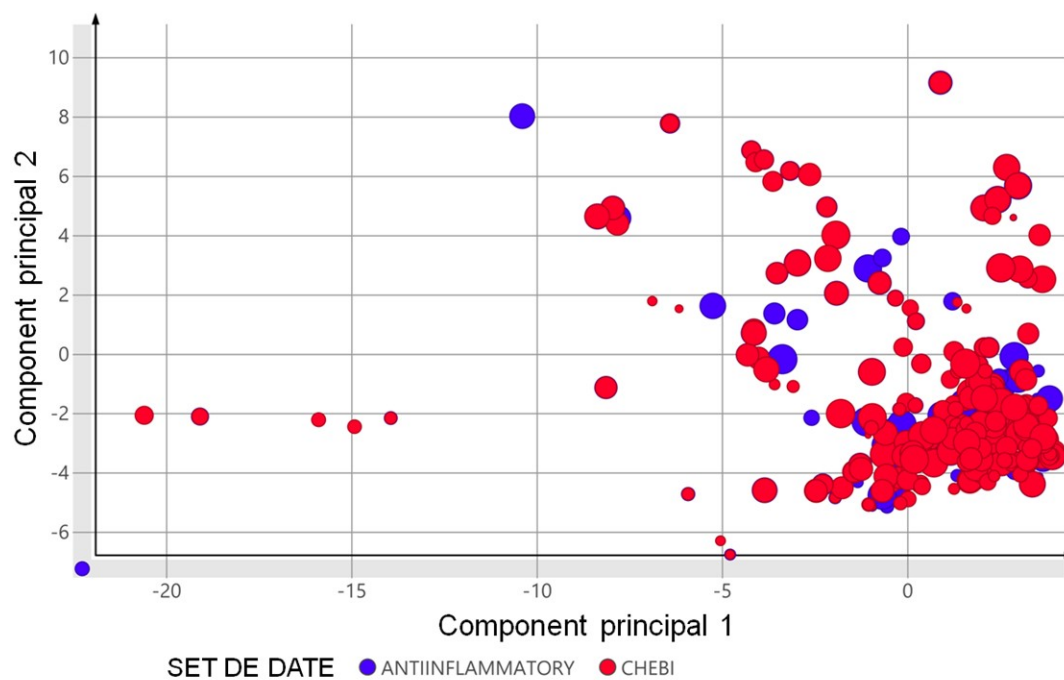


Figura 6.6. Analiza componentelor principale în vecinătatea Y, în urma calculului similarității utilizând scaffold-ul molecular și FragFp (conform rezultatului din DataWarrior), cu o varianță 23.2% a componentului principal 1 și 24.6% a componentului principal 2.

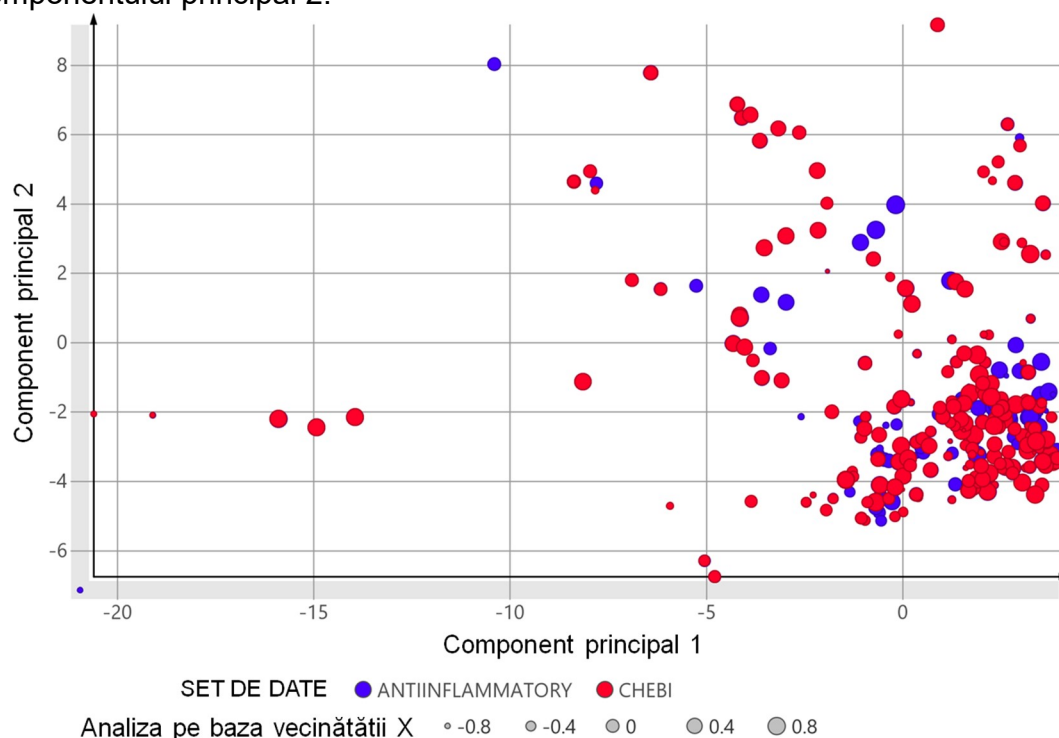


Figura 6.7. Analiza componentelor principale în vecinătatea X, în urma calculului similarității utilizând scaffold-ul molecular și FragFp (conform rezultatului din DataWarrior), cu o varianță 35.6% a componentului principal 1 și 40.2% a componentului principal 2.

Urmărind cele două figuri (Figura 6.6 și Figura 6.7), se poate spune că există o similaritate semnificativă între molecule și că în raport cu acestea există un număr mic de molecule care nu sunt similare. Considerând stereochemia, prin descriptorul SkelSpheres s-au numărat fragmentele în duplicat și s-au codificat heteroatomii calculând similaritatea pe baza acestui descriptor (Figura 6.8).

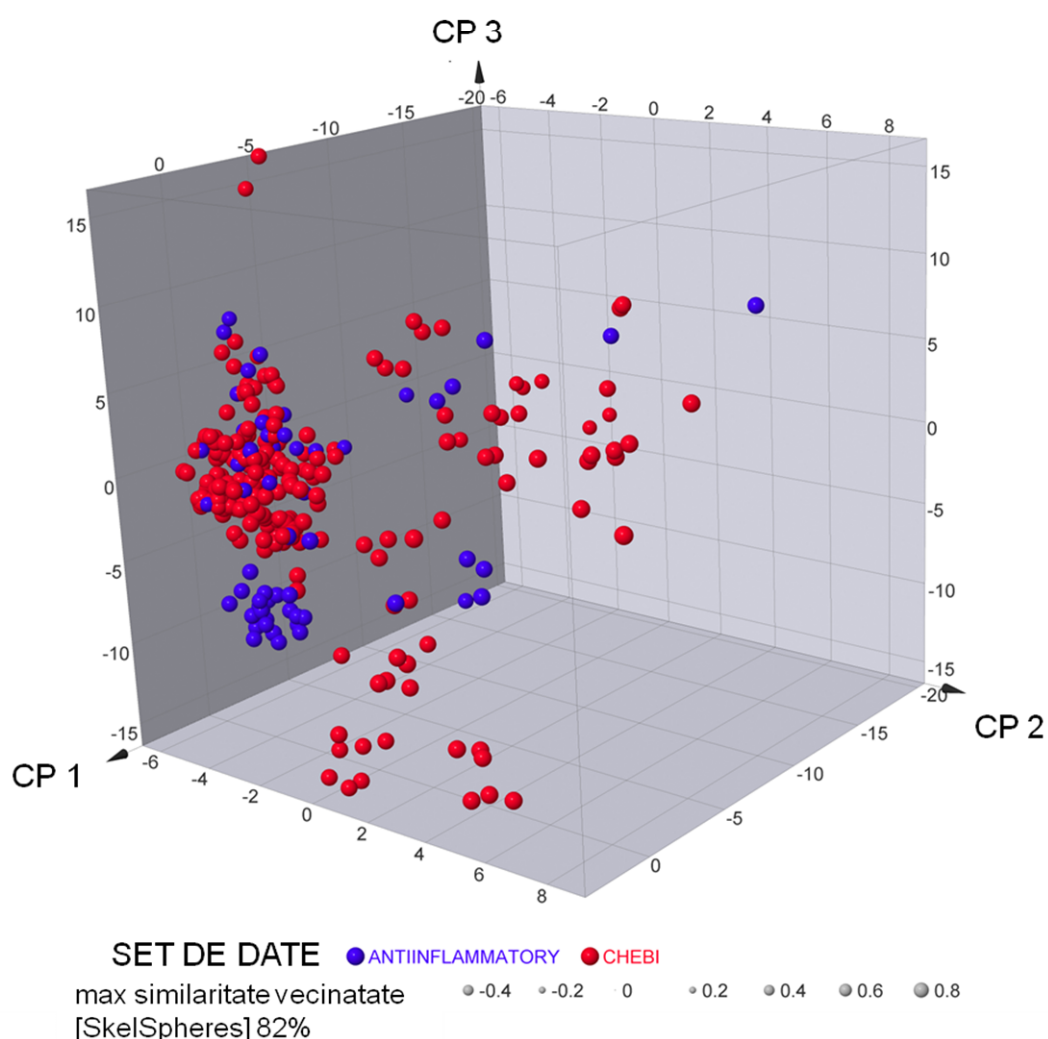


Figura 6.8. Vizualizarea 3D a similarității moleculelor celor două seturi de date (ANTIINFLAMMATORY - albastru; CHEBI - roșu) și evaluarea acestora prin prisma analizei componentelor principale (CP) utilizând descriptorul SkelSpheres în spațiul 3D.

În cadrul analizei componentelor principale (PCA) în spațiul 3D, relația de similaritate dintre componenți este confirmată prin graficul rezultat în Figura 6.8. Similaritatea este astfel redată prin semnificația statistică dată de rezultatul scorului de similaritate de 82% cu valori între 0.6 și 0.8.

6.4. Concluzii

- O serie de compuși naturali din setul de date colectat, sunt similari cu structurile compuşilor medicamentoși din setul luat ca referință;
- Un compus chimic și derivatul său pot fi găsiți într-o diversitate de extracte de plante;
- Rezultatele contribuie la confirmarea și recentralizarea periodică a compuşilor găsiți ca având activități antiinflamatoare.

CAPITOLUL 7. MODEL DE RELAȚIE CANTITATIVĂ STRUCTURĂ-PROPRIETATE ÎN CLASE DE STEROIDE

7.1. Materiale și metode

Tabel 7.1. Cei 40 de derivați ai moleculei de 7 β -hidroxisteroid

Nr.	PubChem ID	logP	Nr.	PubChem ID	logP	Nr.	PubChem ID	logP	Nr.	PubChem ID	logP
1	12760132	10.2	11	76310266	8.2	21	57390981	5.2	31	76322257	10.7
2	70682679	7.1	12	56663807	6.4	22	16758147	8.5	32	76325907	3.8
3	70682680	6.5	13	56847117	6.2	23	22213946	6.2	33	76327928	3.8
4	70688976	6.2	14	70686910	6	24	16759984	5.9	34	76333144	4.2
5	70693211	6.2	15	70691082	7.1	25	16758161	4.2	35	371617	6.1
6	70697302	6.5	16	11647965	8.4	26	76336739	11.2	36	313039	8
7	12836861	4.7	17	52947587	4.9	27	57396177	3	37	9922115	4.2
8	24867469	4.2	18	24982302	3.8	28	57399636	3	38	9924252	5.4
9	16082386	8.1	19	49823443	5.9	29	57401396	3	39	11551321	3
10	12358742	5.2	20	22216291	3	30	76322252	9.8	40	11957457	4.2

7.1.1. Optimizarea geometriei

Geometriile moleculare au fost optimizate în programul HyperChem la nivelul semi-empiric PM3 (Parametrizat Model nr. 3). Fișierele log-date cu colectarea datelor au fost extrase utilizând programul utilitar JSCHEM (Harsa et al., 2014).

7.1.2. Construcția hipermoleculei

Pe baza fiecărei poziție a atomilor ce formează hipermolecula (vezi Figura 7.1), au fost calculați vectorii binari și fragmentele de masă corespunzătoare pentru fiecare moleculă din set.

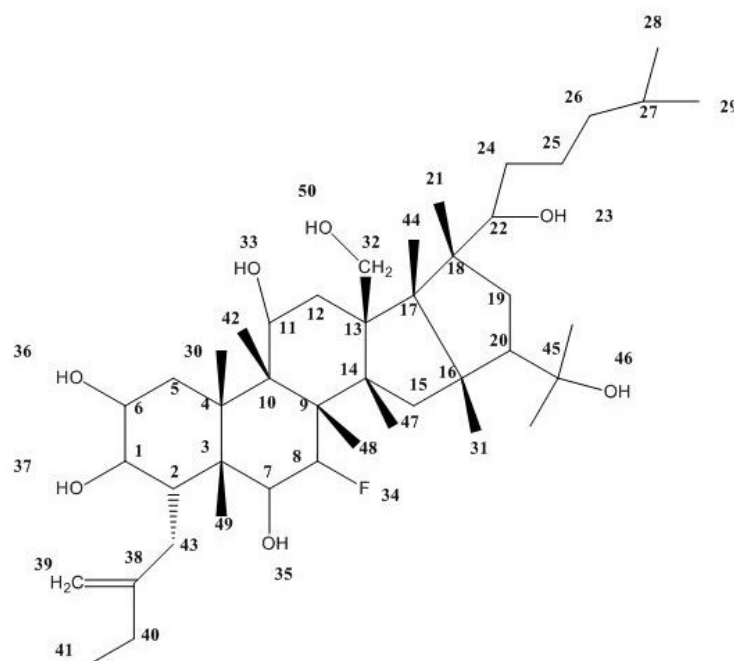


Figura 7.1. Hipermolecula formată în urma suprapunerii a patruzeci de liganzi. Reprezentarea grafică 2D și numerotarea grafului (Pruteanu et al., 2016) realizate în ChemBioDraw („ChemBioDraw 14.0 User Guide”, 2016).

7.1.3. Descriptorii topologici

Tabel 7.2. Indicii topologici calculați pentru derivații 7 β -hidroxisteroid din Tabelul 7.1

Mol.	AD	CON	Di	D3D	De	CjDi	CjDe	CFDi	CFDe
1	35	35	2522	4138.48	7313	4573.5	1750.5	5062	1809
2	34	34	2670	3980.77	6965	4467	1923.5	4918	1966.5
3	33	33	2369	3510.96	6451	4047	1665	4487	1706.5

Mol.	AD	CON	Di	D3D	De	CjDi	CjDe	CFDi	CFDe
4	33	33	2342	3522.35	6424	4019.5	1638	4460	1679.5
5	33	33	2342	3355.66	6424	4019.5	1638	4460	1679.5
6	33	33	2369	3520.97	6451	4047	1665	4487	1706.5
7	25	25	926	1354.38	3297	1729.5	546.5	1983	572.5
8	24	24	802	1163.2	2969	1504.5	463	1739	486.5
9	33	33	2335	3350.73	6511	4029	1648.5	4550.5	1693
10	26	26	1052	1868.14	3627	1956.5	632	2229	660.5
11	35	35	2237	3659.41	8399	4455.5	1302	5281	1361
12	33	33	2345	3436.73	6348	3996	1647.5	4362	1687.5
13	33	33	2342	3309.13	6424	4019.5	1638	4460	1679.5
14	32	32	2098	3117.36	5967	3657.5	1436.5	4086	1476.5
15	34	34	2560	3759.51	6855	4349.5	1813.5	4808	1856.5
16	36	36	2435	3417.31	8962	4849.5	1424.5	5718	1485
17	24	24	796	1153.47	2949	1490.5	461	1732	488
18	27	27	1149	1681.56	3971	2130	715.5	2485	747.5
19	23	23	699	988.57	2662	1310	402	1516	421.5
20	25	25	887	1279.08	3279	1655	535.5	1966	562.5
21	26	26	1052	1530.66	3627	1956.5	632	2229	660.5
22	36	36	2436	3497	9036	4811.5	1459.5	5703.5	1521.5
23	33	33	2342	3355.66	6424	4019.5	1638	4460	1679.5
24	23	23	699	1023	2662	1310	402	1516	421.5
25	24	24	802	1158.41	2969	1504.5	463	1739	486.5
26	35	35	2944	4219.49	7659	5068.5	2080.5	5527	2130
27	25	25	887	1279.08	3279	1655	535.5	1966	562.5
28	25	25	887	1279.08	3279	1655	535.5	1966	562.5
29	25	25	887	1279.08	3279	1655	535.5	1966	562.5
30	34	36	2668	3785.72	7103	4609	1864.5	5048.5	1911.5
31	34	34	2668	3828.26	7103	4609	1864.5	5048.5	1911.5
32	27	27	1172	1646	3967	2178	728	2483	763
33	27	27	1172	1654.23	3967	2177.5	728	2482.5	762.5
34	24	24	802	1169.48	2969	1504.5	463	1739	486.5
35	24	24	732	1086.05	2660	1437	483.5	1707.5	501.5
36	32	32	2276	3204.88	6075	3832	1619	4173	1656.5
37	24	24	802	1169.48	2969	1504.5	463	1739	486.5
38	28	30	1412	2008.12	4395	2535	914	2829	944.5
39	25	25	887	1279.08	3279	1655	535.5	1966	562.5
40	24	24	802	1169.48	2969	1504.5	463	1739	486.5

7.1.4. Pozițiile semnificative

Tabel 7.3. Pozițiile semnificative statistic, corelate cu masa fragmentelor

Mol.	p17	p18	p26	p28	p33	p34	p35	p36	p37	p40	p43	p50
1	12.011	12.011	12.011	0	0	0	0	0	17.007	0	12.011	0
2	12.011	12.011	12.011	12.011	0	0	17.007	0	17.007	0	0	0
3	12.011	12.011	12.011	12.011	0	0	17.007	0	17.007	0	0	0
4	12.011	12.011	12.011	0	0	0	17.007	0	17.007	0	0	0
5	12.011	12.011	12.011	0	0	0	17.007	0	17.007	0	0	0
6	12.011	12.011	12.011	12.011	0	0	17.007	0	17.007	0	0	0
7	12.011	17.007	0	0	0	0	0	0	17.007	0	0	0
8	12.011	17.007	0	0	0	0	0	0	17.007	0	0	0
9	12.011	12.011	12.011	0	0	19	17.007	0	17.007	0	0	0
10	12.011	12.011	0	0	0	0	0	0	17.007	0	0	0
11	12.011	12.011	0	0	0	0	0	0	17.007	0	12.011	0
12	12.011	12.011	12.011	0	0	0	0	0	17.007	0	0	0
13	12.011	12.011	12.011	0	0	0	17.007	0	17.007	0	0	0
14	12.011	12.011	12.011	0	0	0	17.007	0	17.007	0	0	0
15	12.011	12.011	12.011	12.011	0	0	17.007	0	17.007	0	0	0
16	12.011	12.011	0	0	0	0	0	0	17.007	0	12.011	0
17	12.011	0	0	0	0	0	0	17.007	17.007	0	0	0
18	12.011	12.011	0	0	17.007	0	0	0	17.007	0	0	0
19	12.011	0	0	0	0	0	0	0	17.007	0	0	0
20	12.011	17.007	0	0	17.007	0	0	0	17.007	0	0	0
21	12.011	12.011	0	0	0	0	0	0	17.007	0	0	0
22	12.011	12.011	0	0	0	0	17.007	0	0	0	0	0
23	12.011	12.011	12.011	0	0	0	17.007	0	17.007	0	0	0
24	12.011	0	0	0	0	0	0	0	17.007	0	0	0
25	12.011	17.007	0	0	0	0	0	0	17.007	0	0	0
26	12.011	12.011	12.011	0	0	0	0	0	17.007	12.011	12.011	0
27	12.011	17.007	0	0	17.007	0	0	0	17.007	0	0	0
28	12.011	17.007	12.011	0	17.007	0	0	0	17.007	0	0	0
29	12.011	17.007	0	0	17.007	0	0	0	17.007	0	0	0
30	12.011	12.011	12.011	0	0	0	0	0	17.007	0	12.011	0
31	12.011	12.011	12.011	0	0	0	0	0	17.007	12.011	12.011	0
32	12.011	12.011	0	0	0	0	0	0	17.007	0	0	17.007
33	12.011	12.011	0	0	0	0	0	0	17.007	0	0	17.007
34	12.011	17.007	0	0	0	0	0	0	17.007	0	0	0
35	0	0	0	0	0	0	17.007	17.007	17.007	0	0	0
36	12.011	12.011	12.011	17.007	0	0	0	0	17.007	0	0	0
37	12.011	17.007	0	0	0	0	0	0	17.007	0	0	0
38	12.011	12.011	0	0	0	0	0	0	17.007	0	0	0
39	12.011	17.007	0	0	17.007	0	0	0	17.007	0	0	0
40	12.011	17.007	0	0	0	0	0	0	17.007	0	0	0

Proprietatea logP a fost modelată astfel folosind fragmente de masă ca și caracteristici structurale ale moleculelor din setul ales. Modelul a fost validat prin procedurile "leave-one-out" și "training vs. test".

7.2. Rezultate și discuții

7.2.1. Model de regresie obținut pe baza pozițiilor semnificative

A fost obținut un model semnificativ de regresie prin care s-au identificat șapte variabile ca poziții semnificative (Ec.1 este reprezentată în Tabelul 7.4).

$$\log P = 36.2431 + 0.0180 \cdot \text{CjDi} - 1.6780 \cdot \text{AD} - 0.0353 \cdot \text{Di} + 0.0228 \cdot \text{CjDe} - 0.0605 \cdot \text{P18} - 0.0542 \cdot \text{P33} - 0.0495 \cdot \text{P35} \quad (\text{Ec.1})$$

$$R^2 = 0.9610, R^2_{\text{adj}} = 0.9525, Q^2 = 0.9413; s = 0.4808, n = 40$$

F-statistica (p-value) = 113 ($1.02 \cdot 10^{-20}$) unde,

R^2 = coeficientul de determinare,

R^2_{adj} = coeficientul de determinare ajustat,

Q^2 = coeficientul de determinare în analiza leave-one-out,

s = eroarea standard estimată,

n = mărimea probei; F-statistica = statistica Fisher,

p-value = probabilitatea de a obține un model semnificativ.

Table 7.4. Pozițiile semnificative și coeficienții lor de regresie

Variabile	Coeficienții	Eroarea standard	t Stat (p-value)
Intercepția	36.2431	4.4755	8.10 ($3.01 \cdot 10^{-9}$)
CjDi	0.0180	0.0020	8.82 ($4.41 \cdot 10^{-10}$)
AD	-1.6780	0.2383	-7.04 ($5.53 \cdot 10^{-8}$)
Di	-0.0353	0.0060	-5.85 ($1.69 \cdot 10^{-6}$)
CjDe	0.0228	0.0043	5.35 ($7.12 \cdot 10^{-6}$)
P18	-0.0605	0.0197	-3.07 ($4.37 \cdot 10^{-3}$)
P33	-0.0542	0.0157	-3.44 ($1.63 \cdot 10^{-3}$)
P35	-0.0495	0.0137	-3.61 ($1.04 \cdot 10^{-3}$)

CjDi = distanța Cluj; AD = Adiacența;

D_i = Distanța; $CjDe$ = Cluj detur;
 $P18$ = Poziția 18; $P33$ = Poziția 33; $P35$ = Poziția 35

Modelul cu cel mai mic număr de predictori a fost ales ca fiind modelul cu cea mai mare putere explicativă. Acesta a fost obținut prin aplicarea succesivă și repetată a metodei pas cu pas pentru setul de descriptori din tabelele 7.3 și 7.4.

7.2.2. Validarea modelului prin procedura leave-one-out

Pentru validarea modelului a fost efectuată o analiză de tip leave-one-out cu un coeficient de determinare în analiza leave-one-out $Q^2 = 0.9413$ (vezi Ec.1).

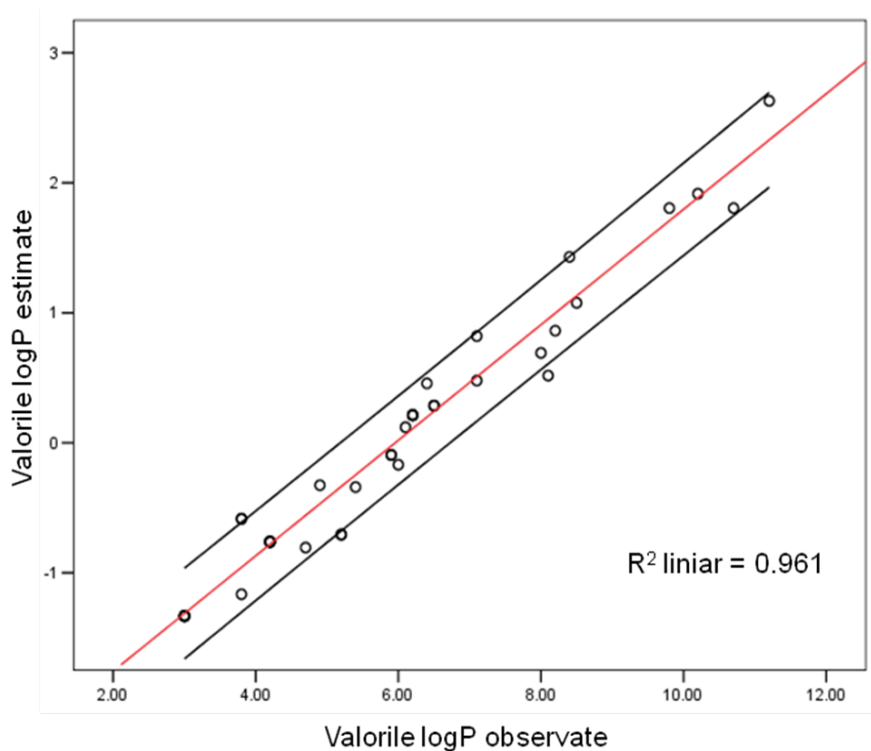


Figure 7.2. Modelul estimat (linia roșie - modelul obținut iar liniile negre reprezintă intervalul de încredere de 95%)

7.2.3. Validarea modelului prin procedura training vs. test

Modelul rezultat a fost validat și prin procedura de tip training vs. test pe setul de cele 40 de molecule analizate (Tabel 7.1), pentru setul de descriptori.

Setul de 40 de molecule a fost împărțit în câte două seturi, 24 de molecule în setul training și restul de 16 molecule în setul de testare.

Următoarele molecule au fost alese aleator pentru a fi parte din setul training: 57396177, 49823443, 16082386, 16758147, 22216291, 9922115, 70688976, 70682680, 22213946, 11647965, 57390981, 12358742, 313039, 76325907, 57401396, 76327928, 76322252, 16759984, 24982302, 52947587, 12760132, 76336739, 76310266 și 70697302.

Ecuția de regresie obținută cu setul de molecule training a fost utilizată pentru a prezice valorile logP pentru restul moleculelor din setul test:

$$\log P_{\text{train}} = 27.014 - 0.08755 \cdot p18 - 0.03625 \cdot p33 - 0.06206 \cdot p35 - 1.2697 \cdot AD - 0.001154 \cdot DI + 0.006573 \cdot CjDi \text{ (Ec.2)}$$

$$R^2_{\text{train}} = 0.9337$$

$$F_{\text{train}} = 40 \text{ (} p_F < 5 \cdot 10^{-9} \text{)}$$

$$R^2_{\text{test}} = 0.873$$

$$F_{\text{test}} = 9 \text{ (} p_F < 2.6 \cdot 10^{-3} \text{) unde,}$$

R^2 = coeficientul de determinare (pentru setul training respectiv setul test),

p-value = probabilitatea de a obține un model semnificativ.

7.4. Concluzii

- Validarea modelului conform procedurii leave-one-out;
- Totodată a fost confirmat faptul că o dată cu scăderea numărului de variabile, semnificația statistică a unui model, scade.
- Pe baza modelului obținut, se poate spune că pozițiile 18, 33 și respectiv 35 sunt cele datorită cărora semnificația statistică scade (toate aceste poziții având un efect negativ asupra valorii coeficienților logP).

CAPITOLUL 8. EFECTUL TERAPEUTIC AL ARSENAȚILOR ASUPRA LINIILOR CELULARE DE CANCER DE SÂN

8.1. Materiale și metode

Cu scopul de a evalua efectul arsenatilor ($2 \text{ HNO}_3 + \text{As}_2\text{O}_3 + 2 \text{ H}_2\text{O} \rightarrow 2 \text{ H}_3\text{AsO}_4 + \text{N}_2\text{O}_3$) asupra a trei linii celulare de cancer de sân (MCF-7 dublu pozitiv, Hs578T triplu negativ, MDA-MB-231 triplu negativ) și o linie celulară normală HUMEC (normal human mammary epithelial cells) pentru control, s-au urmărit modificările la nivel transcriptomic, în special, cu privire la modularea proceselor de apoptoză, autofagie și proliferare celulară.

8.1.3. Analiza bioinformatică a datelor de microarray

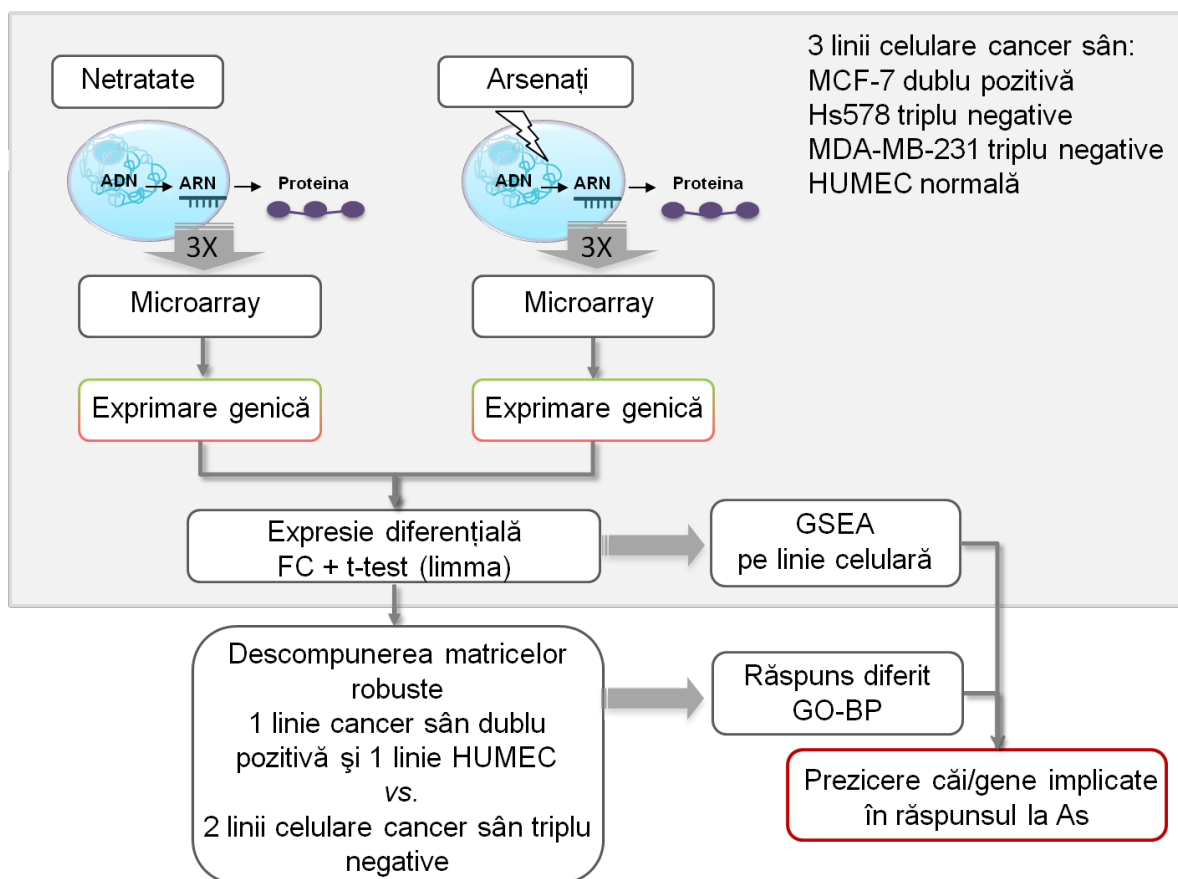


Figura 8.1. Algoritmul analizei la nivel transcriptomic utilizând datele de microarray

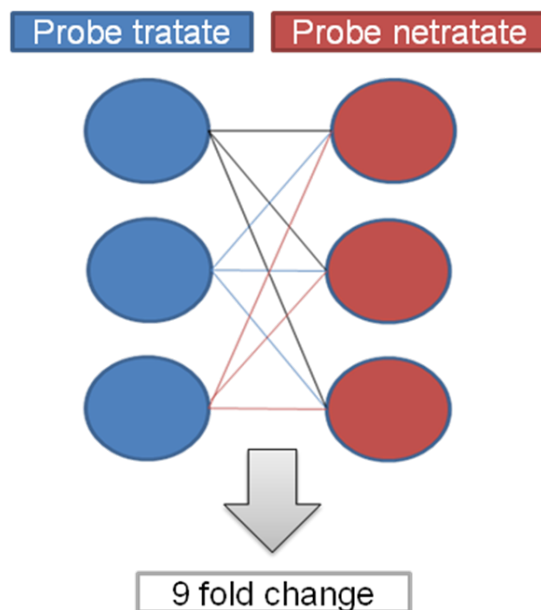


Figura 8.2. Reprezentare calcul fold change

Pentru fiecare din probe, tratate și netratate, a fost calculat nivelul de exprimare genică ($\log_2$ fold change), pentru fiecare linie celulară rezultând astfel o valoare de exprimare de 9 (fold change) pe fiecare linie celulară (Figura 8.2).

Tabel 8.1. Predicția în funcție de diferiți parametri în analiza RDA

Valoare Alpha	Valoare Delta									
	0	0.33	0.7	1	1.333	1.667	2	2.333	2.667	3
0	2	2	1	1	5	7	18	18	18	18
0.11	0	2	10	10	12	10	11	18	18	18
0.22	0	0	9	10	11	11	10	11	18	18
0.33	0	0	9	10	10	12	11	10	11	11
0.44	0	0	9	9	10	11	12	11	10	11
0.55	0	0	2	9	10	10	10	11	12	11
0.66	0	0	0	5	9	9	10	10	10	11
0.77	0	0	0	0	3	9	9	9	10	10
0.88	0	0	0	0	0	0	0	1	3	7
0.99	0	0	0	0	0	0	0	0	0	0

În tabelul 8.1 se pot regăsi și observa toate datele matricelor robuste calculate și rezultate, prin evidențierea acelor valori luate în considerare. Valoarea aleasă marcată în portocaliu, este dimensiunea aleasă în analiza RDA.

8.2. Rezultate și discuții

8.2.1. Evaluarea efectului arsenatilor în cultura de matrigel

Testul de scurtă durată *in vitro* pe cultura de matrigel a fost efectuat cu scopul de a urmări efectele arsenatilor la nivelul organizării celulare (Figura 8.3).

Se arată că în toate liniile celulare, arsenatii modulează procesele implicate în elongarea celulară și reduce interacțiunile intercelulare și a capacității de a forma rețele celulare. Acest efect este mult evidențiat în celulele de cancer de sân triplu negativ Hs578T și MDA-MB-231 (Figura 8.3).

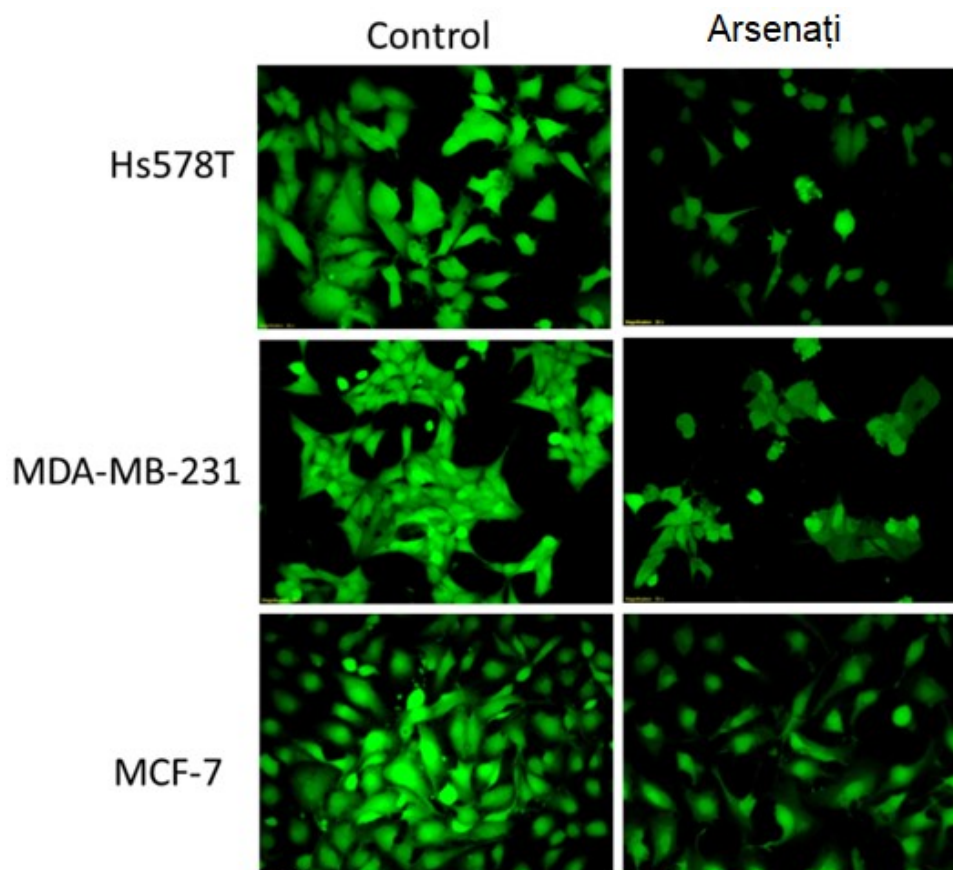


Figura 8.3. Testul de scurtă durată *in vitro* pe cultura de matrigel. Vizualizare microscopică celule netratate (Control) vs. celule tratate cu arsenatii.

8.2.2. Evaluarea efectului arsenaților asupra autofagiei și apoptozei

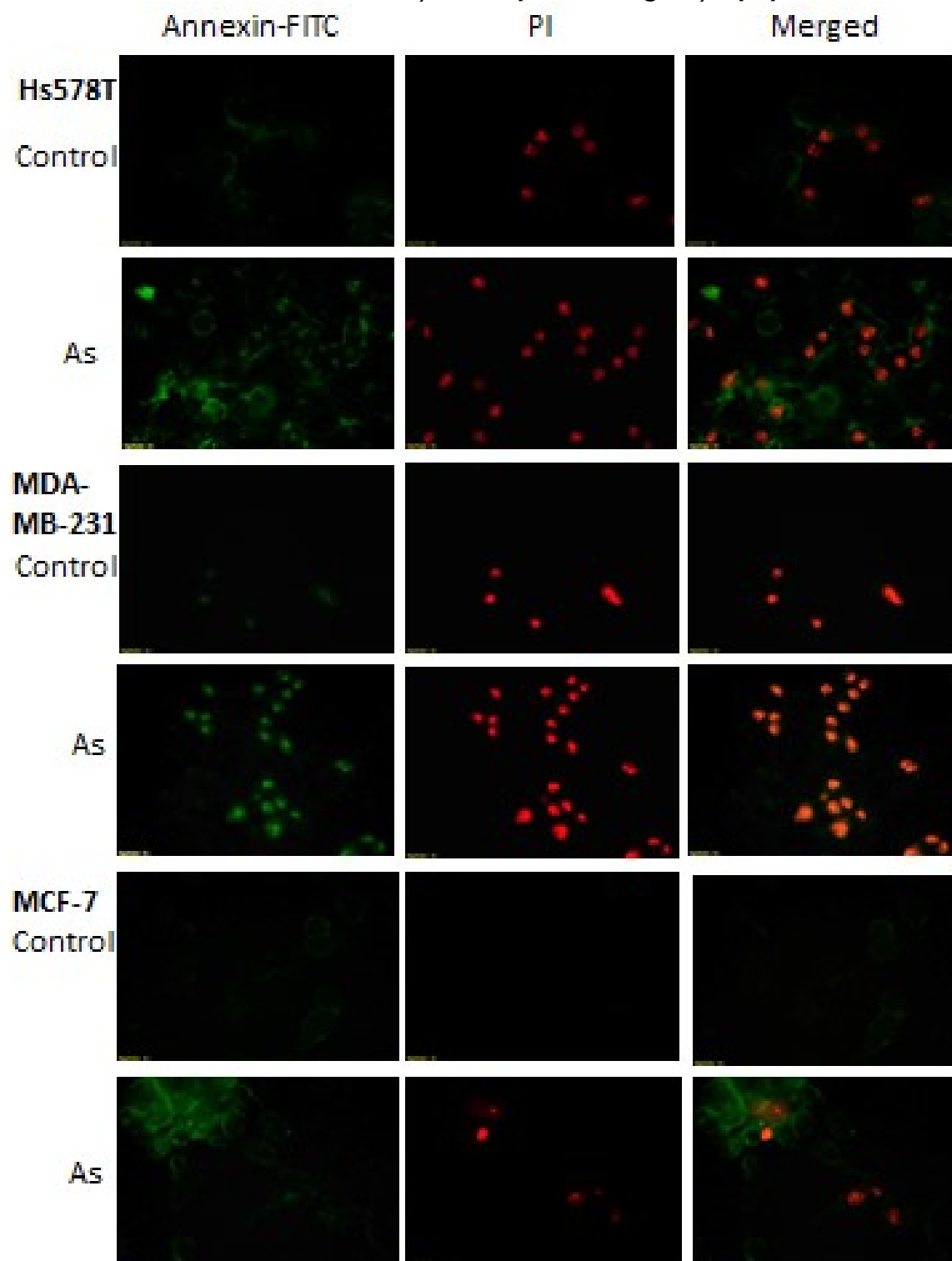


Figura 8.4.a. Evaluarea procesului de apoptoză prin microscopie de fluorescență (20X magnitudine) ca efect al expunerii celulelor la arsenați (As – notație arsenați în imagine) 50 nM.

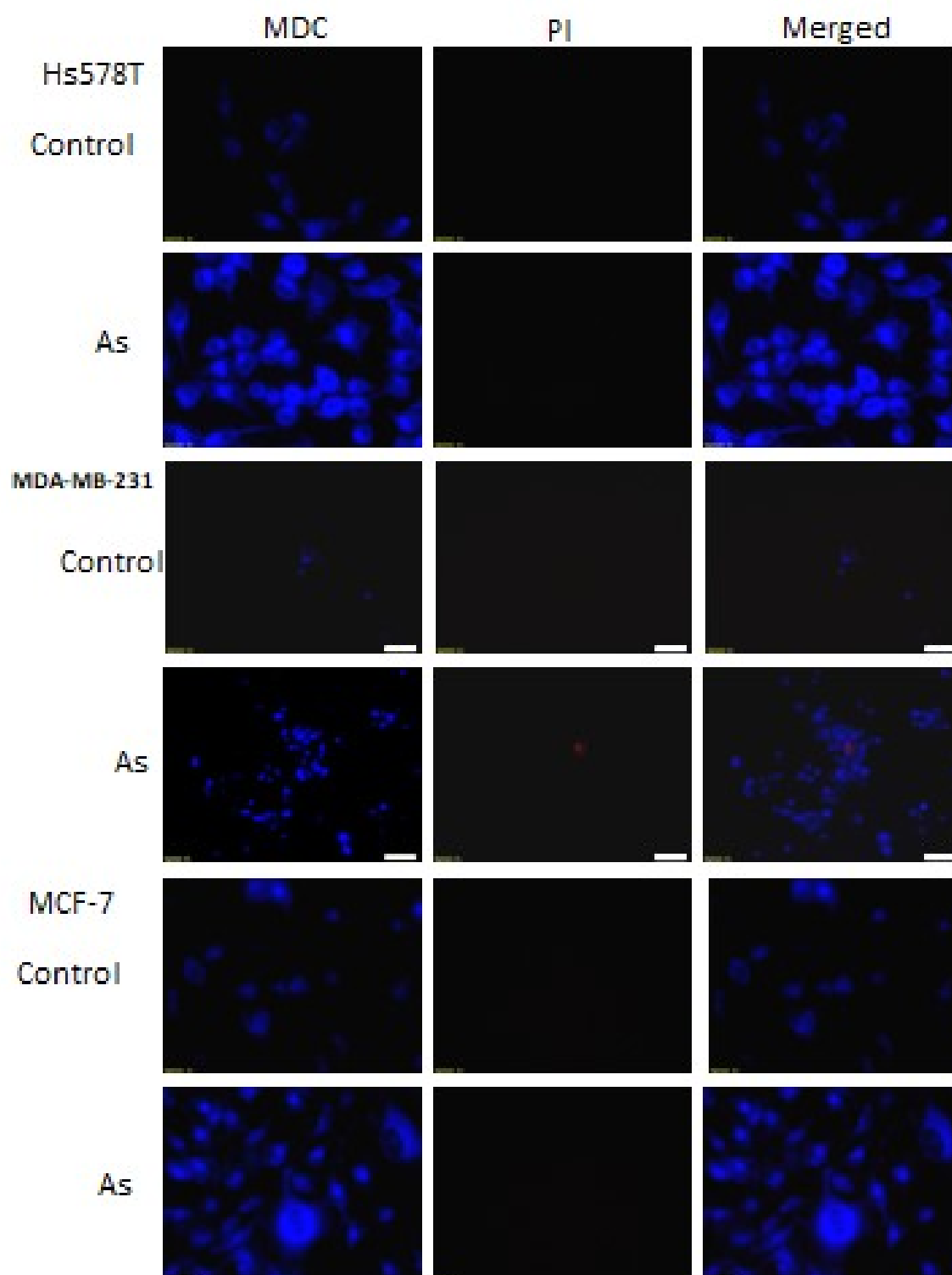


Figura 8.4.b. Evaluarea procesului de autofagie prin microscopie de fluorescență (20X magnitudine) ca efect al expunerii celulelor la arsenazi (As – notație arsenazi în imagine) 50 nM.

8.2.3. Vizualizarea datelor în funcție de răspunsul celulelor în urma tratamentului

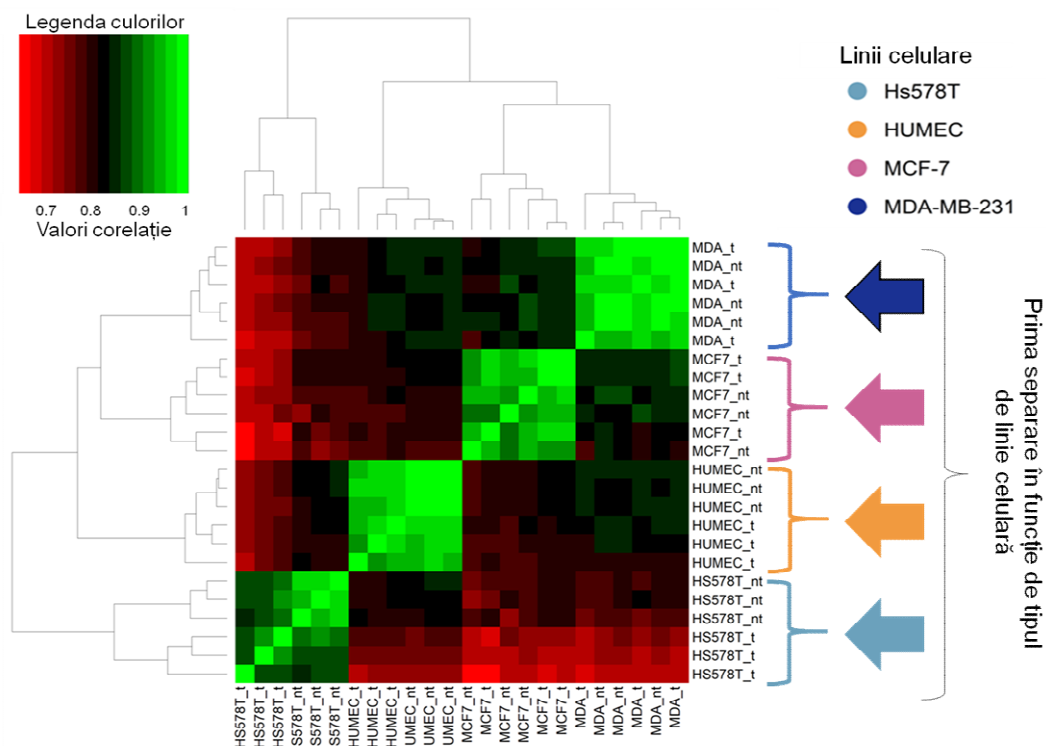


Figura 8.5. Arsenatii modulează răspunsuri diferite în cele patru linii celulare distincte

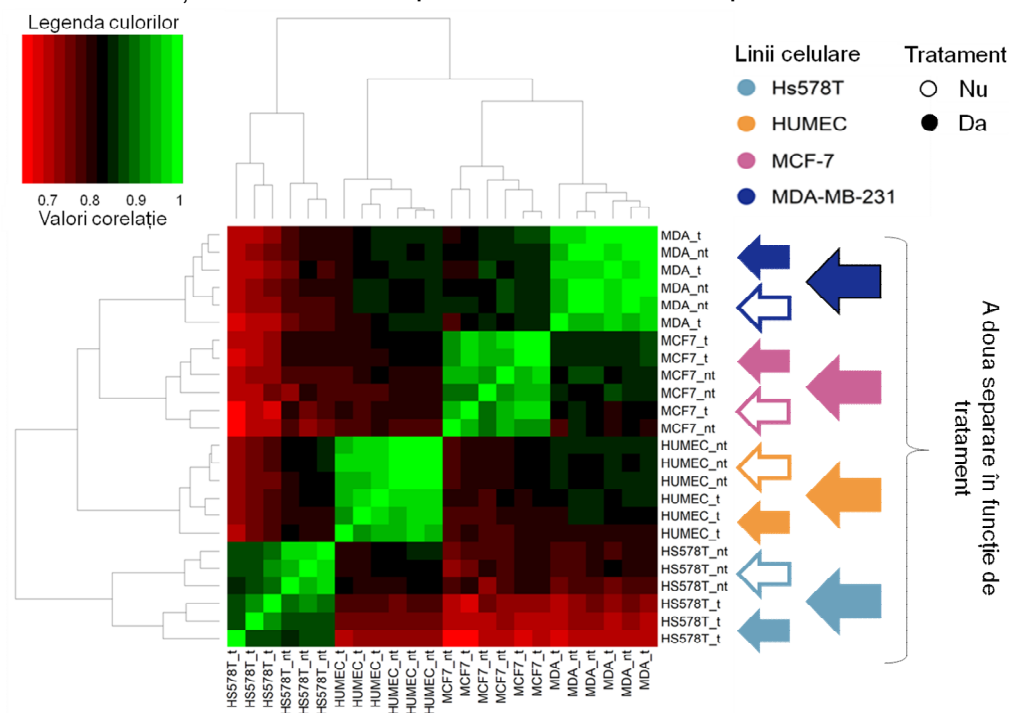


Figura 8.6. Răspunsuri diferite în cele două condiții (tratate-t, netratate-nt)

8.2.4. Analiza componentelor principale

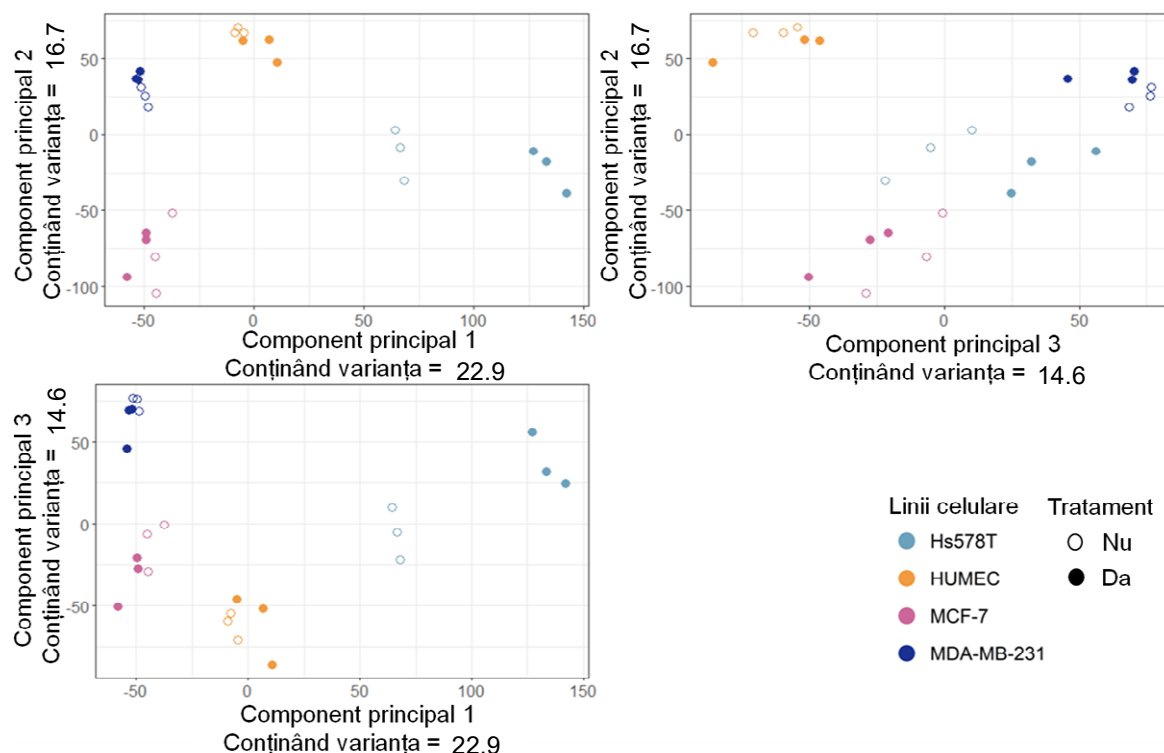


Figura 8.7. Fiecare linie celulară este evidențiată printr-o zonă distinctă în spațiul de exprimare genică, considerând cele trei componente principale conținând o varianță de 22.9% (CP 1), 16.7% (CP2) și 14.6% (CP3).

8.2.5. Analiza setului semnificativ de gene exprimate diferențial

Tabel 8.2. Procesele biologice comune exprimate, ca răspuns la tratamentul cu arsenafi în cele patru linii celulare, considerate semnificative statistic cu un FDR <0.15 (False discovery rate)

Denumirea procesului biologic	FDR valoarea q	
	Hs578T/ MDA-MB-231	MCF-7/ HUMEK
Replicarea ADN-ului dependentă de organizarea nucleozomului	0 0	0 0
Cromatina ce nu se exprimă la rADN	0 0	0 0
Heterotetramerizarea proteinei	0 0	0.00018 0
Cromatina ce nu se exprimă (tăcută)	0 0.00064	0 0

Denumirea procesului biologic	FDR valoarea q	
	Hs578T/ MDA-MB-231	MCF-7/ HUMEC
Reglarea negativă a diferențierii celulelor progenitoare	0 0.01886	0.00027 0
Reglarea negativă a diferențierii megacariocitelor	0 0.00284	0 0
Reglarea negativă a expresiei genice epigenetice	0 0.00083	0.00015 0
Reglarea pozitivă a expresiei genei epigenetic	0 0.04965	0.01009 0
Complex de asamblare al Beta cateninei TCF	0 0.00326	0.03176 0
Heterooligomerizarea proteinei	0 0.04438	0.00748 0
Organizarea telomerilor	0 0.04186	0.00747 0
Gene silențioase de ARN	0 0.01242	0.12874 0.00026
Reglarea genelor silențioase	0.00942 0.00100	0.00363 0.02088

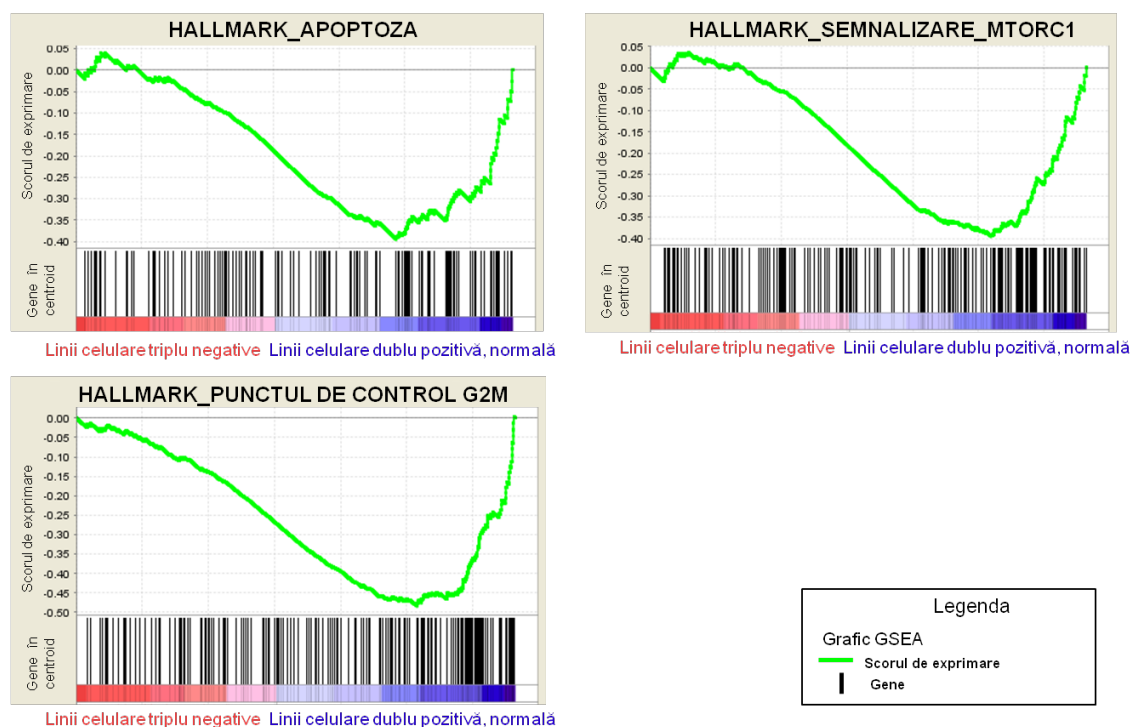


Figure 8.8. Genele sunt perturbate diferit în procesele biologice ce implică apoptoza, MTORC1, punctul de control G2M ($q < 0.15$). Pe axa X se pot observa valorile de centroid care reprezintă unde anume genele sunt cele mai perturbate.

Valorile centroidului de date reprezintă răspunsul comparativ între liniile celulare triplu negative (roșu) *versus* dublu negativă și normală (albastru) (Figure 8.8).

8.3. Concluzii

- Ca și efect global în urma tratamentului celulelor canceroase cu arsenaiți, s-a arătat că prin aceasta reglarea epigenetică a fost perturbată.
- Procesele de apoptoză și autofagie au fost afectate în liniile celulare de cancer de sân triplu negativ (Hs578T și MDA-MB-231).
- Căile biologice ce implică mecanismele de reparare al ADN-ului, organizarea cromatinei și reglarea epigenetică au fost modulate la nivel ridicat de către arsenaiți.
- Considerând doza de arsenaiți (50nM) la care celulele au fost supuse și efectele asupra procesele biologice menționate, se poate concluziona că arsenaiții pot fi atribuiți unui tratament alternativ anti-tumoral pentru acele tumori care implică perturbarea procesului de reparare a ADN-ului.
- Datele transcriptomice rezultate, oferă o perspectivă a evidențelor cu privire la efectul arsenaiților asupra proceselor și mecanismele complexe ce implică inhibarea sau activarea proceselor tumorale.

CAPITOLUL 9. EFECTUL TERAPEUTIC AL LACTOBACILILOR ASUPRA CELULELOR INTESTINALE EPITELIALE PORCINE LA NIVEL DE EXPRESIE GENICĂ

9.1. Materiale și metode

9.1.4. Extracția ARN-ului total pentru analiza microarray

ARN-ul total, atât cel provenit de la celulele IPEC-1 tratate cu lactobacili, cât și cel provenit de la celulele IPEC-1 netratate, a fost izolat utilizând kit-ul QiagenRNeasy midi (QIAGEN GmbH), după protocolul recomandat de furnizor (Pistol et al., 2014). Calitatea și integritatea probelor au fost verificate utilizând bio-analizatorul Agilent

2100 și nano kit-ul Agilent RNA 6000 (Agilent Technologies). Scorul RIN (RNA integrity number) a fost găsit în intervalul 8-10. Probele de ARN purificate au fost păstrate la -80°C până la utilizare.

9.1.6. Analiza statistică a datelor de microarray

Toate secvențele genice a speciei porcine *Sus scrofa* au fost extrapolate la omologii lor umani utilizând *Homology Based Annotation* din baza de date NCBI (www.ncbi.nlm.nih.gov) și BLAST (Braicu et al., 2016).

Analiza de tip *Ingenuity Pathway* (IPA; <http://www.ingenuity.com>) a fost aplicată.

9.1.7. Validarea datelor de exprimare genică

Validarea datelor de exprimare genică a fost efectuată prin analiza cantitativă Real-time PCR (RT-qPCR). Nivelurile de expresie a patru gene (IL-1 β , TLR6, TLR4, IL-10) (Tabel 9.1) selectate aleator au fost măsurate prin RT-qPCR, în toate probele considerate, conform protocolului prezentat în (Taranu et al., 2015).

Tabel 9.1. Validarea datelor de exprimare genică considerând nivelurile de exprimare a patru gene (IL-1 β , TLR6, TLR4, IL-10) cu media geometrică (Geomean) și Fold change (FC), după tratamentul cu lactobacili (LB).

Căi celulare	Gene	Geomean_Fold_LB	FC
Răspunsul inflamator	TLR6	1.03	2.04
	TLR4	1.46	2.75
Citokineza	Il-1b	1.45	2.73
	IL10	1.8	3.48

9.2. Rezultate și discuții

9.2.2. Clasificarea funcțională a genelor exprimate diferențial

Genele care au fost găsite ca fiind exprimate diferențial în urma tratamentului, au fost supuse în continuare unei analize de tip grupare (cluster) și clasificate în opt categorii funcționale și căi de semnalizare astfel: semnalizare, semnalizare celulară, proliferare, factor de transcripție, factori de creștere, citokine, interleukine, răspunsul inflamator (Tabelul 9.2).

Tabel 9.2. Genele supra-(up-) (roșu) și sub-(down-) (verde) exprimate implicate în cele opt categorii funcționale și căi de semnalizare

Denumire căi	Up regulated	Down regulated	Total găsite
Factor de transcripție	597	60	657
Semnalizare	1735	76	1811
Semnalizare celulară	132	5	137
Proliferare	632	60	692
Citokine	224	10	234
Interleukine	186	14	200
Răspunsul inflamator	205	13	218
Factor de creștere	398	13	411

76 de gene dintre cele 1811 au fost găsite ca fiind down-regulated, o supresie genică semnificativă fiind observată pentru genele RGS2 cu un nivel de expresie de -6.67 și p-value = 0.017 și OR1L8 cu un nivel de expresie de -5.26 și p-value = 0.259 (Tabelul 9.3).

Tabel 9.3. Lista genelor supra- (up-) și sub-exprimate (down-) implicate în semnalizarea celulară în celulele epiteliale porcine IPEC-1

ID Gene	Simbol gene	Descrierea genelor	FC	Expresia
Semnalizare				
GACC01000361	<i>nf1</i>	NF1 (neurofibromin 1), varianta de transcriere 1	10.20	Up
AK349266	<i>fuz</i>	FUZ (fuzzy homolog) (<i>Drosophila</i>), varianta de transcriere 1	10.63	Up
			10.70	Up
XM-001928433	<i>or2m3</i>	receptorul olfactiv, familia 2, subfamilia M, membru 3 (OR2M3)	11.63	Up
AK345382	<i>entpd1</i>	ENTPD1 1 (ectonucleoside triphosphate diphosphohydrolase), varianta de transcriere 1	12.13	Up
			13.45	Up
XM-003482962	<i>axin2</i>	AXIN 2 (axis inhibition protein)	13.93	Up
AB530146	<i>rgs2</i>	Reglator al semnalizării proteinei G 2, 24kDa	-6.67	down

ID Gene	Simbol gene	Descrierea genelor	FC	Expresia
XM-001925049	<i>or1l8</i>	Receptorul olfactiv, familia 1, subfamilia L, membru 8	-5.26 -4.17	down down
NM-001001861	<i>cxc12</i>	Chemochine (C-X-C motif) ligand 2	-4.00	down
NM-214376	<i>areg</i>	Amfiregulina	-3.85	down
NM-214376	<i>areg</i>	Amfiregulina	-3.85	down
AY609724	<i>tcf21</i>	Factor de transcripție 21	-3.70	down

Un număr semnificativ (692 gene) de gene exprimate diferențial implicate în proliferarea celulară au fost observate, dintre acestea 632 au fost găsite ca fiind supra-exprimate (dintre care un număr de 59 de gene au un nivel de expresie 4, ex. NF1 cu un 10.20 FC) iar 60 gene sub-exprimate precum AREG cu un nivel de expresie de -3.85 FC, IL1a cu un nivel de expresie de -3.33 FC (Tabelul 9.4).

Tabel 9.4. Lista genelor supra- (up-) și sub-exprimate (down-) implicate în proliferarea celulară în celulele epiteliale porcine IPEC-1

ID Gene	Simbol gene	Descrierea genelor	FC	Expresia
<u>Proliferare</u>				
GACC01000361	<i>nf1</i>	NF1 (neurofibromin 1), varianta de transcriere 1	10.20 10.70 13.45	Up
XM-003482962	<i>axin2</i>	AXIN 2 (axis inhibition protein)	13.93	Up
NM-214376	<i>areg</i>	Amfiregulina	-3.85 -3.45	down
NM-214029	<i>il1a</i>	Interleuchina 1, alpha (IL1A)	-3.33	down
AY610314	<i>ube2v2</i>	Ubiquitin-enzima de conjugare E2 varianta de transcriere 2	-3.33	down

Un grup de gene (657) implicate în factorii de transcripție au fost găsite ca fiind exprimate diferențial astfel, 597 sunt supra-exprimate (cu un nivel de expresie 12.21 și p-value = 0.058 a fost găsită gena TSHZ2, NF1 cu un nivel de expresie de 10.20 și p-value = 0.133, EMX1 cu un nivel de expresie de 9.00 și p-value = 0.133) iar 60 de gene sunt sub-exprimate (spre exemplu gena PKNOX2 cu un nivel de expresie de -8.33 și p-value = 0.015) (Tabelul 9.5).

Tabel 9.5. Lista genelor supra- (up-) și sub-exprimate (down-) implicate în factorii de transcripție în celulele epiteliale porcine IPEC-1

ID Gene	Simbol gene	Descrierea genelor	FC	Expresia
Factorii de transcripție				
XM-003125031	<i>emx1</i>	EMX1 (empty spiracles homeobox 1)	9.00	Up
GACC01000361	<i>nf1</i>	NF1 (neurofibromin 1), varianta de transcriere 1	10.20	Up
AK347929	<i>tshz2</i>	TSHZ2 (teashirt zinc finger homeobox 2), varianta de transcriere 1	12.21	Up
XM-003361490	<i>pknox2</i>	PBX/knotted 1 homeobox 2	-8.33	down
AY609724	<i>tcf21</i>	ref <i>Homo sapiens</i> factor de transcripție 21 (TCF21), varianta de transcriere 2	-3.70	down

Tabel 9.6. Lista genelor supra- (up-) și sub-exprimate (down-) implicate în răspunsul inflamator în celulele epiteliale porcine IPEC-1

ID Gene	Simbol gene	Descrierea genelor	FC	Expresia
Răspunsul inflamator				
XM-003131278	<i>prkca</i>	Protein kinaza C, alpha	5.13 5.58	Up
AK396677	<i>pla2g7</i>	Fosfolipaza A2, grup VII (platelet-activating factor acetylhydrolase, plasma)	5.82	Up
XM-001929161	<i>osm</i>	Oncostatina M	6.19	Up
XM-003130465	<i>il20</i>	Interleuchina 20	6.23	Up
AY669080	<i>bmp2</i>	Proteina morfogenetică osoasă 2	6.87	Up
AK232615	<i>serpina3</i>	Inhibitor peptidaza serpin, CLADE A (alpha-1 antiproteinase, antitrypsin), membru U 3	9.85	Up
AK345252	<i>cxcl2</i>	Chemochine (C-X-C motif) ligand 2	-6.25	down
XM-003129107	<i>cxcl2</i>	Chemochine (C-X-C motif) ligand 2	-6.25	down
XM-003126166	<i>cxcl2</i>	Chemochine (C-X-C motif) ligand 2	-6.25	down
NM-001001861	<i>cxcl2</i>	Chemochine (C-X-C motif) ligand 2	-4.00	down
AY577905	<i>cxcl2</i>	Chemochine (C-X-C motif) ligand 2	-3.45	down

ID Gene	Simbol gene	Descrierea genelor	FC	Expresia
NM_214029	<i>il1a</i>	Interleuchina 1, alpha	-3.33	down

9.2.3. Validarea prin Real-Time PCR

Rezultatele de microarray au fost validate prin analiza qRT-PCR pentru nivelurile de expresie a patru gene (IL-1 β , TLR6, TLR4, IL-10). Nivelurile de expresie a genelor selectate au arătat un grad apropiat de supra-exprimare atât în cazul analizei microarray (IL-1 β cu 2.73 FC, TLR-6 cu 2.04 FC, TLR-4 cu 1.46 FC și IL-10 cu 3.48 FC) cât și în cazul analizei qRT-PCR (IL-1 β cu 2.32 FC, TLR-6 cu 1.84 FC, TLR-4 cu 1.74 FC și IL-10 cu 3.77 FC) (Tabel 9.1). Din ambele cazuri se poate concluziona că s-a obținut o bună corelație între rezultate. Metoda având potențialul de a elimina variabilitatea care ar putea influența cantitativ nivelurile de expresie al genelor.

9.2.4. Analiza căilor celulare

Dintre căile canonice implicate amintim Wnt/ β -catenina și mecanismul molecular implicat în dezvoltarea cancerului (Tabel 9.7), funcțiile celulare, creșterea și proliferarea celulară, diviziunea celulară, moartea (apoptoza) și supraviețuirea celulară (Tabel 9.8 și 9.9).

Tabel 9.7. Căi canonice semnificativ exprimate, implicate asociate cu tratamentul cu *lactobacilli* în urma analizei IPA

Denumire	Tratament LB	
	<i>p-value</i>	Rația
Rolul macrofagelor, fibroblaștilor și celulelor endoteliale în artrita reumatoidă	9.33E-21	56/309 (0.181)
Mecanismul molecular al cancerului	4.29E-20	61/374 (0.163)
Wnt/ β -catenina	1.32E-18	39/169 (0.231)
Pluripotența celulelor stem embrionare umane	1.33E-18	36/143 (0.252)

Tabel 9.8. Gene semnificativ exprimate și funcții celulare implicate asociate cu tratamentul cu *lactobacilli* în urma analizei IPA

Denumire	Tratament LB	
	<i>p-value</i>	Molecule

Creșterea și proliferarea celulară	2.79E-20	-1.53E-124	598
Diviziunea celulară	2.79E-20	-1.13E-117	572
Exprimarea genică	1.83E-33	-1.33E-117	439
Mișcarea celulară	3.35E-20	-2.04E-95	392
Moartea și supraviețuirea celulară	4.95E-20	-5.24E-86	478

Tabel 9.9. Gene semnificativ exprimate, implicate în afecțiuni și funcții biologice asociate cu tratamentul cu lactobacilli în urma analizei IPA

Denumire	Tratament LB		
	<i>p-value</i>		Molecule
Cancer	2.41E-20	-1.11E-58	849
Leziuni și anomalii	2.41E-20	-1.11E-58	860
Afecțiuni gastro-intestinale	2.38E-20	-1.53E-42	761
Tulburări de dezvoltare	3.43E-21	-5.85E-42	282
Răspunsul inflamator	5.00E-33	-2.08E-43	362

Tabel 9.10. Număr de gene (Scor) semnificativ exprimate, implicate în rețele funcționale asociate cu tratamentul cu lactobacilli în urma analizei IPA

Rețele funcționale asociate	Scor
Semnalizarea și interacțiunea intercelulară, ciclul celular	52
Exprimarea genică, diviziunea celulară, dezvoltarea și funcționarea sistemului digestiv	41
Exprimarea genică, tulburări musculare și scheletice, dezvoltarea și funcționalitatea sistemului scheletic și muscular	37
Mișcarea celulară, dezvoltarea și funcționalitatea sistemului hematologic, traficul imunitar	37
Exprimarea genică, dezvoltarea și funcționalitatea sistemului hematologic, morfologia tisulară	37

9.4. Concluzii

- Rezultatele obținute în urma analizei transcriptomice, indică faptul că mixul de cele trei tulpini de lactobacilli (*L. rhamnosus*, *L. plantarum* și *L. paracasei*) de concentrație 1×10^8 CFU /mL modulează diferențial exprimarea genică, având un efect benefic asupra barierei epiteliale funcționale, asupra proliferării celulare, inflamației și răspunsului imun (citokine, chemokine) în celulele epiteliale intestinale IPEC-1.
- Efectul predominant produs de probioticele testate, a fost activarea genică, genele implicate în căile de semnalizare fiind cele mai afectate: 95% supra- (up-) exprimate. Multe dintre genele exprimate sunt implicate în căi celulare și funcții biologice importante. În comparație cu lista de gene control (13950), au fost găsite 12678 gene supra-exprimate și 1272 sub- (down-) exprimate implicate în efectul predominant.
- Majoritatea genelor modulate (1811) au fost asociate cu căile de semnalizare dintre care 121 gene up- exprimate cu un cut off de 2 și nivelul de expresie de referință (FC) mai mare decât 10. Mixul de lactobacilli a avut un efect semnificativ asupra căii de semnalizare Wnt/ β -catenin, pentru care gena AXIN2 a fost găsită ca supra- exprimată cu un fold change de 13.93, un regulator negativ al β -cateninei care joacă un rol important în tumorile canceroase umane.
- Rezultatele obținute în urma analizei de microarray evidențiază efectele mixului de lactobacilli asupra proliferării celulare și transcripției. Gena NF1 care codifică proteina neurofibrominei, un supresor tumoral care împiedică proliferarea necontrolată a celulelor a înregistrat un nivel de expresie mai mare de 10 FC.
- Inducerea genelor precum SERPINA 3, IL-20, OSM, GM-CSF, precum și suprimarea genelor CXCL-2 (MCP-1) și RGS2 și a citokinei pro-inflamatorii IL-18, evidențiază rolul protector al lactobacililor în funcția barierei epidermice împotriva inflamației și în activarea răspunsului imun.

CONCLUZII ȘI PERSPECTIVE VIITOARE

- S-a arătat că un set de compuși anorganici (Capitolul 5) tind să formeze diferit aranjamente în echilibru ca rezultat al interacțiunii cu moleculele de apă iar în absența altor ioni (la o diluție infinită) aranjamentele pot avea simetrie alterată. Totodată, în soluții concentrate cu 8 atomi de litiu și 4 atomi de potasiu, aranjamentele dodecaedrice formate s-au dovedit a fi simetrice și stabile, pe când dodecaedrele conținând 6 atomi de sodiu prezintă instabilitate și simetrie alterată.
- În privința similarității în clasele de compuși cu proprietăți antiinflamatoare (Capitolul 6), s-a arătat că o serie de compuși naturali din setul de date colectat, sunt similari cu structurile compușilor medicamentoși administrați în tratamentele bolilor inflamatorii. De asemenea, pe baza similarității s-a arătat că un compus natural și/sau derivații săi pot fi găsiți într-o diversitate de alte extracte de plante înrudite pe baza filogeniei.
- Urmărind în continuare comportamentul și aranjamentele moleculelor organice, reprezentate de setul de derivați steroidici (Capitolul 7), s-a constatat că în urma suprapunerii lor prin intermediul construcției hipermoleculi a fost posibilă estimarea unui model de relație structură-proprietate, considerând masa fragmentelor și proprietatea de lipofilicitate. Relația dintre caracteristicile structurale și proprietatea considerată a fost găsită ca fiind de o semnificație statistică ridicată, fapt arătat și prin validarea modelului prin procedura leave-one-out, cu un coeficient de determinare $Q^2 = 0.9413$ în analiza leave-one-out și $R^2 = 0.961$ în modelul estimat.
- În urma tratamentului celulelor canceroase cu arsenaiți (Capitolul 8), s-a arătat că reglarea epigenetică a fost perturbată iar procesele de apoptoză și autofagie au fost afectate în liniile celulare de cancer de sân triplu negativ (Hs578T și MDA-MB-231) ceea ce duce implicit la afirmarea faptului că arsenaiții acționează asupra proceselor și mecanismele complexe ce au ca efect inhibarea proceselor tumorale. De asemenea, în urma tratamentului, arsenaiții s-au dovedit ca fiind

implicați în căile biologice ce implică mecanismele de reparare al ADN-ului, organizarea cromatinei și reglarea epigenetică, modulându-le la un nivel ridicat.

- Mixul soluției de probiotice (Capitolul 9), aplicat ca și tratament asupra celulelor epiteliale intestinale porcine, a modulat diferențial exprimarea genică demonstrând efecte benefice asupra barierei epiteliale funcționale, asupra proliferării celulare, inflamației și răspunsului imun.
- Înțelegerea proprietăților structurale și caracterizarea comportamentului molecular (în apă și *in silico*), atât a compușilor anorganici cat și a celor organici, poate conduce la explicabilitatea activităților biologice ale acestora, *in vitro*. Privind prin prisma similarității, aceasta poate aduce o contribuție în alegerea și selecția compușilor pentru a fi testați *in vitro* în scop terapeutic, în funcție de afecțiunea pentru care se dorește tratamentul.

Referințe (Selecție)

- Albrecht, B., Grant, G.H., Richards, W.G., 2004. Evaluation of structural similarity based on reduced dimensionality representations of protein structure. *Protein engineering, design & selection* 17, 425-432. doi:10.1093/protein/gzh049
- Alexander-Dann, B., Pruteanu, L.L., Oerton, E., Sharma, N., Berindan-Neagoe, I., Módos, D., Bender, A., 2018. Understanding and predicting compound-induced toxicity from transcriptomic data. *Molecular Omics*. doi:10.1039/c8mo00042e
- Allen, B.C.P., Grant, G.H., Richards, W.G., 2001. Similarity calculations using two-dimensional molecular representations. *Journal of Chemical Information and Computer Sciences* 41, 330-337. doi:10.1021/ci0003956
- Anjum, A., Jaggi, S., Varghese, E., Lall, S., Bhowmik, A., Rai, A., 2016. Identification of differentially expressed genes in RNA-seq data of *Arabidopsis thaliana*: A compound distribution approach. *Journal of Computational Biology* 23, 239-247. doi:10.1089/cmb.2015.0205
- Auer, P.L., Doerge, R.W., 2010. Statistical design and analysis of RNA sequencing data. *Genetics* 185, 405-416. doi:10.1534/genetics.110.114983
- Bender, A., 2005. Studies on molecular similarity. PhD thesis.
- Bender, A., Jenkins, J.L., Scheiber, J., Sukuru, S.C.K., Glick, M., Davies, J.W., 2009. How similar are similarity searching methods? A principal component analysis of molecular descriptor space. *Journal of Chemical Information and Modeling* 49, 108-119. doi:10.1021/ci800249s
- Bero, S.A., Muda, A.K., Choo, Y.H., Muda, N.A., Pratama, S.F., 2017. Similarity measure for molecular structure: A brief review. *Journal of Physics: Conference Series* 892, 1-8. doi:10.1088/1742-6596/892/1/012015
- Biella, C.D. a, Salvador, M.J., Dias, D.A., Dias-Baruffi, M., Pereira-Crott, L.S., 2008. Evaluation of immunomodulatory and anti-inflammatory effects and phytochemical screening of *Alternanthera tenella Colla* (Amaranthaceae) aqueous extracts. *Memorias do Instituto Oswaldo Cruz* 103, 569-577. doi:10.1590/S0074-02762008000600010
- Bleakley, K., Yamanishi, Y., 2009. Supervised prediction of drug-target interactions using bipartite local models. *Bioinformatics* 25, 2397-2403. doi:10.1093/bioinformatics/btp433
- Bose, B., 2013. Systems biology: A biologist's viewpoint. *Progress in Biophysics and Molecular Biology* 113, 358-368. doi:10.1016/j.pbiomolbio.2013.07.001
- Bourdon-Lacombe, J.A., Moffat, I.D., Deveau, M., Husain, M., Auerbach, S., Krewski, D., Thomas, R.S., Bushel, P.R., Williams, A., Yauk, C.L., 2015. Technical guide for applications of gene expression profiling in human health risk assessment of environmental chemicals. *Regulatory Toxicology and Pharmacology* 72, 292-309. doi:10.1016/j.yrtph.2015.04.010
- Braicu, C., Cojocneanu-Petric, R., Jurj, A., Gulei, D., Taranu, I., Gras, A.M., Marin, D.E., Berindan-Neagoe, I., 2016. Microarray based gene expression analysis of *Sus Scrofa duodenum* exposed to zearalenone: Significance to human health. *BMC Genomics* 17, 1-10. doi:10.1186/s12864-016-2984-8
- Breitling, R., Armengaud, P., Amtmann, A., Herzyk, P., 2004. Rank products: A simple, yet powerful, new method to detect differentially regulated genes in

- replicated microarray experiments. *FEBS Letters* 573, 83-92. doi:10.1016/j.febslet.2004.07.055
- Burnham, C.J., Petersen, M.K., Day, T.J.F., Iyengar, S.S., Voth, G.A., 2006. The properties of ion-water clusters. II. Solvation structures of Na⁺, Cl⁻, and H⁺ clusters as a function of temperature. *The Journal of Chemical Physics* 124, 24327-1-24327-9. doi:10.1063/1.2149375
- ChemBioDraw 14 . 0 User Guide, 2016.
- Chen, X., Yan, C.C., Luo, C., Ji, W., Zhang, Y., 2015. Constructing lncRNA functional similarity network based on lncRNA-disease associations and disease semantic similarity. *Nature Publishing Group* 1-12. doi:10.1038/srep11338
- Clegg, S.L., Brimblecombe, P., 1989. Solubility of ammonia in pure aqueous and multicomponent solutions. *The Journal of Physical Chemistry* 93, 7237-7248. doi:10.1021/j100357a041
- Çöltekin, A., Fabrikant, S.I., Lacayo, M., 2010. Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings. *International Journal of Geographical Information Science* 24, 1559-1575. doi:10.1080/13658816.2010.511718
- Corney, D.C., Basturea, G.N., Jefferson, T., 2013. RNA-seq using next generation sequencing. *Materials and Methods* 3, 1-18. doi:dx.doi.org/10.13070/mm.en.3.203
- DeBlase, A.F., Wolke, C.T., Weddle, G.H., Archer, K.A., Jordan, K.D., Kelly, J.T., Tschumper, G.S., Hammer, N.I., Johnson, M.A., 2015. Water network-mediated, electron-induced proton transfer in [C₅H₅N·(H₂O)_n]- clusters. *Journal of Chemical Physics* 143, 1-8. doi:10.1063/1.4931928
- Di Palma, T.M., Bende, A., 2013. Vacuum ultraviolet photoionization and ab initio investigations of methyl tert-butyl ether (MTBE) clusters and MTBE-water clusters. *Chemical Physics Letters* 561-562, 18-23. doi:10.1016/j.cplett.2013.01.025
- Ding, H., Takigawa, I., Mamitsuka, H., Zhu, S., 2013. Similarity-based machine learning methods for predicting drug-target interactions: A brief review. *Briefings in Bioinformatics* 15, 734-747. doi:10.1093/bib/bbt056
- Diudea, M.V., Vizitiu, A.E., Janezic, D., 2007. Cluj and related polynomials applied in correlating studies. *Journal of Chemical Information and Modeling* 47, 864-874. doi:10.1021/ci600482j
- Diudea, M. V., Gutman, I., Jäntschi, L., 2001. *Molecular topology*, Nova Science Publishers. doi:10.1007/BF01151278
- Dodge, S., 2011. Exploring movement using similarity analysis.
- Dong, J., Cao, D.S., Miao, H.Y., Liu, S., Deng, B.C., Yun, Y.H., Wang, N.N., Lu, A.P., Zeng, W. Bin, Chen, A.F., 2015. ChemDes: An integrated web-based platform for molecular descriptor and fingerprint computation. *Journal of Cheminformatics* 7, 1-10. doi:10.1186/s13321-015-0109-z
- Doucet, J.-P., Weber, J., 1996. Molecular similarity. *Computer-Aided Molecular Design* 328-362. doi:10.2174/1568026043451186
- Egieyeh, S., Syce, J., Christoffels, A., Malan, S.F., 2016. Exploration of scaffolds from natural products with antiplasmodial activities, currently registered antimalarial drugs and public malarial screen data. *Molecules* 21, 1-14. doi:10.3390/molecules21010104

- Fournier, R., 2008. Geometric structure of the magic number clusters $\text{Li}_n\text{Na}_{8-n}$, $\text{Na}_n\text{K}_{8-n}$, and $\text{K}_n\text{Li}_{8-n}$. *Journal of Computational Methods in Science and Engineering* 8, 331-342.
- Galashev, A.E., 2013. A computer study of ammonium adsorption on water clusters. *Russian Journal of Physical Chemistry B* 7. doi:10.1134/S1990793113050047
- Grayson, N.E., Taormina, A., Twarock, R., 2009. DNA duplex cage structures with icosahedral symmetry. *Theoretical Computer Science* 410, 1440-1447. doi:10.1016/j.tcs.2008.12.005
- Guerra, D., David, J., Restrepo, A., 2014. Hydrogen bonding in the binary water/ammonia complex. *Journal of Computational Methods in Sciences and Engineering* 14, 93-102. doi:10.3233/JCM-140487
- Hardcastle, T.J., Kelly, K.A., 2010. baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics* 11, 422. doi:10.1186/1471-2105-11-422
- Harsa, A.M., Harsa, T.E., Bolboacă, S.D., Diudea, M. V, 2014. QSAR in flavonoids by similarity cluster prediction. *Current Computer-Aided Drug Design* 10, 115-128. doi:10.1007/978-94-007-6697-1_14
- Hettne, K.M., Boorsma, A., van Dartel, D.A., Goeman, J.J., de Jong, E., Piersma, A.H., Stierum, R.H., Kleinjans, J.C., Kors, J.A., 2013. Next-generation text-mining mediated generation of chemical response-specific gene sets for interpretation of gene expression data. *BMC medical genomics* 6, 2. doi:10.1186/1755-8794-6-2
- Hong, F., Breitling, R., McEntee, C.W., Wittner, B.S., Nemhauser, J.L., Chory, J., 2006. RankProd: A bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* 22, 2825-2827. doi:10.1093/bioinformatics/btl476
- Hvelplund, P., Kurtén, T., Støchkel, K., Ryding, M.J., Nielsen, S.B., Uggerud, E., 2010. Stability and structure of protonated clusters of ammonia and water, $\text{H}^+(\text{NH}_3)_m(\text{H}_2\text{O})_n$. *The Journal of Physical Chemistry A* 114, 7301-7310. doi:10.1021/jp104162k
- Janeiro-Barral, P.E., Mella, M., 2006. Study of the structure, energetics, and vibrational properties of small ammonia clusters $(\text{NH}_3)_n$ ($n = 2-5$) using correlated ab initio methods. *Journal of Physical Chemistry A* 110, 11244-11251. doi:10.1021/jp063252g
- Jäntschi, L., 2000. Predicția proprietăților fizico-chimice și biologice cu ajutorul descriptorilor matematici. Universitatea „Babeș-Bolyai” din Cluj-Napoca.
- Jäntschi, L., Pruteanu, L.L., Cozma, A.C., Bolboacă, S.D., 2015. Inside of the linear relation between dependent and independent variables. *Computational and Mathematical Methods in Medicine* 2015, 1-11. doi:10.1155/2015/360752
- Jiang, Y., Gerhold, D.L., Holder, D.J., Figueroa, D.J., Bailey, W.J., Guan, P., Skopek, T.R., Sistare, F.D., Sina, J.F., 2007. Diagnosis of drug-induced renal tubular toxicity using global gene expression profiles. *Journal of Translational Medicine* 5, 47. doi:10.1186/1479-5876-5-47
- Kadota, K., Nakai, Y., Shimizu, K., 2008. A weighted average difference method for detecting differentially expressed genes from microarray data. *Algorithms for Molecular Biology* 3, 1-12. doi:10.1186/1748-7188-3-8
- Kochev, N., Monev, V., Bangov, I., 2003. 6. Searching chemical structures. *Chemoinformatics: A Textbook* 291-318. doi:10.1021/c160012a009

- Kukurba, K.R., Montgomery, S.B., 2016. RNA sequencing and analysis. Cold Spring Harbor Protocol 2015, 951-969. doi:10.1101/pdb.top084970.RNA
- Laage, D., Hynes, J.T., 2007. Reorientational dynamics of water molecules in anionic hydration shells. *Proceedings of the National Academy of Sciences* 104, 11167-11172. doi:10.1073/pnas.0701699104
- Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.-P., Subramanian, A., Ross, K.N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S., Haggarty, S.J., Clemons, P., Wei, R., Carr, S., 2006. The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929-1935. doi:10.1126/science.1132939
- Legendre, A.M., 1805. Nouvelles méthodes pour la détermination des orbites des comètes [New methods for the determination of the orbits of comets] (in French). Paris: F. Didot.
- Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15, 550. doi:10.1186/s13059-014-0550-8
- Ma, C., Wang, L., Xie, X.Q., 2011. GPU accelerated chemical similarity calculation for compound library comparison. *Journal of Chemical Information and Modeling* 51, 1521-1527. doi:10.1021/ci1004948
- Milne, G.W.A., 2010. Computer software review of ChemBioDraw 12.0. *Journal of Chemical Information and Modeling*. doi:10.1021/ci100385n
- Minkin, V.I., 1999. Glossary of terms used in theoretical organic chemistry (IUPAC Recommendations 1999). *Pure and Applied Chemistry* 71, 1919-1981. doi:10.1351/pac199971101919
- Minkowski, H., 1953. *Geometrie der Zahlen*. Chelsea Pub. Co., New York.
- Mishra, P., Tripathi, V., Yadav, B.S., 2010. In silico QSAR modeling and drug development process. *GERF Bulletin of Biosciences* 1, 37-40.
- Nidhi, Glick, M., Davies, J.W., Jenkins, J.L., 2006. Prediction of biological targets for compounds using multiple-category bayesian models trained on chemogenomics databases. *Journal of Chemical Information and Modeling* 46, 1124-1133. doi:10.1021/ci060003g
- Orozco, A.A., Audouze, K., Brunaka, S., Taboureaux, O., 2016. In silico systems pharmacology to assess drug's therapeutic and toxic effects. *Bentham Science Publishers* 0, 1-10. doi:10.1174/1381611811666160907093115
- Pearson, K., 1895. On the mathematical theory of evolution. VII. Mathematical Contributions to the theory of evolution— III. Regression, heredity, and Panmixia. *Philosophical Transactions of the Royal Society of London* 253-318.
- Pistol, G.C., Gras, M.A., Marin, D.E., Israel-Roming, F., Stancu, M., Taranu, I., 2014. Natural feed contaminant zearalenone decreases the expressions of important pro-and anti-inflammatory mediators and mitogen-activated protein kinase/NF-kB signalling molecules in pigs. *British Journal of Nutrition* 111, 452-464. doi:10.1017/S0007114513002675
- Pruteanu, L., Ersali, S., Bolboacă, S., 2016. A QSPR model for steroids. *Studia Universitatis Babes-Bolyai Chimia* 61, 71-78.
- Pruteanu, L.L., Jäntschi, L., Ungureșan, M.L., Bolboacă, S.D., 2016. Models of monovalent ions dissolved in water. *Studia Universitatis Babes-Bolyai Chimia* 61, 151-162.

- Rarey, M., Dixon, J.S., 1998. Feature trees: a new molecular similarity measure based on tree matching. *Journal of computer-aided molecular design* 12, 471-490. doi:10.1023/A:1008068904628
- Reynolds, C.A., Burt, C., Graham Richards, W., 1992. A linear molecular similarity index. *Quantitative Structure-Activity Relationships* 11, 34-35. doi:10.1002/qsar.19920110106
- Rininger, J.A., DiPippo, V.A., Gould-Rothberg, B.E., 2000. Differential gene expression technologies for identifying surrogate markers of drug efficacy and toxicity. *Drug Discovery Today* 5, 560-568. doi:10.1016/S1359-6446(00)01597-X
- Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2009. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-140. doi:10.1093/bioinformatics/btp616
- Rogers, J., Tanimoto, T.T., 1960. A computer program for classifying plants. *Science* 132, 1115-1118. doi:10.1126/science.132.3434.1115
- RStudio, 2014. Markdown. Rstudio 4-5. doi:10.1002/wics.1348
- Russell, J.D., 2006. NIST Computational chemistry comparison and benchmark database. NIST standard reference database number 101.
- Ryding, M.J., Ruusuvuori, K., Andersson, P.U., Zatula, A.S., McGrath, M.J., Kurtén, T., Ortega, I.K., Vehkamäki, H., Uggerud, E., 2012. Structural rearrangements and magic numbers in reactions between pyridine-containing water clusters and ammonia. *Journal of Physical Chemistry A* 116, 4902-4908. doi:10.1021/jp3021326
- Sander, T., Freyss, J., Kor, M. Von, Rufener, C., 2015. DataWarrior: An open-source program for chemistry aware data visualization and analysis. *Journal of Chemical Information and Modeling* 55, 460-473. doi:10.1021/ci500588j
- Shi, W., Bugrim, A., Nikolsky, Y., Nikolskya, T., Brennan, R.J., 2008. Characteristics of genomic signatures derived using univariate methods and mechanistically anchored functional descriptors for predicting drug- and xenobiotic-induced nephrotoxicity. *Toxicology Mechanisms and Methods* 18, 267-276. doi:10.1080/15376510701857072
- Singh, J., Singh, S., Shaik, B., Deeb, O., Sohani, N., Agrawal, V.K., Khadikar, P. V., 2008. Mutagenicity of nitrated polycyclic aromatic hydrocarbons: A QSAR investigation. *Chemical Biology and Drug Design* 71, 230-243. doi:10.1111/j.1747-0285.2008.00629.x
- Sirci, F., Napolitano, F., Pisonero-Vaquero, S., Carrella, D., Medina, D.L., di Bernardo, D., 2017. Comparing structural and transcriptional drug networks reveals signatures of drug activity and toxicity in transcriptional responses. *npj Systems Biology and Applications* 3, 23. doi:10.1038/s41540-017-0022-3
- Stears, L., Martinskt, T., Schena, M., 2003. Trends in microarray analysis. *Nature Medicine* 9, 140-145.
- Taranu, I., Braicu, C., Marin, D.E., Pistol, G.C., Motiu, M., Balacescu, L., Beridan Neagoe, I., Burlacu, R., 2015. Exposure to zearalenone mycotoxin alters in vitro porcine intestinal epithelial cells by differential gene expression. *Toxicology Letters* 232, 310-325. doi:10.1016/j.toxlet.2014.10.022
- Taranu, I., Marin, D.E., Braicu, C., Pistol, G.C., Sorescu, I., Pruteanu, L., Beridan-Neagoe, I., Vodnar, D.C., 2018. In vitro transcriptome response to Probiotic Lactobacilli mixture in Intestinal Porcine Epithelial Cell line. *International Journal*

- of Molecular Sciences 19, 1-16.
- Tarca, A.L., Romero, R., Draghici, S., 2006. Analysis of microarray experiments of gene expression profiling. *American Journal of Obstetrics and Gynecology* 195, 373-388. doi:10.1016/j.ajog.2006.07.001
- Taylor, R.C., Acquah-Mensah, G., Singhal, M., Malhotra, D., Biswal, S., 2008. Network inference algorithms elucidate Nrf2 regulation of mouse lung oxidative stress. *PLoS Computational Biology* 4. doi:10.1371/journal.pcbi.1000166
- Terfloth, L., 2003. *Chemoinformatics: A Textbook*. 8. Calculation of structure descriptors. Wiley-VCH Verlag.
- Thomson, W., Kelvin, Lord, Nikolova, N., Jaworska, J., 2003. Approaches to measure chemical similarity - a Review. *QSAR & Combinatorial Science* 22, 1006-1026. doi:10.1002/qsar.200330831
- Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., Pachter, L., 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature biotechnology* 31, 46-53. doi:10.1038/nbt.2450
- Trapnell, C., Williams, B. a, Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., Pachter, L., 2011. Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nature Biotechnology* 28, 511-515. doi:10.1038/nbt.1621. Transcript
- Velkoborsky, J., Hoksza, D., 2016. Scaffold analysis of PubChem database as background for hierarchical scaffold - based visualization. *Journal of Cheminformatics* 8, 1-14. doi:10.1186/s13321-016-0186-7
- Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10, 57-63. doi:10.1038/nrg2484
- Willett, P., 2011. Similarity-based data mining in files of two-dimensional chemical structures using fingerprint measures of molecular resemblance. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1, 241-251. doi:10.1002/widm.26
- Willett, P., Barnard, J.M., Downs, G.M., 1998. Chemical similarity searching. *Journal of Chemical Information and Computer Sciences* 38, 983-996. doi:10.1021/ci9800211
- Yona, G., Dirks, W., Rahman, S., Lin, D.M., 2006. Effective similarity measures for expression profiles. *Bioinformatics* 22, 1616-1622. doi:10.1093/bioinformatics/btl127
- Yu, Z.D., Bawden, C.S., Henderson, H.V., Nixon, A.J., Gordon, S.W., Pearson, A.J., 2006. Microarrays as a discovery tool in wool genomics. *Proceedings of the New Zealand Society of Animal Production* 66, 129-133. doi:10.1079/BJN19660078
- Zhang, B., Horvath, S., 2005. A general framework for weighted gene Co-Expression network analysis. *Statistical applications in genetics and molecular biology* 4. doi:10.2202/1544-6115.1128

Lista publicațiilor

- Jäntschi, L., **Pruteanu**, L.L., Cozma, A.C., Bolboacă, S.D., 2015. Inside of the linear relation between dependent and independent variables. **Computational and Mathematical Methods in Medicine** 2015, 1-11. doi:10.1155/2015/360752.
- Pruteanu**, L.L., Jäntschi, L., Ungureșan, M.L., Bolboacă, S.D., 2016. Models of monovalent ions dissolved in water. **Studia Universitatis Babes-Bolyai Chemia** 61, 151-162.
- Pruteanu**, L.L., Ersali, S., Bolboacă, S.D., 2016. A QSPR model for steroids. **Studia Universitatis Babes-Bolyai Chemia** 61, 71-78.
- Ungureșan, M.L., **Pruteanu**, L.L., Jäntschi, L., 2016. Correlation study among boiling temperature and heat of vaporization. **Studia Universitatis Babes-Bolyai Chemia** 61, 223-234.
- Jäntschi, L., Bálint, D., **Pruteanu**, L.L., Bolboacă, S.D., 2016. Elemental factorial study on one-cage pentagonal face nanostructure congeners. **Materials Discovery** 5, 14-21. doi:10.1016/j.md.2016.12.001.
- Pruteanu**, L.L., Jäntschi, L., Crăciunaș, C., 2017. Molecular Level Investigation of Staphylococci's Resistance Mechanisms to Antibiotics. **Notulae Scientia Biologicae** 9, 307. doi:10.15835/nsb9310153.
- Alexander-Dann, B., **Pruteanu**, L.L., Oerton, E., Sharma, N., Berindan-Neagoe, I., Módos, D. and Bender, A., 2018. Understanding and predicting compound-induced toxicity from transcriptomic data. **Molecular Omics**. doi:10.1039/c8mo00042e - **in press**.
- Taranu, I., Marin, D.E., Braicu, C., Pistol, G.C., Sorescu, I., **Pruteanu**, L., Berindan-Neagoe, I., Vodnar, D.C., 2018. *In vitro* transcriptome response to Probiotic *Lactobacilli* mixture in Intestinal Porcine Epithelial Cell line. **International Journal of Molecular Sciences**, 19. - **acceptat**.