

## Anexa 4. Familii de descriptori moleculari

### 4.1 FPIF - Fragmental Property Index Family

FPIF [1] este o metodă matriceală bazată pe matrice pătratice de proprietăți derivate din structură. Astfel folosește: topologia moleculară ( $d_A(b,c)$  - distanța topologică în structura  $A$  de la atomul  $b$  la atomul  $c$ );  $\delta_A(b,c)$  - deturul - i.e. cea mai lungă cale - topologică în structura  $A$  de la atomul  $b$  la atomul  $c$ ;  $W_A(b,c)$  - mulțimea drumurilor în structura  $A$  de la atomul  $b$  la atomul  $c$ ;  $P_A(b,c)$  - mulțimea căilor în structura  $A$  de la atomul  $b$  la atomul  $c$ ;  $D_A(b,c)$  - mulțimea căilor distanță în structura  $A$  de la atomul  $b$  la atomul  $c$ ;  $\Delta_A(b,c)$  - mulțimea căilor detur în structura  $A$  de la atomul  $b$  la atomul  $c$ ;  $A \setminus p$  - structura ce rezultă din înlăturarea atomilor conținuți în calea  $p$  și legăturilor pe care aceștia le formează în structura  $A$  exceptând capetele căii) pentru a obține pentru fiecare pereche de atomi un set de atomi vecini utilizând în acest scop un set de 6 criterii ( $F_C$  - criteriu fragmentare):  $Sz$ :  $SzDi - Sz_{i,j} = \{x \mid x \text{ Atom}\}$  unde  $d_{\text{Moleculă}}(x,i) < d_{\text{Moleculă}}(x,j)$ ;  $SzDe - Sz_{i,j} = \{x \mid x \text{ Atom}\}$  unde  $\delta_{\text{Moleculă}}(x,i) < \delta_{\text{Moleculă}}(x,j)$ ;  $Cj$ :  $Cj_{i,j,p} = \{x \mid x \text{ Atom}\}$  unde  $d_{\text{Moleculă}}(x,i) < d_{\text{Moleculă}}(x,j)$  și  $\exists w \in W_{\text{Moleculă}}(x,i) \mid w \cap p = \{i\}$ ;  $p \in P_{\text{Moleculă}}(i,j)$ ;  $CjDi - p \in D_{\text{Moleculă}}(i,j)$ ;  $CjDe - p \in \Delta_{\text{Moleculă}}(i,j)$ ;  $Cf$ :  $Cf_{i,j,p} = \{x \mid x \text{ Atom}\}$  unde  $G_p = \text{Moleculă} \setminus p$ ,  $p \in P_{\text{Moleculă}}(i,j)$  și  $d_{G_p}(x,i) < d_{G_p}(x,j)$ ;  $CfDi - p \in D_{\text{Moleculă}}(i,j)$ ;  $CfDe - p \in \Delta_{\text{Moleculă}}(i,j)$ ; patru proprietăți atomice ( $A_P$  - proprietate atomică):  $M$  - masa atomică relativă;  $E$  - electronegativitatea;  $C$  - cardinalitatea (numărul de atomi);  $Q$  - sarcina electrică parțială; opt descriptori de proprietate ( $P_D$  - descriptor de proprietate):  $p$  - proprietatea;  $d$  - distanța;  $1/p$ ;  $1/d$ ;  $pd$ ;  $p/d$ ;  $p/d^2$ ;  $p^2/d^2$ ; cinci modele de suprapunere ( $S_M$  - model de suprapunere):  $S$ : sumă;  $P$ : produs;  $A$ : medie aritmetică;  $G$ : medie geometrică;  $H$ : medie armonică; două modele de interacțiune ( $I_M$  - model de interacțiune):  $R$  - rar (se presupune că proprietatea tuturor atomilor fragmentului este concentrată în centrul de proprietate, a cărui poziție se obține și se folosește în obținerea descriptorului de fragment folosind descriptorul de interacțiune);  $D$  - dens (se obține efectul contribuției fiecărui atom al fragmentului după care se realizează suprapunerea vectorială a acestora); două metrici de distanță ( $D_M$  - metrica de distanță):  $T$  - topologică;  $G$  - geometrică; patru tipuri de indici pe matrice pătratice ( $M_I$  - indici pe matrice):  $IP1$  - semi-suma elementelor din matrice;  $IP2$  - semi-suma pătratelor elementelor din matrice;  $IE1$  - semi-suma elementelor din matricea ce rezultă din înmulțirea matricei de proprietate cu matricea de adiacență;  $IE2$  - semi-suma pătratelor elementelor din matricea ce rezultă din înmulțirea matricei de proprietate cu matricea de

---

[1] Lorentz JĂNTSCHI. 2000. Predicția proprietăților fizico-chimice și biologice cu ajutorul descriptorilor matematici. Teză de doctorat în domeniul chimie specializarea chimie organică - coordonator Prof. Dr. Mircea V. Diudea, Universitatea "Babeș-Bolyai" din Cluj-Napoca, Facultatea de Chimie și Inginerie Chimică, Ianuarie 2000.

adiacență; operator de linearizare ( $L_O$ ): I - funcția identitate  $f(x)=x$ ; R - reciproca  $f(x)=1/x$ ; L - logaritmul  $f(x)=\ln(x)$ ;

Astfel, familia FPIF este constituită dintr-un număr de membrii egal cu înmulțirea tuturor posibilităților de alegere de mai sus ( $2 \cdot 2 \cdot 4 \cdot 8 \cdot 6 \cdot 5 \cdot 4 \cdot 3 = 46080$ ), astfel:

- ÷  $FPIF = I_M \times D_M \times A_P \times P_D \times F_C \times S_M \times M_I \times L_O$
- ÷  $I_M = \{R, D\}$ ;
- ÷  $D_M = \{T, G\}$ ;
- ÷  $A_P = \{M, E, C, Q\}$ ;
- ÷  $P_D = \{p, d, 1/p, 1/d, pd, p/d, p/d^2, p^2/d^2\}$ ;
- ÷  $F_C = \{SzDi, SzDe, CjDi, CjDe, CfDi, CfDe\}$ ;
- ÷  $S_M = \{S, P, A, G, H\}$ ;
- ÷  $M_I = \{IP1, IP2, IE1, IE2\}$ ;
- ÷  $L_O = \{I, R, L\}$ .

#### 4.2 MDF - Molecular Descriptors Family

MDF [2,3] este o metodă bazată pe fragmente moleculare obținute din perechi de vârfuri. În mod similar cu metoda FPIF, metoda MDF folosește doi operatori de distanță ( $D_O$ ): topologică (t) și geometrică (g), șase proprietăți atomice ( $A_P$ ): cardinalitate (C), număr de atomi de hidrogen direct legați (H), masa atomică relativă (M), electronegativitatea (E), electronegativitatea de grup (G) și sarcina atomică parțială (Q), douăzeci și patru de descriptori de interacțiune ( $I_D$ ):  $D(d)$ ,  $d(1/d)$ ,  $O(p_1)$ ,  $o(1/p_1)$ ,  $P(p_1p_2)$ ,  $p(1/p_1p_2)$ ,  $Q(\sqrt{p_1p_2})$ ,  $q(1/\sqrt{p_1p_2})$ ,  $J(p_1d)$ ,  $j(1/p_1d)$ ,  $K(p_1p_2d)$ ,  $k(1/p_1p_2d)$ ,  $L(d\sqrt{p_1p_2})$ ,  $l(1/d\sqrt{p_1p_2})$ ,  $V(p_1/d)$ ,  $E(p_1/d_2)$ ,  $W(p_1^2/d)$ ,  $w(p_1p_2/d)$ ,  $F(p_1^2/d^2)$ ,  $f(p_1p_2/d^2)$ ,  $S(p_1^2/d^3)$ ,  $s(p_1p_2/d^3)$ ,  $T(p_1^2/d^4)$ ,  $t(p_1p_2/d^4)$ , șase modalități de interacțiune ( $I_M$ ): `R` și `r` - modele rare, `M` și `m` - modele medii, și `D` și `d` - modele dense - fiecare dintre ele relativ la primul atom al fragmentului și respectiv la atomul referință exterior fragmentului, patru metode de fragmentare ( $F_C$ ): `m` - minimale, `M` - maximale, `D` - Szeged, `P` - Cluj pe căi, nouăsprezece modalități de suprapunere globală a interacțiunii fragmentelor ( $S_F$  - formula de suprapunere): grupul de mărimi ( `m` - selectează cea mai mică valoare; `M` - cea mai mare valoare; `n` - cea mai mică valoare absolută; `N` - cea mai mare valoare absolută); grupul de medii ( `S` - suma; `A` - media aritmetică după numărul de proprietăți de fragmente; `a` - media aritmetică după numărul de fragmente; `B` - media aritmetică după numărul de atomi; `b` - media aritmetică după numărul de legături);

[2] Lorentz JÄNTSCHI. 2004. MDF - A New QSAR/QSPR Molecular Descriptors Family. *Leonardo Journal of Sciences* 3(4):68-85.

[3] Lorentz JÄNTSCHI. 2005. Molecular Descriptors Family on Structure Activity Relationships 1. Review of the Methodology. *Leonardo Electronic Journal of Practices and Technologies* 4(6):76-98.

grupul geometric ( $\hat{P}$  - multiplicare;  $\hat{G}$  - medie geometrică după numărul de proprietăți de fragmente;  $\hat{g}$  - medie geometrică după numărul de fragmente;  $\hat{F}$  - medie geometrică după numărul de atomi;  $\hat{f}$  - media geometrică după numărul de legături); grupul armonic ( $\hat{s}$  - suma armonică;  $\hat{H}$  - medie armonică după numărul de proprietăți de fragmente;  $\hat{h}$  - medie armonică după numărul de fragmente;  $\hat{T}$  - medie armonică după numărul de atomi;  $\hat{i}$  - media armonică după numărul de legături), și șase operatori de linearizare ( $L_O$ ):  $\hat{I}$  - identitate ( $f(x)=x$ ),  $\hat{i}$  - inversa ( $f(x)=1/x$ ),  $\hat{A}$  - valoare absolută ( $f(x)=|x|$ ),  $\hat{a}$  - inversul valorii absolute ( $f(x)=1/|x|$ ),  $\hat{L}$  - logaritm ( $f(x)=\ln(x)$ ), și  $\hat{l}$  - logaritmul valorii absolute ( $f(x)=\ln(|x|)$ ).

Astfel, familia MDF este constituită dintr-un număr de membrii egal cu înmulțirea tuturor posibilităților de alegere de mai sus ( $2 \cdot 6 \cdot 6 \cdot 24 \cdot 4 \cdot 19 \cdot 6 = 787968$ ), astfel:

$$\div \text{MDF} = D_M \times A_P \times I_D \times I_M \times F_C \times S_M \times L_O$$

$$\div D_M = \{t, g\};$$

$$\div A_P = \{C, H, M, E, G, Q\};$$

$$\div P_D = \{d, 1/d, p_1, 1/p_1, p_1 p_2, 1/p_1 p_2, \sqrt{p_1 p_2}, 1/\sqrt{p_1 p_2}, p_1 d, 1/p_1 d, p_1 p_2 d, 1/p_1 p_2 d, d \sqrt{p_1 p_2}, 1/d \sqrt{p_1 p_2}, p_1/d, p_1/d_2, p_1^2/d, p_1 p_2/d, p_1^2/d^2, p_1 p_2/d^2, p_1^2/d^3, p_1 p_2/d^3, p_1^2/d^4, p_1 p_2/d^4\};$$

$$\div I_M = \{r, R, m, M, d, D\}$$

$$\div F_C = \{m, M, D, P\};$$

$$\div S_F = \{m, M, n, N, S, A, a, B, b, P, G, g, F, f, s, H, h, I, i\};$$

$$\div L_O = \{I, i, A, a, L, l\}.$$

### 4.3 MDFV - Molecular Descriptors Family Vertex

MDFV [4,5] este o metodă bazată pe vârfuri în locul perechilor de vârfuri, în care sunt implementate două valori pentru operatorul de distanță ( $D_O$ ), șapte proprietăți atomice ( $A_P$ ), cincizeci și opt de descriptori de interacțiune ( $I_D$ ), șapte metode de suprapunere la nivel de fragment ( $S_F$ ) și la nivel de moleculă ( $S_M$ ), zece tipuri de interacțiune ( $I_T$ ), două unități de exprimare ( $E_U$  - unitate de descriptor molecular  $\hat{D}$  și unitate de distanță (coordonată centru) descriptor molecular  $\hat{d}$ ), și trei operatori de linearizare ( $L_O$ : I - identitate, R - reciproc, L - logaritm).

Astfel, familia MDF este constituită dintr-un număr de membrii egal cu înmulțirea tuturor posibilităților de alegere de mai sus ( $2 \cdot 7 \cdot 58 \cdot 7 \cdot 7 \cdot 10 \cdot 2 \cdot 3 = 2387280$ ), astfel:

$$\div \text{MDFV} = D_O \times A_P \times I_D \times S_F \times S_M \times I_T \times E_U \times L_O;$$

[4] Lorentz JĂNTSCHI (PI), Sorana D BOLBOACĂ (CoI). 2009. De la Chimia Matematică la Chimia Cuantică și Chimia Medicală: ID1051 - Lucrare în extenso 2009.

[http://ori.academicdirect.org/research/grants/ID1051/PCE\\_ID\\_1051\\_Extenso\\_2009.pdf](http://ori.academicdirect.org/research/grants/ID1051/PCE_ID_1051_Extenso_2009.pdf)

[5] Sorana D BOLBOACĂ, Lorentz JĂNTSCHI. 2009. Comparison of QSAR Performances on Carboquinone Derivatives, *TheScientificWorldJOURNAL* 9(10):1148-1166. <http://dx.doi.org/10.1100/tsw.2009.131>

- ÷  $D_O = \{T, G\}$ ;
- ÷  $A_P = \{C, H, M, E, Q, L, A\}$ ;
- ÷  $I_D = \{J, j, O, o, P, p, Q, q, R, r, K, k, L, l, M, m, N, n, W, w, X, x, Y, y, Z, z, S, s, T, t, U, u, V, v, F, f, G, g, H, h, I, i, A, a, B, b, C, c, D, d, 0, 1, 2, 3, 4, 5, 6, 7\}$ ;
- ÷  $S_F = \{A, a, I, i, F, P, C\}$ ;
- ÷  $S_M = \{A, a, I, i, F, P, C\}$ ;
- ÷  $I_T = \{f, F, c, C, p, P, a, A, i, I\}$ ;
- ÷  $E_U = \{D, d\}$ ;
- ÷  $L_O = \{I, R, L\}$

#### **4.4 SAPF - Structural Atomic Property Family**

SAPF [6] este o metodă ce cumulează proprietăți atomice la nivel molecular. Se folosește o localizare a centrului molecular folosind o modalitate (din trei posibile) în acest sens ( $C_F$ ), folosind o metrică (din două posibile) în acest sens ( $M_D$ ), o proprietate atomică (din mai multe posibile, 5 definite în programul de calcul) în acest sens ( $A_P$ ), o putere a distanței în expresia descriptorului de efect atomic (cu 7 valori posibile) în acest sens ( $D_P$ ), o putere a proprietății atomice în expresia descriptorului de efect atomic (cu aceleași 7 valori posibile) în acest sens ( $P_P$ ), o modalitate de cumulare a proprietății atomice la nivel molecular (cu două valori posibile, sumă sau medie la numărul de atomi) în acest sens ( $O_M$ ) a cărei expresie se construiește dintr-o expresie de cumulare (cu 7 valori posibile) în acest sens ( $G_M$ ) iar rezultatul obținut este supus unei operații de linearizare (cu 7 valori posibile) în acest sens ( $L_O$ ).

Astfel, familia SAPF este constituită dintr-un număr de membrii egal cu înmulțirea tuturor posibilităților de alegere de mai sus ( $6 \cdot 6 \cdot 2 \cdot 7 \cdot 7 \cdot 5 \cdot 2 \cdot 3 = 144060$ ), astfel:

- ÷  $MDFV = C_F \times M_D \times A_P \times D_P \times P_P \times O_M \times G_M \times L_O$ ;
- ÷  $C_F = \{D, P, C\}$ ;
- ÷  $M_D = \{T, G\}$ ;
- ÷  $A_P = \{C, H, M, E, A\}$ ;
- ÷  $D_P = \{-2, -1, -0.5, 0, 0.5, 1, 2\}$ ;
- ÷  $P_P = \{-2, -1, -0.5, 0, 0.5, 1, 2\}$ ;
- ÷  $O_M = \{S, M\}$ ;
- ÷  $G_M = \{I, E, H, G, A, Q, S\}$ ;
- ÷  $L_O = \{I, A, S, T, Q, R, L\}$ ;

---

[6] SAPF: ©2010-2011, Lorentz JÄNTSCHI & Sorana D BOLBOACĂ; nedocumentată