Electronic supplementary information for

Chemical applications of neural networks: Aromaticity of pyrimidine derivatives

Mercedes Alonso,*[a] Carlos Miranda,[a] Nazario Martín[b] and Bernardo Herradón*[a]

Structures of the studied compounds.

Figure S1 depicts the structures of benzene (**1**), pyrimidine (**2**) and substituted pyrimidines studied in the article.
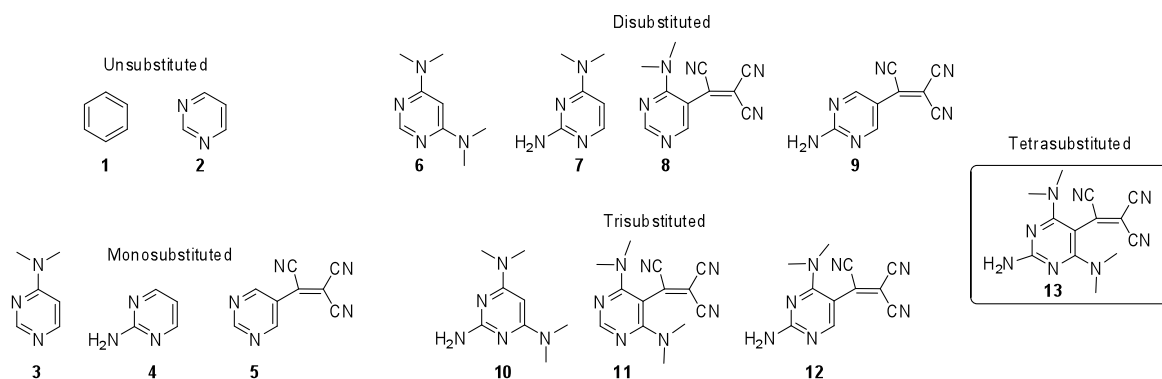

Figure S1

Structural Analysis of the atropoisomers in tricycnovinyl derivatives.

We have found that the tricyanovinyl group causes a high distortion in the pyrimidine ring. The potential conformers were identified by potential surface energy scans and further optimized at the B3LYP/6-311+G** theory level as indicated in the main article. We have identified three conformers in the dihydropyridine derivative used as refence (Figure S2). To quantify the geometrical similarities between the conformers, we carried out root mean square deviation (RMSD) analysis. Table S-1 collects the data.

[a] *Instituto de Química Orgánica General, C.S.I.C., Juan de la Cierva 3, 28006 Madrid, Spain. E-mail: mercuea@iqog.csic.es, herradon@iqog.csic.es; Fax: 34 915644853; Tel: 34 915618806.*
[b] *Departamento de Química Orgánica, Facultad de Química, Universidad Complutense, 28040 Madrid.*
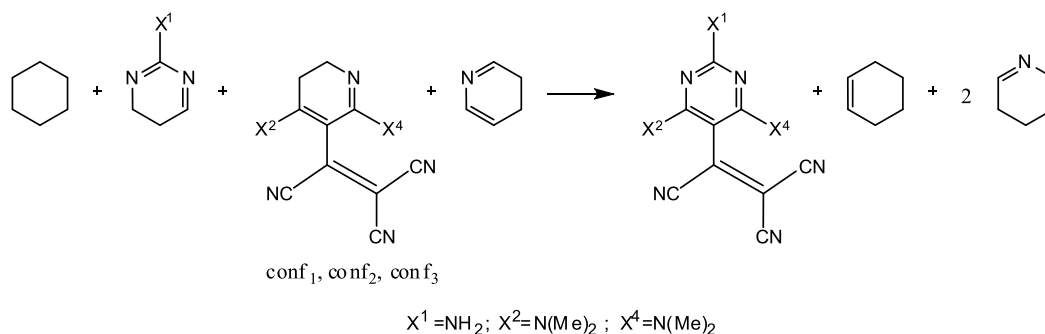
$X^1 = NH_2$; $X^2 = N(Me)_2$; $X^4 = N(Me)_2$

Figure S2

The values indicated in Table S1 are the following: average of the dihedral angles between adjacent atoms in the pyrimidine ring ($|CCCC|_{prom}$); RMSD of the ring atoms compared to the benzene ring ($RMSD_{bz}$) and to the unsubstituted pyrimidine ($RMSD_{Pyr}$), respectively; RMSD between the reference conformers ($conf_1$, $conf_2$, $conf_3$) and the corresponding pyrimidine, considering all the heavy atoms in the molecules ($RMSD_{global}$) and just the heavy atoms that all molecules have in common, excluding the $-(CH_2)_2-$ fragment ($RMSD_{common}$).

Table S1

Compound	$|CCCC|_{prom}$	$RMSD_{Bz}$	$RMSD_{Pyr}$	$RMSD_{common}$			$RMSD_{global}$
				$conf_1$	$conf_2$	$conf_3$	
1	0,0	0,000	0,056	-	-	-	-
2	0,0	0,056	0,000	-	-	-	-
3	0,6	0,057	0,016	-	-	-	-
4	0,3	0,054	0,007	-	-	-	-
5a	1,5	0,057	0,012	0,081	0,695	0,915	0,169
5b	1,5	0,057	0,012	0,083	0,660	**0,978**	0,159
6	1,1	0,056	0,015	-	-	-	-
7	0,8	0,055	0,015	-	-	-	-
8a	8,3	0,082	0,060	**0,142**	0,815	1,226	0,871
8b	4,6	0,066	0,038	0,130	0,992	**1,187**	0,863
9a	1,8	0,057	0,018	0,193	0,557	**1,062**	0,247
9b	1,8	0,057	0,018	0,142	0,600	1,003	0,197
10	1,2	0,054	0,013	-	-	-	-
11a	19,3	0,140	0,127	**0,703**	0,951	1,206	1,497
11b	4,5	0,063	0,036	0,378	0,848	**1,133**	0,796
12a	8,7	0,082	0,063	**0,111**	0,844	1,226	1,085
12b	4,1	0,063	0,035	0,106	0,980	**1,196**	1,083
13a	19,7	0,146	0,133	**0,701**	0,962	1,211	2,332
13b	3,9	0,060	0,034	0,318	0,810	**1,177**	1,937

Comments on the chosen reference reaction and the aromaticity descriptor derived

As indicated in the main article, the choice of the reference reaction to quantify ASE and Λ has been given a big importance in the literature. At the outset of our work, we chose equation 2 (R2); but after the advice of Professor Schleyer, we changed to the equation 1 (R1), which have been discussed in the main article. Although both reaction are homodesmotic, an inconvenience pointed out by Professor Scheler in R2 is the presence of the 3,4-dihydropyrimidine and the 3,4,5,6-tetrahydropyrimidine as reference compounds since the amidinio functionality is not present in the final product (the pyrimidine) (Figure S-3).

We consider that the SOM methodology is an excellent methodology to evaluate the influence of different schemes for quantifying ASE and Λ in the overall descriptor of aromaticity (namely the Euclidean distance between neurons). The results on the determination of ASE and Λ as well as the Euclidean distances (d_j) are presented herein (Table S2).

We have found that the influence of the reference reaction scheme is minimized on using the neural network (NN) methodology. Although a clear explanation on this fact cannot be given with the limited results obtained (we will analyze this issue in further applications), we can anticipate that is a consequence of the nature of the NN. The NN is able to work out a response from different class of input data, apprehending the multidimensional character of a phenomenon. Additionally, the method works mostly finding similarity patterns instead of absolute values of a property.

Figure S3 shows the two homodesmotic reactions used. We are aware that the ASE values can be perturbed by additional effects, such us strain, heteroatom interactions, interactions, hybridization, hyperconjugation, etc. Both reaction schemes are homodesmotic since the number of bonds between given atoms in each state of hybridization is equal in both products and reactants. In addition the number of hydrogen atoms joined to the atoms in given states of hybridization match. All reference compounds are six-membered rings computed in their most stable state. The energies were corrected by B3LYP/6-311+G(d,p) zero-point energies.
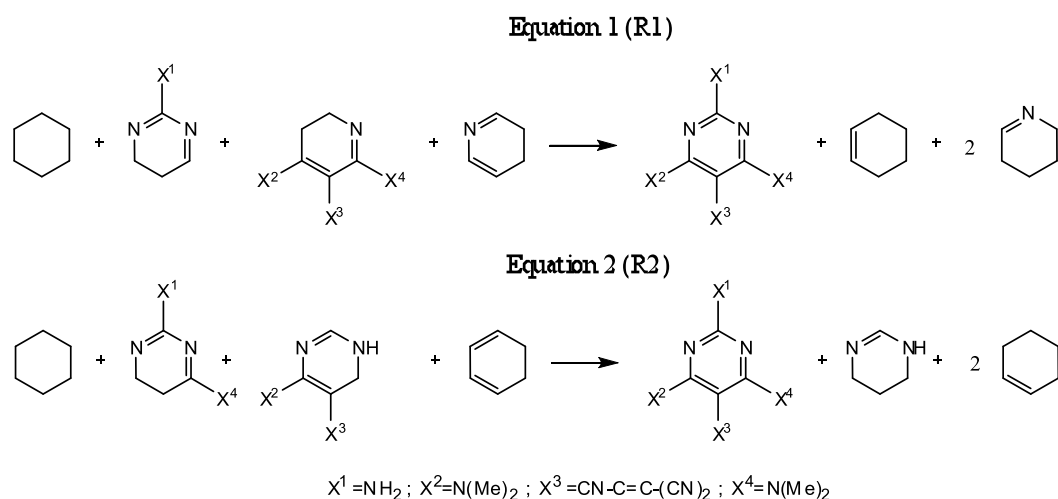
Equation 1 (R1)

Equation 2 (R2)

$X^1 = NH_2$; $X^2 = N(Me)_2$; $X^3 = CN-C=C-(CN)_2$; $X^4 = N(Me)_2$

Figure S-3. Homodesmotic reactions used to evaluate the aromatic stabilization energies (ASE) and magnetic susceptibility exaltation (Λ) of substituted pyrimidines **1-13**. Equation 1 (R1) is discussed in the main article.

Table S-2. Calculated ASE (kcal mol^{-1}), Λ (ppm cgs), NICS, NICS(1) and NICS$_{zz}$(1) (ppm), HOMA and d_j for compounds **1-13**.[a]

Compound	ASE$_{en}$ (eq 1)	ASE$_{en}$ (eq 2)	Λ_{en} (eq 1)	Λ_{en} (eq 2)	d_j (eq 1)	d_j (eq 2)
1	32.69	32.69	-17.28	-17.28	0.0	0.0
2	28.24	27.39	-11.63	-7.34	8.9	10.9
3	33.17	22.66	-8.20	-3.00	15.3	19.5
4	29.07	28.21	-7.40	-3.08	14.3	13.1
5a	24.06	18.26	-16.13	-13.17	13.0	19.7
5b	24.06	18.46	-16.22	-14.53	13.0	19.7
6	36.22	27.16	-6.51	-9.34	18.0	22.1
7	33.03	23.77	-5.04	-0.15	18.1	23.8
8a	25.14	12.57	-13.41	-9.04	15.8	30.3
8b	24.95	12.58	-10.74	-10.31	15.9	30.3
9a	29.89	24.29	-11.72	-10.01	16.3	20.6
9b	29.9	24.10	-11.67	-8.74	16.3	20.6
10	35.34	27.54	-4.18	-7.33	23.7	24.4
11a	25.67	4.49	-5.23	-5.33	27.6	37.5
11b	23.98	2.80	-3.79	-1.32	23.8	38.6
12a	29.17	18.05	-9.85	-8.46	21.4	22.1
12b	28.84	17.52	-7.33	-5.94	24.4	23.8
13a	29.59	9.66	-2.92	-3.34	30.8	37.5
13b	25.53	5.60	-1.53	0.63	26.5	40.8
Mean	28.66	18.06	-8.53	-6.94	19.1	24.0
Esd.	3.87	8.38	4.40	4.47	5.9	10.3

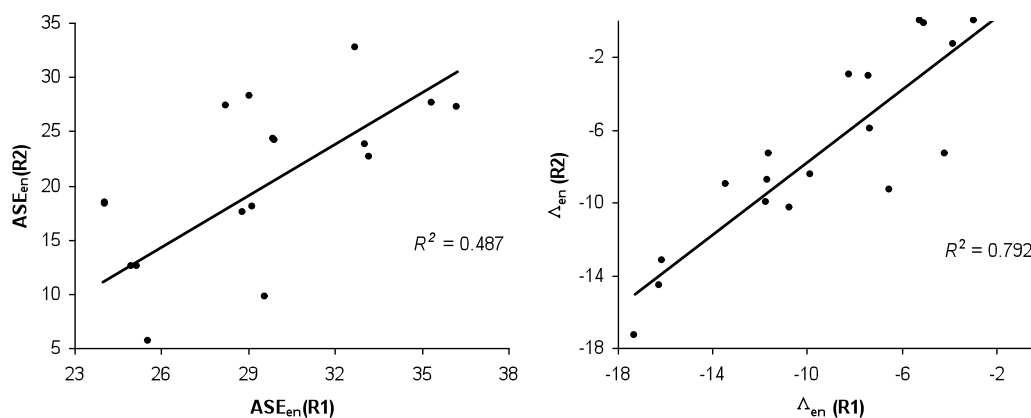[a] Values calculated with the most stable conformers.

Figure S4. Dependences between aromatic stabilization energies and exaltation values calculated using equations 1 and 2 for substituted pyrimidines **1-13**.

We have compared the ASE given by the two different homodesmotic models (Table S2). Figure S4 shows considerable scatter in the ASE(R1) vs. ASE(R2) plot. The correlation coefficient R is only 0.698. This correlation corresponds to what is considered to be a significant linear dependence with a probability > 99%. However, according to the determination coefficient R^2 (0.487), the common variance do not exceed 49% for this set of compounds and, therefore, they are not statistically equivalent.

A closer inspection of data in Table S2 reveals many subtle deviations between the data compared. According to ASE(R1), the electron-accepting tricyanovynil group reduces the aromaticity of the pyrimidine ring, while the electron-donating NMe_2 and NH_2 substituents apparently increase it. Consequently, compounds **3**, **4**, **6**, **7**, **9, 10, 12** and **13** are more aromatic than pyrimidine. Contrarily to ASE(1), the structural and magnetic indices point out that the aromaticity of pyrimidine (**2**) decreases upon substitution irrespective of the electron-donating or electron-accepting character of the substituent. ASE(2) describe a similar trend: as substitution increases, pyrimidine ring is energetically destabilized. ASE(2) are significant smaller than ASE(1) and the highest difference is found for pyrimidines **11a** and **11b** rising to 21 kcal mol^{-1}.

Despite the high differences between ASE values derived from equations 1 and 2, both provide a similar order of aromaticity for substituted pyrimidines **3-13** depending on the kind and position of the substituent (Figure S5). According to ASE(1) and ASE(2), the

least aromatic substituted pyrimidine is **11b**. Both reaction schemes point out that the non-planar conformers are less aromatic than the planar ones.

The exaltation values calculated from the homodesmotic equations 1 and 2 are more similar (Figure S4). The variation of $\Lambda(1)$ and $\Lambda(2)$, described by the standard deviation, is 4.1 in both cases. The correlations with these indices are much better, with $R^2 = 0.8$. The highest differences are found for pyrimidines bearing a NMe$_2$ group.
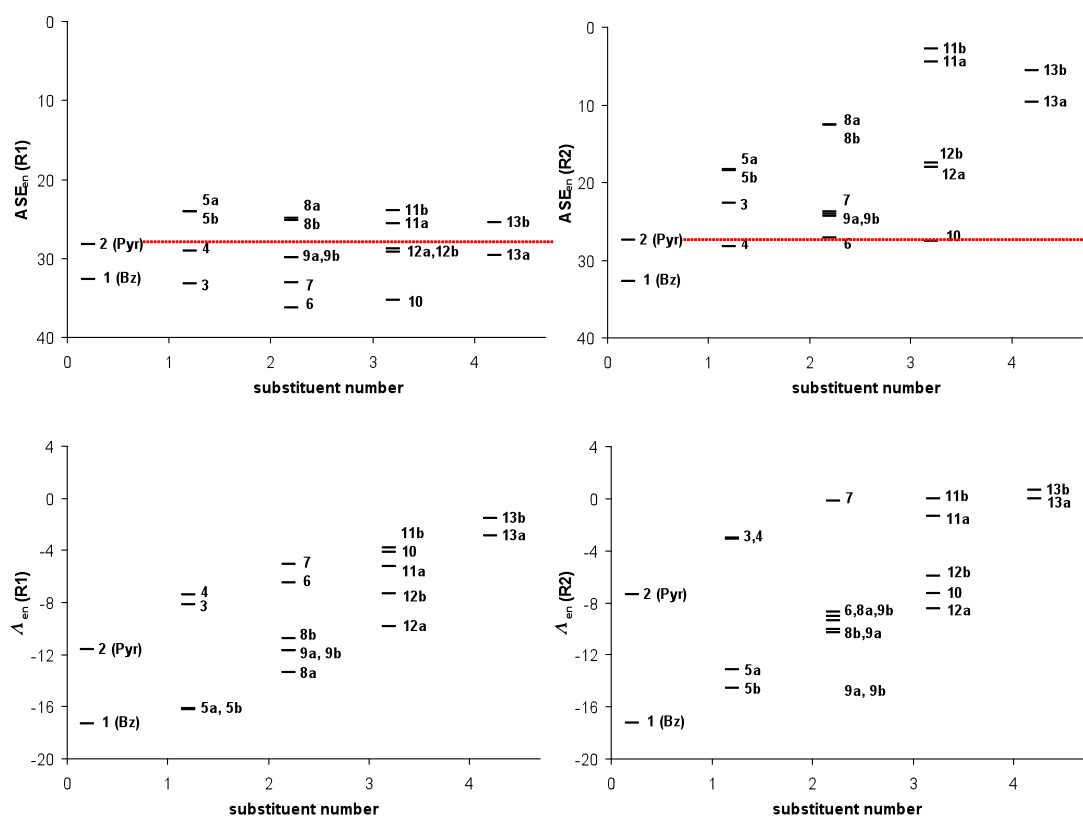

Figure S5. Variation of ASE(R1), ASE(R2), Λ(R1) and Λ(R2) with the number and position of substituents for the substituted pyrimidines **1-13**.

Table S3. Correlation coefficients between several descriptors of aromaticity for substituted pyrimidines **1-13**.

	ASE(1)	Λ(1)	ASE(2)	Λ(2)	NICS$_{zz}$(1)	HOMA	d_j(1)	d_j(2)
ASE(1)	1							
Λ(1)	0,207	1						
ASE(2)	**0,698**[a]	-0,420	1					
Λ(2)	-0,420	-0,309	**0,468**[b]	1				
NICS$_{zz}$(1)	0,066	**0,799**[a]	**-0,551**[b]	**-0,494**[b]	1			
HOMA	0,247	**-0,489**[b]	0,705	**0,853**	**-0,746**[a]	1		
d_j(1)	-0,049	**0,860**[a]	**-0,659**[a]	**-0,500**[b]	**0,967**[a]	**-0,746**[a]	1	
d_j(2)								1

[a] Correlation is significant at the 0.01 level (2-tailed). [b] Correlation is significant at the 0.05 level (2-tailed).
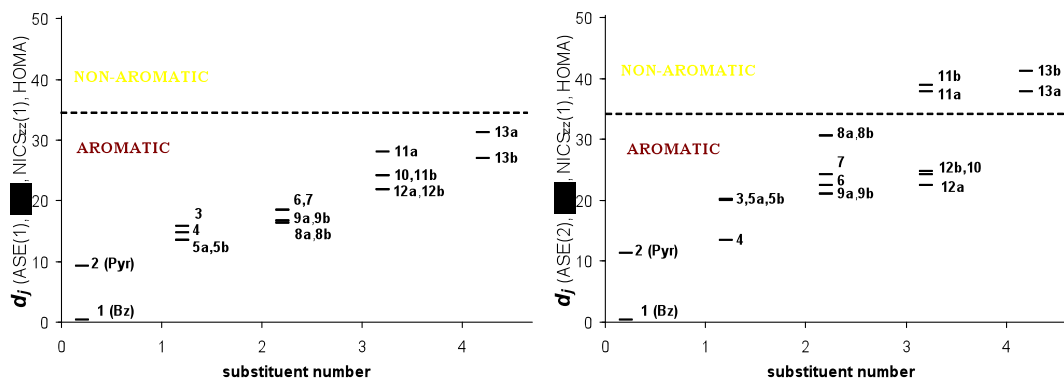

Figure S6. Variation of d_j with the number and position of substituents for the substituted pyrimidines **1-13**, using different values for the ASE and Λ calculated from homodesmotic reactions 1 and 2.

The high complexity and non-linear character of the relathionship between aromaticity and the different criteria makes neural networks the best tool to generate a model to quantify aromaticity of a comprehensive set of compounds. Using ASE, Λ, NICS$_{zz}$(1) and HOMA as molecular descriptors, we have applied the self-organazing both to classify compounds **1-13** according to their aromatic character. Interestingly, the aromaticity of pyrimidine ring decreases upon substitution irrespective of using ASE(1)/Λ(1) or ASE(2)/Λ(2) as molecular descriptors. Although the position of compounds **2-13** in the bidimensional map changes using different values for the

aromatic stabilization energy and the exaltation calculated from homodesmotic reactions, the differences in d_j are smaller than for ASE. Using ASE(2)/Λ(2), the derivatives bearing the tricyanovinyl group and two adjacent dimethylamino groups **11** and **13** are classified as non-aromatic compounds, close to the neurons activated by phosphole and arsole; and employing ASE(1)/Λ(1), they are placed in the aromatic region but near the non-aromatic zone. In general, d_j(eq 2) values are slightly higher than d_j(eq 1), but the relative ordering of aromaticity is essentially the same.

Overall, these results reinforce the need of using more than a single index to analyze aromaticity since the energetic, structural and magnetic descriptors are not statistical equivalent. This fact is indicative of the multidimensional character of aromaticity, justifying the necessity to use appropriate algorithms, such as neural networks, which takes into account the different physical manifestations of the phenomenon and it is expressed as a single value, the Euclidean distance d_j.