



*Dedicated to Professor Zeno Simon
on the occasion of his 80th anniversary*

BIOLOGICAL PROMISCUITY OF HTS HITS EVALUATED VIA CHEMICAL REACTIVITY DESCRIPTORS

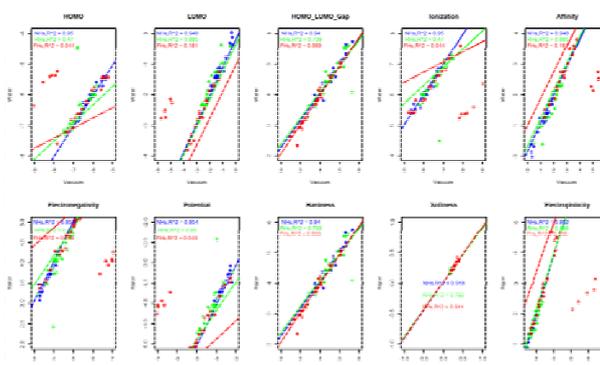
Ramona CURPĂN,^{a,*} Sorin AVRAM,^a Liliana HALIP^a and Cristian BOLOGA^b

^aTimișoara Institute of Chemistry of the Roumanian Academy, 24 Mihai Viteazul Av., 300223 Timișoara, Roumania

^bDepartment of Internal Medicine, University of New Mexico Health Sciences Center, Albuquerque NM 87131, NM, USA

Received August 25, 2014

Understanding the relationship between the biological behavior of high throughput screening hits and their chemical profile is an important task in the drug discovery process. This work presents a methodology for deciphering the biological promiscuity of high throughput screening hits (frequent hitters) via their chemical behavior. To achieve this goal, a series of global reactivity descriptors have been calculated for three sets of compounds: promiscuous, selective, and inactive, respectively compiled from Pubchem database.¹ Water based DFT calculations at B3LYP/6-31+G(d) and B3LYP/6-311++G(2df,2pd) levels have been performed and reactivity indexes have been computed within this formalism. These data have been compared with the previously reported results calculated in vacuum² to assess the environment which better describes the compounds. The results show that water and vacuum based calculations follow similar patterns with LUMO-based descriptors clearly discriminating between the promiscuous or frequent hitters and the other sets, selective or rare hitters and inactive, respectively.



INTRODUCTION

High throughput screening (HTS) has become an important source of new chemical entities in the drug discovery process and in the HTS era, post 1990, successful drugs have emerged from this technology.³⁻⁵ Nonetheless, a large number of the structures identified as hits in HTS show a promiscuous biological behavior being annotated as active in multiple assays and against different

unrelated targets. These structures represent a major source of undesirable false hits in HTS campaigns and their correct annotation early in the process of drug discovery is highly wanted. For this purpose, numerous filters have been developed to flag and remove undesirable moieties⁶⁻⁸ e.g. reactive groups, when libraries are built for HTS. Although large efforts have been dedicated to the design of chemical libraries, the HTS outcomes are still enriched with hit compounds of dubious

* Corresponding author: rcurpan@acad-icht.tm.edu.ro

quality which are not confirmed in the follow-up rounds and need to be disregarded. For this reason considerable efforts have been employed in the process of data analysis and hit selection in the HTS campaigns, but false hits continue to show up leading to unsatisfactory results.

This study is a continuation of our previous work dedicated to the exploration of the biological profile of promiscuous compounds using chemical reactivity descriptors. In our previous paper,² we have performed calculations in vacuum for three sets of compounds: highly active, presumably promiscuous denominated frequent hitters (FHs), selective or rare hitters (RHs) and inactive or neutral hitters (NHs) and identified the global reactivity descriptors which classify and discriminate between sets. In this paper, we are investigating the influence of the environment over the computed electronic descriptors and for this reason we have performed the computations in water at B3LYP/6-31+G(d) and B3LYP/6-311++G(2df,2pd) levels of theory within the DFT formalism. The following global reactivity descriptors have been computed from the energies of the highest occupied molecular orbital, HOMO and the lowest unoccupied molecular orbital, LUMO: ionization potential, IP, electron affinity, EA, electronegativity, χ , hardness, η , softness, S and electrophilicity index, ω . The water-based computed descriptors have been compared with the vacuum-based descriptors for the three sets of compounds. Reasonable to good linear correlations have been found for the RHs and NHs, respectively. A completely different situation has been encountered in the case of FHs, where correlations are practically inexistent with the following exceptions: HOMO-LUMO gap, hardness η , and softness S for which correlation coefficients have values larger than 0.9. Finally, the descriptors which better discriminate between the classes are discussed and similarities with the vacuum-based computations for the same sets are evaluated, in order to assess the environment which better describes the compounds.

METHODS

1. Compound selection

In this study the compounds were selected from Pubchem database (<https://pubchem.ncbi.nlm.nih.gov/>), the detailed selection process being described in

our previous paper.² Shortly, three sets of compounds, each having 25 members, were compiled from Pubchem database. The highly active compounds, identified as promiscuous structures or frequent hitters (FHs) were defined as compounds active in multiple unrelated biological assays; the selective set which comprises compounds active against a limited number of assays or related families of biological targets were denominated as rare hitters (RHs); and the set of compounds which have been tested in a significant number of biological assays but were declared inactive, were denominated as neutral hitters (NHs). The following criteria were applied in the selection process:

a) Primary filter – the selection of the MLSMR (Molecular Libraries Small Molecule Repository) collection,⁹ specially designed for HTS.

b) Biological filter – only assays with at least 20,000 compounds tested were kept and a biological activity matrix was built, and physico-chemical profiling assays (fluorescence, solubility, etc.) were removed. The lines of this matrix are represented by compounds while the columns represent the assays in which the compounds were tested.

c) Chemical criteria – repeated rounds of clustering were employed to select structural dissimilar compounds for each set. The compounds' clusterization process was performed using Mesa Clustering Module.¹⁰

The structures are depicted in Fig. S1 of Supporting Material.

2. Computational details

Geometry optimizations were carried out in water within the DFT formalism using B3LYP hybrid functional with 6-31+G(d) basis set. Also, single point energy calculations were performed at B3LYP/6-311++G(2df,2pd) level of theory. To evaluate if the optimization reached a true minima on the potential energy surface and not a saddle point, vibrational analysis was performed for all compounds at this level of theory. All calculations were carried out with Gaussian09.¹¹

The global reactivity descriptors were computed for all molecules using HOMO and LUMO energies, E_{HOMO} and E_{LUMO} , computed at both levels of theory. According to Koopmans' theorem,^{12,13} E_{HOMO} approximates the ionization potential (IP) while electron affinity (EA) is approximated by E_{LUMO} . The global reactivity descriptors electronegativity (χ), hardness (η),

softness (S) and electrophilicity index (ω) were calculated within the finite difference approximation using IP and EA. Consequently, the following equations were used to compute the descriptors:

$$\text{IP} \approx -E_{\text{HOMO}} \quad (1)$$

$$\text{EA} \approx -E_{\text{LUMO}} \quad (2)$$

$$\mu = -\frac{\text{IP} + \text{EA}}{2} = -\frac{E_{\text{HOMO}} + E_{\text{LUMO}}}{2} \quad (3)$$

$$\chi = \frac{\text{IP} + \text{EA}}{2} = -\frac{E_{\text{HOMO}} + E_{\text{LUMO}}}{2} \quad (4)$$

$$\eta = \text{IP} - \text{EA} = E_{\text{LUMO}} - E_{\text{HOMO}} \quad (5)$$

$$S = \frac{1}{\eta} \quad (6)$$

$$\omega = \frac{\mu^2}{2\eta} \quad (7)$$

In order to assess the influence of the environment, correlations between vacuum and water based descriptors were performed with package *cluster* available in R statistical software (version 2.14.2).¹⁴ In addition, the statistically significant water computed descriptors which would discriminate between FHs, RHs and NHs were determined using the two-sample Mann-Whitney U test. Also, the area under the receiver operating curve (AUC)¹⁵ was computed to classify the water-based descriptors and identify the best ranking descriptors.

RESULTS AND DISCUSSION

Global reactivity descriptors have been successfully employed to study and explain the correlations between the chemical features of the molecules and their biological behavior. The frontier molecular orbitals, HOMO and LUMO, are related with the tendency of the chemical system to donate or accept electronic density and can be used to predict the chemical behavior of a compound when involved in a nucleophilic or electrophilic attack. Similarly, IP and EA, computed in the context of Koopman's theorem, are two quantities which describe the chemical behavior of compounds, the HOMO energy being related with IP while EA is estimated by LUMO

energy. The chemical potential, μ , electronegativity, χ , and chemical hardness, η , are fundamental quantities evaluated directly from IP and EA (eq. 3-5), being extensively used to assess chemical reactivity of compounds.

The chemical potential, μ (eq. 3), the negative value of electronegativity, χ , characterizes the propensity of electrons to escape from the compound's ground state,¹⁶ while χ (eq. 4) describes the tendency of a molecule to attract electrons.¹⁷ Accordingly, these opposite properties can be used to describe chemical reactivity of molecular systems. In this context, hardness η , related with the HOMO-LUMO energy gap¹⁸ can be computed with eq. 5 and it is a direct measure the chemical stability of a compound. The larger of the value of η for a specific compound, the higher is its chemical stability and the lower its reactivity. Also, another measure of chemical stability and reactivity is provided by softness, S , (eq. 6), the opposite of η . Large values of S indicate increase in chemical reactivity and decrease in stability.

The global reactivity descriptors described above were computed from the frontier orbitals energies, E_{HOMO} and E_{LUMO} , employing eq(s). 1-7. The calculated results are collected in Table S1 of the Supporting Information.

Before exploring the water-based reactivity descriptors which discriminate between sets and performing any correlation with vacuum-based descriptors, we examined the basis set dependence of the HOMO and LUMO energies, and implicitly, of the computed descriptors. The energy minimizations were performed with 6-31+G(d) basis set followed by single point energy calculations using the B3LYP functional with a larger basis set, 6-311++G(2df,2pd). The comparative analysis of the results show good correlations between the smaller and the larger basis sets of computed values for HOMO and LUMO. As clearly revealed in Fig. 1a and 1b, all HOMO and LUMO energies show excellent linear correlations with R^2 values of 0.99. These results show that the economical basis set, 6-31+G(d), provides good results for the frontier orbital energies similar to those obtained by the higher level basis set, 6-311++G(2df,2pd). Thus, the smaller basis set can be sufficient for calculating the HOMO and LUMO energies of the molecules at DFT level.

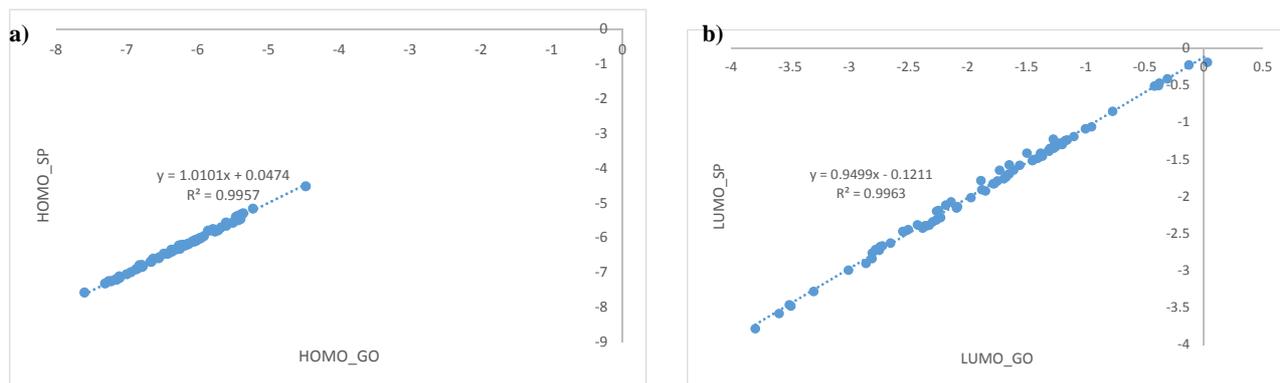


Fig. 1 – Plots of the HOMO (a) and LUMO (b) values calculated using the 6-31+G(d) basis set (GO – geometry optimization) versus those calculated using the larger basis set, B3LYP/6-311++G(2df,2pd) (SP – single point).

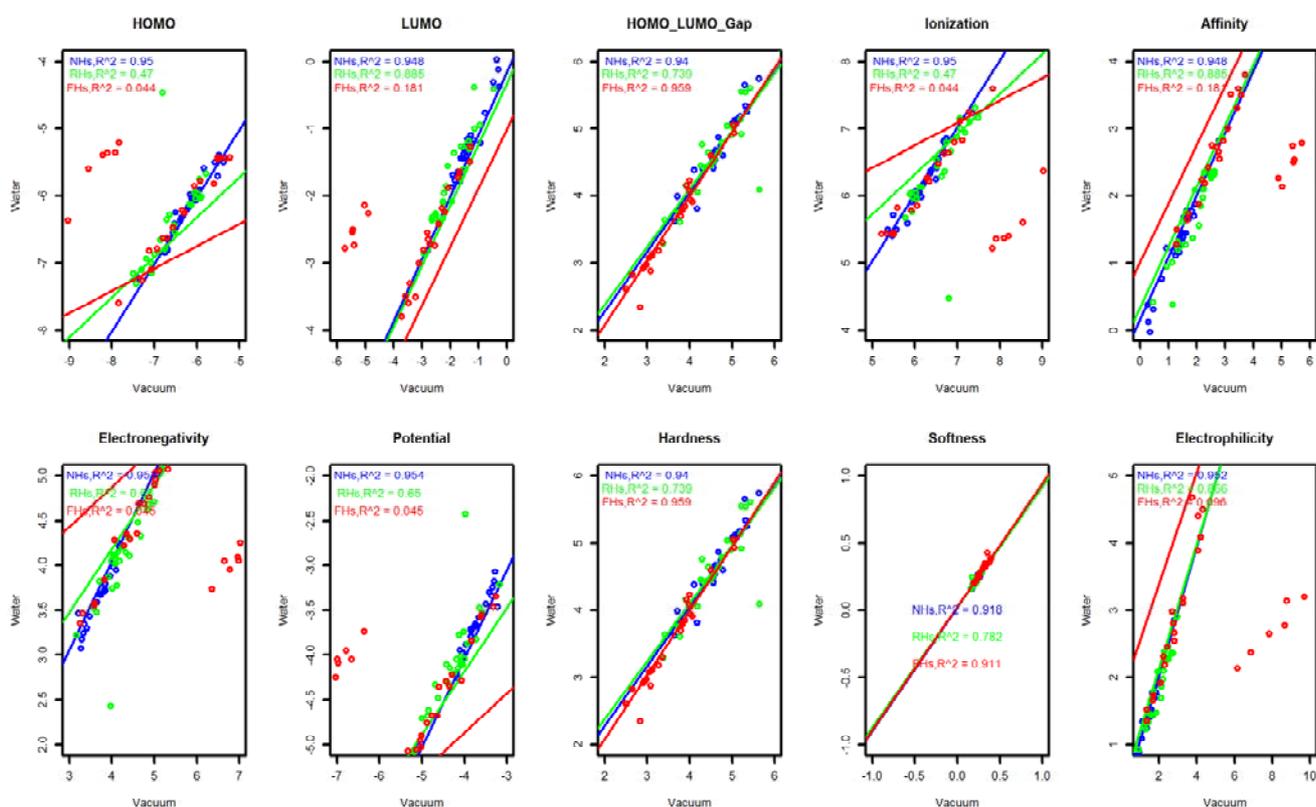


Fig. 2 – Correlations between FHs, RHs, NHs sets for the DFT-based descriptors in water and vacuum.

Good linear correlations between vacuum and water-based descriptors were obtained for NHs (shown in blue in Fig. 2) with correlation coefficients, R^2 , around 0.95. In the case of RHs, satisfactory correlations were noticed for LUMO, EA, hardness, η , softness, S , and electrophilicity index, ω with R^2 ranging between 0.7 and 0.885, while poor correlations were obtained for HOMO, IP, electronegativity, χ and chemical potential, μ . As seen in Fig. 2, for FHs the correlations are practically inexistent between the descriptors

computed in vacuum and those computed in water. Nevertheless, there are few notable exceptions: the HOMO-LUMO gap or hardness, η and softness, S , for which R^2 is around 0.9 (see Fig. 2). Also, these trends were confirmed by Pearson (Fig. 3a) and Spearman rank (Fig. 3b) correlations. These results show without a doubt that HOMO-LUMO gap or hardness, η and softness, S could be reliably used to describe chemical reactivity of the compounds irrespective of the environment used for calculations.

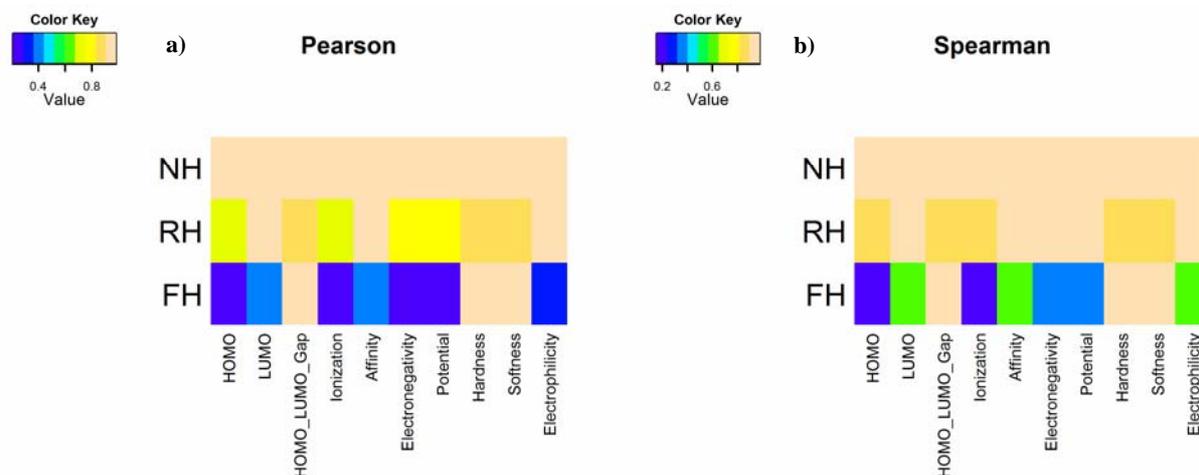


Fig. 3 – Pearson correlation (a) and Spearman rank correlation (b) between vacuum and water computed descriptors.

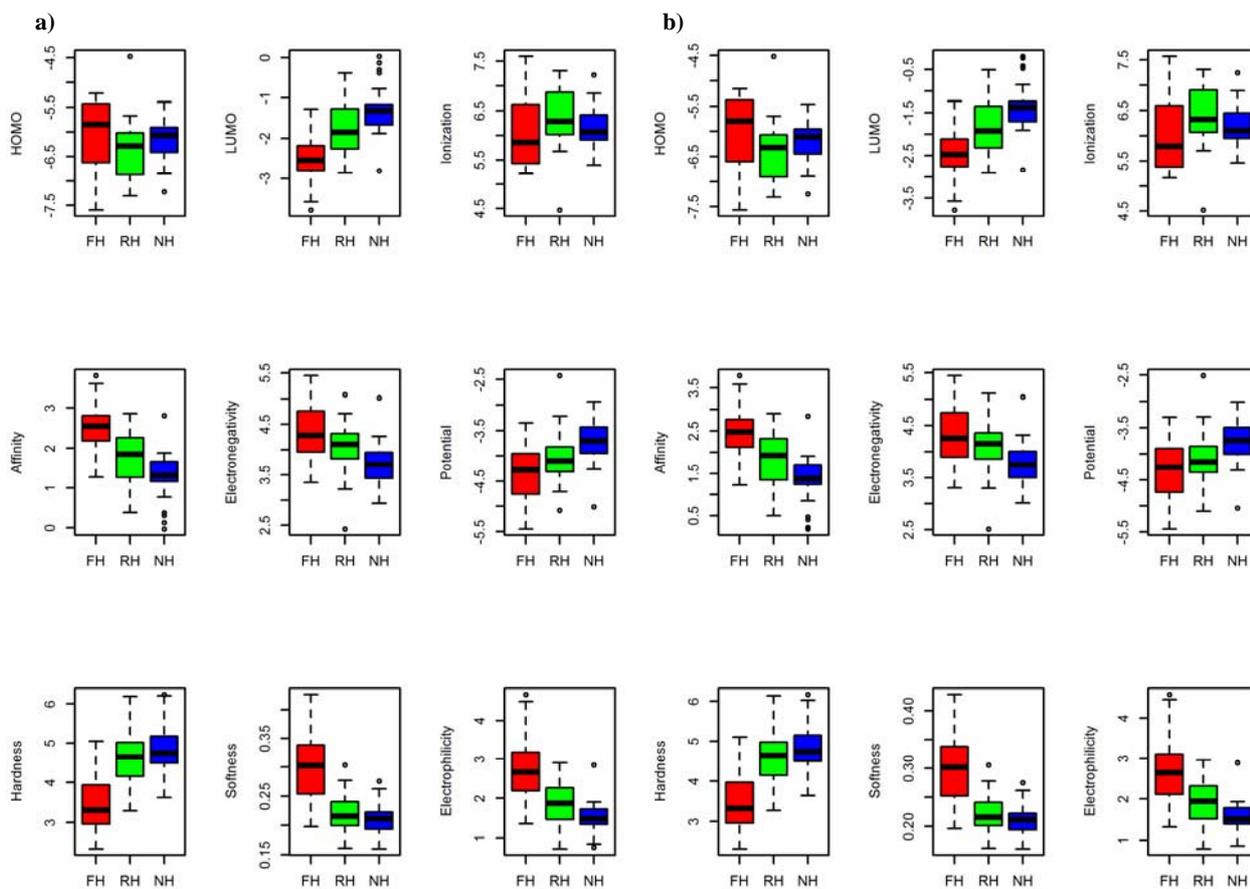


Fig. 4 – Box plot representations of the distributions of descriptors for FHs (red), RHs (green) and NHs (blue) from geometry optimization (panel a) and single point calculations (panel b).

In the following, all of the presented results were obtained based on the values for the HOMO and LUMO energies computed in water with 6-31+G(d) basis set.

The distributions of the descriptors computed from water-based geometries for the three sets of compounds, FHs, RHs and NHs are presented in Fig 4. The results are clearly showing obvious

differences between sets. Thus, E_{LUMO} , EA, hardness, η and softness, S computed for the structures optimized with 6-31+G(d) undoubtedly discriminate between sets, showing significantly lower values for FHs when compared with RHs and NHs, respectively (Fig. 4a). As expected, same trend has been noticed for the single point energy calculations performed with 6-311++G(2df,2pd) basis set (Fig. 4b). These results are in good agreement with the previously reported data obtained for calculations performed in vacuum² and they are practically showing that FHs show a higher tendency of accepting electron density than the RHs and NHs, being able to act as electrophile agents. Therefore, the biological promiscuity of FHs could be a direct consequence of their chemical profile.

The capacity of descriptors to discriminate between each two of the three classes of compounds, i.e., FHs, RHs and NHs was also investigated by computing area under the receiver operating curve (AUC), a classification method. The results indicate generally high AUC values for FHs irrespective of the compounds set used for comparison. Thus, the data are showing that EA, the negative value of LUMO, hardness, η , and electrophilicity index, ω are able to efficiently separate classes of compounds characterized by large hit rates (Fig. 5).

CONCLUSIONS

Global reactivity descriptors calculated in vacuum and water show dependence with the environment but the observed differences are fairly uniform and the same patterns occur for the three sets of molecules irrespective of the environment. Thus, LUMO, hardness, softness and EA descriptors computed from water calculations clearly discriminate between the promiscuous compounds (FHs) on one hand, and the active (RHs) and inactive (NHs) sets on the other hand. The same results have been obtained for the calculations performed in vacuum. It is noteworthy to mention that although the correlations are poor between vacuum and water computed descriptors for the FHs, there is a good correlation between them for the HOMO-LUMO gap, indicating a constant shift of their absolute values which gets canceled when the gap between them is computed. The results demonstrate that electronic descriptors, especially EA (the negative of LUMO) and hardness (HOMO-LUMO gap) can be used to flag structures with high hit rates and to help optimize chemical libraries for HTS.

Acknowledgements: This work was supported by UEFISCDI-CNCSIS project PN-II-RU-PD_502/2010. LH and SA gratefully acknowledges Institute of Chemistry Timișoara of Roumanian Academy, Project no 1.2/2014.

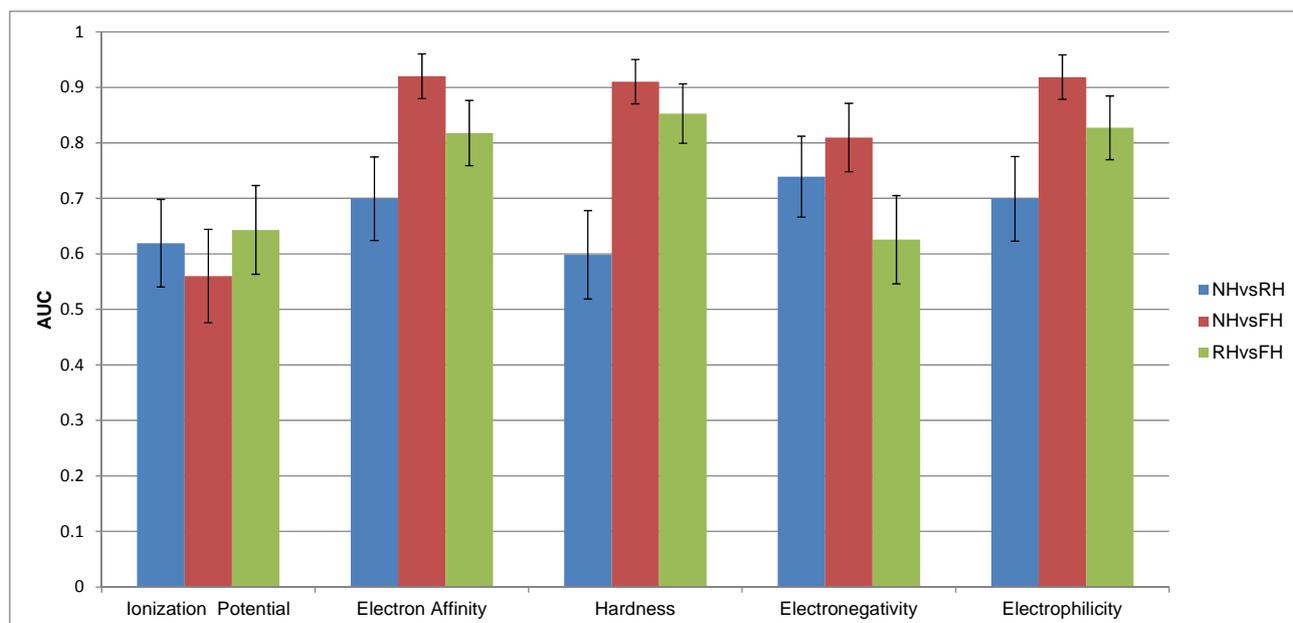


Fig. 5 – Class discrimination by DFT-based descriptors computed in water.

REFERENCES

1. The Pubchem Project, <http://pubchem.ncbi.nlm.nih.gov>; Y. Wang, J. Xiao, T. O. Suzek, J. Zhang, J. Wang, Z. Zhou, L. Han, K. Karapetyan, S. Dracheva, B. A. Shoemaker, E. Bolton, A. Gindulyte and S. H. Bryant, *Nucleic Acids Res.*, **2012**, *40*, D400-D412.
2. R. Curpăn, S. Avram, R. Vianello and C. Bologa, *Bioorg. Med. Chem.*, **2014**, *22*, 2461-2468.
3. L. M. Mayr and D. Bojanic, *Curr. Opin. Pharmacol.*, **2009**, *9*, 580-588.
4. J. G. Houston, M. N. Banks, A. Binnie, S. Brenner, J. O'Connell and E. W. Petrillo, *Drug Discov. Today*, **2008**, *13*, 44-51.
5. A. M. Davis, D. J. Keeling, J. Steele, N. P. Tomkinson and A. C. Tinker, *Curr. Top. Med. Chem.*, **2005**, *5*, 421-439.
6. O. Roche, P. Schneider, J. Zuegge, W. Guba, M. Kansy, A. Alanine, K. Bleicher, F. Danel, E.-M. Gutknecht, M. Rogers-Evans, W. Neidhart, H. Stalder, M. Dillon, E. Sjögren, N. Fotouhi, P. Gillespie, R. Goodnow, W. Harris, P. Jones, M. Taniguchi, S. Tsujii, W. von der Saal, G. Zimmermann and G. Schneider, *J. Med. Chem.*, **2002**, *45*, 137-142.
7. J. B. Baell and G. A. Holloway, *J. Med. Chem.*, **2010**, *53*, 2719-2740.
8. J. Che, F. J. King, B. Zhou and Y. J. Zhou, *J. Chem. Inf. Model.*, **2012**, *52*, 913-926.
9. MLSMR Compounds <http://mli.nih.gov/mli/compound-repository/mlsmrcompounds/>.
10. MESA Analytics & Computing Santa Fe; www.mesaac.com.
11. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09*, Revision B.01; Gaussian: Wallingford, CT, 2009.
12. Z. B. Maksić and R. Vianello, *J. Phys. Chem. A*, **2002**, *106*, 6515-6520.
13. P. Politzer and F. Abu-Awwad, *Theor. Chem. Acta*, **1998**, *99*, 83-87.
14. R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria, **2012**, ISBN 3-900051-07-0. <http://www.R-project.org/>.
15. J. A. Hanley and B. J. McNeil, *Radiology*, **1982**, *143*, 29-36.
16. R. G. Pearson, *J. Chem. Sci.*, **2005**, *117*, 369-377.
17. R. G. Parr, R. A. Donnelly, M. Levy and W. E. Palke, *J. Chem. Phys.*, **1978**, *69*, 3801-3807.
18. R. G. Parr and R. G. Pearson, *J. Am. Chem. Soc.*, **1983**, *105*, 7512-7516.

