

Visualization and Analysis of Complex Reaction Data: The Case of Tautomeric Equilibria

Glavatskikh M., Madzhidov T., Baskin I., Horvath D., Nugmanov R., Gimadiev T., Marcou G., Varnek A.

Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2018 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim Generative Topographic Mapping (GTM) approach was successfully used to visualize, analyze and model the equilibrium constants (KT) of tautomeric transformations as a function of both structure and experimental conditions. The modeling set contained 695 entries corresponding to 350 unique transformations of 10 tautomeric types, for which KT values were measured in different solvents and at different temperatures. Two types of GTM-based classification models were trained: first, a “structural” approach focused on separating tautomeric classes, irrespective of reaction conditions, then a “general” approach accounting for both structure and conditions. In both cases, the cross-validated Balanced Accuracy was close to 1 and the clusters, assembling equilibria of particular classes, were well separated in 2-dimensional GTM latent space. Data points corresponding to similar transformations measured under different experimental conditions, are well separated on the maps. Additionally, GTM-driven regression models were found to have their predictive performance dependent on different scenarios of the selection of local fragment descriptors involving special marked atoms (proton donors or acceptors). The application of local descriptors significantly improves the model performance in 5-fold cross-validation: RMSE=0.63 and 0.82 logKT units with and without local descriptors, respectively. This trend was as well observed for SVR calculations, performed for the comparison purposes.

<http://dx.doi.org/10.1002/minf.201800056>

Keywords

data visualization, Generative Topographic Mapping, QSPR, Support Vector machine, tautomeric equilibria

References

- [1] H. A. Gaspar, G. Marcou, D. Horvath, A. Arault, S. Lozano, P. Vayer, A. Varnek, *J. Chem. Inf. Model.* 2013, 53, 3318–3325
- [2] N. Kireeva, S. L. Kuznetsov, A. Y. Tsivadze, *Ind. Eng. Chem. Res.* 2012, 51, 14337–14343
- [3] H. A. Gaspar, Baskin, I., G. Marcou, D. Horvath, A. Varnek, *Mol. Inf.* 2015, 34, 348–356
- [4] P. Sidorov, H. Gaspar, G. Marcou, A. Varnek, D. Horvath, *J. Comput.-Aided Mol. Des.* 2015, 29, 1087–1108
- [5] J. R. Greenwood, D. Calkins, A. P. Sullivan, J. C. Shelley, *J. Comput.-Aided Mol. Des.* 2010, 24, 591–604
- [6] T. Clark, *J. Comput.-Aided Mol. Des.* 2010, 24, 605–611
- [7] P. Pospisil, P. Ballmer, L. Scapozza, G. Folkers, *J. Recept. Signal Transduction* 2003, 23, 361–371

- [8] F. Oellien, J. Cramer, C. Beyer, W. D. Ihlenfeldt, P. M. Selzer, *J. Chem. Inf. Model.* 2006, 46, 2342-2354
- [9] Y. C. Martin, *J. Comput.-Aided Mol. Des.* 2009, 23, 693-704
- [10] W. A. Warr, *J. Comput.-Aided Mol. Des.* 2010, 24, 497-520
- [11] R. A. Sayle, *J. Comput.-Aided Mol. Des.* 2010, 24, 485-496
- [12] ChemAxon, TautomerizationPlugin, <http://www.chemaxon.com/marvin/help/calculations/tautomers.html>
- [13] F. Milletti, L. Storchi, G. Sforza, S. Cross, G. Cruciani, *J. Chem. Inf. Model.* 2009, 49, 68-75
- [14] T. R. Gimadiev, T. I. Madzhidov, R. I. Nugmanov, I. I. Baskin, I. S. Antipin, A. Varnek, *J. Comput.-Aided Mol. Des.* 2018
- [15] M. Glavatskikh, T. Madzhidov, V. Solov'ev, G. Marcou, D. Horvath, J. Graton, J.-Y. Le Questel, A. Varnek, *Mol. Inf.* 2016, 35, 70-80
- [16] M. Glavatskikh, T. Madzhidov, V. Solov'ev, G. Marcou, D. Horvath, A. Varnek, *Mol. Inf.* 2016, 35, 629-638
- [17] F. Ruggiu, V. Solov'ev, G. Marcou, D. Horvath, J. Graton, J. Y. Le Questel, A. Varnek, *Mol. Inf.* 2014, 33, 477-487
- [18] C. M. Bishop, M. Svensen, C. K. I. Williams, *Neural Comput.* 1998, 10, 215-234
- [19] N. Kireeva, I. I. Baskin, H. A. Gaspar, D. Horvath, G. Marcou, A. Varnek, *Mol. Inf.* 2012, 31, 301-312
- [20] H. A. Gaspar, I. I. Baskin, G. Marcou, D. Horvath, A. Varnek, *J. Chem. Inf. Model.* 2015, 55, 84-94
- [21] H. A. Gaspar, I. I. Baskin, G. Marcou, D. Horvath, A. Varnek, *Mol. Inf.* 2015, 34, 348-356
- [22] H. A. Gaspar, I. I. Baskin, G. Marcou, D. Horvath, A. Varnek, *J. Chem. Inf. Model.* 2015, 55, 2403-2410
- [23] P. Sidorov, H. Gaspar, G. Marcou, A. Varnek, D. Horvath, *J. Comput.-Aided Mol. Des.* 2015, 29, 1087-1108
- [24] H. A. Gaspar, I. I. Baskin, A. Varnek, in *Frontiers in Molecular Design and Chemical Information Science - Herman Skolnik Award Symposium 2015: Jürgen Bajorath*, Vol. 1222, American Chemical Society, 2016, pp. 243-267
- [25] H. A. Gaspar, P. Sidorov, D. Horvath, I. I. Baskin, G. Marcou, A. Varnek, in *Frontiers in Molecular Design and Chemical Information Science - Herman Skolnik Award Symposium 2015: Jürgen Bajorath*, Vol. 1222, American Chemical Society, 2016, pp. 211-241
- [26] A. Smola, V. Vapnik, *Adv. Neural Inform. Process. Syst.* 1997, 9, 155-161
- [27] A. J. Smola, B. Schölkopf, *Stat. Comput.* 2004, 14, 199-222
- [28] V. A. Palm, *Tables of Rate and Equilibrium Constants of Heterolytic Organic Reactions*, VINITI, Moscow, 1978
- [29] Standardizer, 6.1.5, ChemAxon (<http://www.chemaxon.com>), 2013
- [30] A. Varnek, D. Fourches, F. Hoonakker, V. Solov'ev, *J. Comput.-Aided Mol. Des.* 2005, 19, 693-703
- [31] A. Varnek, D. Fourches, D. Horvath, O. Klimchuk, C. Gaudin, P. Vayer, V. Solov'ev, F. Hoonakker, I. V. Tetko, G. Marcou, *Curr. Comput.-Aided Drug Des.* 2008, 4, 191-198
- [32] A. Varnek, D. Fourches, F. Hoonakker, V. P. Solov'ev, *J. Comput.-Aided Mol. Des.* 2005, 19, 693-703
- [33] I. I. Baskin, M. I. Skvortsova, I. V. Stankevich, N. S. Zefirov, *J. Chem. Inf. Comput. Sci.* 1995, 35, 527-531
- [34] I. Baskin, A. Varnek, in *Chemoinformatics Approaches to Virtual Screening* (Eds.: A. Varnek, A. Tropsha), RSC Publisher, Cambridge, 2008, pp. 1-43
- [35] I. Baskin, A. Varnek, *Comb. Chem. High Throughput Screening* 2008, 11, 661-668
- [36] N. I. Zhokhova, I. I. Baskin, V. A. Palyulin, A. N. Zefirov, N. S. Zefirov, *Dokl. Chem.* 2007, 417, 282-284
- [37] ISIDA Fragmentor2017, Laboratory of Cheminformatics, UMR 7140, University of Strasbourg, France, 2017
- [38] T. I. Madzhidov, P. G. Polishchuk, R. I. Nugmanov, A. V. Bodrov, A. I. Lin, I. I. Baskin, A. A. Varnek, I. S. Antipin, *Russ. J. Org. Chem.* 2014, 50, 459-463
- [39] T. I. Madzhidov, A. V. Bodrov, T. R. Gimadiev, R. I. Nugmanov, I. S. Antipin, A. A. Varnek, *J. Struct. Chem.* 2015, 56, 1227-1234
- [40] J. Catalán, V. López, P. Pérez, R. Martín-Villamil, J.-G. Rodríguez, *Liebigs Ann.* 1995, 1995, 241-252
- [41] J. Catalán, C. Díaz, *Liebigs Ann.* 1997, 1997, 1941-1949
- [42] J. Catalán, C. Díaz, V. López, P. Pérez, J.-L. G. De Paz, J. G. Rodríguez, *Liebigs Ann.* 1996, 1996, 1785-1794
- [43] R. W. Taft, M. J. Kamlet, *J. Am. Chem. Soc.* 1976, 98, 2886-2894
- [44] M. J. Kamlet, R. W. Taft, *J. Am. Chem. Soc.* 1976, 98, 377-383
- [45] M. J. Kamlet, J. L. Abboud, R. W. Taft, *J. Am. Chem. Soc.* 1977, 99, 6027-6038
- [46] C.-C. Chang, C.-J. Lin, *ACM Trans. Intell. Syst. Technol.* 2011, 2, 1-27
- [47] D. Horvath, J. Brown, G. Marcou, A. Varnek, *Challenges* 2014, 5, 450
- [48] H. A. Gaspar, Baskin, II, G. Marcou, D. Horvath, A. Varnek, *J. Chem. Inf. Model.* 2015, 55, 84-94
- [49] S. I. Ovchinnikova, A. A. Bykov, A. Y. Tsivadze, E. P. Dyachkov, N. V. Kireeva, *J. Cheminf.* 2014, 6
- [50] M. Mathea, W. Klingspohn, K. Baumann, *Mol. Inf.* 2016, 35, 160-180

[51] A. Varnek, D. Fourches, D. Horvath, O. Klimchuk, C. Gaudin, P. Vayer, V. Solov'ev, F. Hoonakker, I. V. Tetko, G. Marcou, *Curr. Comput.-Aided Drug Des.* 2008, 4, 191-198