

Public (Q)SAR Services, Integrated Modeling Environments, and Model Repositories on the Web: State of the Art and Perspectives for Future Development

Tetko I., Maran U., Tropsha A.

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

Abstract

© 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim Thousands of (Quantitative) Structure-Activity Relationships (Q)SAR models have been described in peer-reviewed publications; however, this way of sharing seldom makes models available for the use by the research community outside of the developer's laboratory. Conversely, on-line models allow broad dissemination and application representing the most effective way of sharing the scientific knowledge. Approaches for sharing and providing on-line access to models range from web services created by individual users and laboratories to integrated modeling environments and model repositories. This emerging transition from the descriptive and informative, but "static", and for the most part, non-executable print format to interactive, transparent and functional delivery of "living" models is expected to have a transformative effect on modern experimental research in areas of scientific and regulatory use of (Q)SAR models.

<http://dx.doi.org/10.1002/minf.201600082>

Keywords

chemoinformatics, model repositories, on-line modeling environments, QSAR, QSPR, web-based models

References

- [1] A. Cherkasov, E. N. Muratov, D. Fourches, A. Varnek, I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y. C. Martin, R. Todeschini, V. Consonni, V. E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, A. Tropsha, *J. Med. Chem.* 2014, 57, 4977-5010.
- [2] I. Baskin, D. Winkler, I. V. Tetko, *Expert Opin. Drug Discov.* 2016, 11, 785-795.
- [3] J. C. Dearden, M. T. Cronin, K. L. Kaiser, *SAR QSAR Environ. Res.* 2009, 20, 241-266.
- [4] T. R. Stouch, J. R. Kenyon, S. R. Johnson, X. Q. Chen, A. Doweiko, Y. Li, *J. Comput. Aided. Mol. Des.* 2003, 17, 83-92.
- [5] S. R. Johnson, *J. Chem. Inf. Model.* 2008, 48, 25-26.
- [6] O. Spjuth, E. L. Willighagen, R. Guha, M. Eklund, J. E. Wikberg, *J. Cheminform.* 2010, 2, 5.
- [7] P. Gramatica, *QSAR Comb. Sci.* 2007, 26, 694-701.
- [8] E. Benfenati, M. Clook, S. Fryday, A. Hart, in *Quantitative Structure-Activity Relationships (QSAR) for Pesticide Regulatory Purposes*, Elsevier, Amsterdam, 2007, pp. 1-57.
- [9] I. V. Tetko, D. J. Livingstone, A. I. Luik, *J. Chem. Inf. Comput. Sci.* 1995, 35, 826-833.
- [10] V. Ruusmann, U. Maran, *J. Comput. Aided. Mol. Des.* 2013, 27, 583-603.

- [11] A. Tropsha, *Mol. Inf.* 2010, 29, 476-488.
- [12] Guidance document on the validation of (quantitative)structure-activity relationships [(Q)SAR] models. [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono\(2007\)2](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2007)2)
- [13] A. G. Artemenko, E. N. Muratov, V. E. Kuz'min, N. N. Muratov, E. V. Varlamova, A. V. Kuz'mina, L. G. Gorb, A. Golius, F. C. Hill, J. Leszczynski, A. Tropsha, *SAR QSAR Environ. Res.* 2011, 22, 575-601.
- [14] D. Fourches, E. Muratov, A. Tropsha, *J. Chem. Inf. Model.* 2010, 50, 1189-1204.
- [15] R. Mannhold, G. I. Poda, C. Ostermann, I. V. Tetko, *J. Pharm. Sci.* 2009, 98, 861-893.
- [16] C. A. Azencott, A. Ksikes, S. J. Swamidass, J. H. Chen, L. Ralaivola, P. Baldi, *J. Chem. Inf. Model.* 2007, 47, 965-974.
- [17] A. Varnek, D. Fourches, D. Horvath, O. Klimchuk, C. Gaudin, P. Vayer, V. Solov'ev, F. Hoonakker, I. V. Tetko, G. Marcou, *Curr. Comp. Aid. Drug Design* 2008, 4, 191-198.
- [18] I. Sushko, S. Novotarskyi, R. Korner, A. K. Pandey, A. Cherkasov, J. Li, P. Gramatica, K. Hansen, T. Schroeter, K. R. Muller, L. Xi, H. Liu, X. Yao, T. Oberg, F. Hormozdiari, P. Dao, C. Sahinalp, R. Todeschini, P. Polishchuk, A. Artemenko, V. Kuz'min, T. M. Martin, D. M. Young, D. Fourches, E. Muratov, A. Tropsha, I. Baskin, D. Horvath, G. Marcou, C. Muller, A. Varnek, V. V. Prokopenko, I. V. Tetko, *J. Chem. Inf. Model.* 2010, 50, 2094-2111.
- [19] K. Mansouri, A. Abdelaziz, A. Rybacka, A. Roncaglioni, A. Tropsha, A. Varnek, A. Zakharov, A. Worth, A. M. Richard, C. M. Grulke, D. Trisciuzzi, D. Fourches, D. Horvath, E. Benfenati, E. Muratov, E. B. Wedebye, F. Grisoni, G. F. Mangiatordi, G. M. Incisivo, H. Hong, H. W. Ng, I. V. Tetko, I. Balabin, J. Kancherla, J. Shen, J. Burton, M. Nicklaus, M. Cassotti, N. G. Nikolov, O. Nicolotti, P. L. Andersson, Q. Zang, R. Politi, R. D. Beger, R. Todeschini, R. Huang, S. Farag, S. A. Rosenberg, S. Slavov, X. Hu, R. S. Judson, *Environ. Health Perspect.* 2016, 124, 1023-1033.
- [20] M. N. Drwal, P. Banerjee, M. Dunkel, M. R. Wettig, R. Preissner, *Nucleic Acids Res.* 2014, 42, W53-58.
- [21] M. N. Drwal, V. B. Sramshetty, P. Banerjee, A. Goede, R. Preissner, M. Dunkel, *Frontiers Environ. Sci.* 2015, 3.
- [22] F. Cheng, W. Li, Y. Zhou, J. Shen, Z. Wu, G. Liu, P. W. Lee, Y. Tang, *J. Chem. Inf. Model.* 2012, 52, 3099-3105.
- [23] I. V. Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V. A. Palyulin, E. V. Radchenko, N. S. Zefirov, A. S. Makarenko, V. Y. Tanchuk, V. V. Prokopenko, *J. Comput. Aided. Mol. Des.* 2005, 19, 453-463.
- [24] S. Brandmaier, W. Peijnenburg, M. K. Durjava, B. Kolar, P. Gramatica, E. Papa, B. Bhatarai, S. Kovarich, S. Cassani, P. P. Roy, M. Rahmberg, T. Oberg, N. Jeliazkova, L. Golsteijn, M. Comber, L. Charochkina, S. Novotarskyi, I. Sushko, A. Abdelaziz, E. D'Onofrio, P. Kunwar, F. Ruggiu, I. V. Tetko, *Altern. Lab. Anim.* 2014, 42, 13-24.
- [25] I. Sushko, S. Novotarskyi, R. Korner, A. K. Pandey, M. Rupp, W. Teetz, S. Brandmaier, A. Abdelaziz, V. V. Prokopenko, V. Y. Tanchuk, R. Todeschini, A. Varnek, G. Marcou, P. Ertl, V. Potemkin, M. Grishina, J. Gasteiger, C. Schwab, I. I. Baskin, V. A. Palyulin, E. V. Radchenko, W. J. Welsh, V. Kholodovych, D. Chekmarev, A. Cherkasov, J. Aires-de-Sousa, Q. Y. Zhang, A. Bender, F. Nigsch, L. Patiny, A. Williams, V. Tkachenko, I. V. Tetko, *J. Comput. Aided. Mol. Des.* 2011, 25, 533-554.
- [26] A. Abdelaziz, Y. Sushko, S. Novotarskyi, R. Korner, S. Brandmaier, I. V. Tetko, *Comb. Chem. High Throughput Screen.* 2015, 18, 420-438.
- [27] R. S. Judson, K. A. Houck, R. J. Kavlock, T. B. Knudsen, M. T. Martin, H. M. Mortensen, D. M. Reif, D. M. Rotroff, I. Shah, A. M. Richard, D. J. Dix, *Environ. Health Perspect.* 2010, 118, 485-492.
- [28] F. Sanz, P. Carrió, O. López, L. Capoferri, D. P. Kooi, N. P. E. Vermeulen, D. P. Geerke, F. Montanari, G. F. Ecker, C. H. Schwab, T. Kleinöder, T. Magdziarz, M. Pastor, *Mol. Inf.* 2015, 34, 477-484.
- [29] B. Hardy, N. Douglas, C. Helma, M. Rautenberg, N. Jeliazkova, V. Jeliazkov, I. Nikolova, R. Benigni, O. Tcheremenskaia, S. Kramer, T. Girschick, F. Buchwald, J. Wicker, A. Karwath, M. Gutlein, A. Maunz, H. Sarimveis, G. Melagraki, A. Afantitis, P. Sopasakis, D. Gallagher, V. Poroikov, D. Filimonov, A. Zakharov, A. Lagunin, T. Glorizova, S. Novikov, N. Skvortsova, D. Druzhilovsky, S. Chawla, I. Ghosh, S. Ray, H. Patel, S. Escher, *J. Cheminform.* 2010, 2, 7.
- [30] A. A. Toropov, A. P. Toropova, I. Raska, Jr., D. Leszczynska, J. Leszczynski, *Comput Biol Med* 2014, 45, 20-25.
- [31] O. Nicolotti, E. Benfenati, A. Carotti, D. Gadaleta, A. Gissi, G. F. Mangiatordi, E. Novellino, *Drug Discov. Today* 2014, 19, 1757-1768.
- [32] A. A. Lagunin, R. K. Goel, D. Y. Gawande, P. Pahwa, T. A. Glorizova, A. V. Dmitriev, S. M. Ivanov, A. V. Rudik, V. I. Konova, P. V. Pogodin, D. S. Druzhilovsky, V. V. Poroikov, *Nat. Prod. Rep.* 2014, 31, 1585-1611.
- [33] Protection Agency EPIU.S. Environmental Suite v 4.11. <http://www.epa.gov/opptintr/exposure/pubs/episuitedl.htm>
- [34] E. Benfenati, A. Manganaro, G. GiniCEUR Workshop Proceedings, in, Vol. 1107, 2013, pp. 21-28.
- [35] G. Patlewicz, N. Jeliazkova, R. J. Safford, A. P. Worth, B. Aleksiev, *SAR QSAR Environ. Res.* 2008, 19, 495-524.
- [36] P. Gramatica, S. Cassani, N. Chirico, *J. Comput. Chem.* 2014, 35, 1036-1044.
- [37] I. V. Tetko, T. I. Aksenova, V. V. Volkovich, T. N. Kasheva, D. V. Filipov, W. J. Welsh, D. J. Livingstone, A. E. P. Villa, *SAR QSAR Environ. Res.* 2000, 11, 263-280.
- [38] I. V. Tetko, V. Y. Tanchuk, T. N. Kasheva, A. E. Villa, *J. Chem. Inf. Comput. Sci.* 2001, 41, 246-252.
- [39] T. I. Aksyonova, V. V. Volkovich, I. V. Tetko, *Syst. Anal. Model. Sim.* 2003, 43, 1331-1339.

- [40] I. V. Tetko, *J. Chem. Inf. Comput. Sci.* 2002, 42, 717–728.
- [41] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, WILEY-VCH, Weinheim, 2000.
- [42] I. V. Tetko, A. Facius, A. Ruepp, H. W. Mewes, *BMC Bioinformatics* 2005, 6, 82.
- [43] D. C. Whitley, M. G. Ford, D. J. Livingstone, *J. Chem. Inf. Comput. Sci.* 2000, 40, 1160–1168.
- [44] S. Sild, U. Maran, A. Lomaka, M. Karelson, *J. Chem. Inf. Model.* 2006, 46, 953–959.
- [45] S. Sild, U. Maran, M. Romberg, B. Schuller, E. Benfenati, in *Lecture Notes in Computer Science*, Vol. 3470, 2005, pp. 464–473.
- [46] B. Schuller, B. Demuth, H. Mix, K. Rasch, M. Romberg, S. Sild, U. Maran, P. Bała, E. Del Grosso, M. Casalegno, N. Piclin, M. Pintore, W. Sudholt, K. K. Baldridge, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 4854 LNCS, 2008, pp. 82–93.
- [47] T. Walker, C. M. Grulke, D. Pozefsky, A. Tropsha, *Bioinformatics* 2010, 26, 3000–3001.
- [48] I. V. Tetko, S. Novotarskyi, I. Sushko, V. Ivanov, A. E. Petrenko, R. Dieden, F. Lebon, B. Mathieu, *J. Chem. Inf. Model.* 2013, 53, 1990–2000.
- [49] I. V. Tetko, Y. Sushko, S. Novotarskyi, L. Patiny, I. Kondratov, A. E. Petrenko, L. Charochkina, A. M. Asiri, *J. Chem. Inf. Model.* 2014, 54, 3320–3329.
- [50] I. V. Tetko, D. Lowe, A. J. Williams, *J. Cheminform.* 2016, 8, 2.
- [51] A. E. Villa, I. V. Tetko, P. Dutoit, Y. De Ribaupierre, F. De Ribaupierre, *J. Neurosci. Methods* 1999, 86, 161–178.
- [52] I. V. Tetko, *Methods Mol. Biol.* 2008, 458, 185–202.
- [53] I. V. Tetko, P. Bruneau, *J. Pharm. Sci.* 2004, 93, 3103–3110.
- [54] C. Cortes, V. Vapnik, *Machine Learn.* 1995, 20, 273–297.
- [55] N. I. Zhokhova, I. I. Baskin, V. A. Palyulin, A. N. Zefirov, N. S. Zefirov, *Dokl. Chem.* 2007, 417, 282–284.
- [56] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, *SIGKDD Explorations* 2009, 11.
- [57] S. Vorberg, I. V. Tetko, *Mol. Inf.* 2014, 33, 73–85.
- [58] I. Sushko, E. Salmina, V. A. Potemkin, G. Poda, I. V. Tetko, *J. Chem. Inf. Model.* 2012, 52, 2310–2316.
- [59] E. S. Salmina, N. Haider, I. V. Tetko, *Molecules* 2016, 21, 1.
- [60] K. Schorpp, I. Rothenaigner, E. Salmina, J. Reinshagen, T. Low, J. K. Brenke, J. Gopalakrishnan, I. V. Tetko, S. Gul, K. Hadian, *J. Biomol. Screen.* 2014, 19, 715–726.
- [61] J. K. Brenke, E. S. Salmina, L. Ringelstetter, S. Dornauer, M. Kuzikov, I. Rothenaigner, K. Schorpp, F. Giehler, J. Gopalakrishnan, A. Kieser, S. Gul, I. V. Tetko, K. Hadian, *J. Biomol. Screen.* 2016, 21, 596–607.
- [62] Y. Sushko, S. Novotarskyi, R. Korner, J. Vogt, A. Abdelaziz, I. V. Tetko, *J. Cheminform.* 2014, 6, 48.
- [63] I. V. Tetko, P. Bruneau, H. W. Mewes, D. C. Rohrer, G. I. Poda, *Drug Discov. Today* 2006, 11, 700–707.
- [64] A. Varnek, C. Gaudin, G. Marcou, I. Baskin, A. K. Pandey, I. V. Tetko, *J. Chem. Inf. Model.* 2009, 49, 133–144.
- [65] I. Oprisiu, S. Novotarskyi, I. V. Tetko, *J. Cheminform.* 2013, 5, 4.
- [66] J. Gasteiger, *J. Med. Chem.* 2006, 49, 6429–6434.
- [67] V. Potemkin, M. Grishina, *Drug Discov. Today* 2008, 13, 952–959.
- [68] I. V. Tetko, V. P. Solov'ev, A. V. Antonov, X. Yao, J. P. Doucet, B. Fan, F. Hoonakker, D. Fourches, P. Jost, N. Lachiche, A. Varnek, *J. Chem. Inf. Model.* 2006, 46, 808–819.
- [69] I. V. Tetko, I. Sushko, A. K. Pandey, H. Zhu, A. Tropsha, E. Papa, T. Oberg, R. Todeschini, D. Fourches, A. Varnek, *J. Chem. Inf. Model.* 2008, 48, 1733–1746.
- [70] I. V. Tetko, H. P. Varbanov, M. Galanski, M. Talmaciu, J. A. Platts, M. Ravera, E. Gabano, *J. Inorg. Biochem.* 2016, 156, 1–13.
- [71] I. V. Tetko, I. Jaroszewicz, J. A. Platts, J. Kuduk-Jaworska, *J. Inorg. Biochem.* 2008, 102, 1424–1437.
- [72] I. Sushko, S. Novotarskyi, R. Körner, A. K. Pandey, V. V. Kovalishyn, V. V. Prokopenko, I. V. Tetko, *J. Chemom.* 2010, 24, 202–208.
- [73] S. Novotarskyi, I. Sushko, R. Korner, A. K. Pandey, I. V. Tetko, *J. Chem. Inf. Model.* 2011, 51, 1271–1280.
- [74] S. Novotarskyi, A. Abdelaziz, Y. Sushko, R. Korner, J. Vogt, I. V. Tetko, *Chem. Res. Toxicol.* 2016, 29, 768–775.
- [75] A. Abdelaziz, H. Spahn-Langguth, K. Werner-Schramm, I. V. Tetko, *Frontiers Environ. Sci.* 2016, 4, 2.
- [76] I. V. Tetko, K. W. Schramm, T. Knepper, W. J. Peijnenburg, A. J. Hendriks, J. M. Navas, I. A. Nicholls, T. Oberg, R. Todeschini, E. Schlosser, S. Brandmaier, *Altern. Lab. Anim.* 2014, 42, 7–11.
- [77] I. V. Tetko, O. Engkvist, U. Koch, J. L. Reymond, H. Chen, *Mol. Inf.* 2016, DOI: 10.1002/minf.201600073
- [78] B. R. Brooks, R. E. Brucoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comp. Chem.* 1983, 4, 187–217.
- [79] I. E. Weidlich, Y. Pevzner, B. T. Miller, I. V. Filippov, H. L. Woodcock, B. R. Brooks, *J. Comput. Chem.* 2015, 36, 62–67.

- [80] A. Palczewska, X. Fu, P. Trundle, L. Yang, D. Neagu, M. Ridley, K. Travis, *Int. J. Inform. Manag.* 2013, 33, 567-582.
- [81] X. Fu, A. Wojak, D. Neagu, M. Ridley, K. Travis, *J. Cheminform.* 2011, 3, 24.
- [82] V. Ruusmann, S. Sild, U. Maran, *J. Cheminform.* 2014, 6, 25.
- [83] S. Sild, G. Piir, D. Neagu, U. Maran, in *Big Data in Predictive Toxicology*, (Eds.: D. Neagu, A. Richarz), Royal Society of Chemistry, 2016, in press.
- [84] V. Ruusmann, S. Sild, U. Maran, *J. Cheminform.* 2015, 7, 32.
- [85] C. Steinbeck, Y. Han, S. Kuhn, O. Horlacher, E. Luttmann, E. Willighagen, *J. Chem. Inf. Comput. Sci.* 2003, 43, 493-500.
- [86] W. M. Berhanu, G. G. Pillai, A. A. Oliferenko, A. R. Katritzky, *ChemPlusChem* 2012, 77, 507-517.
- [87] T. Fujita, D. A. Winkler, *J. Chem. Inf. Model.* 2016, 56, 269-274.
- [88] J. Dearden, *Int. J. Quant. Struct. Prop. Relat.* 2016, 1, 1-14.
- [89] G. Melagraki, A. Afantitis, *Comb. Chem. High Throughput Screen.* 2016, 19, 260-261.
- [90] D. Fourches, E. Muratov, A. Tropsha, *Nat. Chem. Biol.* 2015, 11, 535.
- [91] I. V. Tetko, *J. Comput. Aided. Mol. Des.* 2012, 26, 135-136.