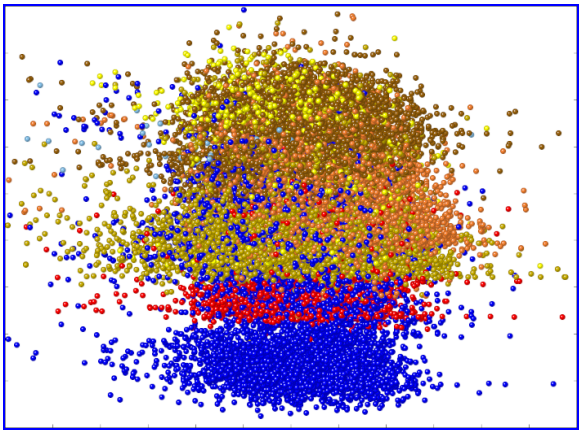
A Web Tool for Calculating Substituent Descriptors Compatible with Hammett Sigma Constants

Peter Ertl

Novartis Institutes for BioMedical Research, Novartis Campus, CH-4056 Basel, Switzerland
peter.ertl@novartis.com
https://peter-ertl.com

Electron donating or accepting power of organic substituents is an important parameter affecting many properties of parent molecules, most notably their reactivity and pKa of ionizable groups. These substituent properties are usually described by Hammett sigma constants obtained by measuring ionization of substituted benzoic acids. Although values of these constants have been measured for the most common functional groups, data for many important substituents are not available. Some time ago we reported a method to calculate substituent descriptors compatible with Hammett sigma constants using quantum chemically derived parameters. The present publication revisits the older study by applying more sophisticated methodology and a larger training data set, as well as introduces a free web tool allowing to calculate substituent descriptors compatible with Hammett sigma constants available at https://bitly.com/getsigmas.



Keywords: cheminformatics, organic substituents, medicinal chemistry, Hammett sigma constants, Hansch pi hydrophobicity constants, QSAR, web tools

Introduction

The concept of electron-withdrawing and electron-donating power of organic substituents characterised by parameters obtained by measuring ionization of substituted benzoic acids introduced by Hammett revolutionized the field of physical organic chemistry. It helped to better understand the effect of substituents on reactivity, pKa of ionizable groups and also on the biological activity of parent molecules. This idea, extended and introduced into the QSAR field by Hansch and Fujita [1–3] followed up by the decision schemes of Craig [4,5] and Topliss [6] proved to be the major breakthroughs in our understanding of the relationship between electronic and hydrophobic substituent properties and the biological activity of the parent molecules.

In the early days of QSAR the experimental Hammett sigma constants were used directly as descriptors in correlation equations. Soon, however, it became evident that this procedure has several drawbacks. Its major disadvantage, known to everybody who tried this approach, is the fact that the values of Hammett sigmas are simply for many important substituents not available. According to our analysis out of 200 most common substituents extracted from the bioactive molecules in the ChEMBL database [7] the experimental Hammett sigma constants [8] are available only for less than half, namely 89. This lack of data led to several attempts to replace experimental values of Hammett sigma constants by calculated parameters.

In our previous study we documented very good correlation between experimental Hammett sigmas and AM1 calculated atomic charges on a simple group attached to the substituents (r^2_{cv} = 0.87 for σ_{meta} and 0.92 for σ_{para}, respectively).[9] These calculated parameters have been successfully used in characterization of substituents in bioisosteric design, [10,11] and generative chemistry applications.[12]

Several other authors also reported good correlation between Hammett substituent constants and quantum chemically derived parameters. Genix et al. [13] has shown that calculated charges and orbital coefficients for a set of substituted benzoic acids correlate well with the experimental Hammett sigma constants. Gross and Seybold [14] performed similar analysis for a set of substituted anilines. Takahata and Chong [15] correlated Hammett sigmas with the core-electron binding energies obtained by density-functional calculations for 12 substituted benzenes. And recently Sessa et al. [16] introduced a quantum chemically derived parameter termed Q descriptor, that he suggested to use as a quantifier of substituent effects and chemical reactivity.

Methodology

In the present study the substituents descriptors compatible with the Hammett sigma constants were calculated in a similar way as in our previous analysis [9], several important improvements, however, were introduced to this methodology. The basic principle of the method consists of attaching a simple group to the substituent in question, calculating atomic charges on this attachment and then correlating these charges with the

experimental Hammett sigma constants for a training set of substituents. The resulting equation can then be used to calculate Hammett descriptors for any arbitrary substituent. As already mentioned, several enhancements were introduced to the method. While in the earlier study data for 63 substituents were used for the training in the current study this set was increased to 89 - the substituents that have known experimental sigma constants out of the 200 most frequent ones. The substituents used for the training are depicted in Figure 1. In the previous study the atomic charges were calculated on a butadienyl group attached to the substituents, in the present case the phenyl group was used as an attachment, providing slightly better correlation then that for butadienyl. The biggest improvement is the way how the atomic charges themselves are calculated. In the previous study this was done by the AM1 method for gemetries created by CORINA without geometry optimization. In the present study the xtb [17], an extremely fast and reliable semiempirical quantum-mechanical method, is used to calculate atomic charges, the 3D geometry is created from SMILES by the Galaxy package [18] and the geometry is subsequently fully optimised by xtb. The calculated charges on the carbons of the phenyl attachment are then collected. Since, with exception of the most simple substituents, the charges on the ortho- and meta- carbon pairs are not symmetrical, they are averaged. In the training process the 4 atomic charges: C_1 (atom to which the actual substituent is attached) and C_o, C_m and C_p (charges on carbons in ortho-, meta- and para- positions) were then correlated with the experimental Hammett constants. The best correlations for σ_{meta} and σ_{para} were selected using the leave-one-out crossvalidation.

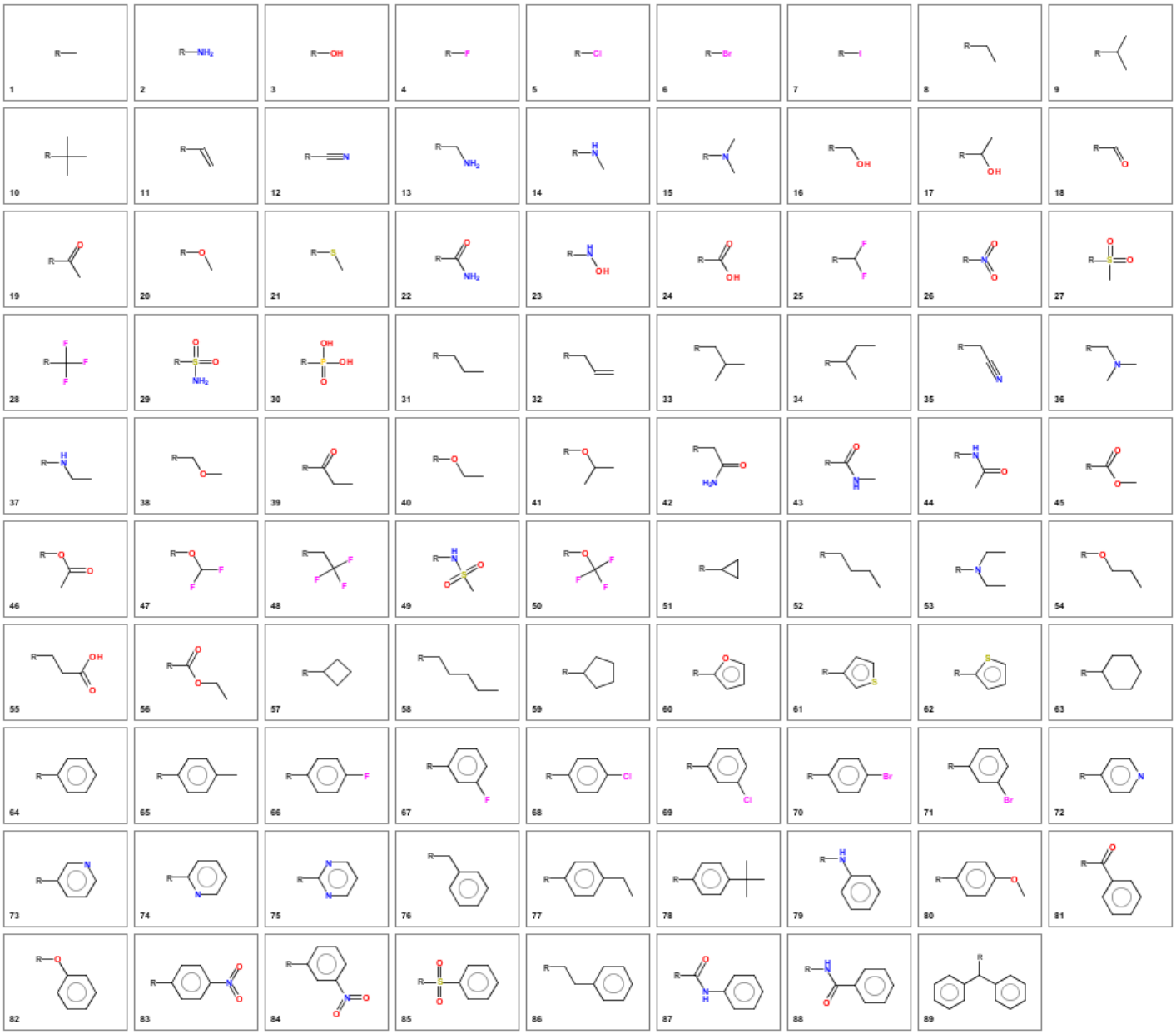


Figure 1. 89 substituents used in this study.

Results and Discussion

The procedure described in the previous section provided good correlation between atomic charges on the phenyl group attached to the substituent and the Hammett σ_{meta} and σ_{para} constants (see also Fig. 2).

σ_{meta} = 1.9381 +0.9273 * q_1 +43.9107 * q_m +8.6174 * q_p

n = 89, r^2_{cv} = 0.873, r^2 = 0.889, avrg. abs. error = 0.053

σ_{para} = 2.0849 +0.2074 * q_1 +28.4679 * q_m +28.9006 * q_p

n = 89, r^2_{cv} = 0.915, r^2 = 0.926, avrg. abs. error = 0.068

The crossvalidation procedure suggested the equations with 3 parameters as the best - containing the charges on the atom to which the substituent is attached (C_1) and on the carbons in the meta and para positions of the phenyl (C_m and C_p). These equations shown above may be used to calculate parameters characterizing the electron-donating and -withdrawing power for substituents, for which the experimental Hammett sigma constants are not available. We applied this method to calculate Hammett sigmas for substituents with up to 15 atoms present in at least 10 bioactive molecules in the ChEMBL database, altogether for about 35,000 substituents. The results, together with the calculated substituent hydrophobicity parameter (Hansch π parameter) are shown on Figure 3. This figure is actually a Craig plot based on *in silico* calculated descriptors.

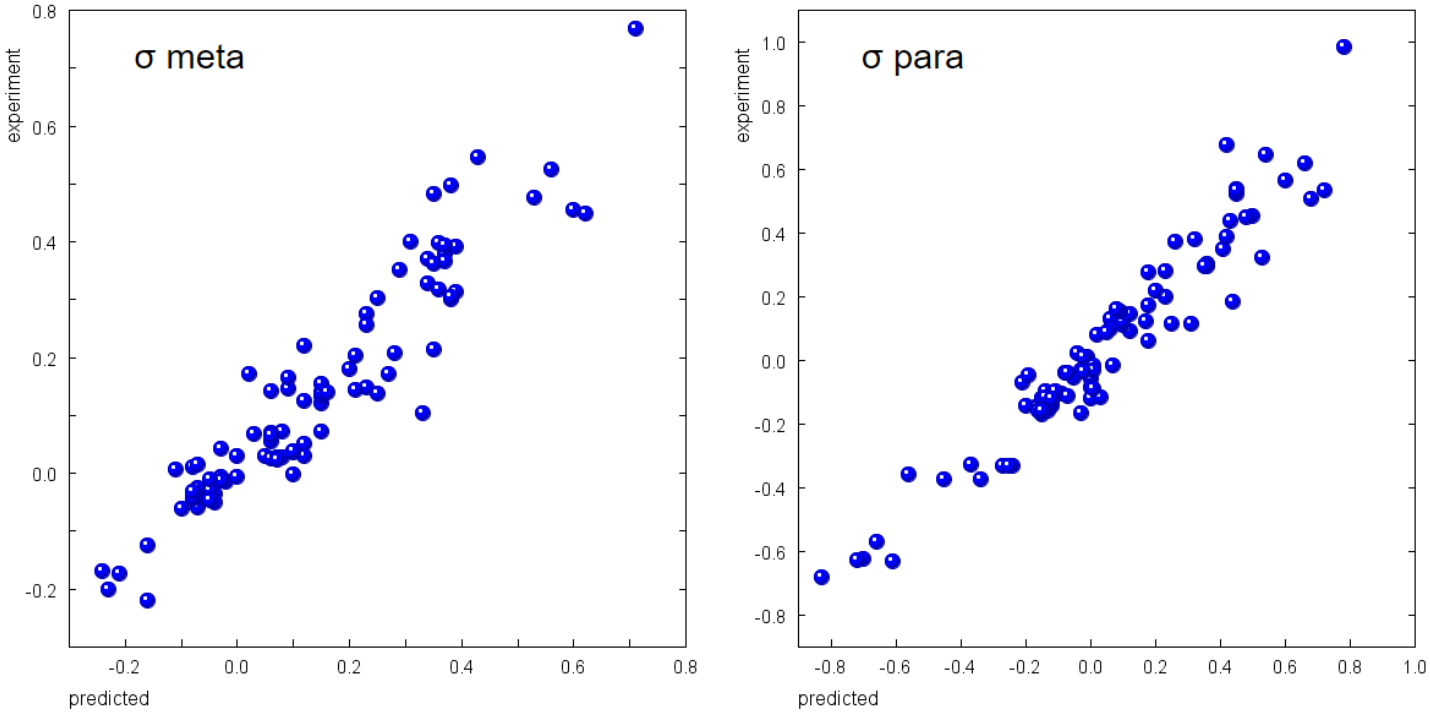


Figure 2. Correlation between experimental and calculated sigma constants for 89 substituents.

Another advantage of the updated workflow is the fact that it allows estimation of Hammett sigma constants also for inorganic and organometallic substituents, since the xtb method used to calculate atomic charges contains parameters for all elements up to radon. To test this procedure also for non-organic substituents the sigma constants for a set of representative inorganic and organometallic substituents for which experimental sigma constants were available [8] were calculated, the results being shown in Table 2. The agreement with experiment is satisfactory, even for complex organometallic structures like ferrocenyl or benzene-chromium-tricarbonyl-yl. The average absolute error for all values listed in Table 2 is 0.086, only slightly worse than that for simple organic substituents. One should also keep in mind that the listed values of experimental sigma constants for such complex structures are less reliable than for those of simple organic substituents.

In the case of organometallic structures like ferrocene or benzene-chromium-tricarbonyl it was not possible to use a simple workflow that was used for organic substituents, i.e. automatically generate 3D structure starting from SMILES, run xtb and then extract atomic charges by a Python script. In such cases it was necessary to manually build the complexes (this was done using the Maestro software [19]), optimize them by xtb and again manually extract atomic charges for calculation of substituent constants.

Table 1. SMILES codes and experimental and calculated σ_{meta} and σ_{para} values for 89 substituents.

n	SMILES	σ_{meta} exp.	σ_{meta} calc.	σ_{para} exp.	σ_{para} calc.	n	SMILES	σ_{meta} exp.	σ_{meta} calc.	σ_{para} exp.	σ_{para} calc.
1	[R]C	-0.07	-0.03	-0.17	-0.15	46	[R]OC(C)=O	0.39	0.31	0.31	0.12
2	[R]N	-0.16	-0.12	-0.66	-0.57	47	[R]OC(F)F	0.31	0.40	0.18	0.17
3	[R]O	0.12	0.05	-0.37	-0.33	48	[R]CC(F)(F)F	0.12	0.22	0.09	0.15
4	[R]F	0.34	0.37	0.06	0.13	49	[R]NS(C)(=O)=O	0.2	0.18	0.03	-0.11
5	[R]Cl	0.37	0.38	0.23	0.20	50	[R]OC(F)(F)F	0.38	0.50	0.35	0.30
6	[R]Br	0.39	0.39	0.23	0.28	51	[R]C1CC1	-0.07	0.01	-0.21	-0.07
7	[R]I	0.35	0.36	0.18	0.28	52	[R]CCCC	-0.08	-0.04	-0.16	-0.16
8	[R]CC	-0.07	-0.06	-0.15	-0.17	53	[R]N(CC)CC	-0.23	-0.20	-0.72	-0.63
9	[R]C(C)C	-0.04	-0.04	-0.15	-0.12	54	[R]OCCC	0.10	0.04	-0.25	-0.33
10	[R]C(C)(C)C	-0.10	-0.06	-0.20	-0.14	55	[R]CCC(O)=O	-0.03	-0.01	-0.07	-0.11
11	[R]C=C	0.06	0.07	-0.04	0.02	56	[R]C(=O)OCC	0.37	0.39	0.45	0.53
12	[R]C#N	0.56	0.53	0.66	0.62	57	[R]C1CCC1	-0.05	-0.01	-0.14	-0.11
13	[R]CN	-0.03	-0.01	-0.11	-0.10	58	[R]CCCCC	-0.08	-0.04	-0.15	-0.16
14	[R]NC	-0.21	-0.17	-0.70	-0.62	59	[R]C1CCCC1	-0.05	-0.05	-0.14	-0.13
15	[R]N(C)C	-0.16	-0.22	-0.83	-0.68	60	[R]c1ccco1	0.06	0.14	0.02	0.08
16	[R]CO	0.00	0.03	0.00	-0.06	61	[R]c1ccsc1	0.03	0.07	-0.02	0.01
17	[R]C(C)O	0.08	0.03	-0.07	-0.04	62	[R]c1cccs1	0.09	0.15	0.05	0.09
18	[R]C=O	0.35	0.48	0.42	0.68	63	[R]C1CCCCC1	-0.05	-0.03	-0.15	-0.12
19	[R]C(C)=O	0.38	0.30	0.50	0.46	64	[R]c1ccccc1	0.06	0.06	-0.01	0.01
20	[R]OC	0.12	0.03	-0.27	-0.33	65	[R]c1ccc(C)cc1	0.06	0.03	-0.03	-0.03
21	[R]SC	0.15	0.07	0.00	-0.12	66	[R]c1ccc(F)cc1	0.12	0.12	0.06	0.10
22	[R]C(N)=O	0.28	0.21	0.36	0.30	67	[R]c1cccc(F)c1	0.15	0.14	0.01	0.12
23	[R]NO	-0.04	-0.05	-0.34	-0.37	68	[R]c1ccc(Cl)cc1	0.15	0.12	0.12	0.09
24	[R]C(O)=O	0.37	0.37	0.45	0.52	69	[R]c1cccc(Cl)c1	0.15	0.13	0.10	0.11
25	[R]C(F)F	0.29	0.35	0.32	0.38	70	[R]c1ccc(Br)cc1	0.15	0.15	0.12	0.14
26	[R]N(=O)=O	0.71	0.77	0.78	0.98	71	[R]c1cccc(Br)c1	0.09	0.17	0.08	0.16
27	[R]S(C)(=O)=O	0.60	0.46	0.72	0.53	72	[R]c1ccncc1	0.27	0.17	0.44	0.18
28	[R]C(F)(F)F	0.43	0.55	0.54	0.65	73	[R]c1cccnc1	0.23	0.15	0.25	0.12
29	[R]S(N)(=O)=O	0.53	0.48	0.60	0.57	74	[R]c1ccccn1	0.33	0.10	0.17	0.12
30	[R]P(O)(O)=O	0.36	0.32	0.42	0.39	75	[R]c1ncccn1	0.23	0.26	0.53	0.32
31	[R]CCC	-0.06	-0.04	-0.13	-0.16	76	[R]Cc1ccccc1	-0.08	0.01	-0.09	-0.10
32	[R]CC=C	-0.11	0.01	-0.14	-0.10	77	[R]c1ccc(CC)cc1	0.07	0.02	-0.02	-0.04
33	[R]CC(C)C	-0.07	-0.04	-0.12	-0.14	78	[R]c1ccc(C(C)(C)C)cc1	0.07	0.03	0.01	-0.03
34	[R]C(C)CC	-0.08	-0.03	-0.12	-0.12	79	[R]Nc1ccccc1	-0.02	-0.01	-0.56	-0.36
35	[R]CC#N	0.16	0.14	0.18	0.06	80	[R]c1ccc(OC)cc1	0.05	0.03	-0.08	-0.04
36	[R]CN(C)C	0.00	-0.01	0.01	-0.09	81	[R]C(=O)c1ccccc1	0.34	0.33	0.43	0.44
37	[R]NCC	-0.24	-0.17	-0.61	-0.63	82	[R]Oc1ccccc1	0.25	0.14	-0.03	-0.17
38	[R]COC	0.08	0.07	0.01	-0.01	83	[R]c1ccc(N(=O)=O)cc1	0.25	0.30	0.26	0.37
39	[R]C(=O)CC	0.38	0.31	0.48	0.45	84	[R]c1cccc(N(=O)=O)c1	0.21	0.20	0.20	0.22
40	[R]OCC	0.10	0.04	-0.24	-0.33	85	[R]S(=O)(=O)c1ccccc1	0.62	0.45	0.68	0.51
41	[R]OC(C)C	0.10	0.00	-0.45	-0.37	86	[R]CCc1ccccc1	-0.07	-0.03	-0.12	-0.14
42	[R]CC(N)=O	0.06	0.07	0.07	-0.01	87	[R]C(=O)Nc1ccccc1	0.23	0.28	0.41	0.35
43	[R]C(=O)NC	0.35	0.21	0.36	0.30	88	[R]NC(=O)c1ccccc1	0.02	0.17	-0.19	-0.04
44	[R]NC(C)=O	0.21	0.14	0.00	-0.09	89	[R]C(c1ccccc1)c2ccccc2	-0.03	0.04	-0.05	-0.05
45	[R]C(=O)OC	0.36	0.40	0.45	0.54						

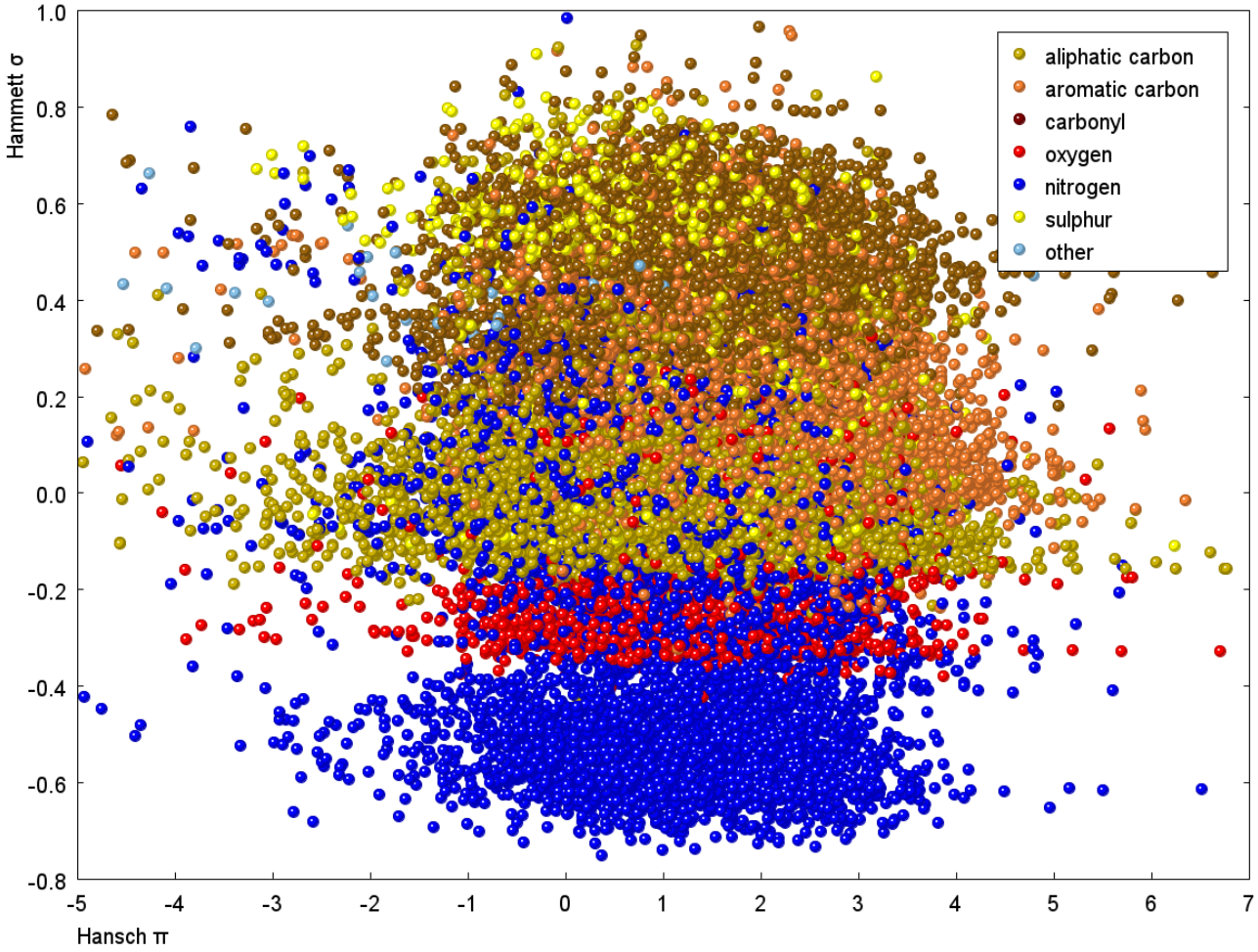


Figure 3. Plot of calculated Hammett σ_{para} parameters *vs.* calculated hydrophobicity π parameters (Craig plot) for 35,000 substituents. The substituents are color coded according to their root atom.

Table 2. Comparison of experimental [8] and calculated Hammett sigma constants for representative inorganic and organometallic substituents.

Substituent	**σ_{meta} exp.**	**σ_{meta} calc.**	**σ_{para} exp.**	**σ_{para} calc.**
ferrocenyl	-0.15	-0.06	-0.18	-0.20
benzene-Cr(CO)3-yl	0.29	0.23	0.14	0.21
GePh3	0.05	0.04	0.08	0.07
GeCl3	0.71	0.54	0.79	0.59
GeF3	0.85	0.93	0.97	0.99
GeMe3	0.00	-0.09	0.00	-0.06
SnMe3	0.00	-0.05	0.00	-0.03
AsPh2	0.03	0.01	0.09	-0.01
AsEt2	0.22	-0.08	0.00	-0.10
BF2	0.32	0.38	0.48	0.62
B(OH)2	-0.01	-0.10	0.12	-0.01

Web tool for calculating substituents descriptors

To offer an opportunity to get calculated Hammett sigma constants to the broad cheminformatics community a web tool providing this functionality was developed. A query substituent may be entered with help of the JSME JavaScript molecule editor [20]. The JSME works also on handheld devices, therefore this tool may be used also from tablets and smartphones. The query SMILES is sent to the server, where the actual calculations are performed using a Python-based workflow. The calculated descriptors are then returned and displayed. In addition to the Hammett sigma descriptors also a substituent hydrophobicity parameter (Hansch pi) [21] is provided. The more detailed information about the web tool and the usage instructions are available directly online on the About page. The web tool is freely available at https://bit.ly/getsigmas.

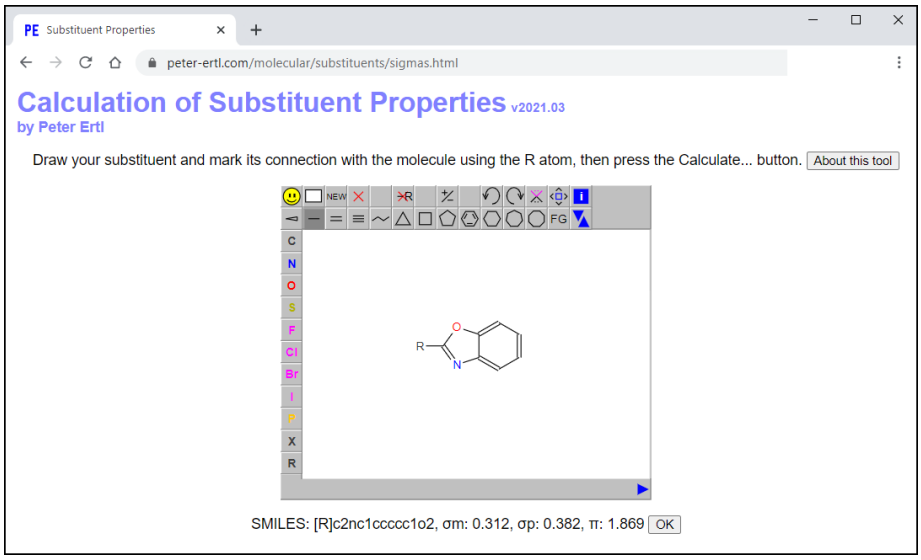


Figure 4. Web tool for calculating substituent properties. The calculated properties Hammett σ_{meta} and σ_{para} parameters, and Hansch hydrophobicity π parameter are displayed below the editor window together with the substituent SMILES.

Conclusions

A method for calculating descriptors characterising electron-donating and electron-withdrawing power of substituents compatible with the Hammett sigma constants is described here. The method provides values that are within 0.06 - 0.07 of the experimental Hammett sigmas. The calculations are possible not only for common organic, but also for inorganic and organometallic substituents. A web tool that allows easy calculation of these substituent descriptors is also made available. The author hopes that this service will be useful as a source of descriptors for QSAR studies and will help to better understand the effect of substituents on pKa, reactivity and biological activity.

Acknowledgements

I want to thank Prof. Grimme for releasing his excellent xtb program to the public as well as to Molinspiration Cheminformatics for allowing me to use their Galaxy SMILES to 3D converter and the mib cheminformatics engine to calculate substituent hydrophobicity.

References

[1] C. Hansch, P. P. Maloney, T. Fujita, and R. M. Muir, “Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients,” *Nature*, vol. 194, no. 4824, Art. no. 4824, Apr. 1962, doi: 10.1038/194178b0.

[2] C. Hansch and T. Fujita, “p-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure | Journal of the American Chemical Society,” *J Am Chem Soc*, vol. 86, no. 8, pp. 1616–1626.

[3] C. Hansch, A. Leo, and D. Hoekman, *Exploring QSAR. Vol. 1: Fundamentals and applications in chemistry and biology*. Washington, DC: American Chemical Society, 1995.

[4] P. N. Craig, “Interdependence between physical parameters and selection of substituent groups for correlation studies,” *J Med Chem*, vol. 14, pp. 680–684, 1971, doi: 10.1021/jm00290a004.

[5] P. Ertl, “Craig plot 2.0: an interactive navigation in the substituent bioisosteric space,” *J. Cheminformatics*, vol. 12, no. 1, p. 8, Jan. 2020, doi: 10.1186/s13321-020-0412-1.

[6] J. G. Topliss, “Utilization of operational schemes for analog synthesis in drug design,” *J Med Chem*, vol. 15, pp. 1006–1011, 1972, doi: 10.1021/jm00280a001.

[7] D. Mendez *et al.*, “ChEMBL: towards direct deposition of bioassay data,” *Nucleic Acids Res.*, vol. 47, no. D1, pp. D930–D940, Jan. 2019, doi: 10.1093/nar/gky1075.

[8] C. Hansch, A. Leo, and R. W. Taft, “A Survey of Hammett Substituent Constants and Resonance and Field Parameters,” *Chem. Rev.*, vol. 91, pp. 165–195, 1991.

[9] P. Ertl, “Simple Quantum Chemical Parameters as an Alternative to the Hammett Sigma Constants in QSAR Studies,” *Quant. Struct.-Act. Relatsh.*, vol. 16, no. 5, pp. 377–382, 1997, doi: https://doi.org/10.1002/qsar.19970160505.

[10] P. Ertl, “World Wide Web-based system for the calculation of substituent parameters and substituent similarity searches,” *J. Mol. Graph. Model.*, vol. 16, no. 1, pp. 11–13, Feb. 1998, doi: 10.1016/S1093-3263(98)00012-6.

[11] P. Ertl, “Cheminformatics Analysis of Organic Substituents: Identification of the Most Common Substituents, Calculation of Substituent Properties, and Automatic Identification of Drug-like Bioisosteric Groups,” *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 2, pp. 374–380, Mar. 2003, doi: 10.1021/ci0255782.

[12] P. Ertl and R. Lewis, “IADE: a system for intelligent automatic design of bioisosteric analogs,” *J. Comput. Aided Mol. Des.*, vol. 26, no. 11, pp. 1207–1215, Nov. 2012, doi: 10.1007/s10822-012-9609-3.

[13] P. Genix, H. Jullien, and R. L. Goas, “Estimation of Hammett sigma constants from calculated atomic charges using partial least squares regression,” *J. Chemom.*, vol. 10, no. 5–6, pp. 631–636, 1996, doi: https://doi.org/10.1002/(SICI)1099-128X(199609)10:5/6<631::AID-CEM451>3.0.CO;2-Z.

[14] K. C. Gross, P. G. Seybold, Z. Peralta-Inga, J. S. Murray, and P. Politzer, “Comparison of Quantum Chemical Parameters and Hammett Constants in Correlating pKa Values of Substituted Anilines,” *J. Org. Chem.*, vol. 66, no. 21, pp. 6919–6925, Oct. 2001, doi: 10.1021/jo010234g.

[15] Y. Takahata and D. P. Chong, “Estimation of Hammett sigma constants of substituted benzenes through accurate density-functional calculation of core-electron binding energy shifts,” *Int. J. Quantum Chem.*, vol. 103, no. 5, pp. 509–515, 2005, doi: https://doi.org/10.1002/qua.20533.

[16] F. Sessa, M. Olsson, F. Söderberg, F. Wang, and M. Rahm, “Experimental Quantum Chemistry: a Hammett-inspired Fingerprinting of Substituent Effects,” *ChemPhysChem*, vol. n/a, no. n/a, doi: https://doi.org/10.1002/cphc.202001053.

[17] C. Bannwarth, S. Ehlert, and S. Grimme, “GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions,” *J. Chem. Theory Comput.*, vol. 15, no. 3, pp. 1652–1671, Mar. 2019, doi: 10.1021/acs.jctc.8b01176.

[18] *Galaxy v2020.10, Molinspiration Cheminformatics, Slovensky Grob, Slovakia*.

[19] *Schrödinger Release 2020-4: Maestro, Schrödinger, LLC, New York, NY, 2020.* .

[20] B. Bienfait and P. Ertl, “JSME: a free molecule editor in JavaScript,” *J. Cheminformatics*, vol. 5, no. 1, p. 24, May 2013, doi: 10.1186/1758-2946-5-24.

[21] *mib v2021.01, Molinspiration Cheminformatics, Slovensky Grob, Slovakia*.