Software News and Update

PaDEL-Descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints

CHUN WEI YAP
Department of Pharmacy, Pharmaceutical Data Exploration Laboratory, National University of Singapore, Singapore

Received 17 May 2010; Revised 22 August 2010; Accepted 12 October 2010
DOI 10.1002/jcc.21707
Published online 17 December 2010 in Wiley Online Library (wileyonlinelibrary.com).

Introduction: PaDEL-Descriptor is a software for calculating molecular descriptors and fingerprints. The software currently calculates 797 descriptors (663 1D, 2D descriptors, and 134 3D descriptors) and 10 types of fingerprints. These descriptors and fingerprints are calculated mainly using The Chemistry Development Kit. Some additional descriptors and fingerprints were added, which include atom type electrotopological state descriptors, McGowan volume, molecular linear free energy relation descriptors, ring counts, count of chemical substructures identified by Laggner, and binary fingerprints and count of chemical substructures identified by Klekota and Roth.
Methods: PaDEL-Descriptor was developed using the Java language and consists of a library component and an interface component. The library component allows it to be easily integrated into quantitative structure activity relationship software to provide the descriptor calculation feature while the interface component allows it to be used as a standalone software. The software uses a Master/Worker pattern to take advantage of the multiple CPU cores that are present in most modern computers to speed up calculations of molecular descriptors.
Results: The software has several advantages over existing standalone molecular descriptor calculation software. It is free and open source, has both graphical user interface and command line interfaces, can work on all major platforms (Windows, Linux, MacOS), supports more than 90 different molecular file formats, and is multithreaded.
Conclusion: PaDEL-Descriptor is a useful addition to the currently available molecular descriptor calculation software. The software can be downloaded at http://padel.nus.edu.sg/software/padeldescriptor.

 J Comput Chem 32: 1466–1474, 2011

Key words: molecular descriptor; open source; software

Introduction

A molecular descriptor "is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into an useful number or the result of some standardized experiment".[1] Molecular descriptors are calculated for chemical compounds and used to develop quantitative structure activity relationship (QSAR) models for predicting the biological activities of novel compounds.

There are currently a large number of molecular descriptors, which can be classified into three broad categories: 1-, 2-, and 3D descriptors that encode chemical composition, topology, and 3D shape and functionality, respectively.[2] A descriptor can be simple, like molecular volume, which encode only one feature of a compound, or can be complex, like 3D-MoRSE, which encode multiple physicochemical and structural properties of a compound. A useful reference for molecular descriptors is the "Molecular Descriptors for Chemoinformatics" by Todeschini and Consonni.[2] Table 1 lists some of the common types of molecular descriptors.

Currently, there are a number of commercial and freely available software for calculation of molecular descriptors. Some of these were developed mainly or solely for the calculation of molecular descriptors (Table 2), while others were QSAR software which had descriptor calculation as one of their features (e.g., CODESSA Pro, Discovery Studio, Sybyl, MOE). In addition, there are some open source libraries, such as JOELib,[48] Chemistry Development Kit (CDK)[49] and Chemical Descriptors Library,[50] which provide molecular descriptor calculation functionality.

A good descriptor calculation software should possess most of the following features:

Correspondence to: C. W. Yap; e-mail: phayapc@nus.edu.sg

Additional Supporting Information may be found in the online version of this article.

Table 1. Common Molecular Descriptors

Constitutional
Functional groups
Molecular weight
Simple counts e.g., number of atoms, bonds, rings
Topological
Atom-pairs[5]
Balaban index[6]
BCUT[7]
Information content indices[8]
Kappa shape indices[9]
Kier and Hall connectivity indices[10]
Kier flexibility index[11]
Kier shape indices[11]
Molecular walk counts[12]
Randic indices[13]
Wiener index[14]
Geometric
Gravitation index[20]
Molecular surface area
Molecular volume[21]
Shadow indices[22]
Solvent accessible molecular surface area
Electrostatic
Charged polar surface area[33]
Galvez topological charge indices[34]
Hydrogen bonding capacities
Maximum and minimum partial charges[35]
Molecular polarizabilities[36]
Fingerprints
Daylight[37]
MDL keys[38]
UNITY[39]
Hydrophobic
Aromaticity indices[3]
Hansch substituent constant[4]
Log D
Log P
Steric
Charton steric parameter[15]
Molar refractivity[16]
Parachor[17]
Taft steric parameter[18]
Quantum chemical[19]
Charges
HOMO and LUMO energies
Orbital electron densities
Superdelocalizabilities
Atom-atom polarizabilities
Molecular polarizabilites
Dipole moments and polarity indices
Energies
Combination
3D-MoRSE[23]
Electrotopological state indices[24]
GETAWAY
LSER[26]
MolSurf[27]
Moreau-Broto topological autocorrelation[28]
Randic molecular profiles[29]
RDF[30]
VolSurf[31]
WHIM[32]

1. Free or cheap to purchase so that it is easily available to researchers.
2. Open source so that researchers could add in their own descriptor calculation algorithms.
3. Has a graphical user interface (GUI) for easy usage and a command line version to allow the software to run in computer clusters through a software job scheduler.
4. Able to work on multiple platforms such as Windows, Linux, and MacOS.
5. Accepts multiple molecular file formats.
6. Able to calculate many types of descriptors.

It can be seen from Table 2 that none of the currently available descriptor calculation software possesses all these features.

Methods

Software Design

PaDEL-Descriptor was programmed using the Java language and consisted of two components: a library component and an interface component. The library component is the actual workhorse of the system and is self-contained. This means that it can function without the interface component and can be easily integrated into other QSAR software to provide the descriptor calculation feature. The library component provides wrapper classes around the 43 molecular descriptor algorithms and seven fingerprint algorithms that were implemented in CDK. The library component also implemented four molecular descriptor and three fingerprint algorithms which were not found in CDK. These include atom type electrotopological state descriptors, McGowan volume, molecular linear free energy relation descriptors, ring counts, count of chemical substructures identified by Laggner,[51] and binary fingerprints and count of chemical substructures identified by Klekota and Roth.[52] These algorithms were implemented by extending relevant classes from CDK so that they would be compatible with other CDK descriptor algorithms and could be added to CDK in the future. Information on the molecular descriptors that are calculated by PaDEL-Descriptor is given in Table 3 and the Supporting Information.

Calculation of molecular descriptors is performed in parallel by using a Master/Worker pattern, which consists of a Master thread and one or more worker threads. The Master thread starts the calculation process by reading molecular files and creates a job description for each molecule. A job description consists of the name and structure of the molecule, preprocessing tasks that are to be done before calculation of descriptors (e.g., remove salt, add hydrogens, and/or convert to 3D), and the types of descriptors and fingerprints to calculate. The jobs are added to a shared job queue and each worker thread will retrieve a job from the shared queue and calculates the descriptors and fingerprints for the molecule specified in the job description. The calculated descriptors and fingerprints are then placed in a shared results queue where it will be retrieved by the Master thread to be stored in a comma-separated value (CSV) file. The Master thread also provides functions to set the type of preprocessing tasks to be done, the types of descriptors and fingerprints to calculate, and the number of worker threads. The advantage of the Master/Worker pattern is that it make efficient use of the multiple CPU cores that are present in most

Table 2. Molecular Descriptor Calculation Software

Software	Number of types of descriptors[a] (fingerprints)	Number of descriptors[b]	Advantages	Disadvantages	Additional remarks
BlueDesc[40]	36	174	Free Open source Multiple platform (Windows, Linux, MacOS)	No GUI Accepts only MDL SDF	Uses JOELib2 and CDK to calculate descriptors
CDK Descriptor Calculator GUI v1.0.2[41]	39 (6)	268[c]	Free Open source GUI and command line interfaces Multiple platform (Windows, Linux, MacOS)	Accepts only MDL SDF and SMILES	Uses CDK to read molecular files and calculate most of the descriptors and fingerprints
DRAGON v5.5[42]	127 (2)	1664[d]	GUI and command line interfaces Multiple platform (Windows, Linux) Accept multiple file formats (SMILES, MDL MOL/SDF, Sybyl MOL2, HyperChem, MacroModel)	Commercial Closed source	1800 euro for a permanent personal academic license of DRAGON Plus
MODEL[43]	75	3780	Free Accepts multiple file formats (PDB, MDL MOL, Sybyl MOL2, COR)	Closed source	Web-based
Molconn-Z v4.10[44]	41	About 1063[e]	Multiple platform (Windows, Linux, MacOS) Accept multiple file formats (SMILES, MDL SDF and Sybyl MOL2).	No GUI Commercial Closed source	US$750 for a single user academic license
Mold2[45]	73	777	Free	Closed source No GUI Single platform (Windows) Accepts only MDL SDF	
PreADMET Descriptor[46]	34	1044	GUI	Commercial Closed source Single platform (Windows) Accepts only MDL MOL/SDF	US$2500 for a perpetual single user academic license
VolSurf v4.1.4[47]	18	94	GUI and command line interfaces Accept multiple file formats (MDL SDF, Sybyl mol2, GRID kout, Multi mol2)	Commercial Closed source Single platform (Linux)	500 GBP for a unlimited user annual academic license
PaDEL-Descriptor v2.0	43 (10)	797	Free Open source GUI with command line interfaces Multiple platform (Windows, Linux, MacOS) Accepts multiple file formats (>90 formats)		Uses CDK to read molecular files and calculate most of the descriptors and fingerprints. Employs Java Web Start technology

[a]Approximate numbers based on theoretical considerations. Number in brackets indicates the number of types of fingerprints. May not correspond to the number provided by the software authors.
[b]Does not include the number of fingerprints.
[c]Number does not include Amino acid count (not relevant for chemical compounds) and Ionization potential (could not be calculated).
[d]Number differs from the 3224 reported by the software because their number includes 1560 fingerprint bits.
[e]Number excludes 231 unused descriptor items and assumes 15 descriptor items each in records 77–79.

Table 3. Molecular Descriptors Calculated by PaDEL-Descriptor

Descriptor class	Descriptor type[a]	Number of descriptors	Calculation speed (mol/s)[b]
2D	ALOGP	3	1084
	APol	1	24,738
	Aromatic atoms count	1	16,878
	Aromatic bonds count	1	16,336
	Atom count	13	2127
	Autocorrelation (charge)	5	6215
	Autocorrelation (mass)	5	777
	Autocorrelation (polarizability)	5	741
	BCUT	6	653
	Bond count	5	6014
	BPol	1	23,060
	Carbon types	9	20,327
	Chi chain	10	310
	Chi cluster	8	439
	Chi path	16	310
	Chi path cluster	6	347
	Eccentric connectivity index	1	11,611
	Atom type electrotopological state	482	289
	Fragment complexity	1	27,400
	Hbond acceptor count	1	16,126
	Hbond donor count	1	16,384
	Kappa shape indices	3	2413
	Largest chain	1	10,088
	Largest Pi system	1	13,804
	Longest aliphatic chain	1	10,233
	Mannhold LogP	1	16,328
	McGowan volume	1	546
	Molecular distance edge	19	2194
	Molecular linear free energy relation	6	300
	Petitjean number	1	10,131
	Ring count	34	1757
	Rotatable bonds count	1	11,471
	Rule of five	1	807
	Topological polar surface area	1	3535
	Vertex adjacency information (magnitude)	1	26,160
	Weight	1	24,532
	Weighted path	5	513
	Wiener numbers	2	10,755
	XlogP	1	910
	Zagreb index	1	23,593
3D	Charged partial surface area	29	309
	Gravitational index	9	10,012
	Length over breadth	2	8459
	Moment of inertia	7	9894
	Petitjean shape index	2	8213
	WHIM (atomic masses)	17	8229
	WHIM (atomic polarizabilities)	17	8150
	WHIM (Mulliken atomic electronegativities)	17	8290
	WHIM (unit weights)	17	8396
	WHIM (van der Waals volumes)	17	8072
Fingerprint	CDK fingerprint	1024	203
	CDK extended fingerprint	1024	189
	CDK graph only fingerprint	1024	223
	Estate fingerprint	79	276
	MACCS fingerprint	166	170
	Pubchem fingerprint	881	56
	Substructure fingerprint	307	116
	Substructure fingerprint count	307	113
	Klekota-Roth fingerprint	4860	12
	Klekota-Roth fingerprint count	4860	12

[a]Based on the types of descriptors and fingerprints that can be selected for calculation by PaDEL-Descriptor. It is not equal to the number of types of descriptors given in Table 2 because descriptors like the 3 Autocorrelation descriptors, 5 WHIM descriptors, and 2 HBond descriptors were grouped together and counted as 3 types instead of 10 types in Table 2.

[b]On an Intel® Xeon® E5530 2.40 GHz processor. This is the raw speed of the descriptor calculation algorithms using one worker thread. It excludes the reading time of molecular files and pre-processing time of molecules.

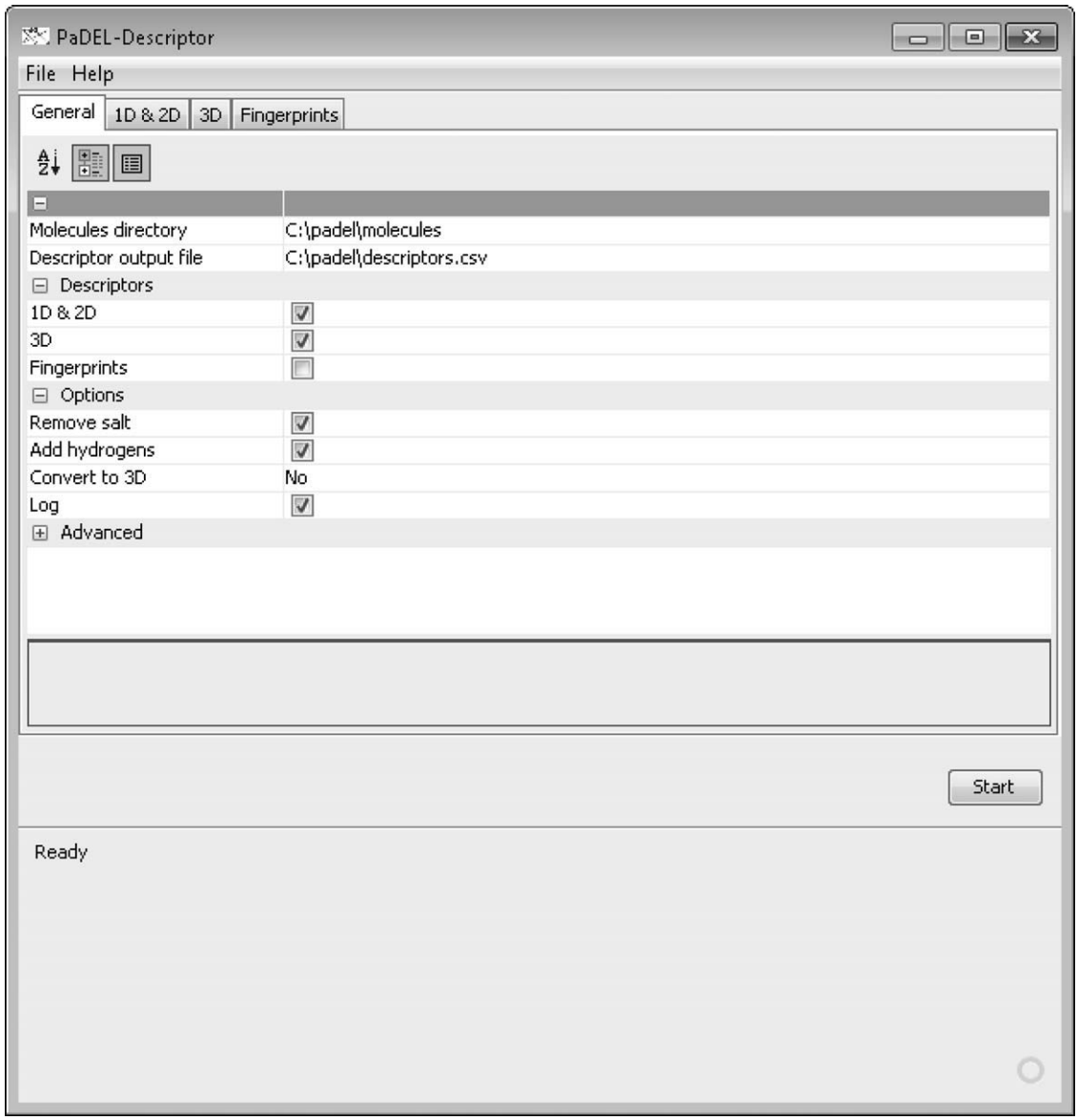

Figure 1. GUI for PaDEL-Descriptor.

modern computers to speed up the calculation of molecular descriptors.

The interface component provides both GUI and command line interfaces for operating the software. Both interfaces allow the user to set the various options that are required by the library component. The GUI, which is shown in Figure 1, was implemented using Swing and l2fprod library. It provides an easy to use interface for users to set the various options and select the individual types of descriptors and fingerprints to calculate. The options can be saved to a configuration file, which can be used to configure the software automatically or manually when the software is run the next time. The selected types of descriptors and fingerprints can also be saved to a XML file, which can be used to ease the reselection of these descriptors and fingerprints.

The command line interface (Fig. 2) was implemented using the Apache Commons CLI library. It is provided to allow the software to run in computer clusters through a software job scheduler. The command line interface is able to use the configuration file and XML file that are generated by the GUI to set the required options and select types of descriptors and fingerprints to be calculated, respectively.

Descriptors Calculation Speed Experiments

All experiments for determining the speed of descriptor calculations were performed on a Dell PowerEdge R710 system with two Intel® Xeon® E5530 2.40 GHz processors and 72 GB RAM. A total of 10,000 compounds with median molecular weight of 214 (range 72–349) were used for the descriptor

Figure 2. Command line interface for PaDEL-Descriptor.

calculations. Each experiment was repeated 30 times and the average of the total time needed to complete the calculation for the 10,000 compounds were computed.

Results and Discussion

Comparison with Other Descriptor Calculation Software

Since PaDEL-Descriptor is a dedicated software for molecular descriptor calculations, comparisons will only be made with other similar dedicated software instead of comparing it with general QSAR software with descriptor calculation features or programming libraries for descriptor calculations. PaDEL-Descriptor has several advantages over existing dedicated molecular descriptor calculation software. Firstly, it is free, open source and is licensed as public domain. Being free will increase the availability of the software to users and being open source allows users to freely inspect the code and modify it to suit their needs. This could potentially improve the detection of bugs and increase the number of features in the software. It is released as public domain so that anyone could freely use the source code without any restrictions. However, those code that uses CDK, Apache Commons CLI, and l2fprod will still be restricted by the respective licenses.

Secondly, PaDEL-Descriptor provides both GUI and command line interfaces, which are available only in CDK Descriptor Calculator GUI, DRAGON and VolSurf. Having both GUI and command line interfaces is important as the GUI will provide ease of use to those who are not IT-savvy while the command line interface caters to those who need to run the software in computer clusters through a software job scheduler.

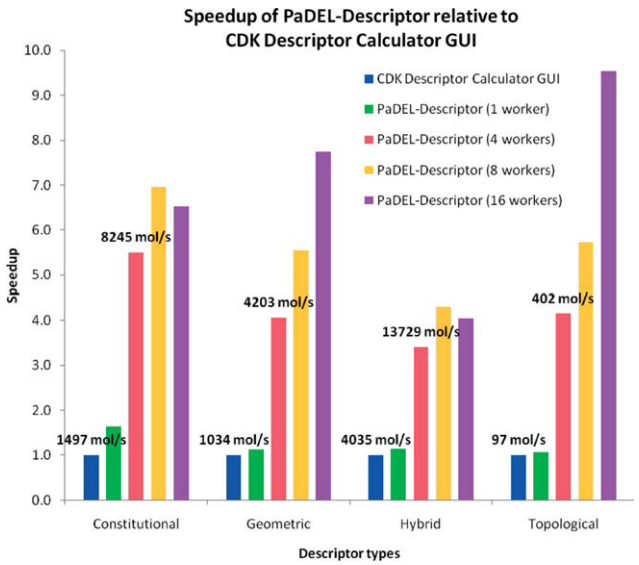

Figure 3. Speedup of PaDEL-Descriptor relative to CDK descriptor calculator GUI. The rate of calculation for CDK descriptor calculator GUI and PaDEL-Descriptor using four worker threads are indicated in the chart.

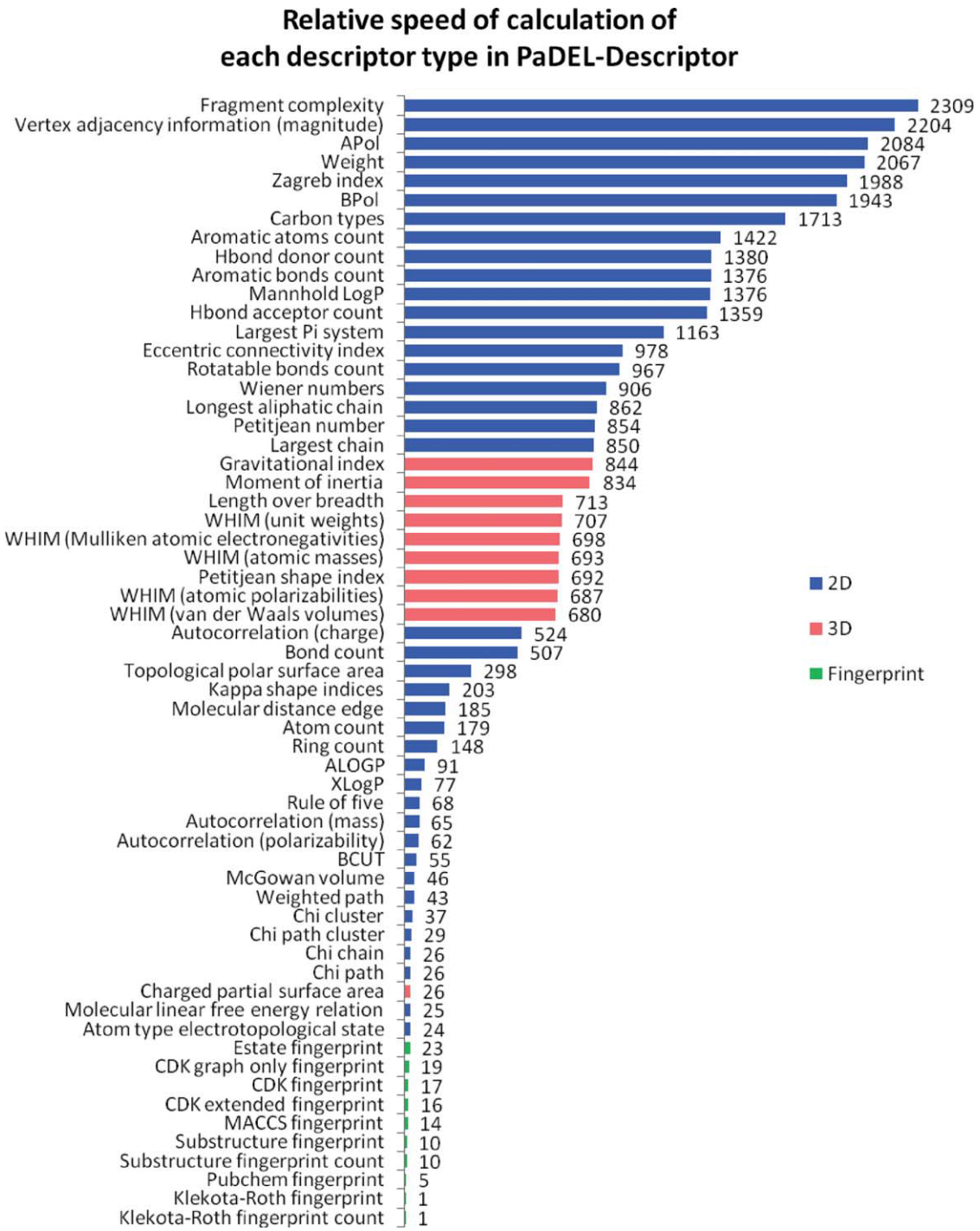

Figure 4. Relative speed of calculation of each descriptor type in PaDEL-Descriptor.

The third advantage of PaDEL-Descriptor is that it can work on any platform that supports Java. This allows it to run on the three major platforms, Windows, Linux, and MacOS, unlike DRAGON, Mold2, PreADMET Descriptor and VolSurf, which supports either one or two platforms only. PaDEL-Descriptor also employs Java Web Start technology. When users make use of this technology to run PaDEL-Descriptor, they will be assured that they are using the most current version of the software.

PaDEL-Descriptor's fourth advantage is that it is the only software that supports more than 90 different molecular file formats. In theory, CDK Descriptor Calculator GUI should also be able to do so since it uses the same CDK library. However, the author of

CDK Descriptor Calculator GUI restricted it to only MDL SDF and SMILES format. The ability to support more file formats will remove the extra conversion step that users need to do when their molecular files are not in the desired format as the descriptor calculation software. It is to be noted that all the software listed in Table 2 are able to support MDL MOL/SDF files, which is a very common file format for storing molecular files.

The last advantage of PaDEL-Descriptor is its speed, especially in multiple CPU cores environment. It is difficult to compare the descriptor calculation speed between the various software in a fair way because the types of descriptors calculated by each software are different. Thus, a relatively fairer comparison can only be made between PaDEL-Descriptor and CDK Descriptor Calculator GUI since both uses the CDK library. The speedup of PaDEL-Descriptor over CDK Descriptor Calculator GUI is shown in Figure 3. The results showed that PaDEL-Descriptor was slightly faster than CDK Descriptor Calculator GUI even when only one worker thread was used. This could be due to the Master/Worker pattern which separates the loading of molecular files and calculation of molecular descriptors into different threads, thus speeding up the calculations. Generally, the amount of speedup increases with the number of worker threads used. However, the speedup is not proportional beyond four worker threads. Hence, it is recommended that four worker threads, which corresponds to the typical number of cores that are found in modern desktops be used in PaDEL-Descriptor.

The only disadvantage of PaDEL-Descriptor is that it does not calculate as many descriptors as some software like DRAGON, MODEL, Molconn-Z, and PreADMET Descriptor. However, it can calculate 10 different types of fingerprints, which is more than these software, and future versions will add more descriptors and fingerprints to the software.

Calculation Speed for Each Descriptor Type in PaDEL-Descriptor

Table 3 and Figure 4 shows the absolute and relative calculation speed of the descriptor types in PaDEL-Descriptor. The results showed that some 2D descriptors can be calculated at a rate of greater than 20,000 molecules per second but there were also some 2D descriptors that were calculated at a rate of less than 1000 molecules per second. Most of the 3D descriptors were calculated at a rate of 8000–10,000 molecules per second. The calculation of fingerprints were the slowest, with a rate of less than 300 molecules per second. The slow calculation of some descriptors and fingerprints is due to the SMARTS pattern matching in these algorithms. Examples of these descriptors and fingerprints are molecular linear free energy relation descriptors, McGowan volume, PubChem fingerprint and Klekota-Roth fingerprint. The slowest algorithms are the Klekota-Roth fingerprint and Klekota-Roth fingerprint count because they are matching 4860 SMARTS patterns for each molecule.

Conclusion

An open source molecular descriptor calculation software, PaDEL-Descriptor, was developed and made readily available at http://padel.nus.edu.sg/software/padeldescriptor. It is a multithreaded software which make full use of the multiple CPU cores in modern desktop computers to increase the speed of calculation of molecular descriptors. The software can be used on all the major platforms (Windows, Linux, MacOS) and can be deployed using Java Web Start technology. PaDEL-Descriptor can be used as a standalone software or can be easily integrated into other QSAR software as the descriptor calculation component.

Acknowledgments

This work was supported by the National University of Singapore (NUS) start-up grant R-148-000-105-133.

References

1. Todeschini, R.; Consonni, V. Handbook of Molecular Descriptors; Wiley-VCH: Weinheim, 2000.
2. Todeschini, R.; Consonni, V. Molecular Descriptors for Chemoinformatics; Wiley-VCH: Weinheim, 2009.
3. Randic, M. Tetrahedron 1975, 31, 1477.
4. Fujita, T.; Iwasa, J.; Hansch, C. J Am Chem Soc 1964, 86, 5175.
5. Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. J Chem Inf Comput Sci 1985, 25, 64.
6. Balaban, A. T. MATCH Commun Math Comp Chem 1986, 21, 115.
7. Pearlman, R. S.; Smith, K. M. J Chem Inf Comput Sci 1999, 39, 28.
8. Basak, S. C.; Magnuson, V. R. Arzn-eim -Forsch Drug Res 1983, 33, 501.
9. Kier, L. B. Med Chem Res 1997, 7, 394.
10. Kier, L. B.; Hall, L. H. Molecular Connectivity in Structure-Activity Analysis; Wiley: New York, 1986.
11. Kier, L. B. In Computational Chemical Graph Theory, Rouvray, D. H., Ed.; Nova Science Publishers: New York, 1990; pp. 151–174.
12. Rücker, G.; Rücker, C. J Chem Inf Comput Sci 1993, 33, 683.
13. Randic, M. Chemom Intell Lab Sys 1991, 10, 213.
14. Nikolic, S.; Trinajstic, N.; Mihalic, Z. Croat Chem Acta 1995, 68, 105.
15. Charton, M. J Am Chem Soc 1975, 97, 1552.
16. Pauling, L.; Pressman, D. J Am Chem Soc 1945, 67, 1003.
17. McGowan, J. C. Nature 1963, 200, 1317.
18. Taft, R. W. J Am Chem Soc 1952, 74, 3120.
19. Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Chem Rev 1996, 96, 1027.
20. Katritzky, A. R.; Mu, L.; Lobanov, V. S.; Karelson, M. J Phys Chem 1996, 100, 10400.
21. Higo, J.; Go, N. J Comput Chem 1989, 10, 376.
22. Rohrbaugh, R.; Jurs, P. C. Anal Chim Acta 1987, 199, 99.
23. Schuur, J. H.; Setzer, P.; Gasteiger, J. J Chem Inf Comput Sci 1996, 36, 334.
24. Kier, L. B.; Hall, L. H. Molecular Structure Description: The Electrotopological State; Academic Press: San Diego, 1999.
25. Consonni, V.; Todeschini, R.; Pavan, M. J Chem Inf Comput Sci 2002, 42, 682.
26. Platts, J. A.; Butina, D.; Abraham, M. H.; Hersey, A. J Chem Inf Comput Sci 1999, 39, 835.
27. Sjoberg, P. In Computer-Assisted Lead Finding and Optimization: Current Tools for Medicinal Chemistry, van de Waterbeemd, H.; Testa, B.; Folkers, G., Eds.; VHCA: Basel, 1997; pp. 83–92.
28. Moreau, G.; Broto, P. Nouv J Chim 1980, 4, 359.
29. Randic, M. New J Chem 1995, 19, 781.
30. Hemmer, M. C.; Steinhauer, V.; Gasteiger, J. Vib Spectrosc 1999, 19, 151.
31. Cruciani, G.; Pastor, M.; Guba, W. Eur J Pharm Sci 2000, 11 (Suppl 2), S29.
32. Todeschini, R.; Lasagni, M.; Marengo, E. J Chemometr 1994, 8, 263.

33. Stanton, D. T.; Jurs, P. C. Anal Chem 1990, 62, 2323.
34. Galvez, J.; Garcia, R.; Salabert, M. T.; Soler, R. J Chem Inf Comput Sci 1994, 34, 520.
35. Kirpichenok, M. A.; Zefirov, N. S. Zh Org Khim 1987, 23, 673.
36. Dewar, M. J. S.; Stewart, J. J. P. Chem Phys Lett 1984, 111, 416.
37. Craig, A. J.; Weininger, D.; Delany, J. In Daylight Theory Manual; Daylight Chemical Information Systems, Inc., 2005. Available at http://www.daylight.com/dayhtml/doc/theory/index.html (last accessed: 15 May 2010).
38. Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. J Chem Inf Comput Sci 2002, 42, 1273.
39. Patterson, D. E.; Cramer, R. D.; Ferguson, A. M.; Clark, R. D.; Weinberger, L. E. J Med Chem 1996, 39, 3049.
40. Hinselmann, G. BlueDesc; 2008. Available at http://www.ra.cs.uni-tuebingen.de/software/bluedesc/welcome_e.html. (last accessed: 15 May 2010).
41. Guha, R. CDK Descriptor Calculator GUI; 2010. Available at http://rguha.net/code/java/cdkdesc.html. (last accessed: 15 May 2010).
42. Talete srl. DRAGON (Software for Molecular Descriptor Calculation). Version 5.5; 2007. Available at http://www.talete.mi.it/. (last accessed: 15 May 2010).
43. Li, Z. R.; Han, L. Y.; Xue, Y.; Yap, C. W.; Li, H.; Jiang, L.; Chen, Y. Z. Biotechnol Bioeng 2007, 97, 389.
44. Hall, L. H.; Kellogg, G. E.; Haney, D. N. Molconn-Z; 2002. Available at http://www.edusoft-lc.com/molconn/ (last accessed 15 May 2010)
45. Hong, H. X. Mold2. Available at http://www.fda.gov/ScienceResearch/BioinformaticsTools/Mold2/default.htm. (last accessed: 15 May 2010).
46. Bioinformatics & Molecular Design Research Center (BMDRC). PreADMET Descriptor; 2010. Available at http://preadmet.bmdrc.org/index.php?option=com_content&view=category&layout=blog&id=63&Itemid=75. (last accessed: 15 May 2010).
47. Cruciani, G.; Crivori, P.; Carrupt, P. A.; Testa, B. J Mol Struc Theochem 2000, 503, 17.
48. Wegner, J. K. JOELib/JOELib2; 2005. Available at http://www-ra.informatik.uni-tuebingen.de/software/joelib/index.html. (last accessed: 15 May 2010)
49. Steinbeck, C.; Hoppe, C.; Kuhn, S.; Floris, M.; Guha, R.; Willighagen, E. L. Curr Pharm Des 2006, 12, 2111.
50. Sykora, V. J.; Leahy, D. E. J Chem Inf Model 2008, 48, 1931.
51. Laggner, C. SMARTS Patterns for Functional Group Classification; 2009. Available at http://code.google.com/p/semanticchemistry/source/browse/wiki/InteLigand.wiki?spec=svn41&r=41. (last accessed: 15 May 2010)
52. Klekota, J.; Roth, F. P. Bioinformatics 2008, 24, 2518.