7-2015

Fast, accurate, and reliable molecular docking with QuickVina 2

Amr ALHOSSARY

Handoko Stephanus Daniel
Singapore Management University, dhandoko@smu.edu.sg

Yuguang MU

Chee-Keong KWOH

DOI: https://doi.org/10.1093/bioinformatics/btv082

Bioinformatics, 31(13), 2015, 2214–2216
doi: 10.1093/bioinformatics/btv082
Advance Access Publication Date: 24 February 2015
Applications Note

OXFORD

Structural bioinformatics

Fast, accurate, and reliable molecular docking with QuickVina 2

Amr Alhossary[1,*], Stephanus Daniel Handoko[2], Yuguang Mu[3] and Chee-Keong Kwoh[1]

[1]School of Computer Engineering, Nanyang Technological University, Singapore 639798, [2]School of Information Systems, Singapore Management University, Singapore 188065, and [3]School of Biological Sciences, Nanyang Technological University, Singapore 637551

*To whom correspondence should be addressed.

Associate Editor: Anna Tramontano

Received on August 16, 2014; revised on January 23, 2015; accepted on February 5, 2015

Abstract

Motivation: The need for efficient molecular docking tools for high-throughput screening is growing alongside the rapid growth of drug-fragment databases. AutoDock Vina ('Vina') is a widely used docking tool with parallelization for speed. QuickVina ('QVina 1') then further enhanced the speed via a heuristics, requiring high exhaustiveness. With low exhaustiveness, its accuracy was compromised. We present in this article the latest version of QuickVina ('QVina 2') that inherits both the speed of QVina 1 and the reliability of the original Vina.

Results: We tested the efficacy of QVina 2 on the core set of PDBbind 2014. With the default exhaustiveness level of Vina (*i.e.* 8), a maximum of 20.49-fold and an average of 2.30-fold acceleration with a correlation coefficient of 0.967 for the first mode and 0.911 for the sum of all modes were attained over the original Vina. A tendency for higher acceleration with increased number of rotatable bonds as the design variables was observed. On the accuracy, Vina wins over QVina 2 on 30% of the data with average energy difference of only 0.58 kcal/mol. On the same dataset, GOLD produced RMSD smaller than 2 Å on 56.9% of the data while QVina 2 attained 63.1%.

Availability and implementation: The C++ source code of QVina 2 is available at (www.qvina.org).

Contact: aalhossary@pmail.ntu.edu.sg

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

AutoDock Vina—also referred to as Vina—employs the iterated local search global optimizer to search for the minimum-energy docking conformations (Trott and Olson, 2010). Subsequently, QuickVina—referred to as QVina 1 hereinafter—enhanced Vina's computation time via heuristics that prevents unnecessary local searches. As proof of concept, QVina 1 needs high exhaustiveness level compared to the original Vina (Handoko *et al.*, 2012). Here, we present the most recent QuickVina—referred to as QVina 2—with improved reliability. QVina 2 avoids the unnecessary local searches in a similar manner to QVina 1. In contrast to QVina 1, however, QVina 2 misses fewer necessary local searches thanks to a new procedure to test if a randomized docked conformation is significant for local search. This improves QVina 2's reliability in discovering the minimum-energy docking conformation.

2 Methods

Vina explores the molecular docking search space by means of global and local optimization, in the forms of Markov chain of modified Monte Carlo algorithm with restart and BFGS method, respectively. The local search is the most time-consuming part of the optimization. QVina 1 restricts the application of local search to those docked conformation candidates deemed to be significant by the *first-order-necessary-condition* heuristics. This is enabled by keeping track of $10N$ last-assessed docked conformations in circular

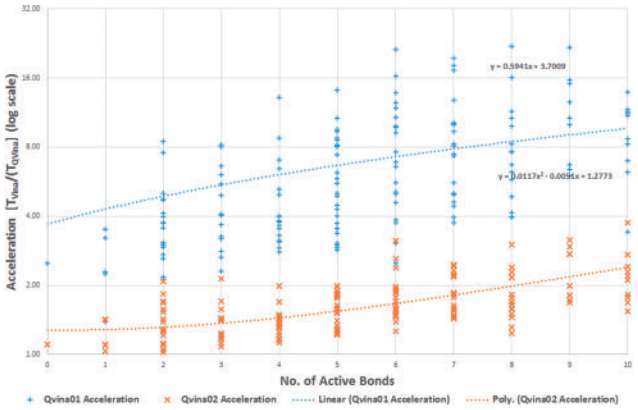

Fig. 1. Acceleration versus no. active bonds. Zoomed view in the range 0–10 Active bonds. The trend lines are drawn according to the full dataset (shown in supplementary data). QVina2 trend shows quadratic uprising tendency with increased number of active bonds

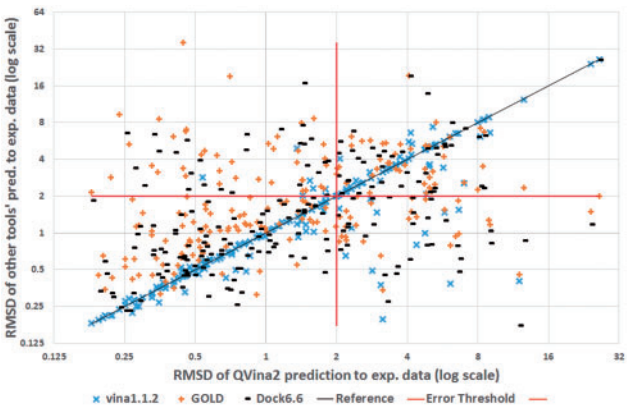

Fig. 2. Comparison of RMSD to experimental data of several tools versus QVina2 RMSD. QVina2 RMSD is on the *X* axis, while other tools' are on the *Y* axis. A value of 2 Å is considered as the prediction binary threshold on both axes

database where N is the number of the design variables. For each newly randomized candidate of docked conformation $\mathbf{p}$, as many as $2N$ nearest (in terms of Euclidean distance) neighbors are retrieved from the database and then the significance test is performed to determine if local search from $\mathbf{p}$ is necessary.

QVina 2 enhances the significance test by introducing the novel *first-order-consistency-check* heuristics. Like in QVina 1, a newly randomized docked conformation $\mathbf{p}$ is deemed as significant for local search if there exists a conformation $\mathbf{q}$ among its $2N$ nearest neighbors such that with respect to each design variable,

$$sign\left\{\left.\frac{\partial f(\mathbf{x})}{\partial x_i}\right|_{\mathbf{x}=\mathbf{p}}\right\}\cdot sign\left\{\left.\frac{\partial f(\mathbf{x})}{\partial x_i}\right|_{\mathbf{x}=\mathbf{q}}\right\}\leq 0 \qquad (1)$$

where $\partial f(\mathbf{x})/\partial x_i|_{\mathbf{x}=\mathbf{y}}$ is the partial derivative of the scoring function f with respect to the design variable x_i at point $\mathbf{y}$. QVina 2 relaxes this condition to minimize the number of necessary local searches that are missed. If $\mathbf{p}$ fails (1) with respect to the design variable x_i, $\mathbf{p}$ is still significant for local search if it passes the following test.

$$sign\left\{\left.\frac{\partial f(\mathbf{x})}{\partial x_i}\right|_{\mathbf{x}=\mathbf{p}}\right\}\cdot sign\{[f(\mathbf{p})-f(\mathbf{q})][p_i-q_i]\}\leq 0 \qquad (2)$$

The rationale behind (2) is that if $\mathbf{p}$'s derivative with respect to x_i is positive and $f(\mathbf{q})$ is higher (or lower) than $f(\mathbf{p})$ while q_i is to the left (or right) of p_i, then there must be a stationary point between p_i and q_i. Reversed relation between the score $f(\mathbf{p})$ and $f(\mathbf{q})$ applies when $\mathbf{p}$'s derivative with respect to x_i is negative. For illustration, see Figure 1 of the Supplementary Materials.

3 Results

The efficacy of QVina 2 was assessed on the core set of the 2014 release of the PDBbind dataset which contains 195 protein–ligand complexes. These are considered as high-quality benchmarks for evaluating various docking/scoring methods (Cheng *et al.*, 2009; Wang *et al.*, 2005).

Vina, QVina 1 and QVina 2 were first compiled on a CentOS release 6.0 x86_64 machine with 16 Intel Xeon 2.27-GHz X7560 CPUs and 98 GB of RAM. The same testing methodology (i.e. the search space definition as well as the RMSD calculation) as that observed in the original Vina (Trott and Olson, 2010) was adopted. For each receptor–ligand complex in the benchmark set, all three docking tools were given the same random number seed and initialized with the same randomized ligand conformation. We then explicitly set both the number of CPUs available for use and the exhaustiveness level to 8. For completeness of comparison, we also docked the same receptor–ligand complexes using GOLD 5.2 and Dock 6.6.

Table 1. Successes/fails of QVina2 versus other tools

		Vina		GOLD		Dock		QVina1	
		Success	Fail	Success	Fail	Success	Fail	Success	Fail
QVina2	Success	116	14	84	39	96	27	79	44
	Fail	7	58	27	45	31	41	9	63

Figure 1 summarizes the accelerations attained by QVina 1 and QVina 2 over the original Vina under different numbers of active rotatable bonds in the range of [0, 10]. Summary on the complete range can be found in the supplementary materials. Figure 2 then shows the RMSD between the actual PDBbind conformations and the predicted ones produced by employing various docking tools. Defining a successful prediction as one with RMSD less than 2 Å, the numbers of successes and failures of the various docking tools are summarized in Table 1. The results indicate that QVina 2 has the highest consistency with Vina among the other docking tools. On the accuracy against the experimental conformations, Dock 6.6 is slightly more accurate than QVina 2 at the expense of long computation time. Dock 6.6 attained 127 successes while QVina 2 achieved 123. Meanwhile, GOLD 5.2 only had 111 successes.

4 Conclusion

QuickVina 2 is a fast and accurate molecular docking tool. Tested against 195 protein–ligand complexes that compose the core set of the 2014 release of the PDBbind using default exhaustiveness level of 8, QVina 2 successfully attained up to 20.49-fold acceleration over Vina. The Pearson's correlation coefficient between Vina's and QVina 2's binding energy was 0.967 for the first predicted mode and 0.911 for the sum of all predicted modes. It is also witnessed that QVina 2 is more accurate than GOLD 5.2 and is only slightly less accurate than Dock 6.6. This shows that QVina 2 has paved the

way for some high-throughput and sufficiently accurate virtual screening of molecular libraries. This in turn brings great value to the fragment-based computer-aided drug design.

Acknowledgements

The authors are thankful to Nafisa Mohamed, Chinh Tran To Su and Xuchang Ouyang for their valuable supports to this work.

Funding

This work was supported by Nanyang Technological University School of Computer Engineering [M060020005-7070680].

Conflict of Interest: none declared.

References

Cheng,T. *et al.* (2009) Comparative assessment of scoring functions on a diverse test set. *J. Chem. Inform. Model.*, **49**, 1079–1093.

Handoko,S.D. *et al.* (2012) QuickVina: accelerating AutoDock Vina using gradient-based heuristics for global optimization, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **9**, 1266–1272.

Trott,O. and Olson,A.J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J. Comput. Chem.*, **31**, 455–461.

Wang,R. *et al.* (2005) The PDBbind database: methodologies and updates. *J. Med. Chem.*, **48**, 4111–4119.