Bioorganic & Medicinal Chemistry Letters 18 (2008) 3509–3512

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

The effect of different electrostatic potentials on docking accuracy: A case study using DOCK5.4

Keng-Chang Tsai [a], Sheng-Hung Wang [b], Nai-Wan Hsiao [c], Minyong Li [d,*], Binghe Wang [d,*]

[a] The Genomics Research Center, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan
[b] Institute of Cellular and Organismic Biology, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan
[c] Institute of Biotechnology, National Changhua University of Education, Changhua 500, Taiwan
[d] Department of Chemistry, Georgia State University, Atlanta, GA 30302-4098, USA

ARTICLE INFO

Article history:
Received 1 March 2008
Revised 6 May 2008
Accepted 7 May 2008
Available online 10 May 2008

Keywords:
Molecular docking
DOCK
Electrostatic potentials
Scoring function
AM1-BCC

ABSTRACT

As a commonly used structure-based approach for virtual screening, molecular design and lead optimization, molecular docking can search the preferred orientation and conformation of a ligand for its optimal binding to a receptor or enzyme active site. In doing so, selecting an appropriate method to calculate the electrostatic potentials is critical. In the current report, nine different semi-empirical and empirical methods, including AM1, AM1-BCC, Del-Re, MMFF, Gasteiger, Hückel, Gasteiger–Hückel, Pullman and formal charges were investigated for their performance on the prediction of docking poses using the DOCK5.4 program. The results demonstrated that the AM1-BCC charges had the highest success rate.

Computational chemistry is playing an increasing role in drug design and discovery. Along this line, there has been a great deal of effort directed toward developing efficient molecular docking methods as tools for the identification of lead compounds.[1–3] Molecular docking is the search for the most energetically favorable binding pose of a ligand to a receptor.[1] During the last decade, considerable progress has been made in using computation methods for the prediction of ligand-target binding modes and activities, and high-throughput virtual screening.[4–8] Several docking programs are readily available including AutoDock,[9,10] GOLD,[11,12] Glide,[13,14] and FlexX.[15,16] The DOCK algorithm uses molecular shape descriptors to position a ligand molecule into a macromolecular receptor and evaluates these poses to generate predicted binding modes for a ligand-receptor complex.[17] The original DOCK program[17] implemented rigid body docking, which allowed users to generate binding mode predictions of ligands. The subsequent versions of the DOCK program have implemented molecular-mechanics force field scoring (DOCK 3.0), energy minimization (DOCK 3.5),[18–20] and ligand conformational flexibility (DOCK 4.0).[21] DOCK 5.4 was developed in a new C++ codebase to maximize the portability and modular nature of the DOCK algorithm.[22] Each major component of the DOCK algorithm has been implemented as a class with a documented interface, allowing these DOCK functions to be modified or replaced easily. DOCK 5.4 features solvation scoring, rigid docking clustering analysis, new ligand conformational search methods, and new minimization methods, and includes support for parallel computing using the Message Passing Interface (MPI) standard. The latest release of DOCK 6.2 is an extension of DOCK 5 but the electrostatic potential in grid calculation is still the same.

Among the most important components of the energy-based scores, such as the default DOCK energy scores, are the proper electrostatic charges that are assigned to the atoms of the ligand. Several charge calculation methods are available and the fundamental differences in their algorithms can result in significant differences in the electrostatic assignments for various atoms. It should be noted that the charge models could have effect on not only the DOCK energy scores, but also the docking conformations, and thus could interfere with the accuracy in docking. So far there has not been a comparative study of the various charge models as applied in docking programs. Herein, we describe our studies of several charge models for their success rate in finding the correct docking conformations and orientations using a standard data set that were derived from experimental results. When charging a small set of ligands, the most accurate ab initio method, such as RESP (Restrained ElectroStatic Potential),[23] could be used. However, due to the time-consuming nature of calculating these ab initio charges, this method was not included in this comparative study. Instead, we focused on nine different semi-empirical and empirical methods, including AM1,[24,25] AM1-BCC,[26,27] Del-Re,[28,29] MMFF,[30–34]

* Corresponding authors. Tel.: +1 404 413 5544; fax: +1 404 413 5543.
E-mail addresses: mli@gsu.edu (M. Li), wang@gsu.edu (B. Wang).

0960-894X/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2008.05.026

Gasteiger,[35,36] Hückel,[37] Gasteiger–Hückel, Pullman[38] and formal charges because they are fast and are widely used.

Again, the proteins and ligands used for the study were extracted from the PDB files as test set reported in literature,[22] which were downloaded from the DOCK website (http://dock.compbio.ucsf.edu/Test_Sets/index.htm, Table 1). The ligands were assigned atom types and bond types manually, and hydrogens were added. Empirical charges were calculated with the method of Del-Re, formal, Gasteiger, Gasteiger–Hückel, Hückel, MMFF and Pullman in the SYBYL 7.2 package.[39] Semi-empirical assignments were performed using the AM1 and AM1-BCC method by the QuACPAC 1.1 program.[40] For proteins, all water molecules, covalently linked sugars, sulfate, and halogens were removed. Co-factors, such as HEME, ATP, and NADPH, were kept, and their atom types and bond types were assigned manually, and Gasteiger–Hückel partial charges were added. Hydrogens were added in protein residues as well as AMBER partial charges and van der Waals parameters. No additional optimization of the protein structures was carried out at this point.

Unless otherwise noted, all studies described in this section involved rigid docking of the ligand to the receptor, both of which were derived from the complex crystal structure. For each case in the test set, the heavy atom RMSD between the top-scoring docked ligand pose and the complex crystal structure ligand pose was evaluated. It should be noted that the RMSD values between the crystal and predicted conformations are widely used as an indicator of whether the correct docking pose is obtained by a program.[41] Usually, an RMSD of 2 Å is considered as the cutoff of correct docking, probably because the resolution in an X-ray crystal structure analysis is often about 2 Å, and higher precision than the resolution of the analysis is not meaningful. Therefore, a DOCK 5 run was considered to be successful if the RMSD between for the top-scoring ligand conformation and the crystal ligand conformation was less than 2.0 Å.[42]

Using the optimized DOCK5 parameters described in literature,[22] rigid and flexible docking experiments were then performed ten times on the entire 114 test sets (Table 1) using different charges. For these nine types of charges, their success rates in prediction and average RMSD values are listed in Table 2. Based on these data, in the case of flexible docking, the AM1-BCC charge model gave the highest success rate (72%, average RMSD = 1.88 Å) followed by Gasteiger–Hückel and MMFF charges. For rigid docking, the AM1-BCC charge model still fared among the lowest average RMSD (1.55 Å) and highest success rate (79%) together with AM1 and Pullman charges. It needs to be noted that in all cases, the formal charge model gave the lowest success rate and the highest average RMSD, presumably because of inaccurate charge assignments.[43]

Figure 1 shows the cumulative percentage of complexes as a function of the RMSD between the predicted conformation and crystal structure results for each docking run. It is clear that for both rigid and flexible dockings the AM1-BCC charges work the best in reproducing the experimentally determined results with the data set studied. No other programs did as well in both flexible and rigid docking, though in each category there are other charge models that gave similar success rates.

Table 1
Complexes used in the test set (total of 114 complexes)

Protein data bank identifier					
1A28	1COM	1FLR	1OKL	1TYL	2MCP
1A6W	1COY	1HAK	1PBD	1UKZ	2PCP
1A9U	1CPS	1HDC	1PDZ	1ULB	2PHH
1ABE	1D3H	1HSL	1PHD	1WAP	2PK4
1ABF	1D4P	1HYT	1PHG	1XID	2TMN
1ACJ	1DBB	1IMB	1PTV	1XIE	2YPI
1ACM	1DBJ	1IVB	1QCF	1YDR	3CPA
1ACO	1DG5	1LAH	1QPE	2AAD	3ERD
1AI5	1DID	1LCP	1QPQ	2ACK	3GPB
1AOE	1DOG	1LDM	1RNT	2ADA	3HVT
1AQW	1DR1	1LST	1ROB	2AK3	4AAH
1AZM	1DWB	1LYL	1RT2	2CHT	4COX
1BYG	1EBG	1MDR	1SNC	2CMD	4CTS
1C5C	1ETT	1MLD	1SRJ	2CPP	4FBP
1C5X	1F0R	1MRG	1TDB	2CTC	4LBD
1C83	1F0S	1MRK	1TNG	2DBL	5ABP
1CBX	1F3D	1MUP	1TNH	2GBP	5CPP
1CIL	1FGI	1NGP	1TNI	2H4N	6RNT
1CKP	1FKI	1NIS	1TNL	2LGS	7TIM

Table 2
Average RMSD and success rate for docking calculation*

Charge	Rigid docking		Flexible docking	
	Average RMSD (Å)	Success rate (%)	Average RMSD (Å)	Success rate (%)
AM1-BCC	1.55	79.0	1.88	71.93
AM1	1.55	74.6	2.33	67.54
Del-Re	1.67	72.8	2.11	67.54
formal	1.87	69.3	2.40	61.40
Gasteiger	1.71	73.7	2.05	70.18
Gasteiger–Hückel	1.65	76.3	1.92	69.30
Hückel	1.63	75.4	2.20	61.40
MMFF	1.59	73.7	1.93	69.30
Pullman	1.52	77.2	2.09	67.54

* All values are averages over ten DOCK runs.

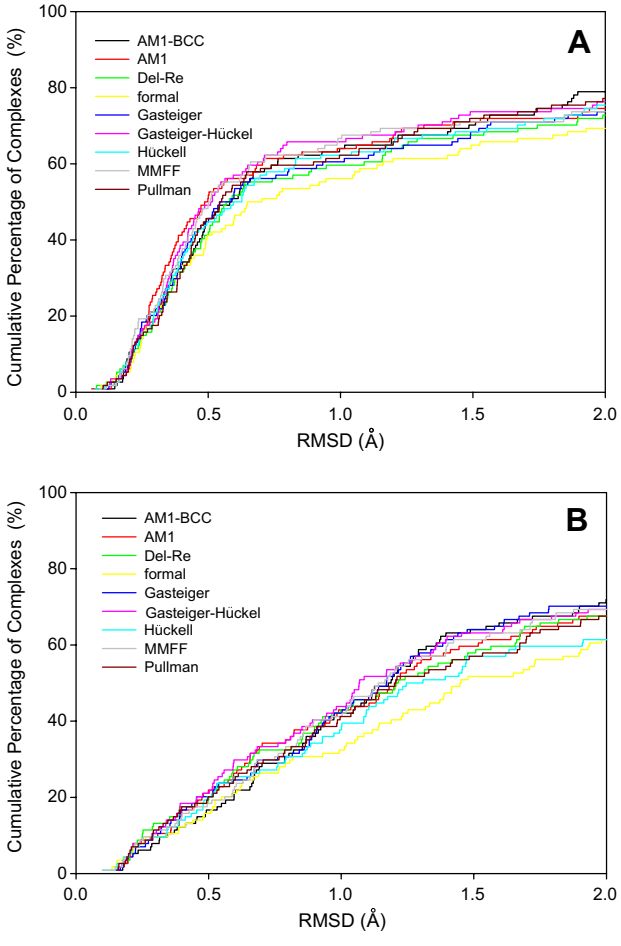

Figure 1. Cumulative percentage of complexes as a function of RMSD in rigid docking (A) and flexible docking (B). All curves are averages over 10 DOCK runs.

The application of AM1-BCC charges improved success rates by about 10% compared to formal charges both in rigid and in flexible docking. Additionally, the success rates were also increased by about 4.5% when compared to AM1 charges. For every charge models, their best prediction and worst prediction are listed in Figure 2.

Herein, we define the best prediction as RMSD < 1 Å, moderate prediction as 1 Å < RMSD < 2 Å, and the worst prediction as RMSD > 2 Å. It is obvious that in both flexible and rigid docking, the AM1-BCC charge model is among the highest in the number of best predictions and lowest in the number of worst predictions.

In order to validate the accuracy of those charges, their energy scores after molecular docking were correlated with binding free energy. DOCK energy scoring function is a classical force field energy function, which sums van der Waals and electrostatic interactions.[21] Among all 114 sets of PDB coordinates, 61 complexes with known dissociation constants (K_d) were retrieved from AffinDB (http://pc1664.pharmazie.uni-marburg.de/affinity/index.php).[44] Their experimental binding free energies are calculated from K_d using the following relationship: $\Delta G_{binding} = RT\ln K_d$, where R is ideal gas constant (1.987 cal/K mol), T is temperature in K (298 K is used in this paper). After linear correlation, the regression results are shown in Table 3.

Among all nine charge methods, AM1-BCC had the highest regression constant (R) in both rigid and flexible docking. Such results are also in agreement with the finding by Shoichet and his coworkers that AM1-BCC can increase the accuracy in the prediction of binding free energies.[45] However, all these correlation coefficients in Table 3 are considered low indicating difficulties in predicting binding free energy with high accuracy using any of the charge models.

Table 3
The correlation between docking energy scores and binding free energies

Charge	Regression coefficient (R)	
	Rigid docking	Flexible docking
AM1-BCC	0.523	0.551
AM1	0.370	0.341
Del-Re	0.130	0.148
formal	0.300	0.311
Gasteiger	0.348	0.382
Gasteiger–Hückel	0.248	0.253
Hückel	0.355	0.300
MMFF	0.318	0.312
Pullman	0.166	0.189

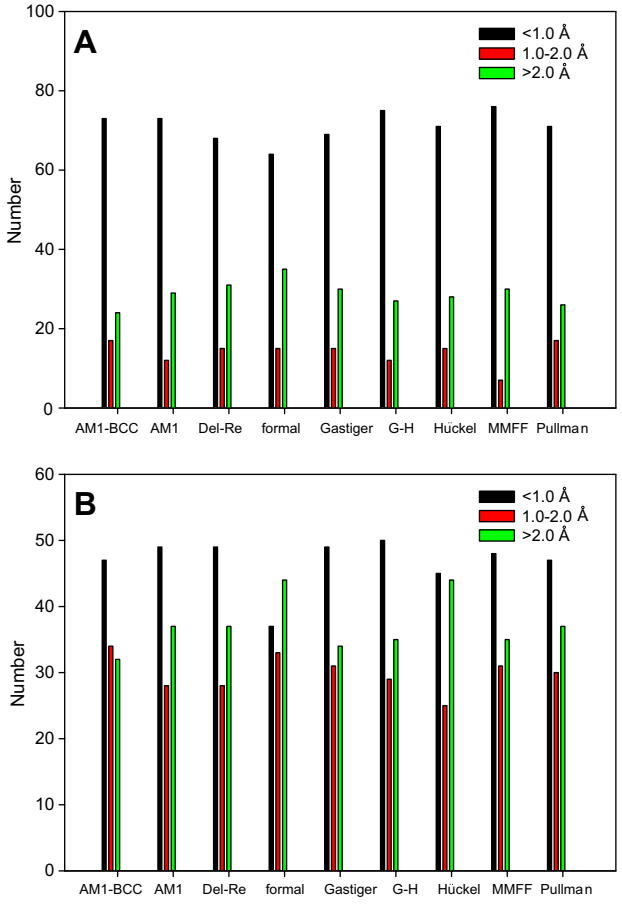

Figure 2. Comparison of different charge methods in terms of docking numbers obtained under 1.0, 1.0–2.0 and below 2.0 Å of RMSD in (A) rigid docking and (B) flexible docking.

It should be noted that AM1-BCC charges quickly and efficiently generates high-quality atomic charges for a variety of polar, nonpolar, and aromatic molecules. AM1-BCC charges start with partial charges derived from the AM1 wave-function. In a second stage, bond-charge corrections (BCC) are applied to the partial charges on each atom to generate the final partial charges. The AM1-BCC method is parameterized to reproduce the HF/6-31G* RESP results, and offer a fast and high quality charge model for organic or biological molecules.[26,27] Dill and collaborators highlighted that the semi-empirical AM1-BCC method for computing charges works almost as well as any of the more computationally expensive ab initio methods for the prediction of small-molecule hydration free energy.[46] Using AM1-BCC charges, we also successfully identified several *Escherichia coli* SecA inhibitors after virtual screening.[47] Therefore, our studies and other available evidences demonstrate that the AM1-BCC charges may offer advantages over the other eight charge models in similar docking studies.

Acknowledgments

We thank OpenEye Scientific Software, Inc. for the use of QuACPAC program. The SYBYL computation was conducted at the National Center for High Performance Computing, Taiwan. We also thank the Georgia Research Alliance and Georgia Cancer Coalition for partial financial support.

Supplementary data

Docking results for all charge models. The supplementary data are available online with the paper in ScienceDirect. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2008.05.026.

References and notes

1. Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins* **2002**, *47*, 409.
2. Brooijmans, N.; Kuntz, I. D. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335.
3. Mohan, V.; Gibbs, A. C.; Cummings, M. D.; Jaeger, E. P.; DesJarlais, R. L. *Curr. Pharm. Des.* **2005**, *11*, 323.
4. Schneider, G.; Bohm, H. J. *Drug Discov. Today* **2002**, *7*, 64.
5. Lyne, P. D. *Drug Discov. Today* **2002**, *7*, 1047.
6. Shoichet, B. K.; McGovern, S. L.; Wei, B.; Irwin, J. J. *Curr. Opin. Chem. Biol.* **2002**, *6*, 439.
7. Hermann, J. C.; Marti-Arbona, R.; Fedorov, A. A.; Fedorov, E.; Almo, S. C.; Shoichet, B. K.; Raushel, F. M. *Nature* **2007**, *448*, 775.
8. Bajorath, J. *Nat. Rev. Drug Discov.* **2002**, *1*, 882.
9. Goodsell, D. S.; Morris, G. M.; Olson, A. J. *J. Mol. Recognit.* **1996**, *9*, 1.
10. Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. *Proteins* **2002**, *46*, 34.
11. Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. *Proteins* **2003**, *52*, 609.
12. Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R. *J. Med. Chem.* **2005**, *48*, 6504.

13. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. *J. Med. Chem.* **2004**, *47*, 1750.
14. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739.
15. Kramer, B.; Rarey, M.; Lengauer, T. *Proteins* **1999**, *37*, 228.
16. Kramer, B.; Rarey, M.; Lengauer, T. *Proteins* **1997**, *Suppl. 1*, 221.
17. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269.
18. Meng, E. C.; Gschwend, D. A.; Blaney, J. M.; Kuntz, I. D. *Proteins* **1993**, *17*, 266.
19. Shoichet, B. K.; Kuntz, I. D. *Protein Eng.* **1993**, *6*, 723.
20. Gschwend, D. A.; Kuntz, I. D. *J. Comput. Aided Mol. Des.* **1996**, *10*, 123.
21. Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. *J. Comput. Aided Mol. Des.* **2001**, *15*, 411.
22. Moustakas, D. T.; Lang, P. T.; Pegg, S.; Pettersen, E.; Kuntz, I. D.; Brooijmans, N.; Rizzo, R. C. *J. Comput. Aided Mol. Des.* **2006**, *20*, 601.
23. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollmann, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620.
24. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
25. Ferenczy, G. G.; Reynolds, C. A.; Richards, W. G. *J. Comput. Chem.* **1990**, *11*, 159.
26. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132.
27. Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623.
28. Del Re, G. *J. Chem. Soc.* **1958**, 4031.
29. Del Re, G.; Pullman, B.; Yonezawa, T. *Biochim. Biophys. Acta* **1963**, *75*, 153.
30. Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490.
31. Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 520.
32. Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 553.
33. Halgren, T. A.; Nachbar, R. B. *J. Comput. Chem.* **1996**, *17*, 587.
34. Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 616.
35. Gasteiger, J.; Marsili, M. *Tetrahedron Lett.* **1978**, 3181.
36. Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219.
37. Purcell, W. P.; Singer, J. A. *J. Chem. Eng. Data* **1967**, *12*, 235.
38. Berthod, H.; Giessner-Prettre, C.; Pullman, A. *Theoret. Chim. Acta* **1967**, *8*, 212.
39. *SYBYL 7.2*; Tripos, Inc.: St. Louis, MO, USA, 2006.
40. *QuACPAC 1.1*; OpenEye Scientific Software, Inc.: Santa Fe, NM, USA, 2007.
41. Kroemer, R. T.; Vulpetti, A.; McDonald, J. J.; Rohrer, D. C.; Trosset, J. Y.; Giordanetto, F.; Cotesta, S.; McMartin, C.; Kihlen, M.; Stouten, P. F. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 871.
42. Onodera, K.; Satou, K.; Hirota, H. *J. Chem. Inf. Model* **2007**, *47*, 1609.
43. Parkin, G. *J. Chem. Educ.* **2006**, *83*, 791.
44. Block, P.; Sotriffer, C. A.; Dramburg, I.; Klebe, G. *Nucleic Acids Res.* **2006**, *34*, D522.
45. Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. *J. Mol. Biol.* **2007**, *371*, 1118.
46. Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. *J. Phys. Chem. B* **2007**, *111*, 2242.
47. Li, M.; Huang, Y. J.; Tai, P. C.; Wang, B. *Biochem. Biophys. Res. Commun.* **2008**, *368*, 839.