

Evaluation of Docking Accuracy and Investigations of Roles of Parameters and Each Term in Scoring Functions for Protein–Ligand Docking Using ArgusLab Software

Akifumi Oda,* Megumi Okayasu, Yukiko Kamiyama, Tomoki Yoshida, Ohgi Takahashi, and Hisao Matsuzaki

Faculty of Pharmaceutical Sciences, Tohoku Pharmaceutical University, 4-4-1 Komatsushima, Aoba-ku, Sendai 981-8558

Received January 17, 2007; E-mail: oda@tohoku-pharm.ac.jp

We evaluated the docking accuracy of ArgusLab, which is a freely available software program for computational docking, using experimentally determined protein–ligand complex structures. Investigations of the roles of parameters used for docking calculations in ArgusLab were carried out in addition to evaluation of the software. The results indicate that one docking engine of ArgusLab, GADock, is superior in terms of accuracy, and that another docking engine, ArgusDock, is advantageous in terms of the computational time it provides. In a pose construction step, larger population sizes generated more accurate poses using GADock. The docking poses constructed by GADock were easier to evaluate than those of ArgusDock in the pose-selection step. Reliable docking calculations are expected to be carried out with ArgusLab using results of this study.

Structure-Based Drug Design (SBDD), in which three-dimensional (3D) structures of drug target biopolymers are used for design and development of novel drugs, is a promising technique of Computer-Aided Drug Design (CADD). The predictions of the 3D structures of target protein–small-ligand complexes serve important roles in SBDD.[1] When the complex structures are determined, several methods for SBDD can be carried out, such as virtual screening and design of compounds that have complementary structures to that of the target. Recently, some computer software programs have been developed for prediction of protein–ligand complex structures, which enable so-called computational docking. Some of those software programs, e.g., GOLD,[2] FlexX,[3] and Glide,[4] have been well tested and then widely used with high reliability. They are all offered commercially both for commercial and for academic users; their users must pay fees. Software DOCK[5] and AutoDock[6] are also used widely among academic users. Although the software is useful at no charge to academic users, commercial users must pay fees. Introducing payware for CADD systems frequently engenders difficulties from the aspect of costs, although support systems are amply established and high reliability can be expected for commercial software. Although academic free software plays an important role in academic studies, software introduction costs cannot be negligible for commercial users. Lower computational costs are desirable when cost-benefit performance is considered. Especially for education of nonspecialists of CADD and for once-over-lightly tests of SBDD in early stages of drug design, research expenditures are frequently limited. Moreover, because numerous licenses are required for large-scale calculations, such as virtual screening of virtual libraries, the number of licenses defines the scale of studies if such payware is used. Although most docking software is also payware, one exception is ArgusLab.[7]

ArgusLab was developed originally as molecular modeling software. It can perform not only computational docking, but also molecular constructions, molecular calculations, and molecular visualization. It has unique features. For example, it can use a UFF force field[8] for molecular calculations. In ArgusLab ver. 4.0.1, computational docking between biopolymers and small ligands has been introduced. It includes two docking engine types: ArgusDock and GADock. In addition, AScore is used as a scoring function.

GUI has been implemented for users in ArgusLab, because the software was developed originally for molecular modeling. Furthermore, because it can run under Windows, ArgusLab is a very easy-to-use docking software program for researchers, who are inexperienced at using UNIX workstations. For those reasons, it appears that ArgusLab is suitable for education of nonspecialists of CADD and for once-over-lightly calculations. Nevertheless, the docking ability of ArgusLab has not been evaluated sufficiently. There is little documentation of ArgusLab, and manuals for the software have not been prepared. The docking feature of ArgusLab has been described in on-line tutorials and presentations made using PowerPoint (Microsoft Corp.) files.[9] Even in that documentation, results of only a few examples are illustrated. The results of pose constructions and of pose selections for unknown complexes have not been discussed sufficiently. Although a research paper describing a comparison of docking efficiencies of GOLD and ArgusLab has been published very recently,[10] only ArgusDock in "High precision" mode is tested in that study; no other settings have been investigated. Furthermore, the other docking engine, GADock, and details of AScores have not yet been evaluated.

In this paper, the docking abilities of ArgusLab were eval-

Published on the web October 12, 2007; doi:10.1246/bcsj.80.1920

uated using re-docking tests of experimentally determined protein–ligand complex structures. There are some types of preset parameters both for ArgusDock and for GADock. However, they were perturbed, and their roles were investigated. Using the results of those investigations, docking accuracy and limitations of ArgusLab were determined, and suitable parameter settings are proposed. The results of this study are expected to provide useful guidelines for cost-effective methods of protein–ligand docking.

Method

In this study, 3D structures of 55 protein–ligand complexes, which were an intersectional part of the Glide test set[4] and Eldridge's test set[11] excluding inadequate complexes for docking tests of drug-like molecules, e.g., immunoglobulins and complexes, in which ligands are bonded covalently with proteins, were used as a test set for evaluations of docking by ArgusLab. Although both the Glide set and Eldridge's set were widely used, the sets include many complexes that have identical target proteins. Using the intersectional test set, bias for target proteins could be eliminated, and a reliable test could be carried out. Although docking calculations of some complexes in the set required much computational time (see the Results and Discussion section), those high-cost complexes were retained, thereby providing a fair test. Experimentally determined protein–ligand complex structures were obtained from the RCSB Protein Data Bank (abbreviated as PDB).[12] The PDB IDs of the complexes included in the set are presented in Table 1. Structures of both protein and ligand molecules were extracted from the PDB data and were used for docking tests. For the docking features of ArgusLab, because hydrogen atoms of proteins are ignored, hydrogens of proteins were not added in this study. In addition, protein structures extracted from PDB data were used without any modification. Consequently, our tests are so-called re-docking tests. On the other hand, because hydrogen atoms of ligands serve important roles in docking, ligand molecules filled the valence using ArgusLab. After adding the hydrogens, ligand molecules were minimized using the UFF force field[8] implemented in ArgusLab.

For the docking tests, both ArgusDock and GADock were evaluated; then, the results were compared. For ArgusDock, both "Regular precision" mode and "High precision" mode were tested as settings of "Docking precision." The default setting of the scoring function and adjusted functions were used for the study. The AScore form is shown below.

Table 1. Test Sets Used for Docking by Using ArgusLab

1aaq	1abe	1abf	1adf	1apb
1apt	1apu	1bap	1bra	1cbx
1dih	1dog	1ebg	1eed	1etr
1ets	1ett	1hbv	1hpv	1hsl
1htf	1mbi	1nnb	1nsc	1nsd
1ppc	1pph	1ppk	1rbp	1tlp
1tmn	1ulb	2cpp	2gbp	2ifb
2phh	2tmn	2xis	2ypi	3ptb
4dfr	4tln	4tmn	5abp	5cpp
5tln	5tmn	6abp	6cpa	6tmn
7abp	7cpa	8abp	9abp	9hvp

$$\Delta G_{\text{bind}} = \Delta G_{\text{vdw}} + \Delta G_{\text{hydrophobic}} + \Delta G_{\text{H-bond}} + \Delta G_{\text{H-bond(chg)}} + \Delta G_{\text{deformation}} + \Delta G_0, \quad (1)$$

where, the right-hand-side terms respectively represent the contributions of van der Waals' force (vdW force), hydrophobic effects, hydrogen-bonding effects for neutral donors and acceptors, hydrogen-bonding effects for charged donors and acceptors, the effects of numbers of rotatable bonds, and the segment value of regression expression. The AScore was based on terms taken from the HPScore piece of XScore;[13] each term has a regression coefficient. The coefficients of AScore were determined to fit the experimental binding free energies derived from Wang test sets.[14] That is, the AScore was generated with considerations for only one role of scoring functions: the usage of virtual screening.[15,16] Although a few tests assessed the remaining two roles, i.e., pose construction and pose selection, details of those roles are not discussed. In Ref. 9, docking tests were carried out only for protein–ligand complexes used for parameterization. Although the results of Ref. 10 are interesting and highly suggestive, only tests of ArgusDock in high-precision mode, which is not an ordinary setting (high-precision mode is an optional and very high-cost mode), are described in that paper. Furthermore, although different parameterizations from those for virtual screening might be suitable for pose selection and pose construction, the optimal settings for the two steps are not discussed. In this study, we evaluated the efficiencies of ArgusLab in terms of the two roles, pose selections and pose constructions, using both ArgusDock and GADock. During the investigations, parameters of AScore, such as regression coefficients and vdW radii, were varied; also, the effects of parameter setting for pose constructions and pose selections were estimated for ArgusDock. For GADock, one parameter of genetic algorithms, i.e., population size, was adjusted. From these two types of docking engine and these parameter changes, 11 types of calculations were carried out in all. The settings are illustrated in Tables 2 and 3. In addition to those 11 settings, settings in which three coefficients of AScore were changed simultaneously were also adopted to discuss the results of settings **iii**, **iv**, **vii**, and **viii**, closely. Those four calculations used: a setting in which hydrophobic, hydrogen bonding, and vdW terms were all decreased, and three settings in which one of those three terms was reduced and the other two terms were increased. The settings are presented in Table 4. For those calculations, docking engine ArgusDock and regular precision mode were used. For docking calculations, all water molecules in PDB data were deleted in this study. In docking calculations, determinations of ligand binding sites play very important roles. In our study, the sites were defined as the collections of amino acids selected by the default setting of ArgusLab concerning bound ligand in PDB structures. Furthermore, the sizes of the binding site bounding box were determined automatically using ArgusLab. The default settings were used, except as described above. In this study, the docking tests were carried out not only for ArgusLab, but also for commercial docking software GOLD,[2] and the results were compared. By comparison, docking abilities of ArgusLab were clearly determined. For the GOLD calculations, the number of poses generated for each test complex

Table 2. Calculation Settings of ArgusDock

	Accuracy	Modification of scoring function		Accuracy	Modification of scoring function
i	Regular	default	**v**	Regular	vdW radii × 0.8
ii	High	default	**vi**	Regular	vdW radii × 1.2
iii	Regular	delete the hydrophobic term	**vii**	Regular	vdW term × 0.5
iv	Regular	H-bond term × 0.5	**viii**	Regular	vdW term × 2.0

Table 3. Calculation Settings of GADock

	Scoring function	Population size
I	default	default (50)
II	default	100
III	default	200

was set to 10. In the "Define site using" section, "Point" was selected, and the coordinates of the site point were set to coordinates of centroids of ligands in crystal structures. "Active site radius" was set to 10.0 Å (default). Other conditions, such as GA parameters, were set to default settings.

The root-mean-square deviations (RMSDs) between the experimental and computational ligand structures were computed to evaluate the accuracy of the calculated poses. The experimental structure that was obtained from the PDB was regarded as the correct answer; calculated poses with heavy-atom RMSDs that were less than or equal to 2.0 Å were defined as reasonable poses. Because it was difficult to obtain an exact match with any of the docking programs and scoring functions, we judged the success of docking and scoring calculations whether a reasonable pose was obtained or not. In addition to docking accuracy, the computation times of respective settings were compared.

ArgusLab calculations were performed using a Windows XP computer with a Pentium D 3.0 GHz processor and 2 GB memory. Calculations by GOLD version 3.1.1 were carried out using a Linux (Red Hat Enterprise Linux WS 3.0) computer with a Pentium D 3.73 GHz processor and 4 GB memory. Although ArgusLab 4.0.1 was used for preparation and execution of docking, post-processes, such as calculations of RMSDs, were performed using Maestro 7.5,[17] which is freely available to academic users.

Results and Discussion

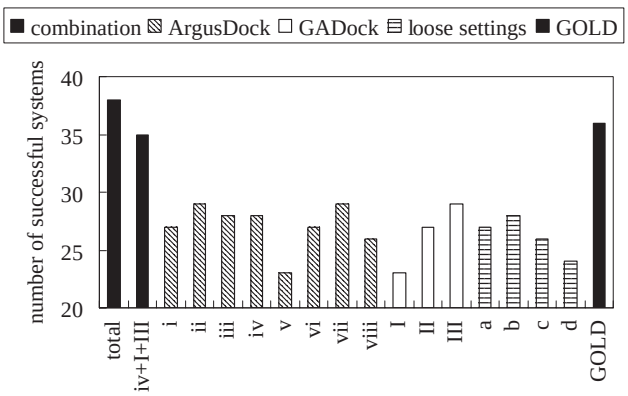
First, the usefulness of a scoring function and docking algorithms for pose construction was elucidated using all settings. The results are illustrated in Fig. 1. In that figure, numbers of complex systems, for which at least one reasonable pose was obtained using the setting, are depicted as a bar chart (shown as the "number of successful systems" in Fig. 1a). The success rates are displayed as a line plot in Fig. 1b. Success rates (ratios of successfully modeled systems to the total number of systems) were evaluated using the following formula.

$$p_{\text{success}} = \frac{n_{\text{success}}}{n_{\text{all}}}, \tag{2}$$

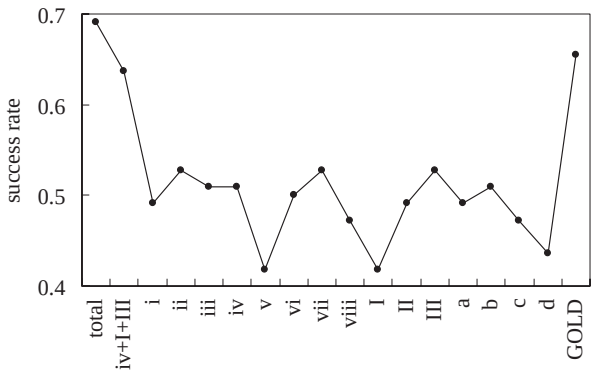
where, p_{success} is the success rate, n_{success} is the number of protein–ligand systems, in which reasonable poses are obtainable using each setting, and n_{all} is the total number of systems (in this study, $n_{\text{all}} = 55$ except for setting **vi**, as described below). For example, for 27 systems of all 55 complexes, reasonable poses were obtained by using setting **i**. Thereby, p_{success} was equal to $27/55 = 0.491$. The "total" value expresses the number and rate of complexes for which reasonable poses are obtainable by at least one setting of 11 settings. One advantage of

Table 4. Calculation Settings to Investigate Loose Conditions

	Modification of scoring function
a	delete hydrophobic, H-bond × 0.5, vdW × 0.5
b	hydrophobic × 2, H-bond × 2, vdW × 0.5
c	hydrophobic × 2, H-bond × 0.5, vdW × 2
d	delete hydrophobic, H-bond × 2, vdW × 2



(a) Number of successful systems



(b) Success rate

Fig. 1. Results of pose constructions. The bar chart (a) shows the number of successful systems; the line plot (b) illustrates the success rate. The "number of successful systems" represents the numbers of complex systems for which at least one reasonable pose was obtained.

ArgusLab is its high-speed calculation capability. The median of computational time of "total," which represents the sum of computational times for settings **i** to **III**, was shorter than that of one trial of GOLD (described below in Tables 5 and 6). Therefore, the pose set of "total" can be generated faster than GOLD poses. When the average (not the median) was considered, "total" was more costly than GOLD. For that reason, the combination **iv** + **I** + **III**, in which all poses generated by setting **iv**, **I**, and **III** were put together, is also shown. The combination **iv** + **I** + **III** gave the best results in all combinations, which were generated faster with ArgusLab than with GOLD in terms of average computational time. In actual drug design trials, when reasonable docking poses, which are consistent with known information such as results of earlier studies, cannot be obtained by one docking study, recalculations are often carried out with some changes in docking settings. Furthermore, using the advantages of ArgusLab (freely available, with very fast calculation), not only a trial-and-error approach but also multiple calculations with various settings, e.g., setting **i** to **III**, can be carried out in parallel from the beginning. The tests of "total" and "**iv** + **I** + **III**" were performed to reveal the abilities of ArgusLab for those situations. Although setting **vi** yielded no poses (neither reasonable nor unreasonable ones) for one protein–ligand complex system (2xis), at least one pose was given for each of the 55 systems by all other settings.

Figure 1 shows that success rates of the majority of independent calculations of ArgusLab were greater than 50%. In addition, the results of combination tests, i.e., "total" and "**iv** + **I** + **III**," were around 70%. Although results of docking tests depend on test sets and test procedures to some degree, an acceptable success rate is generally considered as 50–90%[18] (or more) for the pose construction step. In fact, results of some studies of the literature suggest that even a success rate of 30–50% is reasonable.[19] Thus, the results of combination tests were in the middle of the acceptable range; even a majority of independent tests can give reasonable results. These results suggest that, although ArgusLab is not clearly superior to other docking programs, it is comparable to commercial software and is sufficiently useful for casual docking. In addition to the rule-of-thumb comparisons mentioned above, actual comparisons to GOLD using the same test set are also illustrated in Fig. 1. In the pose construction step, all independent calculations of ArgusLab produced worse results than GOLD; they were consistent with the results of Ref. 10. The results suggest that different software should be used for different purposes, such as docking accuracy, costs, and computational time. On the other hand, the success rate of "total" was greater than that of GOLD, and that of "**iv** + **I** + **III**" was comparable to that of GOLD. In fact, GOLD is one of the best docking software at present.[18] Therefore, these results suggest that ArgusLab can generate appropriate pose sets as top-class commercial software by gathering individual pose sets or by a trial-and-error approach.

The results of settings **v**, **vi**, **vii**, and **viii**, in which the vdW term was changed, differed more from that of default setting **i** than the results of settings **iii** and **iv**, in which hydrophobic and hydrogen-bond terms were modified, respectively. This difference indicates that the vdW term serves an important role in AScore for the pose construction steps of ArgusDock, which is consistent with a previous suggestion by Kitchen et al.[1] that enthalpy terms play important roles in pose constructions. The results of settings **v** and **vi**, in which vdW radii were changed, were not improved over those of the default setting. Therefore, modifications of regression coefficients seem to be more effective than changes in the radii for improvements of the vdW term. The figure shows that a decrease in the regression coefficients of hydrophobic, hydrogen bond and vdW terms (**iii**, **iv**, and **vii**) improved the success rates in comparison to increasing the coefficient (**viii**). The results of **a**–**d** (see Fig. 1), which were additional tests for the detailed studies of the loose settings **iii**, **iv**, and **vii**, show the following: although the results of setting **a**, in which coefficients were decreased, were as good as that of **i** (default), and settings **c** and **d**, in which vdW coefficient was increased and either hydrophobic or hydrogen-bonding coefficients were decreased, cause unreasonable results compared to those of default **i**. These results suggest that the effect of a strict setting for vdW term is more important than effects of loose settings for hydrophobic or hydrogen-bonding terms in the pose construction step. For that reason, strict setting of the vdW term is apparently unfavorable, more so than loose settings are favorable. In addition, these results indicate that the vdW term plays an important role in AScore for the pose construction step and they seem to be additional evidence that suggests the importance of enthalpy effects in pose constructions. In Fig. 2, the average numbers of poses generated are depicted as a bar chart. As shown in that figure, the numbers of poses were always less when the coefficients of vdW terms were increased compared with default without exceptions. The results suggest that the search space became narrower than the default because of strict vdW terms: the narrowed search space can lower the success rate.

On the other hand, for GADock, the success rate was further improved when the population size was increased. The results indicate the importance of the adjustment of GA parameters. Although the default setting of GADock (**I**) yielded worse results than the ArgusDock default (**i**), setting **II**, in which population size was increased to 100, gave an equivalent success rate to that of ArgusDock. The success rate was comparable to that of the high-precision setting of ArgusLab (**ii**) when the population size was increased to 200 (**III**). These results sug-

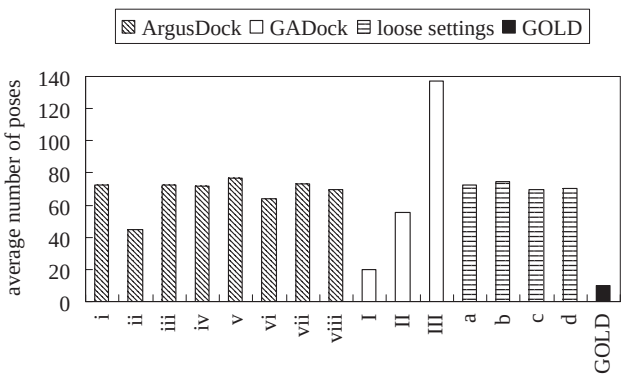


Fig. 2. Number of poses constructed. The bar chart expresses the average number of poses constructed.

Table 5. Computational Time of ArgusDock and GADock

	i	**ii**	**iii**	**iv**	**v**	**vi**	**vii**	**viii**	**I**	**II**	**III**
Average/s	296	4049	160	299	295	414	307	345	22	45	91
Median/s	5	145	6	6	6	6	6	6	20	35	70
Maximum/s	10610	54105	3102	10782	12723	9846	12194	11088	70	134	260

Table 6. Computational Time of Additional Settings and GOLD

	a	**b**	**c**	**d**	Total	**iv + I + III**	GOLD
Average/s	147	343	317	175	6319	412	424
Median/s	5	6	7	6	357	94	414
Maximum/s	3043	13503	11836	3198	123808	11005	923

gest that GADock, with a large population size, is preferable to ArgusDock with regular precision for pose constructions if computer resources can be sufficiently prepared.

Second, the computational time of docking using each setting is shown in Tables 5 and 6. In that table, average, median and maximum values of computational times for 55 complexes (for 54 systems for setting **vi**) are listed. The average, median and maximum of the total times necessary to complete 11 (**i** to **III**) calculations, and those of **iv + I + III** are also shown as "total" and "**iv + I + III**." Furthermore, the computational time of GOLD is provided for comparison. As shown in the table, the computational speeds for all independent settings, except for **ii**, in which the high-precision mode was used, were faster than those of GOLD in terms of both averages and medians. In addition, the combination setting of "total" was faster than GOLD in terms of average, and "**iv + I + III**" was faster than GOLD in terms of both average and median. From the comparison between ArgusDock and GADock shown in the table, ArgusDock with regular precision (settings **i** to **viii** without **ii**) yielded smaller medians than those of GADock, but the averages were larger. Although ArgusDock can perform docking calculations faster than GADock for almost all complex systems, for a minority of exceptional systems, ArgusDock requires extremely long computational times. As shown in the maximum values illustrated in Table 5, GADock calculated all systems in a few minutes without exception, but ArgusDock sometimes required several hours. That delay indicates that ArgusDock has certain problems for some protein–ligand systems and that GADock is more suitable for such cases if computational times must be limited. These results are consistent with a report by Thompson,[9] which warns that some ligand/binding site types still cause problems. The maximum time was quite large, even in comparison with other ArgusDock calculations, when the ArgusDock docking precision was set to high precision (setting **ii**). On the other hand, for GADock, computational times increased with an increase in the population size, indicating that computational times of GADock are easily controlled and that a plan for computational docking is easy to make. However, for most systems, the computational times of ArgusDock were much faster than those of GADock. ArgusDock frequently completed the calculations in a few seconds. In contrast, GADock required several tens of seconds for all systems, suggesting that a small number of test calculations are indispensable for using ArgusLab for large computational dockings. When test results are sufficiently good and ArgusDock can calculate them rapidly, ArgusDock is useful for subsequent full-scale calculations. Otherwise, GADock is desirable.

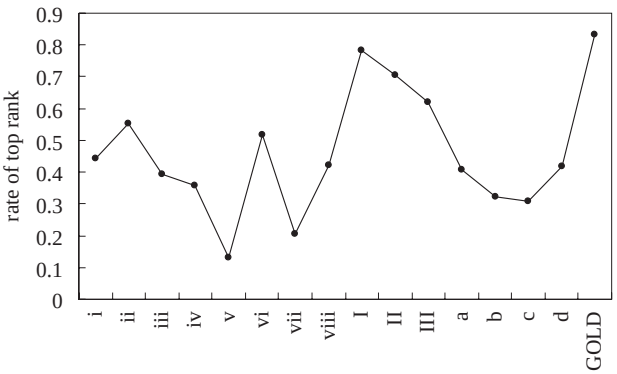


Fig. 3. Results of pose selections. The line plot illustrates the rates of top ranks.

Finally, the efficiency of AScore for pose selections using each setting is illustrated in Fig. 3. AScore calculated the binding free energy between a protein and a ligand for each pose; every pose was subsequently ranked in the order corresponding to the energy. The top-ranked pose was regarded as the most likely docking pose. In this figure, the rates of complexes, for which the top-ranked poses were reasonable among successfully modeled complexes (shown as "rate of top rank" by line plot in Fig. 2), were evaluated as

$$p_{\mathrm{top}} = \frac{n_{\mathrm{top}}}{n_{\mathrm{success}}}, \tag{3}$$

where p_{top} is the rate of the top rank, n_{top} is the number of protein–ligand systems in which the top-ranked pose is the reasonable one at each setting, and n_{success} is the number of complexes in which reasonable poses are obtainable. For pose selection, the rule-of-thumb acceptable p_{top} value mentioned in previous studies is around 60%.[14] From Fig. 3, the results of ArgusLab were around (or more than) 60% with some settings, and only a few settings were less than 30%. Even for commercial scoring functions, the rates were 26–76%.[14] Although AScore is not clearly superior to commercial software, it is useful for pose selection. In fact, as shown in actual comparison with GOLD tests shown in Fig. 3, although p_{top} of all settings were less than those of GOLD, the results of GADock are comparable to those of GOLD. GOLD has been reported to

be among the best software packages for pose selection at present in previous papers.[18] Therefore, the results suggest that AScore is also reliable for that purpose. As shown in Figs. 2 and 3, when the constructed poses were fewer, the rates of top ranks were higher. When the generated poses are fewer, it is assumed that docking poses are easier to rank, because the numbers of unreasonable poses are also smaller. Especially, because the number of constructed poses using setting **I** was smallest, the rate of the top rank was around 80%, which indicates equivalent or better results than those obtained using commercial software.[14,18] In addition to setting **I**, GADock with other settings (**II** and **III**) also gave better results than ArgusDock, even if the number of generated poses was larger than that for ArgusDock. That fact indicates that GADock generates more easy-to-rank poses than ArgusDock generates, i.e., AScore seems to be more suitable for scoring of docking poses constructed by GADock than for scoring of ArgusDock poses. Especially, setting **III** gave comparable results to the high-precision setting of ArgusDock (setting **ii**), not only for pose constructions (shown in Fig. 1) but also for pose selections (shown in Fig. 3), in spite of much shorter computational time than that indicated by setting **ii** (shown in Table 5). On the other hand, the rates of top ranks were, at best, around 50% by using ArgusDock. Although they are reasonable values compared to those of payware,[14,18] more accurate scoring is desirable for practical drug-design trials. These results suggest that some bold modifications, e.g., adding a surface-area term indicated by Wang et al. for PMF,[14] are necessary for AScore to evaluate docking poses constructed by ArgusDock. The results with setting **v**, in which vdW radii were reduced, were worst of all, as with results of pose constructions shown in Fig. 1. These results suggest that the reduction of vdW radii from default values yields unreasonable results for ArgusDock calculations.

Conclusion

In this study, docking features of ArgusLab were evaluated. Suitable settings for it were investigated. The results indicate that one docking engine, GADock, is superior in terms of accuracy and that another docking engine, ArgusDock, offers advantages in terms of low computational times. Docking accuracy became worse when vdW radii were modified. Therefore, adjusting the regression coefficients are inferred to be more effective than changes in vdW parameters for improvement of ArgusLab settings.

Although a salient advantage of docking features of ArgusLab is the ability of high-speed computations, the results suggest that the docking accuracy is also comparable to that of commercial software. We expect that our results will aid low-cost docking computations and that computational docking will become more widely used for drug design and development.

References

1 D. B. Kitchen, H. Decornez, J. R. Furr, J. Bajorath, *Nat. Rev. Drug Discovery* **2004**, *3*, 935.

2 G. Jones, P. Willett, R. C. Glen, *J. Mol. Biol.* **1995**, *245*, 43.

3 M. Rarey, B. Kramer, T. Lengauer, G. Klebe, *J. Mol. Biol.* **1996**, *261*, 470.

4 R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, P. S. Shenkin, *J. Med. Chem.* **2004**, *47*, 1739.

5 I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, T. E. Ferrin, *J. Mol. Biol.* **1982**, *161*, 269.

6 G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, *J. Comput. Chem.* **1998**, *19*, 1639.

7 *ArgusLab 4.0.1*, Planaria Software LLC, Seattle, **2004**. (http://www.planaria-software.com/)

8 A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard, III, W. M. Skiff, *J. Am. Chem. Soc.* **1992**, *114*, 10024.

9 M. Thompson, *Molecular Docking With ArgusLab*, Planaria Software, Seattle, **2004**. (http://www.planaria-software.com/dockingtalk.htm)

10 S. Joy, P. S. Nair, R. Hariharan, M. R. Pillai, *In Silico Biol.* **2006**, *6*, 0053.

11 M. D. Eldridge, C. W. Murray, T. R. Auton, G. V. Paolini, R. P. Mee, *J. Comput. Aided Mol. Des.* **1997**, *11*, 425.

12 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235.

13 R. Wang, L. Lai, S. Wang, *J. Comput. Aided Mol. Des.* **2002**, *16*, 11.

14 R. Wang, Y. Lu, S. Wang, *J. Med. Chem.* **2003**, *46*, 2287.

15 A. Oda, N. Yamaotsu, S. Hirono, *SAR News* **2006**, *11*, 13.

16 A. Oda, K. Tsuchida, T. Takakura, N. Yamaotsu, S. Hirono, *J. Chem. Inf. Model.* **2006**, *46*, 380.

17 *Maestro 7.5*, Schrödinger, LLC, New York, **2006**.

18 E. Kellenberger, J. Rodrigo, P. Muller, D. Rognan, *Proteins: Struct., Funct., Bioinf.* **2004**, *57*, 225.

19 Y. Eguchi, H. Toh, K. Kinoshita, *Encyclopedia of Bioinformatics*, ed. by the Japanese Society for Bioinformatics, Kyoritsu Shuppan, Tokyo, **2006**, pp. 597–659. Descriptions of the range of success rates are on pp. 653–654.