Journal of Computer Science & Systems Biology - Open Access
Research Article JCSB/Vol.1 2008

# Docking Efficiency Comparison of Surflex, a Commercial Package and Arguslab, a Licensable Freeware

**Abdelouahab Chikhi* and Abderrahmane Bensegueni**

Laboratory of Theoretical Chemistry, Department of Chemistry. Faculty of Sciences, Mentouri University, Constantine, Algeria
Laboratory of Microbiological Applications, Department of Biochemistry-Microbiology. Faculty of natural and life Sciences, Mentouri University, Constantine, Algeria

*Corresponding author: Abdelouahab Chikhi,
E-mail: abchikhi@yahoo.fr; Tel: + 213-793-112-547.

Received September 06, 2008; Accepted December 20, 2008; Published December 26, 2008

**Citation:** Abdelouahab C, Abderrahmane B (2008) Docking Efficiency Comparison of Surflex, a Commercial Package and Arguslab, a Licensable Freeware. J Comput Sci Syst Biol 1: 081-086. doi:10.4172/jcsb.1000007



## Abstract

**Structure-based lead optimization approaches are increasingly playing a role in the drug-discovery process. Virtual screening by molecular docking has become a largely used approach to lead discovery in the pharmaceutical industry when a high-resolution structure of the biological target of interest is available. The performance of two docking programs (Arguslab and Surflex), for virtual database screening, is studied. Surflex is well recognized commercial package while Arguslab is distributed freely for Windows platforms by Planaria Software. Comparisons of these docking programs and scoring functions using a large and diverse data set of pharmaceutically interesting targets and active compounds are carried out. We focus on the problem of docking and scoring flexible compounds which are sterically capable of docking into a rigid conformation of the receptor. The three dimensional structures of a carefully chosen set of 300 pharmaceutically relevant protein-ligand complexes were used for the comparative study. The results show that Surflex outperforms largely Arguslab in all tests studied.**

**Keywords**: Docking programs; Drug discovery; Biological target; Comparative study; ArgusLab; Surflex

## Introduction

The development and implementation of a range of molecular docking algorithms, based on different search methods (Taylor et al., 2002; Halperin et al., 2002) was observed in the last few years. This approach has had several recent successes in drug discovery (Sechi et al., 2005; Liu et al., 2005).

In the field of molecular modeling, docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex (Lengauer and Rarey, 1996). Knowledge of the preferred orientation in turn may be used to predict the strength of association or binding affinity between two molecules using for example scoring functions.

Docking is frequently used to predict the binding orientation of small molecule drug candidates to their protein targets in order to in turn predict the affinity and activity of the small molecule. Hence docking plays an important role in the rational design of drugs (Kitchen et al., 2004). Given the biological and pharmaceutical significance of molecular docking, considerable efforts have been directed towards improving the methods used to predict docking.

Evaluation of existing docking algorithms can assist in the choice of the must suitable docking programs for any particular study. Effectively, several studies estimating and comparing the accuracies of protein-ligand programs like Dock, ICM, Gold have been reported (Perola et al., 2004;

Bursulaya et al., 2003).

The goal of this study was to evaluate the ability of ArgusLab, a freely distributed molecular modeling package in which molecular docking is implemented, to reproduce crystallographic binding orientations and to compare its accuracy with that of a widely well established docking package, Surflex.

## Methods

*ArgusLab4.0* has fast become a favorite introductory molecular modeling package with academics mainly because of its user-friendly interface and intuitive calculation menus (Thompson, 2004). The ArgusDock docking engine, implemented in ArgusLab, approximates an exhaustive search method. Flexible ligand docking is possible with ArgusLab, where the ligand is described as a torsion tree and grids are constructed that overlay the binding site. Ligand's root node (group of bonded atoms that do not have rotatable bonds) is placed on a search point in the binding site and a set of diverse and energetically favorable rotations is created. For each rotation, torsions in breadth-first order are constructed and those poses that survive the torsion search are scored. The N-lowest energy poses are retained and the final set of poses undergoes coarse minimization, re-clustering and ranking.

*Surflex1.3* method is based on an incremental construction. There are three steps in performing dock : choosing how to identify the active site of the protein and constructing a docking target to which to match molecules (called a protomol), docking one or many molecules and post-processing the results.

Each of the basic tasks is controlled by a series of user-settable parameters, but the built-in defaults are reasonably robust to many different protein/ligand pairs. All input molecules must be protonated as expected at physiological pH including non-polar hydrogens**.** The protonation state may strongly affect docking. Docking requires a ligand, a protomol, and a protein. Surflex will fragment the molecule, search the fragments, dock the fragments, and construct the molecule in the active site of the protein. The final output is the top ten scoring conformations (Jain, 2003).

### Docking Protocols

In the two algorithms studied here, the receptor is treated as a rigid body and a grid potential is used to evaluate the scoring functions. This simplification allows one to perform docking more efficiently, which is especially crucial in database screening.

Arguslab requires a PDB format file for both ligand and receptor. The binding site was defined from the coordinates of the ligand in the original PDB file. Argusdock exhaustive search docking engine was used, with grid resolution of 0.40 Å. Docking precision was set to 'high precision' and 'flexible ligand docking' mode was employed for each docking run.

In Surflex, the preferred input file format is Sybyl mol2. Molecular output is generally in Sybyl mol2 format as well. MDL mol or sd file (for ligands) and PDB (for proteins) are also acceptable, although PDB files may generate errors or unexpected results, since the format is frequently variable. There are two scores for each docked conformation: an affinity ($-\log K_d$) and a crash score (also $pK_d$ units). The crash score is the degree of inappropriate penetration into the protein by the ligand as well as the degree of internal self-clashing that the ligand is experiencing (which is also reported). Crash scores that are close to 0.0 are favorable. The reported score includes the crash score reported (and the crash score includes the self-clash term). The final docked conformations are reported in descending order of score, and correspond to the output files final-*.mol2. It's possible to optimize the fine position of a molecule that has already been docked by a command and the optimized conformation is stored in "opt.mol2" (Jain, 2004).

## Results and Discussion

The best ranking poses predicted by the two programs Arguslab and Surflex are shown in the figure 1 and their root mean square deviation (RMSD) values from the original crystallographic pose determined. It can be observed that Surflex surpasses Arguslab. For RMSD interval < 2Å, the difference in docking accuracies between the two programs is so important but decrease significantly in RMSD interval < 3Å.

Figure 2 shows the evaluation of the docking algorithms for their sampling accuracy. The percentage of poses with RMSD within 2Å from the experimental structure was 63% for Surflex and only 33% for ArgusLab. This confirms the results reported by earlier studies, Surflex appears to be highly efficient in terms of sampling (Kellenberger et al., 2004). However, under less rigorous conditions, the performance of ArgusLab is vastly improved with 62% of the top ten poses falling within 3Å of the crystallographic pose. This suggests that ArgusLab still gives some biological results and can be used in educational demonstrations.

Another kind of analysis we have carried out is the effect of a ligand parameter on docking accuracy (figure 3). It is a

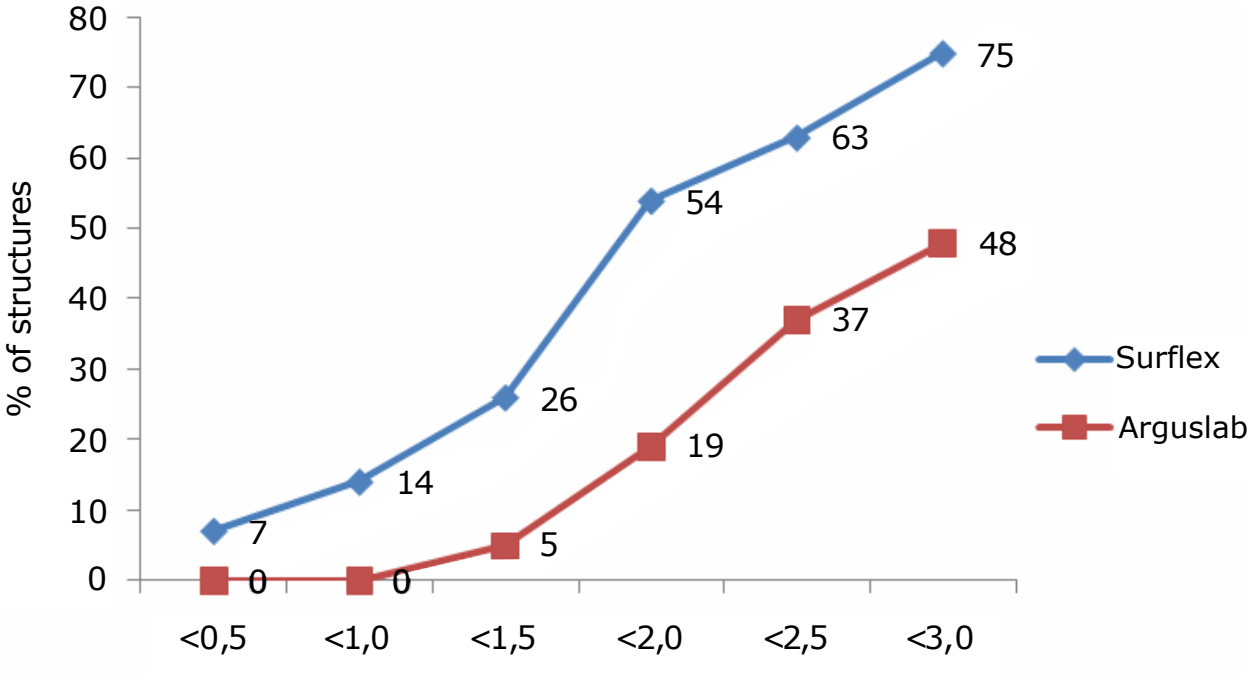


**Figure 1:** Best pose with reference to crystallographic pose.

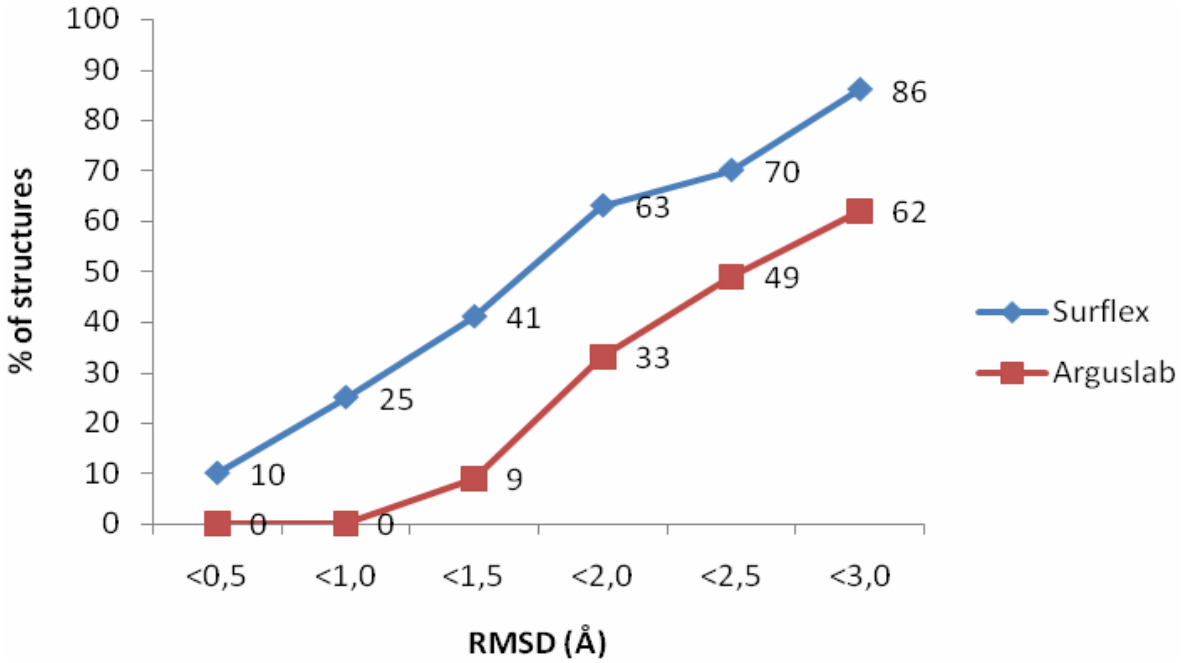


**Figure 2:** Top ten poses with reference to crystallographic pose.

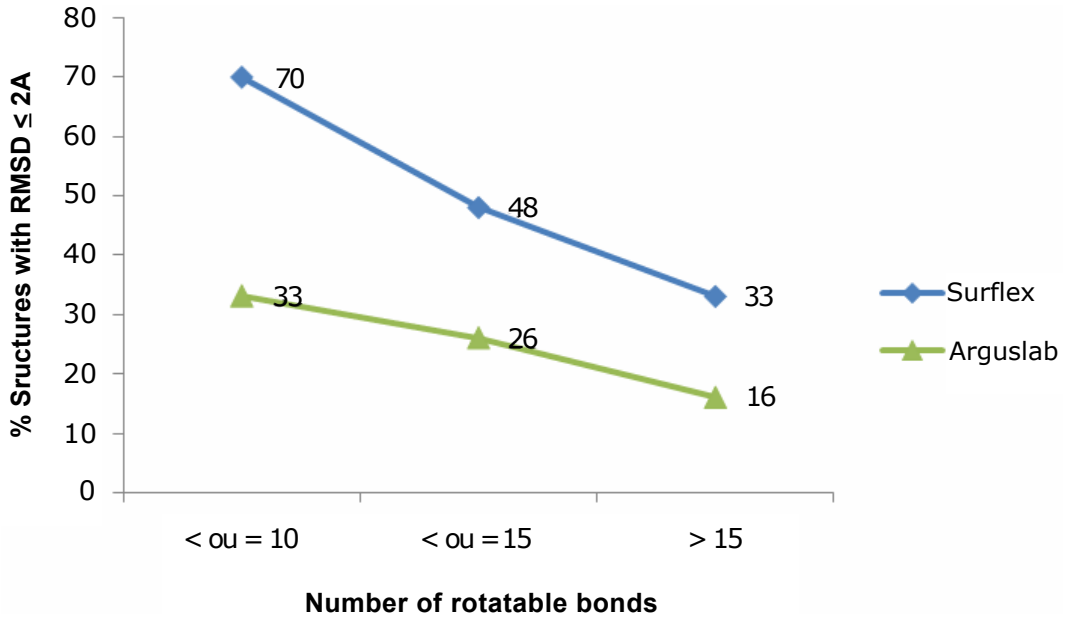


**Figure 3:** Ligand rotatable bonds in relation to docking accuracy.

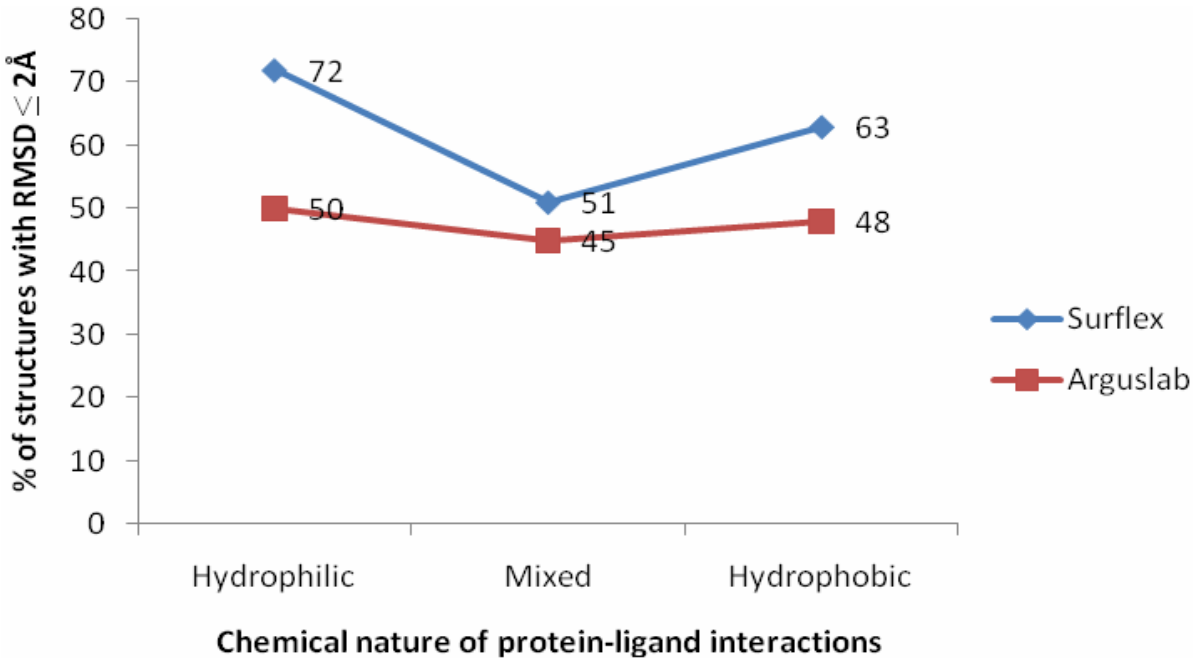


**Figure 4:** % of hydrogen bonding in terms of docking accuracy.

well-known fact that as the number of rotatable bonds of the ligand increases, the docking accuracy falls since a much larger conformational space has to be sampled. The complexes in the present study were divided into three groups, ligands with 1 to ≤10 rotatable bonds, ligands with 11 to ≤ 15 rotatable bonds and those with > 15 rotatable bonds. The results confirm earlier works. Indeed for all algorithms, the docking accuracy decreases when the number of rotatable bonds increases. Also in all cases, accuracy of Surflex is approximately double that of ArgusLab. This decrease is very marked when the number of rotatable bonds exceed 15. Though, an essential remark is that docking time in ArgusLab is typically much shorter than that of Surflex.

To further evaluate these docking programs, another test we have conducted is to study the chemical nature of their protein-ligand interactions and then to check the success rate of each scoring function (figure 4). The classification is aided by using X-Score. For any given protein-ligand complex, if the contribution of the H-bond term in X-Score is 50% larger than the hydrophobic term, it is classified as the "hydrophilic" type. If the contribution of the hydrophobic term is 50% larger than the H-bond term, it is classified as the "hydrophobic" type. Otherwise, the complex is considered to have mixed hydrophilic and hydrophobic factors in the protein-ligand interaction and thus is classified as the "mixed" type. We have used X-Score for this classification process because it is the only one with open source codes, so we can analyze the hydrophobic and the hydrophilic terms conveniently. Hydrogen bond driven complexes are the best results given by Surflex (72%) and also for hydrophobic-burial driven ones (63%). There is no perceptible change in the docking accuracy of ArgusLab with degree of hydrogen bonding.

Studies for determination of $IC_{50}$ and MIC, in specialized laboratory, are needed to confirm these *in silico* results.

## Conclusion

Our results prove that the Surflex program does a rational job in docking and should assist significantly the drug discovery process. Its use for molecular docking appears to be most valuable. This study shows that the commercial package surpasses the freely available docking program in all parameters tested. The study also reveals that, in less scrupulous conditions, ArgusLab can be used for demonstration of molecular docking method to beginners in this area owing to its easiness to use graphical user interface. Moreover, some future advances can be made in this program at the expense of the docking time.

## References

1. ArgusLab 4.0, Mark AT *Planaria Software LLC, Seattle*, http://www.ArgusLab.com.

2. Bursulaya BD, Totrov M, Abagyan R, Brooks CL 3rd (2003) Comparative study of several algorithms for flexible ligand docking. J Comput Aided Mol Des 17: 755-763. » CrossRef » Pubmed » Google Scholar

3. Halperin I, Ma B, Wolfson H, Nussinov R (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins 47: 409-443. » CrossRef » Pubmed » Google Scholar

4. Jain AN (2003) Surflex: Fully Automatic Flexible Molecular Docking Using a Molecular Similarity-Based Search Engine. J Med Chem 46: 499-511. » CrossRef » Pubmed » Google Scholar

5. Jain AN (2004) Virtual screening in lead discovery and optimization. Current Opinion in Drug Discovery & Development 7: 396-403. » CrossRef » Pubmed » Google Scholar

6. Kellenberger E, Rodrigo J, Muller P, Rognan D (2004) Comparative Evaluation of Eight Docking Tools for Docking and Virtual Screening Accuracy. PROTEINS Structure Function and Bioinformatics 57: 225-242. » CrossRef » Pubmed » Google Scholar

7. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Nature reviews Drug discovery 3: 935-949. » CrossRef » Pubmed » Google Scholar

8. Lengauer T, Rarey M (1996) Computational methods for biomolecular docking. Curr Opin Struct Biol 6: 402-406. » CrossRef » Pubmed » Google Scholar

9. Liu Z, Huang C, Fan K, Wei P, Chen H, et al. (2005) Virtual screening of novel noncovalent inhibitors for SARS-CoV 3C-like proteinase. J Chem Inf Model 45: 10-17. » CrossRef » Pubmed » Google Scholar

10. Perola E, Walters WP, Charifson PS (2004) A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. Proteins 56: 235-249. » CrossRef » Pubmed » Google Scholar

11. Sechi M, Sannia L, Carta F, Palomba M, Dallocchio R, et al. (2005) Design of novel bioisosteres of beta-diketo acid inhibitors of HIV-1 integrase. Antivir Chem Chemother 16: 41-46. » CrossRef » Pubmed » Google Scholar

12. Taylor RD, Jewsbury PJ, Essex JW (2002) A review of

protein-small molecule docking methods. J Comput Aided Mol Des 16: 151-166. » CrossRef » Pubmed » Google Scholar

13. Thompson MA (2004) Molecular docking using ArgusLab, an efficient shape-based search algorithm and the AScore scoring function. ACS meeting, Philadelphia 172 CINF 42 PA. » Google Scholar