ELSEVIER

Evaluation of docking strategies for virtual screening of compound databases: cAMP-dependent serine/threonine kinase as an example

Jeffrey W. Godden,* Florence Stahura,* and Jürgen Bajorath*†

**Computational Chemistry and Informatics, MDS Panlabs, Bothell, Washington, USA*
†Department of Biological Structure, University of Washington, Seattle, Washington, USA

In an effort to establish efficient docking routines for computational screening of compound databases on protein structures, cAMP-dependent protein kinase has been selected as a test case and a variety of docking options and scoring functions were compared. These included rigid-body and flexible docking and scoring based on surface complementarity and/or force field energy. Inhibitors were removed from complex crystal structures and added to compound libraries in their binding conformations and, in addition, deliberately modified conformations. Rigid-body docking and contact scoring well reproduced two of three experimental enzyme–inhibitor complexes. Ligand docking with flexible torsional angles failed to do so but anchored search of some inhibitors converged near to experimental structures, however, only when energy scoring was applied. © 1999 by Elsevier Science Inc.

Keywords: computational docking, database screening, protein-inhibitor complexes, ligand flexibility, scoring functions

INTRODUCTION

Computational (virtual) screening of compound databases on three-dimensional (3-D) structures of target proteins is used to aid in the selection of candidate inhibitors for experimental testing.[1] We aim to identify reliable and computationally efficient docking strategies for virtual high-throughput screening of combinatorial libraries. Human cAMP-dependent protein kinase (termed PK)[2,3] was selected as a test case.

Color Plate for this article is on page 165.

Address reprint requests to: Jürgen Bajorath, Computational Chemistry and Informatics, MDS Panlabs, Bothell, Washington 98011-8805, USA.
E-mail: jbajorath@panlabs.com.

PK phosporylates specific Ser/Thr residues in target proteins, using ATP as a phosphate donor. The X-ray structure of PK complexed with MgATP and a peptide inhibitor identified the cofactor and substrate binding sites.[3] The catalytic core is well conserved among members of the Ser/Thr and Tyr kinase families.[4] The 3-D structure is bilobal with a deep active site cleft. X-ray structures of PK in complex with three isoquinoline-sulfonamide derivatives have been determined.[2]

The noncovalent inhibitors studied here bind to the ATP site of PK. The ATP-binding site is largely conserved in the Ser/Thr and Tyr kinase families.[5] Despite conservation of the cofactor-binding site, ATP analog inhibitors were shown to inhibit a variety of protein kinases with distinctly different specificity.[5,6] Thus, the docking strategies investigated in this study should be broadly applicable to identify inhibitors of many different protein kinases.

Using PK as an example, we have compared the ability of different docking routines and scoring functions to reproduce crystallographically determined binding conformations and orientations of known inhibitors. The calculations were carried out using the most recent version of the DOCK program,[7] which includes flexible docking options. For database screening, altered conformations of selected inhibitors were added to a library containing 1 000 compounds. For each docking simulation, the rank of inhibitors among top scoring compounds was determined, and we investigated how the relative ranking changed when docking protocols were modified.

For two of three PK inhibitors, the X-ray complex was well reproduced by rigid-body docking of the bound conformation, and less well reproduced using an anchored search starting from a modified conformation. The score and relative ranking of docked compounds changed significantly when docking parameters were modified. For PK, an effective docking strategy was established.

Journal of Molecular Graphics and Modelling 16, 139–143, 1998

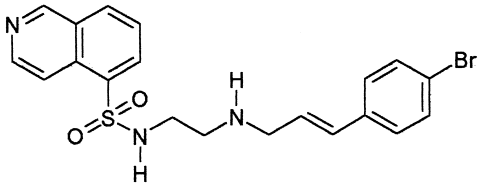
655 Avenue of the Americas, New York, NY 10010

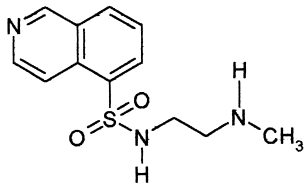
1093-3263/98/$–see front matter
PII S1093-3263(99)00003-0

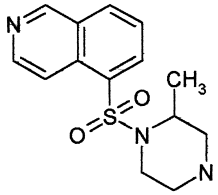
MATERIALS AND METHODS

Molecular graphics and structural manipulations were carried out with MOE (Molecular Operating Environment; Chemical Computing Group, Montreal, Quebec, Canada). All docking calculations were carried out with DOCK 4.0.1.[7] Three-dimensional structures of PK were obtained from the Protein Data Bank.[8] After removal of the inhibitors and water molecules from the coordinate sets, atomic partial charges were calculated with AMBER.[9]

The PK-binding site for docking experiments was created using pdb entry 1yrd after removal of the inhibitor 1-(5-isoquinoline-sulfonyl)-2-methylpiperazine (termed H-7). Twenty-eight residues were selected. These residues corresponded to those within nine Å of the first phosphorus atom in the structure of the ternary complex of PK with MgATP and peptide inhibitor (pdb entry 1atp). Two other PK inhibitors, H-8 (*N*-[2-(methyl-amino)ethyl]-5-isoquinoline-sulfonamide) and H-89 (*N*-[2-(*p*-bromocinamylamino)ethyl]-5-iso-quinoline-sulfonamide), taken from pdb entries 1yds and 1ydt, respectively, were selected. The structures of the different inhibitors are as shown:







Optimization of the dock sphere models was done using MOE to manipulate the first cluster calculated by SPHGEN. A total of 62 PK spheres was generated. The Connolly surface was calculated with the program MS,[10] using a surface density of five surface points per Å^2 and a probe radius of 1.4 Å, which were the default parameters. Dock parameters were chosen to automate the placement of multiple ligands. For rigid-body docking the parameters orient−ligand, score−ligand, multiple−ligand, match−receptor−sites, energy−score, rank−ligands, and automated−matching were all set on (bump−maximum = 0). For flexible docking the parameters flexible−ligand, anchor−search, and multiple−anchors were also selected. All other parameters were set to suggested defaults (http://www.cmpharm.ucsf.edu/kuntz/dock.html). The docking site itself was rigid in all calculations.

A subset of 1 000 compounds was randomly selected from Optiverse, a diverse screening library,[11] and a low-energy 3-D conformation of each compound was generated using MOE. Hydrogen atoms and Gasteiger–Marsili partial charges[12] were added. Hydrogens and partial charges were also added to the three PK inhibitors H-7, H-8, and H-89 in their binding conformations prior to inclusion in the compound library. The three PK inhibitors were also energy minimized and added to the library. In addition, 10 different conformations of each inhibitor, generated by random incremental pulse search (RIPS),[13] were randomly selected and added to Optiverse compounds. RIPS first randomly changes all rotatable bonds in a molecule and then perturbs atomic positions by a specified value, here 1 Å. The conformations are then minimized to correct steric inconsistencies and improve intramolecular contacts. All atom root mean square deviation (rmsd) values after least-squares fitting were calculated for the selected RIPS conformers and the corresponding X-ray structures and ranged from 2.16 to 3.67 Å.

Docking calculations were carried out on an SGI Octane 225-Mhz dual processor workstation. Compounds were docked using rigid-body or flexible docking (either torsional angle or anchored search[14]) and only the best scoring orientation was selected for each molecule. Each docked ligand was ranked on the basis of shape complementarity (contact scoring) and/or AMBER force field energy. The first 500 compounds selected by rigid-body docking and shape complementarity scoring were subjected to flexible docking and re-evaluated by shape complementarity and/or energy scoring. The different ranking positions of the inhibitors were analyzed to determine how the score changed when the docking protocol was modified.

RESULTS AND DISCUSSION

Docking approach

Different docking techniques have been developed.[15] One of our reasons to focus on the use of the DOCK program is the wealth of DOCK applications reported in the literature,[1,15] which help to establish more generally applicable docking protocols. Major limitations of our calculations include the use of molecular mechanics force field energies for scoring and, more importantly, the neglect of solvation effects. Thus, our analysis of protein–ligand interactions is approximate at best. However, since our major goal is to establish efficient docking protocols for virtual screening of large compound databases, the explicit inclusion of solvent in our calculation would be computationally prohibitive. Thus, the calculations represent, by definition, a compromise between accuracy and efficiency.

Structure comparison

In order to analyze how well experimentally known conformations and binding site orientations of inhibitors were reproduced, all atom root mean square deviations were calculated in two different ways. rmsd1 was calculated after least-squares fitting of observed and docked inhibitors, which takes only conformational differences into account (i.e., rmsd1 is 0 for rigid-body docking). rmsd2 was calculated for inhibitor atoms after superposition of the docking sites of the PK X-ray structure, thus taking conformational and positional differences into account.

Rigid-body docking

Using rigid-body docking, two of three crystallographically determined inhibitor positions were well reproduced among top

scoring compounds, H-89 (rmsd2 = 1.13 Å) and H-7 (rmsd2 = 0.74 Å) (Table 1 and Color Plate 1). In contrast, inhibitor H-8 was found in the top 500 list only by shape complementarity-based scoring (rank 111) and had a position very different from the X-ray structure (rmsd2 = 7.84 Å). This may be because this inhibitor is small and can fit many sphere positions within the active site. H-8 was not considered for further docking studies.

Scoring of inhibitors in different conformations

Although the docked H-7 ranked only at position 332 by force field energy scoring, this conformation was closest to the original X-ray structure (rmsd2 = 0.74 Å) (Table 2). This was surprising, since inhibitor H-7 is also a small compound, albeit less flexible than H-8. In contrast, H-7 ranked at position 39 by shape complementarity but did not reproduce the X-ray structure (rmsd2 = 8.02 Å). H-89 gave good results considering its size and flexibility. The original H-89 conformation gave a better result (rmsd2 = 1.13 Å) than the same compound minimized before being docked (rmsd1 = 0.92 Å, rmsd2 = 1.77 Å). Interestingly, the same positions were obtained with force field energy and shape complementarity scoring for original and minimized H-89 but with different relative rankings in both lists (Table 2). The best scoring conformations of inhibitors are the original and minimized X-ray structures. Deliberately modified conformations, generated by RIPS, depart more from the binding conformation and score less well.

Table 1. Root mean square deviation between X-ray structures and modeled inhibitors[a]

A. Rigid-body docking

	H-89-X-ray	H-89-cnt	H-89-nrg	H-89-min-cnt	H-89-min-nrg
H-89-X-ray	0.0	1.126 (0.0)	1.126 (0.0)	1.772 (0.916)	1.772 (0.916)
H-89-cnt		0.0	0.0	1.705 (0.916)	1.705 (0.916)
H-89-nrg			0.0	1.705 (0.916)	1.705 (0.916)

B. Rigid-body docking

	H-7-X-ray	H-7-cnt	H-7-nrg
H-7-X-ray	0.0	8.016 (0.0)	0.742 (0.0)

C. Anchored search docking

	H-89-X-ray	H-89-2-nrg (2.162)	H-89-3-nrg (3.665)	H-89-7-nrg (2.625)
H-89-X-ray	0.0	4.020 (2.740)	3.785 (2.352)	3.687 (2.673)
H-89-2-nrg		0.0	1.614 (1.599)	3.170 (2.660)
H-89-3-nrg			0.0	3.221 (2.836)

D. Anchored search docking

	H-7-X-ray	H-7-10-nrg (2.468)
H-7-X-ray	0.0	3.711 (1.683)

[a] Names of compounds were defined according to the original name of inhibitors: -X-ray for the inhibitors taken from the X-ray structure; -indices (from 1 to 10) for the generated RIPS conformations; -cnt and -nrg for the compounds generated from DOCK 4.0 using shape complementarity and force field energy scoring, respectively. H-89-min-cnt and H-89-min-nrg correspond to the minimized X-ray structures. Rmsd1 (in parentheses) and rmsd2 values, in angstroms, were calculated as described in text using an accessory program of DOCK 4.0. In the top row [for (**C**) and (**D**)], rmsd1 values for X-ray structures and corresponding modified conformations prior to anchored search docking are reported in parentheses.

Table 2. Scoring of the inhibitors using rigid-body docking[a]

Docked compounds	Rank by shape complementarity	Rank by force field energy	Docked compound	Rank by shape complementarity	Rank by force field energy
H-89-X-ray	**267**	**4**	**H-7-X-ray**	39	**332**
H-89-min	**8**	**28**	H-7-min	Not top 500	460
H-89-1	Not top 500	Not top 500	H-7-1	Not top 500	Not top 500
H-89-2	308	267	H-7-2	353	Not top 500
H-89-3	419	232	H-7-3	491	Not top 500
H-89-4	Not top 500	Not top 500	H-7-4	Not top 500	Not top 500
H-89-5	Not top 500	Not top 500	H-7-5	326	452
H-89-6	382	452	H-7-6	Not top 500	Not top 500
H-89-7	278	240	H-7-7	Not top 500	Not top 500
H-89-8	Not top 500	Not top 500	H-7-8	Not top 500	Not top 500
H-89-9	Not top 500	Not top 500	H-7-9	347	Not top 500
H-89-10	220	Not top 500	H-7-10	403	Not top 500

[a] Names of compounds were defined according to the original name of the inhibitors: -X-ray for the inhibitors taken from the X-ray structure; -min for the minimized X-ray structures, and -indices (from 1 to 10) for the top 500 ranked molecules by either shape complementarity or force field energy. The results in boldface indicate the best reproduction of the X-ray structure.

Docking with flexible torsional angles

The top scoring 500 compounds were redocked with flexible torsional angles using the DOCK 4.0.1 standard parameter set. In this case, docking and scoring by shape complementarity and/or force field energy did not produce inhibitor conformations close to the X-ray structures. This protocol was time consuming. Two to 3 days of computing time was required for docking of the top 500 compounds and scoring by shape complementarity and force field energy. On average, rigid-body docking was, for our target and compound database, at least an order of magnitude faster than the corresponding flexible docking calculations.

Anchored search

A so-called anchored search is another form of flexible docking in which ligands are divided, a core is docked, and the remaining fragments progressively added.[14] This search option reproduced the orientation of bound PK inhibitors near to their bound conformations but only on the basis of force field energy scoring. Although the orientations of the inhibitors in the binding site were reproduced, as shown in Color Plate 1, docked inhibitors displayed, in part, significant conformational deviation from the X-ray structures. For H-89, the best fit (rmsd1 = 2.67 Å, rmsd2 = 3.69 Å) was obtained when the search was started from the RIPS-generated conformation (with a rmsd1 value from the binding conformation of 2.63 Å) (Table 1). The relatively large rmsd values for H-98 are due to a flip of the condensated ring moiety, illustrated in Color Plate 1. Similarly, one of the RIPS-generated H-7 conformations converged near to the X-ray structure (rmsd1 = 1.68 Å, rmsd2 = 3.71 Å) and was the top scorer by force field energy. Within the top 500 compounds scored by energy, eight conformers of H-98 were ranked. The worst rank for any of the H-89 starting conformers was 55 for the rigid-body docked X-ray structure. The other conformers were RIPS-generated starting conformations and ranked at positions 54, 52, 30, 17, 16, 8, 7, and 6, respectively. Results obtained for H-7 were not consistent. This is probably because H-7 is less flexible and smaller than H-89. DOCK scores compounds in a cumulative atom-based way and may thus favor larger compounds, dependent on the number of spheres representing the docking site. Regardless, anchored search in conjunction with energy scoring was a suitable tool to "find" PK inhibitors in large compound collections. Thus, similar calculations may identify novel candidate inhibitors for other protein kinases.

CONCLUSIONS

Reproducibility of experimental structures and changes in the relative ranking of compounds were used as criteria to establish a docking strategy for virtual screening. For PK, a first selection of compounds by rigid-body docking and shape complementarity, followed by a re-evaluation of the best scoring compounds using an anchored search with force field energy scoring, was the most effective technique by which to select known inhibitors from a compound database. As discussed, the approach may aid in the identification of novel inhibitors for protein kinases, since their ATP-binding sites are largely conserved.

REFERENCES

1 Kuntz, I.D. Structure-based strategies for drug design and discovery. *Science* 1992, **257,** 1078–1082.

2 Engh, R.A., Girod, A., Kinzel, V., Huber, R., and Bossemeyer, D. Crystal structures of catalytic subunit of cAMP-dependent protein kinase in complex with isoquinolinesulfonyl inhibitors. *J. Biol. Chem.* 1996, **271,** 26157–26164.

3 Zheng, J., Knighton, D.R., Ten Eyck, L.F., Karlsson, R., Xuong, N., Taylor, S., and Sowadski, J.M. Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MgATP and peptide inhibitor. *Biochemistry* 1993, **32,** 2154–2161.

4 Taylor, S., and Radzio-Andzelm, E. Three protein kinase

structures define a common motif. *Structure* 1994, **2,** 345–355.
5 Denny, W.A. ATP site-directed kinase inhibitors. *Drugs* 1998, **1,** 10–12.
6 Gray, N.S., Wodicka, L., Thunnissen, A.W., Norman, T.C., Kwon, S., Espinoza, F.H., Morgan, D.O., Barnes, G., LeClerc, S., Meijer, L., Kim, S. H., Lockhart, D.J., and Schultz, P.G. Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. *Science* 1998, **281,** 533–538.
7 Meng, E.C., Gschwend, D.A., Blaney, J.M., and Kuntz, I.D. Orientational sampling and rigid-body minimization in molecular docking. *Proteins* 1993, **17,** 266–278.
8. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodger, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 1977, **112,** 535–542.
9. Weiner, S. J., Kollman, P. A., Nguyen, D. T., and Case, D. A., An all atom force field for simulation of proteins and nucleic acids. *J. Comp. Chem.* 1986, **7,** 230–252.
10 Connolly, M.L. Analytical molecular surface calculations. *J. Appl. Crystallogr.* 1983, **16,** 548–558.
11 Garr, C.D., Peterson, J.R., Schultz, L., Oliver, A.R., Underiner, T.L., Cramer, R.D., Ferguson, A.M., Lawless, A.S., and Patterson, D.E.J. Solution phase synthesis of chemical libraries for lead discovery. *Biomol. Screening* 1996, **1,** 179–186.
12 Gasteiger, J., and Marsili, M. Iterative partial equalization of orbital electronegativity: Rapid access to atomic charges. *Tetrahedron* 1980, **36,** 3219–3288.
13 Ferguson, D.M., and Raber D.J. A new approach to probing conformational space with molecular mechanics: Random incremental pulse search. *J. Am. Chem. Soc.* 1989, **111,** 4371–4378.
14 Leach, A.R., and Kuntz, I.D. Conformational analysis of flexible ligands in macromolecular receptor sites. *J. Comput. Chem.* 1992, **13,** 730–748.
15 Lengauer, T., and Rarey, M. Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.* 1996, **6,** 402–406.