Exploiting Ordered Waters in Molecular Docking

Niu Huang and Brian K. Shoichet*

Department of Pharmaceutical Chemistry, University of California—San Francisco, 1700 Fourth Street, San Francisco, California 94158-2550

Received May 26, 2008

Abstract: A current weakness in docking is the treatment of water-mediated protein−ligand interactions. We explore switching ordered water molecules "on" and "off" during docking screens of a large library. The method assumes additivity and scales linearly with the number of waters sampled despite the exponential growth in configurations. It is tested for ligand enrichment against 24 targets, exploring up to 256 water configurations. Water inclusion increased enrichment substantially for 12 targets, while most others were largely unaffected.

Ordered water molecules play an important role in protein−ligand recognition, either being displaced on ligand binding or bridging groups to stabilize the complex.[1] In a recent survey of high-resolution structures, over 85% of complexes had one or more water molecules bridging protein and ligand, with an average of 3.5 per complex.[2] Since the identity of the mediating waters can change from ligand to ligand and since many of the waters observed in an apo-structure are displaced by ligand binding, predicting the role of a particular ordered water molecule to ligand binding remains challenging.[3]

The problem of treating ordered waters is acute in molecular docking, which relies on rapid evaluations of discrete states. These rise exponentially with the number of water molecules sampled. Also, it is rarely clear which waters should be treated as displaceable and which should be treated as fixed, despite efforts to categorize them based on environment or crystallographic observation (e.g., thermal factors).[2,4,5] Rarey and colleagues used a "particle concept" to place implicit waters for docking pose prediction.[6] Whereas this improved the fidelity of some ligand geometries, others deteriorated. Other approaches, including GOLD, AUTODOCK, and GLIDE, have also explored incorporating waters (implicitly or explicitly) to predict ligand−protein complexes.[7−13]

Here, we investigate the effect of sampling multiple water positions in docking database screens, which, as far as we are aware, has not been previously investigated. We use this procedure in screens of 24 targets from the DUD database, which contains annotated ligands and property-matched decoys for each target[14] (Table 1). Water molecules within 5 Å of the ligand were obtained from the X-ray structures. All waters that bridged the protein and the ligand, as well as all that formed at least two hydrogen bonds with the protein−ligand complex and with the primary bridging waters, were selected; only those DUD targets that had such water-mediated contacts were included, reducing the targets from 40 to 24. The water hydrogen positions were optimized using the protein local optimization program (PLOP).[15] The resulting orientations were typically unaffected by the presence of the ligand except in two cases (HSP90 and DHFR), consistent with a recent study.[13]

* To whom correspondence should be addressed. Phone: 1-415-514-4126. Fax: 1-415-514-4260. E-mail: shoichet@cgl.ucsf.edu.

Table 1. Docking calculation times with (wat) and without (ref)[a] displaceable waters

			total time (h)		
protein	no. wat	no. config	ref	wat	performance factor
CDK2	7	128	571.7	2078.9	35.2
EGFr	6	64	1681.4	4716.2	22.8
FGFr1	3	8	2088.1	5347.9	4.8
HSP90	6	96	177.4	1479.3	11.5
SRC	6	64	3192.3	9529.9	21.4
TK	5	32	56.5	207.5	8.7
VEGFr2	6	64	2038.4	11188.4	11.7
Fxa	5	32	315.0	1794.5	5.6
Thrombin	5	32	231.5	1495.4	5.0
Trypsin	5	32	710.1	2488.5	9.1
ACE	6	64	119.8	866.7	8.8
ADA	7	128	105.8	885.6	15.3
COMT	2	4	606.7	1535.9	1.6
PDE5	7	128	780.1	3160.1	31.6
DHFR	2	6	960.6	1750.0	3.3
GART	1	2	79.1	140.9	1.1
AChE	8	256	227.0	2009.4	28.9
AmpC	6	216	607.4	4447.1	29.5
GPB	8	256	102.5	1635.0	16.1
HIVPR	4	16	2049.4	5001.1	6.6
HMGR	3	8	505.6	1301.8	3.1
NA	4	16	275.5	1302.1	3.4
PNP	3	8	51.3	130.6	3.1
SAHH	1	2	28.7	37.6	1.5

[a] Total time reported for 1 CPU in hours. For each target, the entire DUD set of over 97 000 molecules was docked.

To sample multiple waters positions, we adapted a flexible-receptor docking method that treats multiple flexible regions of the protein independently, recombining them to generate discrete conformations.[16] Because the method assumes additivity among the regions, it scales linearly rather than exponentially with degrees of freedom. This is, of course, an approximation. We modified the approach to treat individual water molecules as the flexible receptor regions. Each water is represented in the "off" state (displaced) or one of several "on" states (retained, there can be more than one configuration for a retained water molecule), and all waters are treated as equally displaceable; we did not consider the differential energy of water binding. This is a second approximation.

For each water molecule a separate electrostatic and van der Waals potential map is calculated. Every docked molecule is scored against each individual water potential grid, as well as an overall grid representing the rest of the invariant protein. For every water molecule the "off" state or one of the "on" states is chosen on the basis of whether the overall interaction of the protein and waters with the ligand is improved. The optimal water configuration for any docked molecule is assembled from the best state for each water, and the score of the docked molecule is summed from the ligand−protein and ligand−water interactions. For instance, AChE has eight water molecules that were sampled in the docking calculation (Figure 1). In their best orientations, ligand 1 interacts with four waters but displaces four others; ligand 2 chooses five waters displacing three others, while ligand 3 keeps only three waters and displaces the other five.

We assessed the quality of the docking screen by the enrichment of annotated ligands from the vast majority of DUD molecules that are decoys. Enrichment factors measure the prioritization of ligands among the top ranks of a docking-ordered library. For instance, if 10% of the ligands are found

10.1021/jm8006239 CCC: $40.75
Published on Web 08/05/2008

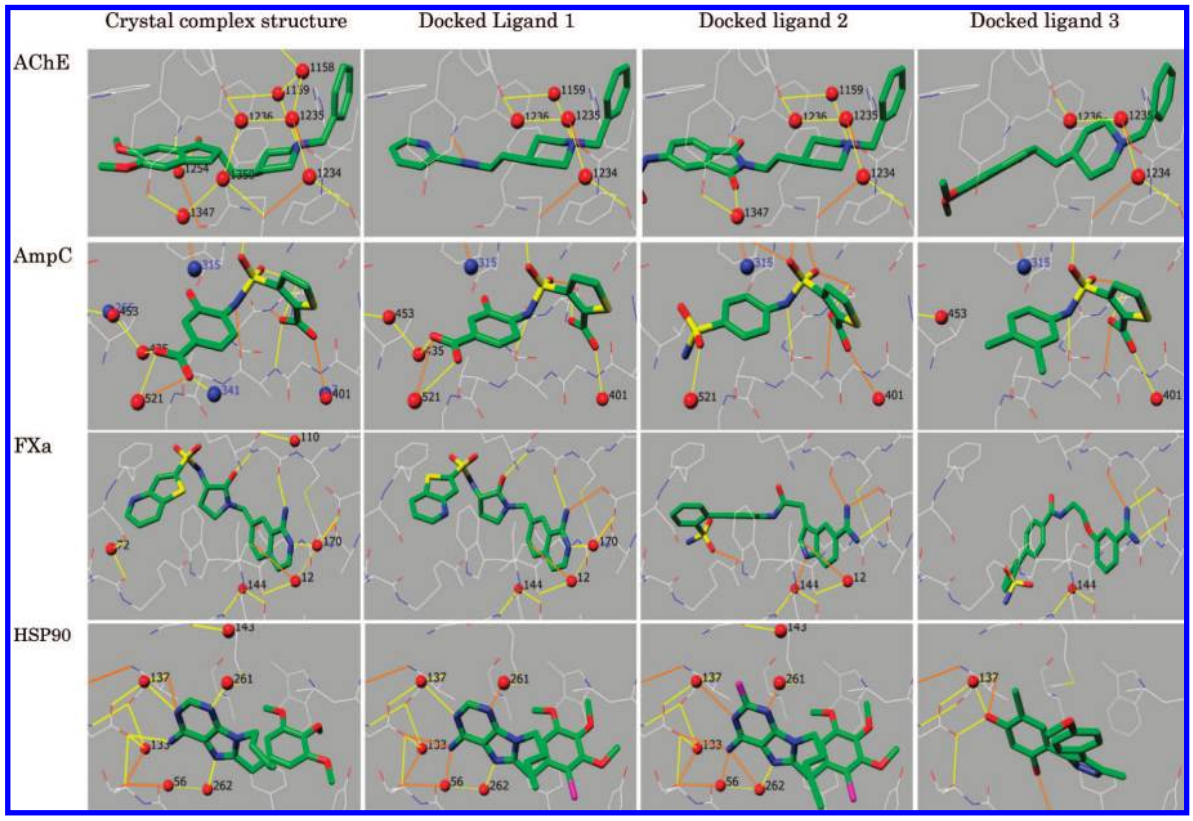

Figure 1. Docking poses of representative ligands (carbons in green) from the library screens showing the roles of different specific water configurations. Note that we combined waters from two different crystal structures for AmpC (waters from 1xgj colored in red and from 1l2s colored in blue). Images generated with Chimera.[17]

among the top 0.1% of a docking-ranked database, the enrichment factor at this point is 100, if 20% of the ligands are found among the top 1% of the ranked database, the enrichment factor is 20. For each of the 24 proteins targeted here there were typically dozens to hundreds of annotated ligands, each with 36 decoys that resembled them physically but differed from them topologically. The overall DUD database consists of 40 targets, 2950 annotated ligands, and 95 316 decoys (for some DUD targets bridging waters were not observed, excluding them from this study). We compared docking screens against the naked protein without ordered water molecules to exactly the same screen where up to eight ordered waters were modeled.

The screens were evaluated by four criteria: How often were the enrichment factors improved by including displaceable waters? Were the docked poses sensible? Did success hinge on the inclusion of particular waters? What was the calculation cost? Between one and eight water molecules, or 2−256 configurations, were modeled in the 24 targets. For 12 targets enrichment increased noticeably (Figure 2). For targets such as CDK2, the improvement was modest, with the enrichment factor at 1% and 20% of the ranked database (EF_1 and EF_{20}[a]) increasing from an EF_1 of 0 (no ligands found at all) without ordered waters to 2.0 when they were included and a corresponding increase at EF_{20} of 0.4−2.2 (Table S1, Supporting Information). For targets like COMT, the improvements were more substantial, from an EF_1 of 8.2 without ordered water molecules to one of 41.2 when waters were modeled. For 11 of the 24 targets enrichment was largely unaffected; for 7 of these, however, enrichment factors were already high even without the waters, leaving little room for improvement. For one target, VEGFr, enrichment was slightly diminished by the inclusion of displaceable water molecules. Ligand enrichment was matched by docked ligand geometries that corresponded well to the crystallographic result (Figure 1).

Are the increased enrichments due to biased choice of waters? Although most were picked from those mediating specific ligand−protein interactions, several considerations moderate this bias. First, the enrichment calculations consider tens to hundreds of ligand molecules, not simply the one observed crystallographically. Second, for four targets, factor Xa, trypsin, AmpC, and neuraminidase, we purposely included apo structure water molecules that are displaced on ligand binding (Figure S1, Supporting Information). The enrichment curves were undiminished relative to using waters only from the holo complex. Third, we repeated the docking screens with the formerly displaceable waters fixed in the site. This diminished enrichment substantially for 15 targets, left it unaffected for 8, and improved for only 1 (Figure S2, Supporting Information). Finally, for AmpC water molecules from two different structures were combined, still resulting in improved enrichment.

The improved enrichment factors were accompanied by a large increase in the degrees of freedom of the system: from a single naked site to one with up to 256 water configurations. If calculation time increased correspondingly, the approach would

[a] Abbreviations: CDK2, cyclin-dependent kinase 2; EGFr, epidermal growth factor receptor; FGFr1, fibroblast growth factor receptor kinase; HSP90, human heat shock protein 90; SRC, tyrosine kinase SRC; TK, thymidine kinase; VEGFr2, vascular endothelial growth factor receptor; FXa, factor Xa; ACE, angiotensin-converting enzyme; ADA, adenosine deaminase; COMT, catechol *O*-methyltransferase; PDE5, phosphodiesterase 5; DHFR, dihydrofolate reductase; GART, glycinamide ribonucleotide transformylase; AChE, acetylcholinesterase; AmpC, AmpC β-lactamase; GPB, glycogen phosphorylase β; HIVPR, HIV protease; HMGR, hydroxymethylglutaryl-CoA reductase; NA, neuraminidase; PNP, purine nucleoside phosphorylase; SAHH, *S*-adenosylhomocysteine; EF_1 and EF_{20}, enrichment factors at 1% and 20%.

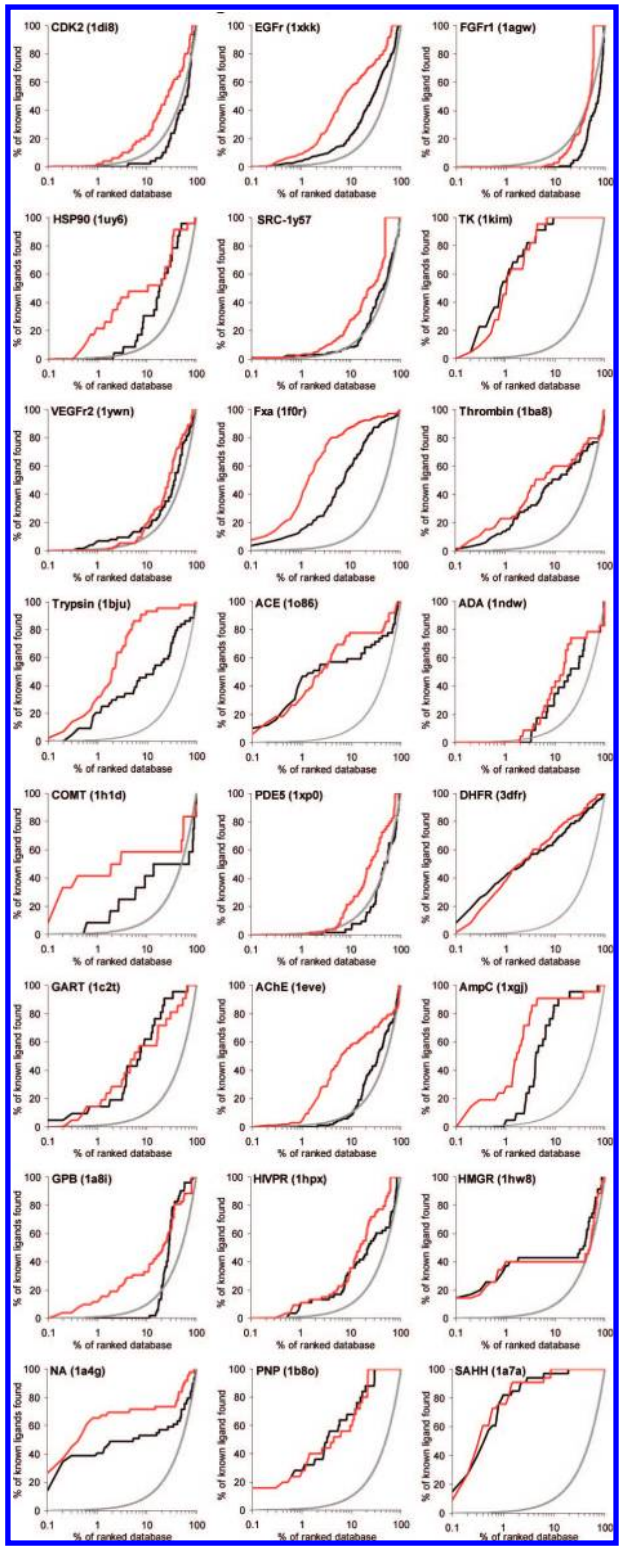

Figure 2. Docking enrichment plots for 24 protein targets. The docking ranked database (*x*-axis) vs the percentage of ligands found (*y*-axis) at any percentage of the ranking. The gray line is random selection; the black line is docking enrichment neglecting the binding site waters, and the red line is the enrichment with displaceable waters. Targets are listed in same order as in Table 1.

be impractical. However, treating each water molecule independently led to run times that scaled linearly with the number of water molecules, not as their exponent. For instance, for HIV protease and neuraminidase (4 water molecules, 16 configurations) the database screen took 2.5- and 4.7-fold longer than the screen against the naked receptor alone (2049 and 275 CPU

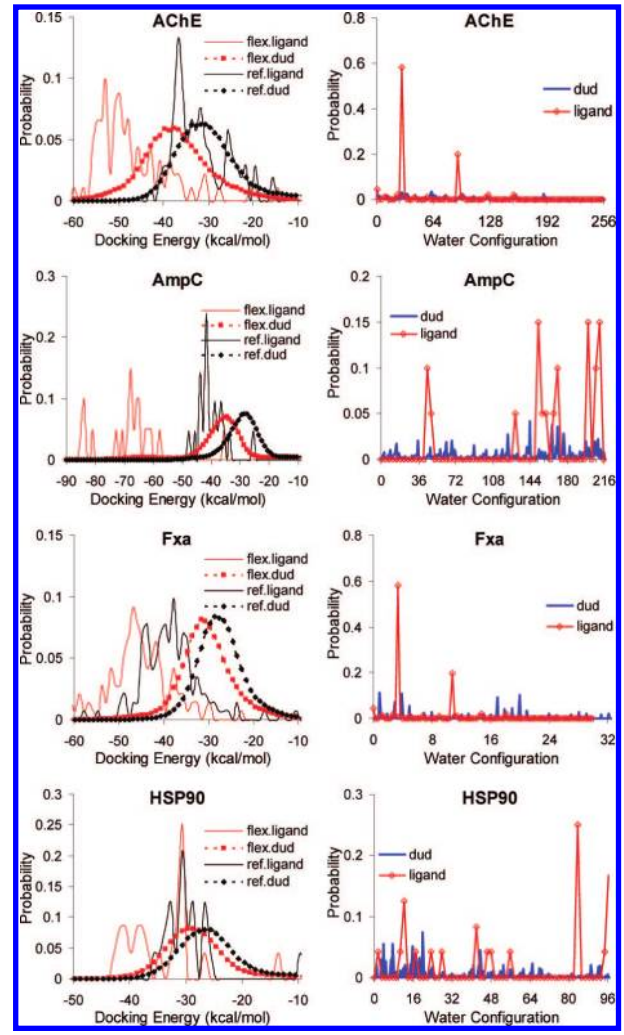

Figure 3. Docking energies for ligands and decoys with and without displaceable waters for four targets. (Left) The distribution of docking energies (*x*-axis) and ligand and decoy ranks expressed as a normalized probability (*y*-axis): ligands (solid line) and decoys (dotted line) with (red) and without (black) displaceable waters. (Right) The presence of particular water configurations for high-scoring ligand (red) and decoy (blue) geometries, numbered sequentially.

hours vs 5001 and 1302 with the waters included). For AChE and GPB (8 water molecules, 256 configurations) the calculation time was 8.9- and 15.9-fold longer than screening the naked receptor (Table 1). Admittedly, neither the linear increase in run time nor the memory allotted to the scoring grids can be extended indefinitely, but for now both are within reason for modern computational systems.

In principle, one can imagine how the approximations in this approach might doom it. The assumption that each water behaves independently can never be wholly true. Still, if the waters are sufficiently far from each other, their cooperativity may be modest.[11] More dubious is the assumption that the displacement energy of every water is identical. Quantifying this differential displacement cost should further improve the screening results.

More generally, increasing degrees of freedom in docking, for instance by sampling alternative protein conformations or relaxing the protein, can diminish docking quality unless they are carefully balanced with internal energy considerations. Here, although sampling ordered waters improves the docking scores for ligands and decoys, the improvement is substantially better for the ligands (Figure 3). Since the waters are displaceable and different waters are exploited by different ligands (Figure 3),

this cannot be explained by the trivial exclusion of all but the true ligands by the mere presence of the waters. Intriguingly, whereas several ligand water configurations are exploited, they are dominated by a few (10 of a possible 216 in the case of AmpC β-lactamase, for instance, Figure 3). For decoys the water configurations are more evenly distributed among the possible configurations (Figure 3).

If there is no first principle reason why this method *should* work, given that it ignores the internal energy differences among the different ordered water structures, the observation that it does work for many targets, and rarely degrades performance for others, may be widely exploited. There is little in the approach that is specific to the docking program used here (DOCK3.5.54), and it may be straightforward to implement in other platforms. Modeling ordered, displaceable water molecules may amplify the specificity for ligands implicit in the structure of the protein without greatly affecting decoy molecules.

Acknowledgment. This work was supported by NIH Grant GM59957. We thank Binqing Wei for algorithmic advice and Sarah Boyce and Michael Mysinger for reading this manuscript.

Supporting Information Available: The affect of modeling waters displaced on ligand binding, the dependence of run time, and enrichment on the number of waters modeled. This material is available free of charge via the Internet at http://pubs.acs.org. Protein structures and water positions are available from http://dud.docking.org/jmc-water.

References

(1) Barillari, C.; Taylor, J.; Viner, R.; Essex, J. W. Classification of water molecules in protein binding sites. *J. Am. Chem. Soc.* **2007**, *129*, 2577–2587.
(2) Lu, Y.; Wang, R.; Yang, C. Y.; Wang, S. Analysis of ligand-bound water molecules in high-resolution crystal structures of protein−ligand complexes. *J. Chem. Inf. Model.* **2007**, *47*, 668–675.
(3) Mancera, R. L. Molecular modeling of hydration in drug design. *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 275–280.
(4) Fitzpatrick, P. A.; Steinmetz, A. C. U.; Ringe, D.; Klibanov, A. M. Enzyme crystal structure in a neat organic solvent. *Proc. Nat. Acad. Sci. U.S.A.* **1993**, *90*, 8653–8657.
(5) Raymer, M. L.; Sanschagrin, P. C.; Punch, W. F.; Venkataraman, S.; Goodman, E. D.; Kuhn, L. A. Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. *J. Mol. Biol.* **1997**, *265*, 445–464.
(6) Rarey, M.; Kramer, B.; Lengauer, T. The particle concept: placing discrete water molecules during protein−ligand docking predictions. *Proteins* **1999**, *34*, 17–28.
(7) Minke, W. E.; Diller, D. J.; Hol, W. G.; Verlinde, C. L. The role of waters in docking strategies with incremental flexibility for carbohydrate derivatives: heat-labile enterotoxin, a multivalent test case. *J. Med. Chem.* **1999**, *42*, 1778–1788.
(8) Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* **2002**, *46*, 34–40.
(9) de Graaf, C.; Pospisil, P.; Pos, W.; Folkers, G.; Vermeulen, N. P. Binding mode prediction of cytochrome p450 and thymidine kinase protein−ligand complexes by consideration of water and rescoring in automated docking. *J. Med. Chem.* **2005**, *48*, 2308–2318.
(10) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R. Modeling water molecules in protein−ligand docking using GOLD. *J. Med. Chem.* **2005**, *48*, 6504–6515.
(11) Corbeil, C. R.; Englebienne, P.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. *J. Chem. Inf. Model.* **2007**, *47*, 435–449.
(12) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein−ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.
(13) Roberts, B. C.; Mancera, R. L. Ligand−protein docking with water molecules. *J. Chem. Inf. Model.* **2008**, *48*, 397–408.
(14) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789–6801.
(15) Huang, N.; Kalyanaraman, C.; Irwin, J. J.; Jacobson, M. P. Physics-based scoring of protein−ligand complexes: enrichment of known inhibitors in large-scale virtual screening. *J. Chem. Inf. Model.* **2006**, *46*, 243–253.
(16) Wei, B. Q.; Weaver, L. H.; Ferrari, A. M.; Matthews, B. W.; Shoichet, B. K. Testing a flexible-receptor docking algorithm in a model binding site. *J. Mol. Biol.* **2004**, *337*, 1161–1182.
(17) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera−a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

JM8006239