

# Predicting Fragment Binding Poses Using a Combined MCSS MM-GBSA Approach

Muhammad K. Haider,<sup>†</sup> Hugues-Olivier Bertrand,<sup>‡</sup> and Roderick E. Hubbard<sup>\*,†,§</sup>

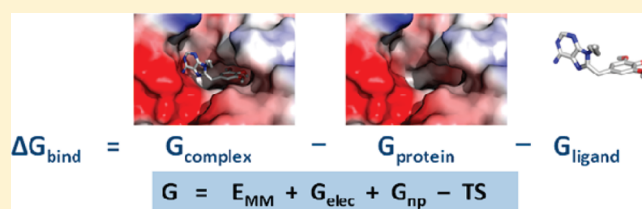
<sup>†</sup>York Structural Biology Laboratory, University of York, Heslington, York YO10 5DD, U.K.

<sup>‡</sup>Accelrys, 20 rue Jean Rostand, 91898 Orsay Cedex, France

<sup>§</sup>Vernalis (R&D) Ltd., Granta Park, Cambridge CB21 6GB, U.K.

 Supporting Information

**ABSTRACT:** Improved methods are required to predict the position and orientation (pose) of binding to the target protein of low molecular weight compounds identified in fragment screening campaigns. This is particularly important to guide initial chemistry to generate structure–activity relationships for the cases where a high resolution structure cannot be obtained. We have assessed the benefit of an implicit solvent method for assessment of fragment binding poses generated by the Multiple Copy Simultaneous Search (MCSS) method in CHARMM. Additionally, the effect of using multiple receptor structures for a flexible receptor is investigated. The original MCSS performance –50% of fragment positions accurately predicted and scored – was increased up to 67% by scoring MCSS energy minima with a Molecular Mechanics Generalized Born approach with molecular volume integration and Surface Area model (MM-GBSA). The same increase in performance (but occasionally for different targets) was observed when using the docking program GOLD followed by MM-GBSA rescoring. The combined results from both methods resulted in a higher success rate emphasizing that a comparison of different docking methods can increase the correct identification of binding poses. For a receptor where multiple structures are available, Hsp90, the average performance on randomly adding receptor structures was also investigated. The results suggest that predictions using these docking methods can be used with some confidence to guide chemical optimization, if the structure of the target either remains relatively fixed on ligand binding, or if a number of crystal structures are available with diverse ligands bound and there is information on the positions of key water molecules in the binding site.



## INTRODUCTION

Fragment-based methods for the discovery of potent inhibitors of drug targets have gained widespread interest in recent years. Several recent reviews have summarized fragment-based lead discovery (FBLD),<sup>1–4</sup> and a number of examples have been published that demonstrate the various medicinal chemistry strategies to develop low-affinity fragments into high-affinity inhibitors against different targets.<sup>5</sup> The size of fragments (molecular weight <250 Da) allows a smaller library of compounds to sample a large chemical space and provide higher hit rates than screening of larger compounds as in high throughput screening (HTS).<sup>6</sup> Due to their size and low affinity, experimental screening of fragments requires biophysical methods such as Nuclear Magnetic Resonance (NMR), Surface Plasmon Resonance (SPR), and X-ray crystallography.<sup>3</sup> Most targets generate a large number of fragment hits, and the primary challenge is therefore to decide on the strategy for choosing and optimizing a fragment to a higher affinity compound. To date, all successful fragment-based lead discovery campaigns have relied on a high resolution structure (from NMR spectroscopy or X-ray crystallography) to reveal the fragment pose and allow the initial rounds of chemical optimization to be designed. In the

absence of such structural guidance, it can be difficult to identify chemical modifications that generate the necessary structure–activity relationships (SAR) to guide further rounds of chemical optimization. It would therefore be extremely useful if computational methods could give some guidance on possible binding modes for those cases where experimental structural information is not available.

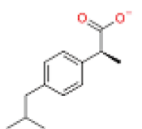
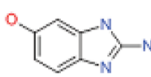
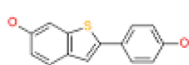
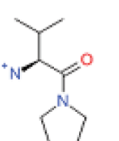
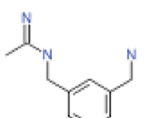
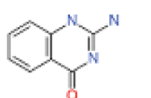
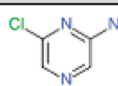
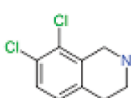
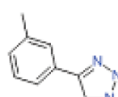
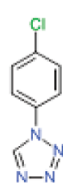

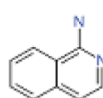
A number of algorithms to predict binding modes and affinity of small-molecule fragments and functional groups in binding sites were developed before FBLD came into practice. The two most notable early efforts were the GRID program by Goodford<sup>7</sup> and the Multiple Copy Simultaneous Search (MCSS) method from Miranker and Karplus.<sup>8</sup> These methods probe the protein active site for energetically favorable positions of polar, nonpolar, and charged functional groups. These positions are then ranked according to an energy function which for MCSS is the CHARMM force-field and for GRID is an empirical energy function. These methods have been combined with different strategies for the construction of ligands from functional group

**Received:** December 2, 2010

**Published:** April 29, 2011

Table 1. List of Complexes in Fragment Docking Data Set

PDB code	receptor	resolution (Å)	reference
1EQG	cyclooxygenase-1 (COX-1)	2.60	Selinsk et al., 2001 <sup>46</sup>
1FV9	urokinase (uPA)	3.00	Hajduk et al., 2000 <sup>47</sup>
1GWQ	estrogen receptor $\alpha$ (ER)	2.45	Warnmark et al., 2002 <sup>34</sup>
1N1M	dipeptidyl peptidase IV (DPP-IV)	2.50	Rasmussen et al., 2003 <sup>48</sup>
1QWC	neuronal nitric oxide synthase (nNOS)	2.30	Fedorov et al., 2003 <sup>49</sup>
1S39	tRNA guanine transglycosylase (TGT)	1.95	Meyer et al., 2004 <sup>50</sup>
1WCC	cyclin dependent kinase 2 (CDK2)	2.20	Hartshorn et al., 1999 <sup>51</sup>
1YZ3	phenylethanolamine N-methyl transferase (PNMT)	2.40	Wu et al., 2005 <sup>52</sup>
2ADU	methionine aminopeptidase (MetAp2)	1.90	Kallander et al., 2005 <sup>35</sup>
2C90	thrombin	2.25	Howard et al., 2006 <sup>53</sup>
2JJC	heat shock protein 90 (Hsp90)	1.95	Congreve et al., 2008 <sup>1</sup>
2OHK	$\beta$ -secretase (BACE-1)	2.20	Murray et al., 2007 <sup>54</sup>

PDB	Fragment	RB	MW
1EQG		4	206
1FV9		0	149
1GWQ		0	242
1N1m		2	171
1QWC		3	177
1S39		0	161
1WCC		0	130
1YZ3		0	202
2ADU		0	161
2C90		0	181
2JJC		0	95
2OHK		0	144

**Figure 1.** List of fragments in the fragment docking data set, RB: rotatable bonds (include bonds that are considered to confer flexibility to the ligand), MW: molecular weight.

positions and binding modes<sup>9,10</sup> and applied to different protein targets such as HIV-1 aspartic proteinase,<sup>9</sup> human  $\alpha$ -thrombin,<sup>11</sup> picornavirus capsid proteins,<sup>12</sup> class II MHC protein HLA-DR4,<sup>13</sup> and others.<sup>14,15</sup>

The binding of small probe molecules such as organic solvents to protein binding sites was also developed in an X-ray crystallographic technique called Multiple Solvent Crystal Structures (MSCS).<sup>16,17</sup> Comparison of the experimental positions of solvent probes with predicted positions from MCSS and/or

GRID was performed for RNase A,<sup>18</sup> thermolysin,<sup>19</sup> and elastase.<sup>20</sup> Good correlation was obtained in some cases, but in others it highlighted the shortcomings of the approach. Most of the poor quality predictions were associated with the lack of appropriate treatment of desolvation, false-positive energy minima, and conformational flexibility of the receptor.<sup>15</sup> Other approaches have been developed for computational mapping such as the SEED (solvation energy for fragment docking) approach for fragment docking based on continuum solvent

approximation,<sup>21</sup> FT-Map which is based on the Fourier domain correlation technique,<sup>22</sup> and the solvent mapping algorithm based on 3D RISM theory.<sup>23</sup>

More rigorous scoring functions can be used to evaluate the energy minima obtained from such mapping algorithms. One example is the estimation of the binding free energy of MCSS minima based on contributions from bonding energy terms, van der Waals interaction, and polar and nonpolar solvation free energy.<sup>11</sup> The polar solvation free energy contribution was further divided into intermolecular shielded electrostatic interactions, protein and ligand desolvation energy which was evaluated by solving the linearized Poisson–Boltzmann (PB) equation. The ranking of energy minima was shown to be in agreement with experimental data. The solvation correction in this way is accompanied by additional computational cost.<sup>11,15</sup>

There have recently been significant developments in the use of implicit solvent methods in biomolecular simulations and estimation of binding affinity for protein ligand complexes.<sup>24</sup> Scoring schemes based on such methods have been employed in docking and virtual screening.<sup>25–28</sup> Semianalytical approximations to the PB equation such as Generalized Born (GB) equation and its variations have been developed to reproduce solvation energies approaching the accuracy of PB equation at less computational cost.<sup>29</sup>

The main aims of the study reported here are to investigate if physics-based scoring methods with a fast implicit solvent model can improve the performance of MCSS in predicting fragment poses. This was achieved by assessing how well MCSS can reproduce the experimental binding pose of a set of experimentally determined protein-fragment complexes, followed by re-evaluation of the results after scoring MCSS generated poses with a Molecular Mechanics and Generalized Born with Surface Area (MM-GBSA) approach. A similar set of calculations was undertaken using the docking program, GOLD.<sup>30</sup> Finally, the effect of using multiple receptor structures on docking performance by MCSS was assessed for a receptor for which multiple crystal structures were available.

## DATA SETS

**Fragment Docking Data Set.** A set of 12 protein-fragment complexes has been reported in a recent review.<sup>1</sup> It consists of targets that have been investigated using fragment based methods and for which X-ray crystallographic structures bound to various fragments are available in the PDB.<sup>31</sup> This data set will be referred to as the ‘fragment docking data set’. The protein targets, their corresponding PDB codes, and literature references are given in Table 1. The chemical structure for each of the fragments, its molecular weight and the number of rotatable bonds are shown in Figure 1.

**Hsp90 Data Set.** The Hsp90 entry in the fragment-docking data set, 2JJC, and an additional set of 10 HSP90 complexes were obtained from the PDB to investigate the use of multiple receptors in the current docking/scoring scheme. The complexes were chosen with the criterion that bound ligands are fragment-sized compounds with less than 20 heavy atoms. This data set will be referred to as the ‘Hsp90 data set’. The protein targets, their corresponding PDB codes, and literature references are listed in Table 2, and the chemical structure for each of the Hsp90 ligands, its molecular weight, and the number of rotatable bonds are also shown in Figure 2.

**Table 2. List of Hsp90-Fragment Complexes Used in the Hsp90 Data Set**

PDB code	resolution (Å)	reference
1QYE	2.10	Soldano et al., 2003 <sup>55</sup>
1ZWH	1.65	Immormino, R. M. (to be published)
2CCS	1.79	Barril et al., 2006 <sup>56</sup>
2JJC	1.95	Congreve et al., 2008 <sup>1</sup>
2QF6	3.10	Huth et al., 2007 <sup>41</sup>
2QFO	1.68	Huth et al., 2007 <sup>41</sup>
2WI1	2.30	Brough et al., 2009 <sup>57</sup>
2WI2	2.09	Brough et al., 2009 <sup>57</sup>
3BM9	1.60	Gopalsamy et al., 2008 <sup>58</sup>
3EKO	1.55	Kung et al., 2008 <sup>45</sup>
3FT5	1.90	Barker et al., 2009 <sup>59</sup>

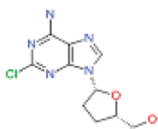
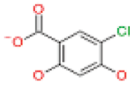
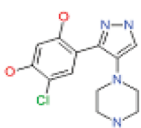

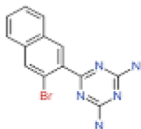
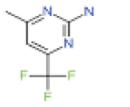
## METHODS

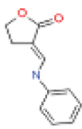
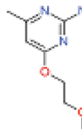
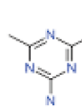
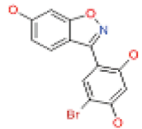
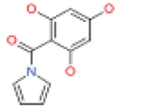
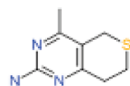
**Preparation of Receptors and Fragments.** All calculations were performed in Discovery Studio 2.5.<sup>32</sup> In order to prepare receptor structures, all ligands were removed from both data sets. For the fragment docking data set, all waters were removed except for Hsp90 (2JJC) and PNMT (1YZ3), where water molecules that bridge interactions between the fragment and the protein binding site were included in the calculations. For the Hsp90 data set, up to four highly conserved water molecules were also included in the binding site of all protein structures, except for 1ZWH, 2CCS, 3BM9, and 3EKO, where one of these water molecules was absent. All ions were removed except for MetAp2 (2ADU) where two cobalt (Co<sup>2+</sup>) ions are involved in key interactions with the bound fragment. Similarly, for 1QWC, the heme prosthetic group was retained in the binding site.

Where two or more copies of the protein-fragment complex were present in the asymmetric unit, the copy with the lowest average B-factors of binding site residues was selected. Only one set of conformations (conformer A) was kept for side chains with alternate conformers.

The CHARMM Momany and Rone force field<sup>33</sup> was used for the protein. Histidine residues were treated as neutral. These include His residues interacting with the fragments (for 1GWQ and 2ADU) in which case it was confirmed from the literature<sup>34,35</sup> that there was no apparent change in protonation state of His residues associated with fragment binding. The binding sites for docking were defined for each receptor as an 8.0 Å sphere from the center of the fragment binding position.

The fragments were prepared as follows. The coordinates were extracted from the original PDB files and visually inspected to correct bond orders and hydrogen atoms added to complete valency. The atom types in the fragment (and cofactor when present) were assigned using the automated typing rules in the Discovery Studio 2.5 package. These use a set of rules to assign types based on local topology such as nearest neighbor connectivity, hybridization, and aromaticity. Parameters were assigned from the CHARMM Momany and Rone forcefield<sup>33</sup> and assigned MMFF94<sup>36</sup> charge. For 1QWC, the heme group in the binding site was typed in the same way but using Gasteiger charges for heme atoms. Missing forcefield parameters were estimated automatically based on similar combinations already available in the parameter list. The fragment structures were minimized with CHARMM<sup>37</sup> (ABNR Gradient: 0.1 kcal/mol/Å).

PDB	Fragment	RB	MW
1QYE		2	270
1ZWH		0	188
2CCS		2	295
2JJC		0	95
2QF6		0	316
2QFOa		1	177

PDB	Fragment	RB	MW
2QFOb		0	189
2WI1		4	183
2WI2		1	156
3BM9		0	322
3EKO		1	219
3FT5		0	181

**Figure 2.** List of fragments in the Hsp90 data set, RB: rotatable bonds (include bonds that are considered to confer flexibility to the ligand), MW: molecular weight.

**MCSS Protocol.** In a typical MCSS protocol, several copies (up to 1000) of a functional group are placed in the binding site and then simultaneously energy-minimized such that copies experience only the field from the protein. If more than one copy converges to the same position, only one copy is retained. This generates a collection of functional group poses, each of which is associated with a position, interaction geometry, and energy score. In this study, 750 copies of each fragment were placed inside the binding site of the receptor structure. A distance-dependent dielectric model was used to approximate the solvent (using default value for EPS = 1).

The energy minimization was carried out by performing an initial 500 steps of steepest descent followed by 300 steps of additional steepest descent and then 20 repetitions of 500 steps of conjugate gradient minimization. At each repetition, copies are checked for convergence and where two copies overlap with an rmsd of less than 0.2 Å, one is removed. At the end of the MCSS run, energy minima within 2.0 Å rmsd of each other were considered as one cluster, and the copy within each cluster with the lowest energy was selected as cluster representative.

For the fragment docking data set, MCSS calculations were performed for each fragment in its own receptor. For the Hsp90 data set set cross-docking calculations were also performed by running MCSS calculations for each fragment against all receptors in the binding site sphere defined by the position of the native fragment.

**Docking with GOLD.** For the fragment docking data set, docking was also performed with GOLD.<sup>30</sup> The binding site sphere defined previously for each receptor structure was used, and 40 docking runs were performed for each fragment. Default values were used for genetic algorithm parameters, and “Generate Diverse Solutions” was set to TRUE (Cluster Size = 2, rmsd = 1); solutions were scored using the GOLDScore fitness function.

**In Situ Minimization of Fragment Poses.** As a postprocessing step, MCSS and GOLD poses were minimized in the context of the target binding site using CHARMm (ABNR, Gradient: 0.1 kcal/mol/Å). The protein was held rigid, while the fragments were fully flexible. The native positions of the fragments (X-ray poses) were also minimized under the same conditions to give the *in situ* minimized X-ray poses.

**MM-GBSA Scoring Scheme.** The binding free energy of energy minima obtained from MCSS and GOLD were evaluated using a variation of the standard MM-GBSA approach.<sup>26</sup> For each minimum the binding free energy is evaluated as

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{protein}} - G_{\text{ligand}}$$

The free energy of each of the above terms is calculated from

$$G = E_{\text{MM}} + G_{\text{elec}} + G_{\text{np}} - TS$$

$E_{\text{MM}}$  is the molecular mechanical energy calculated from the CHARMm force-field which was calculated from the standard



**Table 3. Success Rates of Different Docking and Scoring Schemes Investigated in This Study<sup>a</sup>**

method/reference	rmsd			
	1 Å		2 Å	
	X-ray	X-ray Min	X-ray	X-ray Min
MCSS	50%	50%	67%	67%
MCSS-GBSA	67%	67%	75%	75%
GOLD	58%	58%	67%	67%
GOLD-GBSA	50%	58%	75%	75%
Combined	75%	92%	75%	92%

<sup>a</sup>The success rates are compared at different rmsd cut-offs (1.0 Å or 2.0 Å) and with respect to different reference poses, X-ray and *in situ* minimized X-ray pose (X-ray Min).

CHARMm energy function.  $G_{\text{elec}}$  and  $G_{\text{np}}$  represent the electrostatic and nonpolar components of solvation free energy.  $G_{\text{elec}}$  represents the electrostatic component of solvation energy and was calculated using the Generalized Born method with Molecular Volume integration (GBMV).<sup>24,38,39</sup> The nonpolar contribution ( $G_{\text{np}}$ ) to solvation free energy was calculated based on the Surface Area (SA) model which assumes a linear relationship between  $G_{\text{np}}$  and the solvent accessible surface area.<sup>40</sup> A value of 50 cal/mol.Å<sup>2</sup> was used for the surface tension coefficient in the SA model. The solute entropy term, TS, was assumed to be constant among a set of poses for the same ligand in a binding site. Further details on the calculation are described in the Supporting Information.

## RESULTS

**MCSS Calculations on Fragment Docking Data Set.** The results of the MCSS calculations is a set of poses for each fragment in a binding site, each with an associated score calculated from a CHARMm interaction energy with electrostatics in a Coulombic formalism. The poses are subsequently rescored with MM-GBSA and ranked accordingly. The best scoring pose for each fragment was compared with the original X-ray pose and the *in situ* minimized X-ray pose and the success rate assessed at 1 Å and 2 Å rmsd cut-offs. The *in situ* minimization of the X-ray pose can sometimes lead to significant movements, so it is important to consider both poses as references. Table 3 summarizes the success rates of various docking and scoring schemes for the fragment docking data set. The success rate is defined as the percentage of fragments for which the top-scoring pose was predicted within the rmsd cutoff. It was noticed that the success rate of MCSS was not affected by the choice of the reference pose. After rescoring with MM-GBSA, the success rate was slightly increased at 1 Å rmsd cutoff; however, it did not change at 2 Å rmsd cutoff. An improvement in ranking was observed for individual cases, but this was offset by deterioration in the ranking of some other cases. This resulted in a net success rate of 67% after rescoring MCSS poses with MM-GBSA (Table 3).

Further details of these results are summarized in Table 4 where in each case, the top-scoring pose from MCSS is shown, along with its rmsd from the X-ray pose ( $\text{rmsd}_{\text{X-ray}}$ ) and from the *in situ* minimized X-ray pose ( $\text{rmsd}_{\text{X-ray Min}}$ ). The rmsd between the X-ray pose and the *in situ* minimized X-ray pose is also shown ( $\text{rmsd}_{\text{X-ray|X-ray Min}}$ ). These values lie in the range 0.2 Å to 1.0 Å.

**Table 4. Top-Scoring Poses Obtained from MCSS for Fragments in the Fragment Docking Data Set<sup>a</sup>**

PDB	$\text{rmsd}_{\text{X-ray X-ray Min}}$	$\text{rmsd}_{\text{X-ray}}$	$\text{rmsd}_{\text{X-ray Min}}$
1EQG	0.3	0.3	0.1
1FV9	1.0	2.0	1.9
1GWQ	0.5	1.5	1.5
1N1M	0.5	0.8	0.4
1QWC	0.9	1.2	1.0
1S39	0.3	0.3	0.0
1WCC	0.2	4.8	4.8
1YZ3	0.4	0.4	0.0
2ADU	0.5	0.5	0.1
2C90	0.6	5.2	5.0
2JJC	0.4	6.3	6.4
2OHK	0.5	2.7	2.6

<sup>a</sup>For each case, its rmsd from the X-ray pose ( $\text{rmsd}_{\text{X-ray}}$ ), rmsd from the *in situ* minimized X-ray pose ( $\text{rmsd}_{\text{X-ray Min}}$ ) and the rmsd between the X-ray pose and the *in situ* minimized X-ray pose is shown ( $\text{rmsd}_{\text{X-ray|X-ray Min}}$ ).

**Table 5. Most Favorable MCSS Poses after MM-GBSA Scoring for Fragments in the Fragment Docking Data Set<sup>a</sup>**

PDB	$\text{rmsd}_{\text{X-ray}}$	$\text{rmsd}_{\text{X-ray Min}}$
1EQG	0.3	0.1
1FV9	2.0	1.9
1GWQ	7.1	7.1
1N1M	0.8	0.4
1QWC	7.3	7.6
1S39	0.3	0.0
1WCC	0.2	0.0
1YZ3	0.4	0.0
2ADU	9.1	9.1
2C90	0.6	0.0
2JJC	0.3	0.0
2OHK	2.7	2.6

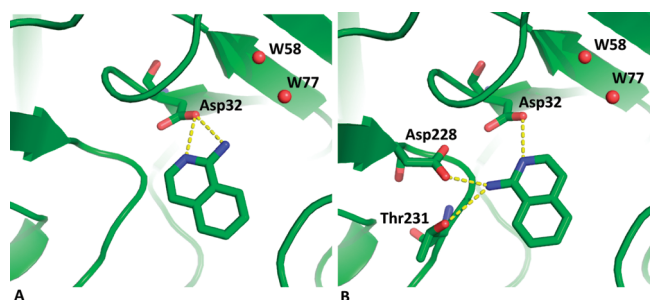
<sup>a</sup> $\text{rmsd}_{\text{X-ray}}$ : rmsd of the top-scoring pose from the X-ray pose,  $\text{rmsd}_{\text{X-ray Min}}$ : rmsd of the top-scoring pose from the *in situ* minimized X-ray pose.

High  $\text{rmsd}_{\text{X-ray|X-ray Min}}$  values (>0.50 Å) are mostly associated with resolution greater than 2.0 Å, as expected.

Further inspection of the cases for which an MCSS top-scoring pose was not within 1.0 Å or 2.0 Å of any of the reference poses revealed that the MCSS protocol had found a correct pose in all cases, but the MCSS score was not able to identify it reliably.

The results after rescoring with MM-GBSA are summarized in Table 5. For each case, the most favorable pose obtained after rescoring is shown along with rmsds against the references. A comparison of Tables 4 and 5 shows that ranking of experimentally relevant poses was improved for 1WCC, 2C90, and 2JJC, where poses within 1.0 Å of the reference poses were picked up by MM-GBSA and scored as the most favorable ones.

In total, four cases showed incorrect ranking after rescoring (1GWQ, 1QWC, 2ADU, and 2OHK). In one of these cases, 2OHK, the top-scoring pose from MCSS, remained the most favorable pose after rescoring and represents an incorrect binding mode. For each of the other three cases, however, MCSS generated a top-scoring pose within 2.0 Å of the X-ray position,



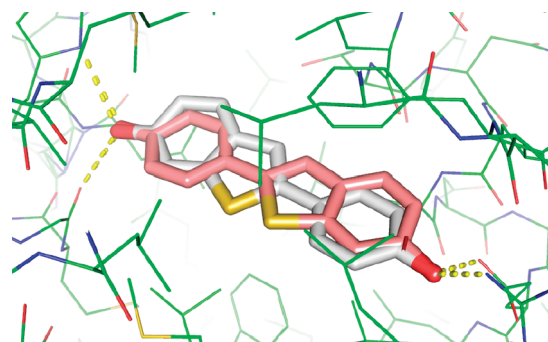
**Figure 3.** Prediction of fragment binding to 2OHK. A. The top-scoring pose deviates significantly from the X-ray pose ( $\text{rmsd}_{\text{X-ray}} = 2.6 \text{ \AA}$ ) and has polar atoms oriented toward a binding site cavity occupied by water molecules (not included in calculations). B. The pose with the lowest  $\text{rmsd}_{\text{X-ray}}$  ( $1.26 \text{ \AA}$ ) ranked 4th despite reproducing key interactions in the binding site.

which failed to obtain a favorable score from MM-GBSA. For 1GWQ, the lowest  $\text{rmsd}_{\text{X-ray}}$  pose was ranked second, after rescoring, at a difference of only 0.7 kcal/mol from the top-scoring pose. Interestingly, for the other two cases, the binding site contained metal ions (heme group iron atom in 1QWC and  $\text{Co}^{2+}$  ions in 2ADU). These four cases are discussed below in more detail.

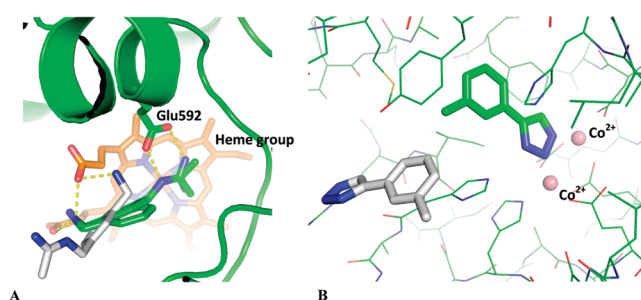
In the case of 2OHK the top ranking pose from MCSS remained the most favorable binding pose after rescoring, with  $\text{rmsd}_{\text{X-ray}}$  and  $\text{rmsd}_{\text{X-rayMin}}$  values  $2.6 \text{ \AA}$  and  $2.7 \text{ \AA}$ , respectively (Table 5). The main difference between the X-ray pose and the top-scoring pose is in the placement of the ligand amino group (Figure 3). In the X-ray pose, it is hydrogen bonded to Asp32 and Asp228, whereas in the top-scoring pose it is placed in an opposite orientation, pointing to a binding site cavity occupied by some water molecules in the X-ray structure. Visual inspection of the binding site indicated that none of these water molecules was present at a hydrogen bonding distance from the ligand amino group (Figure 3A). The docking pose with the lowest rmsd value ( $\text{rmsd}_{\text{X-ray}} 1.26 \text{ \AA}$ ) was ranked fourth after rescoring and reproduced key interactions as observed in the X-ray pose. It therefore appears that the rescoring procedure favored a solvent-oriented position for the amino group over a buried but well hydrogen bonded orientation, assigning a higher desolvation penalty to the latter.

1GWQ was scored correctly by MCSS alone with  $1.5 \text{ \AA}$   $\text{rmsd}_{\text{X-ray}}$  for the top-scoring pose. However, after rescoring, this pose obtained a  $\Delta G$  of binding that was 0.7 kcal/mol higher than that of the most energetically favorable pose which has a  $7.1 \text{ \AA}$   $\text{rmsd}_{\text{X-ray}}$  (Table 5). As can be seen in Figure 4, these two poses represent a flipped binding mode for this pseudo symmetrical fragment, which allows both poses to satisfy their hydrogen bonding potential. The minor difference in the binding energy values of the two poses could arise from the different placement of the sulfur in the thiophene ring.

1QWC and 2ADU represent special cases in the data set as the binding sites in both cases contain metal ions. In 1QWC, the X-ray binding mode indicates key interactions of ligand amino and amidine groups with heme propionate groups and amidine group of the ligand to Glu592 side-chain, respectively. The top-scoring pose from MCSS looks very similar to this pose with  $\text{rmsd}$   $1.2 \text{ \AA}$  and reproduces most of the key interactions, except flipping methyl and amino groups (Figure 5A). After rescoring with MM-GBSA, a pose with a completely different orientation,



**Figure 4.** MCSS generated poses for 1GWQ ranked 1st (magenta C-atoms) and 2nd (gray C-atoms) by MM-GBSA. The 2nd pose which corresponds to the X-ray pose ( $\text{rmsd}_{\text{X-ray}} 1.5 \text{ \AA}$ ) is 0.7 kcal/mol less favorable than the 1st pose which is in completely opposite orientation to the X-ray pose with high  $\text{rmsd}_{\text{X-ray}}$  ( $7.1 \text{ \AA}$ ).

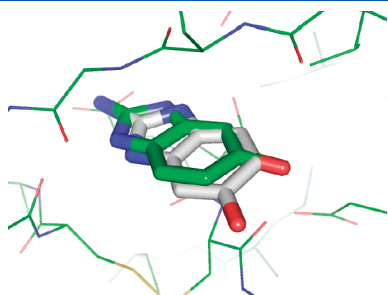


**Figure 5.** MM-GBSA scoring and metal interactions. A. In 1QWC, X-ray and top-scoring MCSS pose (green C-atoms) had similar polar interactions in the binding site. The top-scoring pose after MM-GBSA significantly differed and had a solvent oriented position for amidine group. B. The X-ray pose for the fragment in 2ADU interacts with two  $\text{Co}^{2+}$  ions. The top-scoring pose from MCSS  $\text{rmsd}_{\text{X-ray}} 0.5 \text{ \AA}$  (green C-atoms) did not score well in MM-GBSA scoring and another pose (gray C-atoms) with completely different orientation obtained the highest score.

$\text{rmsd}_{\text{X-ray}} 7.3 \text{ \AA}$ , obtains the most favorable score. The amidine group in this pose is projected toward the solvent, indicating a desolvation penalty assigned by MM-GBSA to its X-ray position.

This was further highlighted in the case of 2ADU. In this case, the fragment interacts with two  $\text{Co}^{2+}$  ions in the binding site. The parameters for  $\text{Co}^{2+}$  ions were missing in the CHARMM Momany and Rone force-field, so these were changed to  $\text{Zn}^{2+}$  ions for calculations. The resulting  $\Delta G$  values obtained for the X-ray and *in situ* minimized X-ray poses were all positive. Similarly, most of the MCSS poses obtained a positive binding energy value. The most favorable pose was located very far from the binding site. Interestingly, MCSS produced a correct solution (Table 4), but the rescoring reduced the quality of prediction, giving the most favorable energy to a different pose that was located very far from the X-ray pose (Figure 5B). It was noticed that all MCSS poses which were interacting with the metal ions received unfavorable binding energy values like the X-ray and X-ray minimized pose. The examples of 1QWC, 2ADU, and 2OHK suggested that MM-GBSA scoring could result in assigning desolvation penalties to polar interactions in the binding site where they seem favorable as indicated from the X-ray positions.

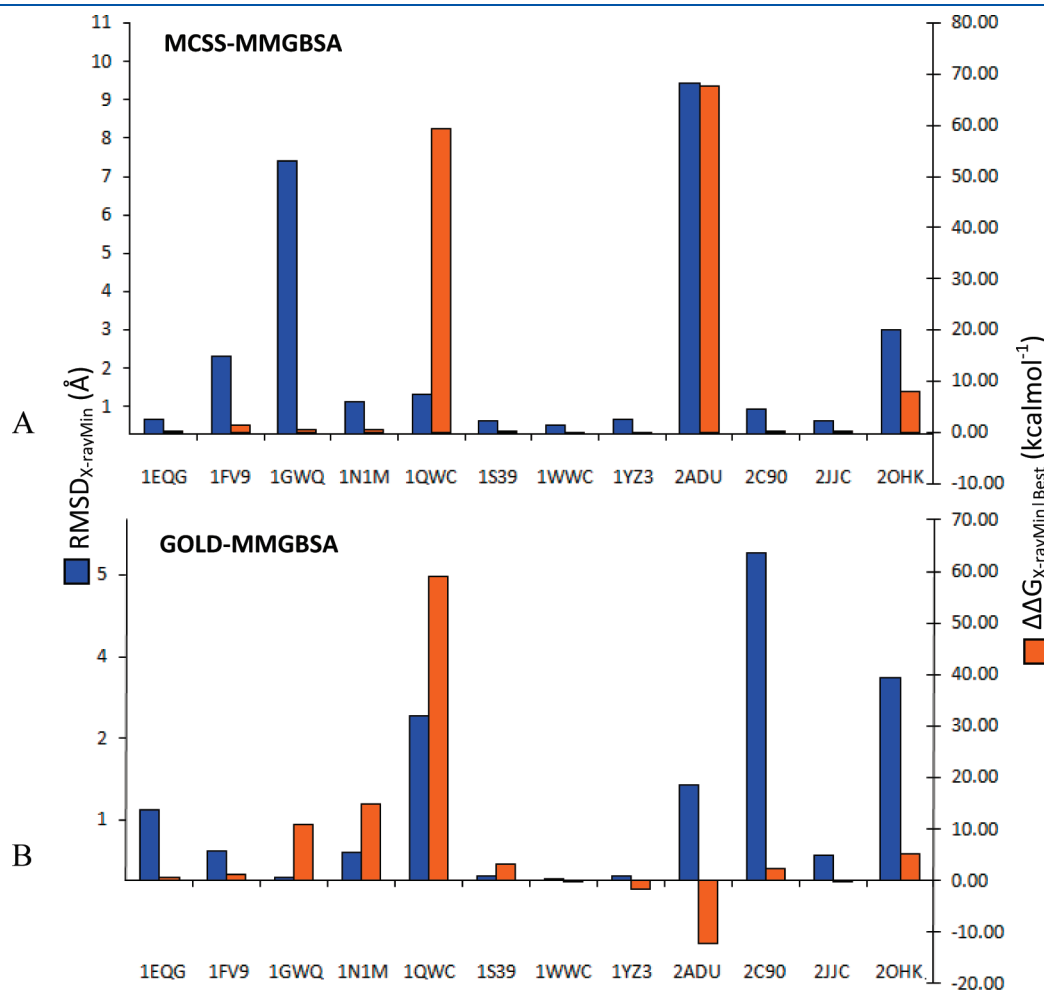
There are two cases that lie between the 1.0 Å and 2.0 Å rmsd cutoff (Table 5). 1FV9 represents a case where the top-scoring pose is about 2.0 Å away from the reference poses. As can be seen in Figure 6, the two poses superpose, but the top-scoring pose is inverted with reference to the position of the hydroxyl group attached to the benzene ring. This highlights the importance of using rather strict criterion when assessing the performance of scoring functions on protein-fragment complexes.



**Figure 6.** Top-scoring MCSS pose for 1FV9 at 2.02 Å rmsd. Predicted pose is shown in gray (C-atoms), and the reference pose (X-ray pose) is shown in green (C-atoms).

There are two reasons why a docking algorithm will not identify the X-ray pose - either it fails to generate the pose (a sampling issue) or it fails to identify that pose as the energetically most favorable (a scoring issue). To test this, Figure 7A plots the difference in free energy calculated with MM/GBSA between the X-ray minimized pose and the top-scoring pose from MCSS for the fragment docking set. (The values for  $\Delta G$  calculated for the X-ray pose, *in situ* minimized X-ray pose, and top-scoring poses are listed in the Supporting Information.) All the values are positive - which means that MCSS is finding a pose which is being scored with a lower energy than the X-ray pose. Also plotted in this figure is the  $\text{rmsd}_{\text{X-rayMin}}$  for this top-scoring pose. For most of the data set, this top-scoring pose is sufficiently close to the experimentally observed pose. The two exceptions have been discussed above (for both 2ADU and 1QWC there are issues with parametrization of metal ions). The other anomaly (1GWQ) was also mentioned above where an equivalent energy MCSS pose is flipped due to pseudosymmetry in the ligand. This indicates that MCSS consistently finds low energy conformers in the binding site; therefore, it is mostly the imperfections in scoring that lead to false positives.

**Comparison with GOLD.** The performance of MCSS and MM-GBSA scoring was compared to that obtained with a more conventional docking program, GOLD, in the place of MCSS.



**Figure 7.** Most favorable MCSS and GOLD poses after MM-GBSA scoring. For each case, the difference in the binding energy of top-scoring pose and the *in situ* minimized X-ray pose is plotted ( $\Delta\Delta G_{\text{X-rayMin|Best}}$ , orange bars) alongside the rmsd from the *in situ* minimized X-ray pose ( $\text{rmsd}_{\text{X-rayMin}}$ ), to show if failure in docking is associated with insufficient sampling of low energy poses.

**Table 6. Top-Scoring Poses Obtained from GOLD with GOLDScore Scoring for Fragments in the Fragment Docking Data Set<sup>a</sup>**

PDB	rmsd <sub>X-ray XrayMin</sub>	rmsd <sub>X-ray</sub>	rmsd <sub>X-rayMin</sub>
1EQG	0.3	0.3	0.3
1FV9	1.0	2.0	1.9
1GWQ	0.5	0.7	0.1
1N1M	0.5	0.5	0.8
1QWC	0.9	2.6	2.4
1S39	0.3	0.2	0.4
1WCC	0.2	6.2	6.2
1YZ3	0.4	0.7	0.9
2ADU	0.5	0.6	0.5
2C90	0.6	5.2	5.0
2JJC	0.4	0.3	0.4
2OHK	0.5	3.2	3.3

<sup>a</sup> For each case, its rmsd from the X-ray pose (rmsd<sub>X-ray</sub>), rmsd from the *in situ* minimized X-ray pose (rmsd<sub>X-ray Min</sub>), and the rmsd between the X-ray pose and the *in situ* minimized X-ray pose is shown (rmsd<sub>X-ray|XrayMin</sub>).

**Table 7. Top-Scoring Poses Obtained from GOLD with MM-GBSA Scoring for Fragments in the Fragment Docking Data Set<sup>a</sup>**

PDB	rmsd <sub>X-ray</sub>	rmsd <sub>X-rayMin</sub>	rmsd <sub>GOLD-MCSSL</sub>
1EQG	1.2	1.1	0.8
1FV9	1.2	0.5	0.8
1GWQ	0.6	0.0	6.5
1N1M	0.8	0.4	0.1
1QWC	4.3	4.5	3.0
1S39	0.4	0.1	0.0
1WCC	0.2	0.0	0.0
1YZ3	0.4	0.1	0.0
2ADU	1.3	1.5	7.8
2C90	5.1	5.2	4.4
2JJC	0.3	0.4	0.0
2OHK	3.3	3.2	0.6

<sup>a</sup> For each case, its rmsd from the X-ray pose (rmsd<sub>X-ray</sub>), rmsd from the *in situ* minimized X-ray pose (rmsd<sub>X-ray Min</sub>). The final column is the rmsd between the top-scoring pose by MM-GBSA for GOLD and MCSSL.

The third and fourth rows of Table 3 show the success rate obtained from GOLD at 1.0 Å and 2.0 Å rmsd cut-offs with respect to X-ray and *in situ* minimized X-ray poses. The success rate of docking and scoring with GOLD did not change with the choice of reference. At 1.0 Å rmsd cutoff the success rate was slightly better than the equivalent MCSSL success rate, but was the same for two methods at 2.0 Å rmsd cutoff.

After rescoring GOLD poses with MM-GBSA, the success rate remained almost the same at 1.0 Å but improved at 2.0 Å which was also slightly better than the equivalent success rate of MCSSL.

The results from GOLD docking and MM-GBSA rescoring of GOLD poses are summarized in Table 6 and Table 7, respectively. The increase in the success rate after rescoring was due to the ranking of correct poses for 1FV9 and 1WCC. In three other cases, 1QWC, 2C90, and 2OHK, rescoring did not improve results.

**Table 8. Success Rate of MCSSL and MCSSL with MM-GBSA Scoring on the Hsp90 Data Set, at Different rmsd Cut-Offs and Considering X-ray and *in Situ* Minimized X-ray Poses As References**

reference	rmsd			
	1 Å		2 Å	
	X-ray	X-ray Min	X-ray	X-ray Min
MCSSL	17%	25%	25%	25%
MCSSL-MM-GBSA	50%	58%	58%	67%

Figure 7B is the plot of  $\Delta\Delta G_{X-rayMin|Best}$  and rmsd<sub>X-rayMin</sub> for the top-scoring poses after rescoring which shows that all four cases with incorrect ranking have binding energy values very close to the *in situ* minimized X-ray pose. GOLD reproduced the correct binding mode for 2ADU which was also the most favorable pose after rescoring. As discussed previously, MCSSL also produced the correct binding mode for 2ADU, but after rescoring it scored less than a solvent oriented incorrect pose. A similar pose was absent from the GOLD generated poses, and rescoring resulted in the correct pose getting the most favorable binding energy. However, all GOLD poses obtained positive energy scores from MM-GBSA, like most of the MCSSL poses, for 2ADU and this is reflected in the  $\Delta\Delta G_{X-rayMin|Best}$  value in Figure 7B. Therefore, it appears that the failures to identify the correct pose were not due to an inadequate search but due to defects in scoring.

Although the performance of MCSSL and GOLD after rescoring is comparable, it can be seen from Figure 7A and B that most of the failures are different in each method. We therefore calculated the combined success rate which is defined as the number of correct poses (under a given rmsd cutoff, 1.0 or 2.0 Å, and a given reference, X-ray or *in situ* minimized pose) in at least one of the scoring schemes. The combined result gives a success rate of 75% and 92%, at 1 Å and 2 Å rmsd considering *in situ* minimized X-ray poses as the reference after MM-GBSA rescoring. In addition, we calculated the rmsd between the top scoring pose from MCSSL and GOLD with MM-GBSA rescoring (final column in Table 7). This shows that if both methods find the same pose, it is the correct pose in 7 out of 8 cases (2OHK is the exception).

**Docking Calculations on the Hsp90 Data Set.** The docking and scoring protocol with MCSSL and MM-GBSA was applied to the Hsp90 data set which contains 11 Hsp90 N-terminal domain structures bound mostly to fragment sized molecules. All the fragments were docked into their native and non-native receptor structures in the data set. Tables 8 and 9 summarize the results obtained from docking fragments into their native structures with MCSSL and scoring based MM-GBSA scoring for each case.

The success rate after rescoring was 58% and 67% at 1 Å and 2 Å rmsd, respectively, using the *in situ* minimized X-ray pose as the reference (Table 8). This was significantly better than the original success rate obtained from MCSSL scoring alone on this data set.

It can be seen from Table 9 that for 7 cases, the predicted pose with the lowest rmsd<sub>X-ray</sub> was assigned the most favorable binding energy. There were five cases where the most favorable binding pose was far from the X-ray pose. For two such cases (1QYE and 2QFOa), the lowest rmsd<sub>X-ray</sub> pose was ranked



**Table 9. Top-Scoring Poses, for MCSS Poses after Docking and MM-GBSA Scoring of Fragments in Native Receptor Structures from the Hsp90 Data Set**

PDB	rmsd <sub>X-ray</sub>	rmsd <sub>X-ray</sub> Min
1QYE	4.1	4.1
1ZWH	2.0	1.1
2CCS	5.9	5.8
2JJC	0.3	0.0
2QF6	0.4	0.1
2QFOa	3.7	3.6
2QFOb	2.3	2.1
2WI1	0.9	0.1
2WI2	1.0	0.9
3BM9	0.3	0.6
3EKO	5.0	4.9
3FT5	0.7	0.7

second, whereas in the other three cases (2CCS, 2QFOb, and 3EKO) the lowest rmsd<sub>X-ray</sub> pose was ranked sixth, seventh, and fourth, respectively. The plots of  $\Delta\Delta G_{X-rayMin|Best}$  and rmsd<sub>X-ray</sub>Min for this data set are provided in the Supporting Information and again confirm that for most systems the issues are with scoring rather than pose sampling.

In 1QYE, the top two poses have rmsd<sub>X-ray</sub> 4.11 Å and 1.14 Å, respectively, with a binding energy difference of about 2 kcal/mol. The main difference in the rmsd<sub>X-ray</sub> came from the position of the adenine moiety of the ligand, which is an adenosine analog. In the X-ray pose, adenine is extensively hydrogen bonded to key Hsp90 binding site residues and conserved water molecules. The top-scoring pose has this moiety placed in a solvent oriented position, probably due to the desolvation penalty arising from GBBSA scoring. However in the second pose, these interactions are almost perfectly reproduced, contributing mainly to the 1.14 Å rmsd<sub>X-ray</sub>.

The initial scoring of docked poses from MCSS can influence the outcome of the rescoring procedure. At the end of the MCSS minimization cycle, docked poses are clustered using an rmsd cutoff (2.0 Å in this study). The best scoring pose within each cluster is selected as the cluster representative. It is however likely that within a cluster that is located at the X-ray binding position of the ligand, a pose with the lowest rmsd to the X-ray pose, does not get selected as cluster representative due to its low MCSS score. This was seen in the case of 2CCS, where the lowest rmsd<sub>X-ray</sub> pose (1.89 Å) was ranked sixth. The MCSS cluster that this pose belonged to contained another pose with even lower rmsd<sub>X-ray</sub> (0.47 Å). Since its MCSS score was lower than other poses in the cluster, it was not present in the rescored poses. When this pose was included in the MCSS output, it obtained the most favorable binding energy. This problem can easily be circumvented either by processing nonclustered output from MCSS or setting a small clustering rmsd to allow more poses in the clustered output.

2QFOa and 2QFOb represent a case of co-operative binding. It was shown that 2QFOb binds more strongly in the presence of 2QFOa.<sup>41</sup> The native binding mode of these fragments also indicates  $\pi$ – $\pi$  stacking interactions between the pyrimidine ring of 2QFOa and the phenyl ring of 2QFOb. In this protocol, these fragments were docked and scored independent of each other; therefore, energetic contributions resulting from the direct

interactions between these two fragments could not be accounted for. This probably contributed to the lowest rmsd<sub>X-ray</sub> poses for 2QFOa and 2QFOb getting rank 2 and 7, respectively. The difference in the top scoring and lowest rmsd<sub>X-ray</sub> poses for 2QFOa was only about 0.5 kcal/mol. In the top scoring pose of 2QFOb, the furanone moiety was accurately superimposed on the X-ray pose, and the major deviation came from the benzene ring which is stabilized by stacking interactions with pyrimidine ring of 2QFOa (not present in 2QFOb docking and scoring).

Among the successfully docked Hsp90 ligands, the most anomalous is 1ZWH where the most favorable pose is on the margin of the 2.0 Å threshold (Table 9) and where there is the largest deviation between the X-ray pose and *in situ* minimized pose. In the X-ray binding mode of the fragment, the carboxylic group attached to the ring is partially exposed to the solvent. The *in situ* minimization (with distance-dependent dielectric model) favored a binding mode where this group formed a hydrogen bond with ASN92 in the binding site. It should be noted that the MCSS minimization is also carried out with a distance-dependent dielectric model. The initial placement of fragments can therefore result in an experimentally relevant pose getting a low MCSS rank. In the case of 1ZWH however, an MCSS pose which reproduced most of the key interactions in the native binding mode obtained the most favorable binding energy after rescoring.

The incorrect ranking in the case of 3EKO is explained in the next section.

**Effect of Multiple Receptor Structures.** The Hsp90 binding site can adopt different conformations in the presence of different ligands.<sup>42</sup> The base of the pocket (containing the purine binding site with conserved solvent structure binding to Asp93) remains relatively static across all known structures. The main change is the rearrangement of a part of the binding site  $\alpha$ -helix which can sometimes open up to form an additional hydrophobic pocket.<sup>43</sup> It is therefore important in docking and scoring of fragments to use the most appropriate target structure. One approach is to use multiple receptor structures obtained either from molecular dynamics simulation or from X-ray crystallography.<sup>44</sup> The fragments in the Hsp90 data set were therefore docked into the set of native and non-native structures and subsequently rescored with MM-GBSA to study how the consideration of multiple structures could affect the outcome of docking.

The results from cross-docking indicated that the success rate obtained using individual receptor structures, in terms of percentage of fragments in the data set correctly docked and scored, varied from 0 to 45% within 2 Å rmsd from the X-ray pose. The correct binding mode for each fragment was reproduced and scored at the top in at least one or more receptor structures, except 2QFOb. We were interested in studying the effect of successively adding multiple receptor structures on the docking result. For this purpose, after docking and rescoring into non-native receptor structures, we randomly picked a receptor structure and noted the success rate, followed by successively adding more structures at random and calculating the total success rate after each addition. The same procedure was repeated 100 times to assess the increase in the average success rate while adding structures at random. Figure 8 is a plot of the average success rate against the number of receptor structures added, showing the average performance increased almost linearly with successive addition of structures at random, leading up to the performance of 91% when using all the structures in the data set. This increase in performance is observed with both the 1 Å and 2 Å rmsd criteria (Figure 8).

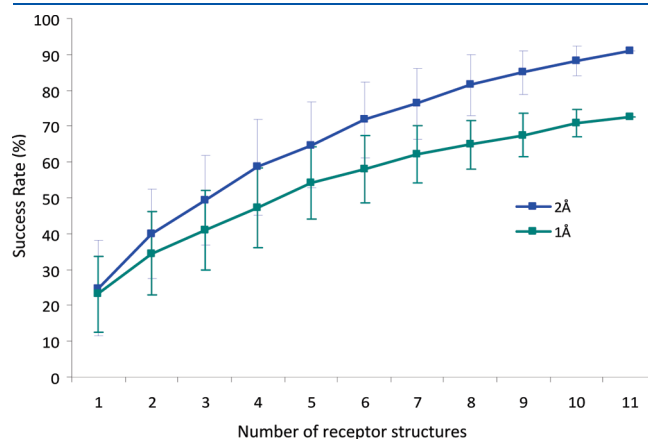
This performance relies on the X-ray crystallographic binding mode of the fragment being reproduced and scored correctly in at least one receptor structure. Different receptor structures in the data set can then either improve or deteriorate the quality of predictions. If a native receptor predicts the X-ray binding mode with the most favorable interaction energy (the highest rank), it is possible that a non-native receptor gives a poor rank to the lowest rmsd pose, thereby deteriorating the quality. Similarly, if a native receptor fails to reproduce and correctly rank the X-ray binding mode, a non-native receptor can improve the performance by correctly identifying the correct binding mode. In order to compare how non-native receptors perform with respect to the native receptor, the improvement or deterioration of the quality of predictions were assessed as follows. The receptor structure is considered 'deteriorating' when a native receptor reproduces a pose with  $\leq 2.0$  Å rmsd<sub>X-ray</sub> with the highest score and the non-

native receptor fails to score a pose with  $\leq 2.0$  Å rmsd<sub>X-ray</sub> within the top 3. The receptor structure is considered 'improving' when a native receptor fails to reproduce the native binding mode and score it correctly and a non-native receptor reproduces a pose with  $\leq 2.0$  Å rmsd<sub>X-ray</sub> within the top 3 poses. In this way for each fragment, the number of non-native receptors that have an improving or deteriorating effect on the quality of predictions can be calculated. This is shown in Figure 9 where for each fragment the number of improving versus deteriorating receptor structures is shown.

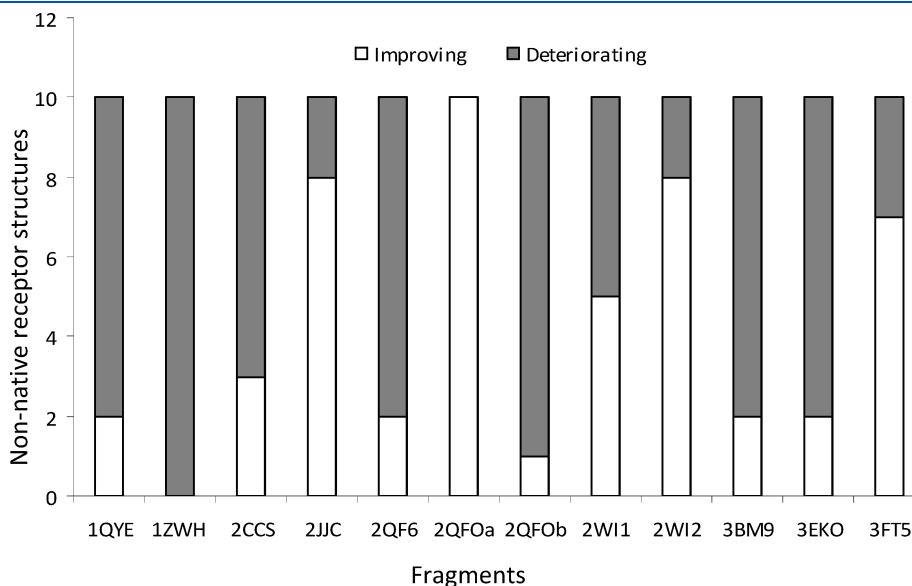
It was previously observed for the Hsp90 data set that in 5 cases prediction in the native structures was not successful (Table 9). It can be seen from Figure 9 that for such cases (1QYE, 2CCS, 2QFOa, 2QFOb, and 3EKO) at least one non-native receptor was able to reproduce the native binding mode with correct scoring, thereby improving the quality of predictions. This was particularly notable for 2QFOa for which all of the non-native receptors reproduced the native binding mode.

For the 7 cases where docking and scoring in the native receptor was successful (1ZWH, 2JJC, 2QF6, 2WI1, 2WI2, 3BM9, and 3FT5), a fraction of the non-native receptors deteriorate the quality of scoring and ranking (Figure 9). The most significant of these was 1ZWH where none of the non-native receptors could reproduce the native binding mode with correct ranking.

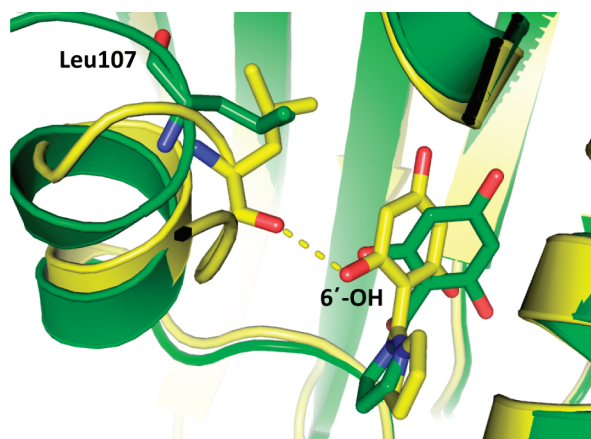
We investigated the cases where a non-native receptor successfully ranked the X-ray pose of a ligand and one such example is described here. As mentioned earlier, four conserved water molecules in the binding site were kept during docking and rescoring calculations. However, Hsp90 ligands also interact with 'nonconserved water molecules' in the binding site that are not consistently seen across multiple structures; one such example is 3EKO, where the 6'-hydroxyl group on the ligand is shown to interact with two water molecules.<sup>45</sup> During docking in the native receptor structure, a pose with rmsd<sub>X-ray</sub> 0.36 Å was reproduced, but it did not obtain a favorable binding energy due to the missing interactions with 6'-hydroxyl group (Figure 10). On the other hand, in a non-native receptor structure (2WI2) with a slightly altered conformation of binding site loop region, a pose with



**Figure 8.** The effect of multiple structures on scoring performance. The scoring is defined as the percentage of correctly docked and scored ligands in at least one receptor structure. Successive addition of structure is based on 100 random selections at each point, and the average performance is plotted against the number of structures with error bars representing standard deviation of success rate in 100 random selections of additional structures.



**Figure 9.** Relative contribution of non-native receptor structures toward scoring of each docked fragment in the Hsp90 data set.



**Figure 10.** Docking in multiple receptor structures. Ligand from 3EKO docked into native (green) and a non-native (yellow) receptor structure, 2WI2. The alternative conformation of Leu107 in non-native receptor structure reproduces some of the interactions that were missing from the native receptor.

$\text{rmsd}_{\text{X-ray}}$  1.05 Å obtained the most favorable binding energy. The alternative conformation of the loop region in this receptor resulted in a hydrogen bond between Leu107 carbonyl group and the 6'-hydroxyl group of the ligand (Figure 10). This should largely compensate for the energetic penalty incurred in the native receptor structure for a similar pose. The position of Leu107 carbonyl oxygen atom in 2WI2 was found close to ( $\sim 2.0$  Å) a water molecule in 3EKO. It is interesting to note that the interactions with 'nonconserved' water molecules are associated with a region in the binding site that shows conformational changes. This example indicates that the use of multiple structures sometimes accounts for what is missing in native receptor structure leading to interesting results.

## DISCUSSION

MCSS is one of the oldest methods to compute energetically favorable positions of functional groups in protein binding sites.<sup>8</sup> As fragment-based lead discovery methods come of age<sup>3</sup> it could provide a valuable computational technique to predict the binding pose for fragments that have been identified to bind to a target but for which a crystal structure cannot be obtained. If a reasonably robust model of fragment binding can be generated, then this can provide essential guidance for the initial stages of chemistry for building a fragment into a more potent hit or lead compound. However, the same challenges apply as in docking of larger compounds - primarily the limitations of scoring functions. For MCSS, a recent review<sup>15</sup> highlighted the need for 1) evaluation of MCSS minima and their ranking for high quality predictions and 2) sampling multiple receptor conformations as the receptor is held rigid in MCSS minimization. In this study we investigated both of these issues by applying MCSS calculations followed by binding energy estimation on a set of 12 different protein-fragment complexes and on a set of 11 Hsp90 protein-fragment complexes that showed some conformational variation.

In the MCSS scoring function the solvent effects are taken into account by a very crude approximation of dielectric screening from a distance-dependent dielectric model. This has led to inaccuracies in scoring as the desolvation penalty accompanying some of the polar interactions between the protein and the functional group are not appropriately taken into account. An

alternative approach is to couple the molecular mechanical energy component with polar and nonpolar solvation free energies derived from implicit solvent formalisms in the MM-GBSA or PBSA method. The sorting of MCSS minima based on MM-PBSA derived free energy estimates has previously shown encouraging results.<sup>11</sup> The MM-PBSA method, considered as a benchmark in implicit solvent methods, is time-consuming and less feasible on a large-scale. Continuing developments in implicit solvent methods have led to semianalytical solutions to PB equation with high accuracy, such as the GB equation.<sup>39</sup> It is therefore timely to investigate and evaluate the performance of MCSS combined with MM-GBSA as a fragment docking and scoring scheme. The calculations based on a set of protein-fragment complexes suggest encouraging results with some limitations.

The assessment of the success rate was complicated by factors such as the choice of reference pose and the rmsd cutoff for comparison with the predicted poses. As MCSS is an energy minimization routine so the X-ray position for each fragment was *in situ* minimized to provide an additional reference for the comparison. Similarly, due to the small size of fragment molecules the rmsd cutoff for determining success rate was set at 1 Å and 2 Å to highlight borderline cases. It was noticed that the choice of reference position was more relevant at lower rmsd cutoff as the success rate converged to similar values at 2 Å regardless of the reference chosen.

Using MCSS primarily as a method to generate poses for subsequent evaluation with a more rigorous scoring function shows significant increase in scoring accuracy as compared to docking and scoring with MCSS alone. The observed increase in the percentage of correctly docked and scored ligands was from 50% to 67%. This was the same amount of improvement that was observed for a standard docking and scoring method, GOLD, combined with MM-GBSA scoring. Interestingly, despite the similar success rate both methods show success and failure for different cases. The combined success rate was 92% and suggests that combining the docking output from different methods and then scoring with MM/GB-SA may be a useful approach. There would be increased confidence where the same pose is found.

There were a few cases where MCSS alone was unable to identify a pose close to any of the reference poses. Most of these failures came from scoring where nonexperimentally relevant energy minima were given favorable binding energy values. The origin of this poor scoring is inadequate treatment of solvent effects. This is largely rectified in the GB-SA scoring step and gives the observed improvement of ranking. Occasionally, the desolvation penalty arising from GB-SA also led to preference of a solvent-oriented position of the ligand, which was not related to the X-ray pose e.g., 2OHK and 2ADU (Figures 3 and 5). As with any other docking program sampling receptor conformations is a challenging aspect in MCSS. Multiple receptor structures from X-ray crystallography or molecular dynamics simulations provide a way of sampling different conformations.<sup>44</sup> Therefore, we studied the feasibility of using this docking and scoring scheme for a flexible target by including multiple ligand-bound conformations. It was noted that random addition of structures to see if a particular fragment can be docked and scored correctly in at least one receptor structures in the data set increases the overall percentage performance. When multiple structures are taken into account, some receptor structures would improve the quality of predictions by reproducing the correct binding mode and interaction energy particularly where the native receptor structure has



failed to do so. On the other hand, some non-native receptor structures produce the incorrect binding mode for the ligand, despite the correct binding mode in the native receptor, thereby reducing the quality of prediction. A way to use this approach would be to benchmark a set of receptor structures based on how effectively they reproduce and score binding modes of non-native ligand/fragments. This can then be used as a preselection criterion for how many and which receptor structures could be used for optimal results. It should be noted that the plastic nature of binding in flexible active sites cannot be fully accounted for in this approach. The ligand could induce novel conformational variations in the binding site and there is the continuing problem of solvent molecules providing bridging and flexibility in the interactions that can be made.

## CONCLUSION

The work described in this study has shown that the success rate obtained by MCSS followed by rescoring with MM-GBSA is high enough that the technique could be a useful method to support early ligand discovery efforts when it is not possible to determine crystal structures of fragments binding to proteins. Using a dual docking procedure (MCSS – GOLD) resulted in higher success rate than obtained for a single method. This supports the idea that a combination of docking strategies has the potential of yielding better results by mutually compensating the shortcomings of different approaches.

The conformational variations in the receptor structure upon binding, water molecules at the binding interface, and interactions involving metal ions pose significant challenges. We show that the use of multiple receptor structures can possibly increase the success rate but at the expense of including receptor structures that also deteriorate quality of predictions. The treatment of highly conserved water molecules as part of the binding results in good quality prediction. Using better parametrization of metal ion interactions with ligands in the binding sites should also give better results. As with all docking and modeling approaches, these findings emphasize that a comprehensive understanding of the target binding site is a valuable prerequisite for using such methods to predict compound binding. The main barrier to using this type of approach to enable the early stages of fragment optimization will be the availability of suitable crystal structures that sample the conformations likely to appear on ligand binding and the positions of key solvent molecules. Where such structures are available, the work presented here suggests that a docking campaign followed by MM-GBSA scoring could give reliable enough predictions of fragment binding pose to guide chemistry for optimization.

## ASSOCIATED CONTENT

**S Supporting Information.** Theoretical background on MM-GBSA methods, tables containing details of top-scoring MCSS and GOLD poses after MM-GBSA scoring for fragment-docking and the Hsp90 data set, including  $\Delta G$  values for X-ray and in situ minimized poses,  $\Delta G$  of top-scoring poses, and rmsds of top-scoring from the X-ray and in situ minimized poses. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*Phone: +44(0) 1904 328267. Fax: +44(0)1904 328266. E-mail: [rod@ysbl.york.ac.uk](mailto:rod@ysbl.york.ac.uk).

## ACKNOWLEDGMENT

This work was supported by a studentship from Higher Education Commission, Pakistan.

## REFERENCES

- (1) Congreve, M.; Chessari, G.; Tisi, D.; Woodhead, A. J. Recent developments in fragment-based drug discovery. *J. Med. Chem.* **2008**, *51*, 3661–3680.
- (2) Erlanson, D. A.; McDowell, R. S.; O'Brien, T. Fragment-based drug discovery. *J. Med. Chem.* **2004**, *47*, 3463–3482.
- (3) Fischer, M.; Hubbard, R. E. Fragment-based ligand discovery. *Mol. Interventions* **2009**, *9*, 22–30.
- (4) Zartler, E. R.; Shapiro, M. J. Fragonomics: fragment-based drug discovery. *Curr. Opin. Chem. Biol.* **2005**, *9*, 366–370.
- (5) de Kloe, G. E.; Bailey, D.; Leurs, R.; de Esch, I. J. P. Transforming fragments into candidates: small becomes big in medicinal chemistry. *Drug Discovery Today* **2009**, *14*, 630–646.
- (6) Hajduk, P. J.; Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discovery* **2007**, *6*, 211–219.
- (7) Goodford, P. J. A Computational-Procedure For Determining Energetically Favorable Binding-Sites On Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
- (8) Miranker, A.; Karplus, M. Functionality Maps Of Binding-Sites - A Multiple Copy Simultaneous Search Method. *Proteins: Struct. Funct. Genet.* **1991**, *11*, 29–34.
- (9) Cafisch, A.; Miranker, A.; Karplus, M. Multiple copy simultaneous search and construction of ligands in binding sites - application to inhibitors of HIV-1 aspartic proteinase. *J. Med. Chem.* **1993**, *36*, 2142–2167.
- (10) Eisen, M. B.; Wiley, D. C.; Karplus, M.; Hubbard, R. E. HOOK: A Program for Finding Novel Molecular Architectures That Satisfy the Chemical and Steric Requirements of a Macromolecule Binding Site. *Proteins: Struct. Funct. Genet.* **1994**, *19*, 199–221.
- (11) Cafisch, A. Computational combinatorial ligand design: Application to human alpha-thrombin. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 372–396.
- (12) Joseph-McCarthy, D.; Hogle, J. M.; Karplus, M. Use of the multiple copy simultaneous search (MCSS) method to design a new class of picornavirus capsid binding drugs. *Proteins: Struct. Funct. Bioinf.* **1997**, *29*, 32–58.
- (13) Evensen, E.; Joseph-McCarthy, D.; Weiss, G.; Schreiber, S.; Karplus, M. Ligand design by a combinatorial approach based on modeling and experiment: application to HLA-DR4. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 395–418.
- (14) Cross, S.; Cruciani, G. Molecular fields in drug discovery: getting old or reaching maturity?. *Drug Discovery Today* **2010**, *15*, 23–32.
- (15) Schubert, C. R.; Stultz, C. M. The multi-copy simultaneous search methodology: a fundamental tool for structure-based drug design. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 475–489.
- (16) Allen, K. N.; Bellamacina, C. R.; Ding, X. C.; Jeffery, C. J.; Mattos, C.; Petsko, G. A.; Ringe, D. An experimental approach to mapping the binding surfaces of crystalline proteins. *J. Phys. Chem.* **1996**, *100*, 2605–2611.
- (17) Mattos, C.; Ringe, D. Locating and characterizing binding sites on proteins. *Nat. Biotechnol.* **1996**, *14*, 595–599.
- (18) Joseph-McCarthy, D.; Fedorov, A. A.; Almo, S. C. Comparison of experimental and computational functional group mapping of an RNase A structure: Implications for computer-aided drug design. *Protein Eng.* **1996**, *9*, 773–780.
- (19) English, A. C.; Groom, C. R.; Hubbard, R. E. Experimental and computational mapping of the binding surface of a crystalline protein. *Protein Eng.* **2001**, *14*, 47–59.
- (20) Silberstein, M.; Dennis, S.; Brown, L.; Kortvelyesi, T.; Clodfelter, K.; Vajda, S. Identification of substrate binding sites in enzymes by computational solvent mapping. *J. Mol. Biol.* **2003**, *332*, 1095–1113.



- (21) Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caffisch, A. Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins: Struct. Funct. Genet.* **1999**, *37*, 88–105.
- (22) Brenke, R.; Kozakov, D.; Chuang, G. Y.; Beglov, D.; Hall, D.; Landon, M. R.; Mattos, C.; Vajda, S. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* **2009**, *25*, 621–627.
- (23) Imai, T.; Oda, K.; Kovalenko, A.; Hirata, F.; Kidera, A. Ligand Mapping on Protein Surfaces by the 3D-RISM Theory: Toward Computational Fragment-Based Drug Design. *J. Am. Chem. Soc.* **2009**, *131*, 12430–12440.
- (24) Feig, M.; Brooks, C. L. Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr. Opin. Struct. Biol.* **2004**, *14*, 217–224.
- (25) Guimaraes, C. R. W.; Cardozo, M. MM-GB/SA rescoring of docking poses in structure-based lead optimization. *J. Chem. Inf. Model.* **2008**, *48*, 958–970.
- (26) Massova, I.; Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discovery Des.* **2000**, *18*, 113–135.
- (27) Michel, J.; Verdonk, M. L.; Essex, J. W. Protein-ligand binding affinity predictions by implicit solvent simulations: A tool for lead optimization?. *J. Med. Chem.* **2006**, *49*, 7427–7439.
- (28) Thompson, D. C.; Humblet, C.; Joseph-McCarthy, D. Investigation of MM-PBSA rescoring of docking poses. *J. Chem. Inf. Model.* **2008**, *48*, 1081–1091.
- (29) Onufriev, A. Continuum Electrostatics Solvent Modeling with the Generalized Born Model. In *Modeling Solvent Environments Applications to Simulations of Biomolecules*; Feig, M., Ed.; WILEY-VCH Verlag GmbH & Co. KGaA: Weinheim, 2010; pp 127–165.
- (30) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
- (31) Berman, H. M.; Bhat, T. N.; Bourne, P. E.; Feng, Z. K.; Gilliland, G.; Weissig, H.; Westbrook, J. The Protein Data Bank and the challenge of structural genomics. *Nat. Struct. Biol.* **2000**, *7*, 957–959.
- (32) Discovery Studio, 2.5; Accelrys: San Diego, CA, 2009.
- (33) Rone, R.; Momany, F. A.; Dygert, M. Conformational studies on vancomycin using QUANTA-CHARMM. Smith, J. a., Rivier, J. E.; Eds.; 1992; pp 299–301.
- (34) Warnmark, A.; Treuter, E.; Gustafsson, J. A.; Hubbard, R. E.; Brzozowski, A. M.; Pike, A. C. W. Interaction of transcriptional intermediary factor 2 nuclear receptor box peptides with the coactivator binding site of estrogen receptor alpha. *J. Biol. Chem.* **2002**, *277*, 21862–21868.
- (35) Kallander, L. S.; Lu, Q.; Chen, W. F.; Tomaszek, T.; Yang, G.; Tew, D.; Meek, T. D.; Hofmann, G. A.; Schulz-Pritchard, C. K.; Smith, W. W.; Janson, C. A.; Ryan, M. D.; Zhang, G. F.; Johanson, K. O.; Kirkpatrick, R. B.; Ho, T. F.; Fisher, P. W.; Mattern, M. R.; Johnson, R. K.; Hansbury, M. J.; Winkler, J. D.; Ward, K. W.; Veber, D. F.; Thompson, S. K. 4-Aryl-1,2,3-triazole: A novel template for a reversible methionine aminopeptidase 2 inhibitor, optimized to inhibit angiogenesis in vivo. *J. Med. Chem.* **2005**, *48*, 5644–5647.
- (36) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.
- (37) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caffisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (38) Feig, M.; Chocholousova, J.; Tanizaki, S. Simulating nucleic acids: Towards implicit and hybrid solvation models. *Biophys. J.* **2005**, *88*, 513A–513A.
- (39) Lee, M. S.; Feig, M.; Salsbury, F. R., Jr; Brooks Iii, C. L. New analytical approximation to the standard molecular volume definition and its application to generalized Born calculations. *J. Comput. Chem.* **2003**, *24*, 1348–1356.
- (40) Sitkoff, D.; Sharp, K. A.; Honig, B. Correlating solvation free energies and surface tensions of hydrocarbon solutes. *Biophys. Chem.* **1994**, *51*, 397–409.
- (41) Huth, J. R.; Park, C.; Petros, A. M.; Kunzer, A. R.; Wendt, M. D.; Wang, X. L.; Lynch, C. L.; Mack, J. C.; Swift, K. M.; Judge, R. A.; Chen, J.; Richardson, P. L.; Jin, S.; Tahir, S. K.; Matayoshi, E. D.; Dorwin, S. A.; Lador, U. S.; Severin, J. M.; Walter, K. A.; Bartley, D. M.; Fesik, S. W.; Elmore, S. W.; Hajduk, P. J. Discovery and design of novel HSP90 inhibitors using multiple fragment-based design strategies. *Chem. Biol. Drug Des.* **2007**, *70*, 1–12.
- (42) Solit, D. B.; Chiosis, G. Development and application of Hsp90 inhibitors. *Drug Discovery Today* **2008**, *13*, 38–43.
- (43) Richter, K.; Buchner, J. hsp90: Twist and Fold. *Cell* **2006**, *127*, 251–253.
- (44) Cozzini, P.; Kellogg, G. E.; Spyraakis, F.; Abraham, D. J.; Costantino, G.; Emerson, A.; Fanelli, F.; Gohlke, H.; Kuhn, L. A.; Morris, G. M.; Orozco, M.; Pertinhez, T. A.; Rizzi, M.; Sottriffer, C. A. Target Flexibility: An Emerging Consideration in Drug Discovery and Design. *J. Med. Chem.* **2008**, *51*, 6237–6255.
- (45) Kung, P. P.; Funk, L.; Meng, J.; Collins, M.; Zhou, J. Z. X.; Johnson, M. C.; Ekker, A.; Wang, J.; Mehta, P.; Yin, M. J.; Rodgers, C.; Davies, J. F.; Bayman, E.; Smeal, T.; Maegley, K. A.; Gehring, M. R. Dihydroxyphenyl amides as inhibitors of the Hsp90 molecular chaperone. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 6273–6278.
- (46) Selinsky, B. S.; Gupta, K.; Sharkey, C. T.; Loll, P. J. Structural analysis of NSAID binding by prostaglandin H-2 synthase: Time-dependent and time-independent inhibitors elicit identical enzyme conformations. *Biochemistry* **2001**, *40*, 5172–5180.
- (47) Hajduk, P. J.; Boyd, S.; Nettesheim, D.; Nienaber, V.; Severin, J.; Smith, R.; Davidson, D.; Rockway, T.; Fesik, S. W. Identification of novel inhibitors of urokinase via NMR-based screening. *J. Med. Chem.* **2000**, *43*, 3862–3866.
- (48) Rasmussen, H. B.; Branner, S.; Wiberg, F. C.; Wagtmann, N. Crystal structure of human dipeptidyl peptidase IV/CD26 in complex with a substrate analog. *Nat. Struct. Biol.* **2003**, *10*, 19–25.
- (49) Fedorov, R.; Hartmann, E.; Ghosh, D. K.; Schlichting, I. Structural basis for the specificity of the nitric-oxide synthase inhibitors W1400 and N-omega-propyl-L-Arg for the inducible and neuronal isoforms. *J. Biol. Chem.* **2003**, *278*, 45818–45825.
- (50) Meyer, E. A.; Furler, M.; Diederich, F.; Brenk, R.; Klebe, G. Synthesis and In Vitro Evaluation of 2-Aminoquinazolin-4(3H)-one-Based Inhibitors for tRNA-Guanine Transglycosylase (TGT). *Helv. Chim. Acta* **2004**, *87*, 1333–1356.
- (51) Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Frederickson, M.; Tickle, I. J.; Jhoti, H. Fragment-based lead discovery using X-ray crystallography. *J. Med. Chem.* **2005**, *48*, 403–413.
- (52) Wu, Q.; Gee, C. L.; Lin, F.; Tyndall, J. D.; Martin, J. L.; Grunewald, G. L.; McLeish, M. J. Structural, mutagenic, and kinetic analysis of the binding of substrates and inhibitors of human phenylethanolamine N-methyltransferase. *J. Med. Chem.* **2005**, *48*, 7243–7252.
- (53) Howard, N.; Abell, C.; Blakemore, W.; Chessari, G.; Congreve, M.; Howard, S.; Jhoti, H.; Murray, C. W.; Seavers, L. C. A.; van Montfort, R. L. M. Application of fragment screening and fragment linking to the discovery of novel thrombin inhibitors. *J. Med. Chem.* **2006**, *49*, 1346–1355.
- (54) Murray, C. W.; Callaghan, O.; Chessari, G.; Cleasby, A.; Congreve, M.; Frederickson, M.; Hartshorn, M. J.; McMenamin, R.; Patel, S.; Wallis, N. Application of fragment screening by X-ray crystallography to beta-secretase. *J. Med. Chem.* **2007**, *50*, 1116–1123.
- (55) Soldano, K. L.; Jivan, A.; Nicchitta, C. V.; Gewirth, D. T. Structure of the N-terminal domain of GRP94 - Basis for ligand specificity and regulation. *J. Biol. Chem.* **2003**, *278*, 48330–48338.
- (56) Barril, X.; Beswick, M. C.; Collier, A.; Drysdale, M. J.; Dymock, B. W.; Fink, A.; Grant, K.; Howes, R.; Jordan, A. M.; Massey, A.;

Surgenor, A.; Wayne, J.; Workman, P.; Wright, L. 4-amino derivatives of the Hsp90 inhibitor CCT018159. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2543–2548.

(57) Brough, P. A.; Barril, X.; Borgognoni, J.; Chene, P.; Davies, N. G. M.; Davis, B.; Drysdale, M. J.; Dymock, B.; Eccles, S. A.; Garcia-Echeverria, C.; Fromont, C.; Hayes, A.; Hubbard, R. E.; Jordan, A. M.; Jensen, M. R.; Massey, A.; Merrett, A.; Padfield, A.; Parsons, R.; Radimerski, T.; Raynaud, F. I.; Robertson, A.; Roughley, S. D.; Schoepfer, J.; Simmonite, H.; Sharp, S. Y.; Surgenor, A.; Valenti, M.; Walls, S.; Webb, P.; Wood, M.; Workman, P.; Wright, L. Combining Hit Identification Strategies: Fragment-Based and in Silico Approaches to Orally Active 2-Aminothieno 2,3-d pyrimidine Inhibitors of the Hsp90 Molecular Chaperone. *J. Med. Chem.* **2009**, *52*, 4794–4809.

(58) Gopalsamy, A.; Shi, M. X.; Golas, J.; Vogan, E.; Jacob, J.; Johnson, M.; Lee, F.; Nilakantan, R.; Petersen, R.; Svenson, K.; Chopra, R.; Tam, M. S.; Wen, Y. X.; Ellingboe, J.; Arndt, K.; Boschelli, F. Discovery of benzisoxazoles as potent inhibitors of chaperone heat shock protein 90. *J. Med. Chem.* **2008**, *51*, 373–375.

(59) Barker, J. J.; Barker, O.; Boggio, R.; Chauhan, V.; Cheng, R. K. Y.; Corden, V.; Courtney, S. M.; Edwards, N.; Falque, V. M.; Fusar, F.; Gardiner, M.; Hamelin, E. M. N.; Hesterkamp, T.; Ichihara, O.; Jones, R. S.; Mather, O.; Mercurio, C.; Minucci, S.; Montalbetti, C.; Muller, A.; Patel, D.; Phillips, B. G.; Varasi, M.; Whittaker, M.; Winkler, D.; Yarnold, C. J. Fragment-based Identification of Hsp90 Inhibitors. *ChemMedChem* **2009**, *4*, 963–966.