Structure-based virtual screening for glycosyltransferase\textsubscript{51}

Min Yang\textsuperscript{a*}, Lu Zhou\textsuperscript{a}, Zhili Zuo\textsuperscript{b}, Xiangyang Tang\textsuperscript{c}, Jian Liu\textsuperscript{a} and Xiang Ma\textsuperscript{a}

\textsuperscript{a}Department of Pharmaceutical Engineering, College of Chemical Engineering, Sichuan University, Sichuan, Chengdu, People’s Republic of China; \textsuperscript{b}Centre for Biomedical & Life Sciences, Singapore Polytechnic, Singapore, Republic of Singapore; \textsuperscript{c}College of Computer Science, Southeast University for Nationalities, Sichuan, Chengdu, People’s Republic of China

(Received 14 March 2008; final version received 25 June 2008)

Glycosyltransferase is an essential and easily accessible drug target for antibiotic-resistance. The crystal structures of glycosyltransferase (GT\textsubscript{51}) provide us with the chance to develop new antibiotics that interrupt a yet unexplored molecular target. Based on the crystal structure of GT\textsubscript{51}, we have carried out computational screening of GT\textsubscript{51} in order to look for novel GT\textsubscript{51} inhibitors. The present study was accomplished by using advance docking and scoring methodology. It is the first example of virtual screening of GT\textsubscript{51} inhibitors. Two docking procedures (Surflex-Dock and FlexX-Pharm dockings) were applied and nine novel potential leads are proposed after thorough examination by a combination of methods.

Keywords: glycosyltransferase\textsubscript{51}; virtual screening; docking; Surflex-Dock; FlexX-Pharm

1. Introduction

Rapidly increasing antibiotic resistance in recent years has rendered the current antibiotics ineffective for treating many microbial infections, resulting in a worldwide health care crisis. Therefore, new targets for developing novel antimicrobial agents are urgently needed for combating the antibiotic crisis. Peptidoglycan, the major structural component of the bacterial cell wall, is essential for bacterial growth. Since the polymer is absent in humans, disruption of peptidoglycan synthesis is an attractive method for eliminating bacteria in a search for new antibacterial agents. In the final stages of its synthesis, the disaccharide units are polymerised by the Glycosyltransferase (GT) and the peptide chains are cross-linked by the transpeptidase (TP). Previous studies [1–3] of the final steps of peptidoglycan biosynthesis have tended to focus on the transpeptidase. However, the inhibition of TP activity is only half the story, and the GT activity of the bifunctional enzymes is an excellent target for the development of new antibiotics [4]. The bifunctional enzymes include PBP1b from \textit{Escherichia coli} and PBP2 from \textit{Staphylococcus aureus} [5]. Despite their importance to bacterial physiology and drug discovery, they have not been studied in detail. Lovering et al. [6] recently reported the first three-dimensional structure of PBP2 in \textit{Science}; they determined the X-ray crystal structure of the GTs family member GT\textsubscript{51}. It provides us an opportunity for the rational design of novel inhibitors targeting the active site of GT\textsubscript{51}. Up to now, there is only one well-characterised inhibitor of the GT\textsubscript{51}, the \textit{Streptomyces} natural product moenomycin (MOE), which is not effective in humans due to poor absorption properties [7]. Therefore, it would be very useful to design a new antibiotic targeting GT\textsubscript{51}. Here, we report the discovery of several compounds as high potency GT\textsubscript{51} inhibitors using the virtual screening method. These novel molecular scaffolds present an opportunity for further optimisation into more potent antibiotics compounds.

Virtual screening may be regarded as a conformational filter that reduces the size of a chemical library to be screened experimentally. It improves the likelihood of finding a good compound as opposed to random screening. There has recently been increasing success in using molecular docking methods [8,9] in the discovery of new chemical entities that bind to structures of therapeutic targets [10,11]. Here, we report how this approach led to the discovery of a new class of potential GT\textsubscript{51} inhibitors.

2. Materials and methods

2.1 Protein structure preparation

GT\textsubscript{51} crystal structures were available at the initiation of this study. The crystal structure of GT\textsubscript{51} was used as the target for virtual screening. The structure of GT\textsubscript{51} with 2.8 Å resolution (from \textit{Staphylococcus aureus}) complexes with inhibitor MOE was retrieved from Protein Data Bank (PDB, 2OLV). The putative catalytic residues of GT\textsubscript{51} are Glu114 and Glu171. The important features that interact with the amino acids Glu114 and Glu171 of GT\textsubscript{51} were identified from previous crystallographic study [6]. A careful docking investigation shows the ligand MOE form the two key hydrogen bonds with these amino acids (E114 and E171).
2.2 Compound database preparation
Two compounds databases (DrugBank and Life Chemical) were used for virtual screening. DrugBank (http://redpoll.pharmacy.ualberta.ca/drugbank/) is a unique bioinformatics and cheminformatics resource that combines detailed drug data with comprehensive drug target information, the database contains about 4100 drug entries [12]. Life Chemical database (http://www.lifechemicals.com/) contains about 170,752 commercially available organic compounds. To save screening time and hit selection time, the raw database was pre-processed by drug-likeness (Lipinski’s rule of five) to omit unreasonable compounds. For Lipinski’s rules [13,14], the maximum molecular weight was set to 500; hydrogen bond donors (OHs and NHs) were set to < 6; hydrogen bond acceptors (Ns and Os) to < 9; $M \log P < 6$. Apart from this, the number of rotatable bonds was limited to eight. The compounds we finally got obey Lipinski’s ‘rule of five’, which making them ideal candidates for drug discovery. All compounds of two databases were minimised with Tripos force field; Gasteiger–Huckel charges were then assigned.

2.3 Docking methodology
One of the most widely used programs for docking is FlexX, and FlexX-Pharm [15] is an extended version of FlexX, in which a set of pharmacophore features in the active site was previously defined to constrain the docking calculation. An empirical scoring function and a patented search engine were used in Surflex-Dock to dock ligands into a protein’s binding site. As reported, it is particularly successful at eliminating false positive results [16] and offering unparalleled enrichments in virtual high throughput screening combined with state-of-the-art speed, accuracy and usability [17,18]. In our work, both of the two docking programs were used for virtual screening against GT51.

For Surflex-Dock set, generation of a pseudo-binding site (called protomol) should be done at first. Protomol construction was based on protein residues proximal to the native ligand and on parameter settings to produce a small and buried docking target. Two parameters (protomol_bloat and protomol_threshold) determine extent of the protomol and the docking performance will depend on the binding site (protomol). Protomol_bloat is a factor determining how far from a potential ligand the site should extend and protomol_threshold is a factor about how deep into the protein the atomic probes used to define the protomol can penetrate (more publications at: http://www.biopharmics.com/publications.html) [19]. We used the default parameter: protomol_threshold is 0.5 and protomol_bloat is zero, the default values are adequate for most datasets. In this work, protomol was generated based on identification of the residues proximal to the native ligand (MOE) in 2OLV. Docking was run with the new whole molecule approach and with default settings for all other parameters. Each docking of a putative ligand returned up to 30 scored poses, with the score consisting of a nominal affinity score in units of $-\log(K_d)^2$ as in Hammerhead.

For FlexX-Pharm docking, active sites of the target were defined as all the residues within the 6.5 Å of the bound ligand and one conserved water molecule was added. The most critical interactions between ligand and target i.e. hydrogen bonds to Glu114, Glu171, Tyr196, Arg167, Lys163 and Lys155 were mapped in the crystal structures as shown in Figure 1 [20–22]. Default parameters of FlexX-Pharm program were used as implemented in SYBYL7.3 with pharmacophore constraints and default protonation states were used as defined in FlexX. Formal charges at protein residues were normally assigned. All stored poses were rescored using the CScore of SYBYL7.3 comprising the following functions: Total score, G_score, D_score, Potential of Mean Force (PMF), Chem_score and Frag_No scoring function [23,24].

3. Result and discussion
3.1 Docking study with known inhibitors
Firstly, re-docking [25] was performed between the target and its co-crystallised ligands. These calculations were completed to validate the docking algorithm. The root mean square distance (RMSD) between the native co-crystallised ligand and its docked copy was calculated. A lower number RMSD indicates a good docking. Values below 2.5 Å are considered acceptable. Despite of large ligand size, the RMSD values of the two re-docking experiments were less than 2.5 Å (for Surflex-Dock RMSD = 1.88 Å and for FlexX-Pharm RMSD = 1.91 Å) for the top-ranking solutions [26]. The superposition of the docked conformations and the crystallographic conformations of MOE is quite good. The binding mode of the ligand-targeting GT51 is displayed in Figure 1.

In order to be assured of the docking experiment accuracy, it was desirable to assess how the docking scores used predict the affinities of known inhibitors (the structures of which as shown in Figures 2 and 3), which have an experimentally determined IC50 values. For this purpose, 12 different experimentally known inhibitors (the first 12 compounds of 16 known inhibitors) representing the active molecules from their respective classes were taken from the available literature [27–29] and were docked into GT51-binding site. The docking scores of the reference molecules together with their IC50 are shown in Table 1. As this table shows, the Surflex-Dock scores vary from 7.55 to 10.02 and the FlexX-Pharm score varies from $-15.08$ to $-28.58$ for the previously reported active inhibitors. Since the inhibitors were reported by different groups, direct comparison of IC50 with the docking scores...
is obviously difficult. Nevertheless, irrespective of this direct correlation comparison, a hit with FlexX-Pharm score higher than the reported inhibitors should ideally present a potential lead. This is because the procedures taken for all reference molecules were identical using the same protein, the same active site, the same software and hardware. Moreover, the different docking scores are all the measures of the free energy and bind affinity of inhibitor-\textit{GT}_{51} binding.

As a further means of validating the docking scores, four active molecules (the last four compounds of known inhibitors) and 10 inactive were taken into the docking experiment using the same methodology. All the active molecules were found with very high fitness scores while the inactive molecules showed low fitness scores. The scores of all 16 known active compounds are displayed together with their IC_{50} values in Table 1. Both Surflex-Dock and FlexX-Pharm scores appeared to parallel experimental IC_{50} quite well. This all shows that Surflex-Dock and FlexX-Pharm scores are reliable enough to make a decision as to which hits to take for further study. Hence, these scores were used as a basis to select compounds that would be expected to bind with a higher affinity.

3.2 Identification of molecules by virtual screening

Virtual screening often produces some false positives and a few false negatives. It is believed that concurrent use of more than one docking program will readily minimise these errors. FlexX-Pharm scores are measures of the free energy of binding and Surflex-Dock scores are measures of binding affinities. The highest scores that could be obtained from the reference molecules for Surflex-Dock is 10.02 and FlexX-Pharm for 28.58. Since there are only two molecules with scores above 10.00 (\(-\log(K_d)^2\)) in the reference structures for Surflex-Dock docking, compounds with a Surflex-Dock score higher than 10.00 were selected. Meanwhile, there are only three molecules with higher scores than \(-28.00\) in the reference structures for FlexX-Pharm, so compounds with a score better than \(-28.00\) were selected. As FlexX-Pharm docking scores are
measures of the free energy and Surflex-Dock docking scores are measures of binding affinities, hits with either scores better than 28.00 for FlexX-Pharm docking or scores better than 10.00 for Surflex-Dock docking were taken to represent novel potential leads. Using these cut-offs, the DrugBank database search yielded 15 hits while 12 were obtained from the Life Chemical search. Apart from this, the scores from other scoring functions were

Table 1. Sulflex-Dock and FlexX-Pharm docking scores of 16 known active compounds of GT51.

<table>
<thead>
<tr>
<th>ID no.</th>
<th>Total score</th>
<th>Crash</th>
<th>Polar</th>
<th>IC50 (µM)</th>
<th>Surflex-Dock scores</th>
<th>FlexX-Pharm scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total score</td>
<td>G_score</td>
</tr>
<tr>
<td>1</td>
<td>M4c</td>
<td>8.13</td>
<td>-3.12</td>
<td>10.03</td>
<td>0.152</td>
<td>-22.10</td>
</tr>
<tr>
<td>2</td>
<td>M4d</td>
<td>8.81</td>
<td>-3.52</td>
<td>10.17</td>
<td>0.089</td>
<td>-21.84</td>
</tr>
<tr>
<td>3</td>
<td>M4e</td>
<td>8.85</td>
<td>-3.4</td>
<td>8.17</td>
<td>0.029</td>
<td>-25.76</td>
</tr>
<tr>
<td>5</td>
<td>M4g</td>
<td>8.67</td>
<td>-3.19</td>
<td>8.54</td>
<td>0.042</td>
<td>-25.32</td>
</tr>
<tr>
<td>6</td>
<td>M6a</td>
<td>8.53</td>
<td>-2.94</td>
<td>7.51</td>
<td>0.031</td>
<td>-23.44</td>
</tr>
<tr>
<td>7</td>
<td>M6b</td>
<td>8.95</td>
<td>-4.35</td>
<td>9.67</td>
<td>0.012</td>
<td>-25.39</td>
</tr>
<tr>
<td>8</td>
<td>M6c</td>
<td>10.01</td>
<td>-2.36</td>
<td>9.29</td>
<td>0.05</td>
<td>-28.55</td>
</tr>
<tr>
<td>9</td>
<td>MOE</td>
<td>9.65</td>
<td>-3.21</td>
<td>13.1</td>
<td>0.0156</td>
<td>-24.10</td>
</tr>
<tr>
<td>10</td>
<td>TS0510</td>
<td>10.02</td>
<td>-4.67</td>
<td>7.86</td>
<td>0.0081</td>
<td>-26.86</td>
</tr>
<tr>
<td>11</td>
<td>TS0511</td>
<td>9.98</td>
<td>-3.17</td>
<td>9.92</td>
<td>0.0086</td>
<td>-28.39</td>
</tr>
<tr>
<td>12</td>
<td>TS0512</td>
<td>9.21</td>
<td>-3.22</td>
<td>6.09</td>
<td>0.016</td>
<td>-25.93</td>
</tr>
<tr>
<td>14</td>
<td>TS30153</td>
<td>7.55</td>
<td>-3.24</td>
<td>6.64</td>
<td>14.3</td>
<td>-15.08</td>
</tr>
<tr>
<td>15</td>
<td>TS30663</td>
<td>7.65</td>
<td>-3.42</td>
<td>8.92</td>
<td>9.8</td>
<td>-15.87</td>
</tr>
<tr>
<td>16</td>
<td>TS30888</td>
<td>7.74</td>
<td>-2.44</td>
<td>8.38</td>
<td>6.7</td>
<td>-16.21</td>
</tr>
</tbody>
</table>

Note: The activity of each compound (IC50) is also listed.
Figure 4. Structures of the nine validated hits.
also requested in the hit verification. In most of the cases, the PMF, Chem_score, G_score and D_score did correlate with the experimental IC50 and hence, these too were used to aid in prioritisation [30]. As can be seen, these docking scores for the hits are mostly better than the reference molecules.

For docking programs and scoring functions, a number of false positives would usually appear in the top ranking list. It is necessary to check and analyse the binding mode of each compound to determine if it has reasonable interaction and geometry fitting. We have investigated all the binding modes and binding sites of the compounds for their shape complementarily and potentiality in forming hydrogen bonds in this intersection. The molecules that did not fit GT51-binding site well were excluded from the list after this inspection. Furthermore, we observed visually how the final hits that remained docked into the GT51-binding site. Hits that docked well were finally chosen and anticipated to be new potential hits. Last, but importantly, the binding modes of the hits were visually analysed. Nine hits (DrugBank database yielded seven hits while Life Chemical database yielded two) were finally

Table 2. Sulflex-Dock and FlexX-Pharm docking scores of the nine hits identified.

<table>
<thead>
<tr>
<th>ID no.</th>
<th>Total score</th>
<th>Crash</th>
<th>Polar</th>
<th>Total score</th>
<th>G_score</th>
<th>PMF</th>
<th>D_score</th>
<th>Chem_score</th>
<th>CS</th>
<th>FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>A4P</td>
<td>10.02</td>
<td>-2.89</td>
<td>5.03</td>
<td>-29.10</td>
<td>-124.35</td>
<td>-49.55</td>
<td>-123.90</td>
<td>-30.27</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>CAA</td>
<td>10.21</td>
<td>-3.50</td>
<td>2.17</td>
<td>-29.84</td>
<td>-175.69</td>
<td>-47.99</td>
<td>-99.59</td>
<td>-35.90</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>BA3</td>
<td>10.26</td>
<td>-3.22</td>
<td>5.16</td>
<td>-28.96</td>
<td>-156.06</td>
<td>-52.30</td>
<td>-196.60</td>
<td>-28.60</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>M439446</td>
<td>10.59</td>
<td>-3.62</td>
<td>5.02</td>
<td>-36.91</td>
<td>-136.27</td>
<td>-59.69</td>
<td>-89.59</td>
<td>-36.32</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>BCA</td>
<td>10.09</td>
<td>-2.19</td>
<td>6.21</td>
<td>-37.32</td>
<td>-150.40</td>
<td>-37.69</td>
<td>-83.59</td>
<td>-36.01</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>M01581</td>
<td>10.03</td>
<td>-2.07</td>
<td>5.94</td>
<td>-33.66</td>
<td>-186.69</td>
<td>-56.01</td>
<td>-79.73</td>
<td>-39.88</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>APR</td>
<td>10.35</td>
<td>-2.35</td>
<td>3.89</td>
<td>-38.39</td>
<td>-151.98</td>
<td>-36.31</td>
<td>-43.59</td>
<td>-30.91</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>ADQ</td>
<td>10.11</td>
<td>-3.02</td>
<td>6.07</td>
<td>-30.55</td>
<td>-156.29</td>
<td>-42.56</td>
<td>-58.87</td>
<td>-36.98</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>AP5</td>
<td>10.15</td>
<td>-2.02</td>
<td>6.81</td>
<td>-34.10</td>
<td>-176.03</td>
<td>-38.69</td>
<td>-120.26</td>
<td>-39.96</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 5. Interactions between M439446 and the active sites of GT51.
found to dock well into the GT\textsubscript{51}-active site. The structures of these nine validated hits identified are shown in Figure 4 while the docking scores are shown in Table 2. The interactions of the two highest scoring hits displayed in Figures 5 and 6 attest that the hits do retain the important common binding mode that is known for the inhibitors so far reported. Chemically, hydrophobic nitrogen heterocycles can be a general feature in the previously reported inhibitors.

3.3 Analyse interactions of M439446 and APR with GT\textsubscript{51}

The interaction of M439446 with GT\textsubscript{51}-active site is shown in Figure 5. As the figure shows, the NH\textsubscript{2} group is seen to make hydrogen-bonding interactions with the carbonyl oxygen of the amide bond of Glu114. The carbonyl group of the amide moiety is noticed to make hydrogen-bonding interactions with the hydroxyl of Glu171. Besides, the carbonyl group of phosphoric acid group is noticed to make hydrogen-bonding interactions with the NH\textsubscript{2} of Tys163, Arg167 and Lys155. The interaction of APR with GT\textsubscript{51}-binding site is shown in Figure 6. Glu114 and Glu171 are interacting via hydrogen bonding with the OH of the hit. The carbonyls of Asp156 are also interacting via hydrogen bonding with the NH of the hit. Finally, the NH of Tyr196 is observed to make favourable hydrogen bonding-interactions with one of the nitrogen heterocycles. These hydrogen-bonding interactions are acknowledged common to all of the previously experimentally known GT\textsubscript{51} inhibitors. In both cases, the hits were able to reproduce this common binding mode seen so far.

4. Conclusions

The present study is the first example for computer aided design of GT inhibitors. It was accomplished by virtual screening for inhibitors of GT\textsubscript{51} advance docking and scoring methodology. In this work, a new class of compounds with potential activity against GT\textsubscript{51} has been selected by means of docking-based virtual screening. The hits have been further investigated through a combination of several standard hit validation measures: visual examination of how well they dock into the GT\textsubscript{51}-binding groove, detailed analysis of their docking scores using different scoring functions, comparative investigation of the docking scores of the hits with that of the thus far known 16 inhibitors of GT\textsubscript{51} and last, but importantly, how the hits retain interactions with the important amino acid residues of GT\textsubscript{51} binding site with which the known GT\textsubscript{51} inhibitors are known to interact. The data obtained provides a strong indication for the usefulness of our docking and scoring approach. These novel molecular scaffolds may present an opportunity for further optimisation into more potent antibiotics compounds.
Furthermore, the present work may further our current knowledge of the molecular basis of activation, inhibition and regulation of this pharmaceutically important enzyme. Future investigations based on the inhibitor molecules identified here will lend needed insight into the mechanism, targeting, and therapeutic potential of this enzyme.

References