Chem-Bio Informatics Journal, Vol. 11, pp.52-62(2011)

Discovery of Novel Antimicrobial Agents Targeting the Bacterial RNA Polymerase by High-Throughput Virtual Screening

Kenji Onodera*, Takayasu Kawasaki, Shunsuke Kamijo

Institute of Industrial Science, University of Tokyo, 4-6-1 Komaba, Meguro, Tokyo 153-8505, Japan.

**E-mail: onodera.kenji@gmail.com*

(Received February 24, 2011; accepted May 17, 2011; published online June 13, 2011)

Abstract

Bacterial RNA polymerase (RNAP) is the least popular target for antibiotics, and currently Rifampicin is only an approved drug for clinical use. However, RNAP is essential for bacterial growth and survival, and it can be a promising target for antimicrobial agents. Thus, we decided to search new antimicrobial agents for RNAP by virtual screening. When virtual screenings are performed, certain compounds repeatedly appears on hits covering a wide range of targets (frequently hitters). Also, the performance of hit generation is important factor in success of the virtual screening. Since we previously developed the optimized docking scores, we examined our scoring methods with rigorous removals of frequent hitters. We used two complex structures for RNAP, and also used two unrelated structures as negative controls to remove frequent hitters. Finally, we selected seven high-scored candidates from hits, and two of them showed the inhibition of Gram-positive bacteria by paper disk agar diffusion assay *in vivo*.

Key Words: Docking, virtual screening, scoring function, score optimization, antibiotics

Area of Interest: Information and Computing Infrastructure for Drug Design and Toxicology

1. Introduction

Bacteria evolve and always find out ways to overcome antibiotics. Usages of antibiotics cause mutations in bacteria that bring drug resistance. One of the solutions we have is to keep developing new drugs all the time. Thus, many new classes of antibiotics have been developed. Currently, antibiotics target at cell envelope, DNA replication and transcription (supercoiling and relaxing of DNA), transcription (RNA synthesis), translation, fatty acid synthesis, and tetrahydrofolate (cofactor) biosynthesis [1]. Among them, most popular target is cell envelops, and one of the most popular antibiotics in the category is β-lactams such as penicillin. On the other hand, the least popular target is transcription (RNA synthesis).

Bacterial RNA polymerase (RNAP) is the enzyme responsible for transcription of DNA into RNA. It is composed of multi-subunits, α2ββ'. The known inhibitor of RNAP is Rifampicin (Fig. 1). It is only approved drug for clinical use targeting RNAP, and used for treatments of tuberculosis and inactive meningitis [2]. Rifampicin binds in a pocket of the RNA polymerase β subunit deep within DNA/RNA channel. It probably blocks the RNA extrusion pathway and stop synthesis of RNA transcripts longer than 2-3 nucleotides in length [3][4]. Streptolydigin and Sorangicin A are also known inhibitors of RNAP (Figure 1). Sorangicin A shares the same β subunit pocket with Rifampicin. All residues that interact with Rifampicin also interact with Sorangicin A [4]. Streptolydigin binds to the site adjacent to but not overlapping Rifampicin active site (Figure 2), and stabilizes a conformation of RNAP with a straight-bridge-helix [5].

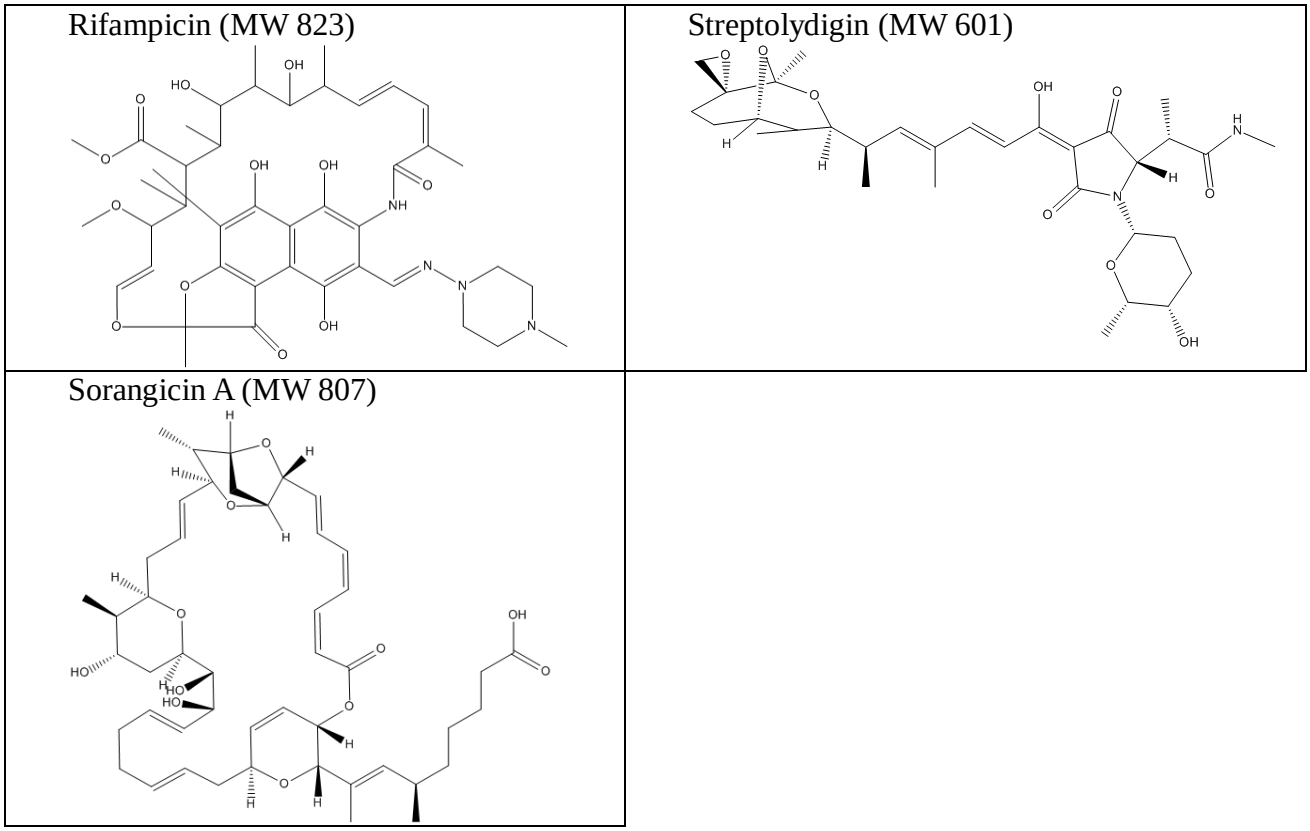


Figure 1. Known antibiotics for RNAP

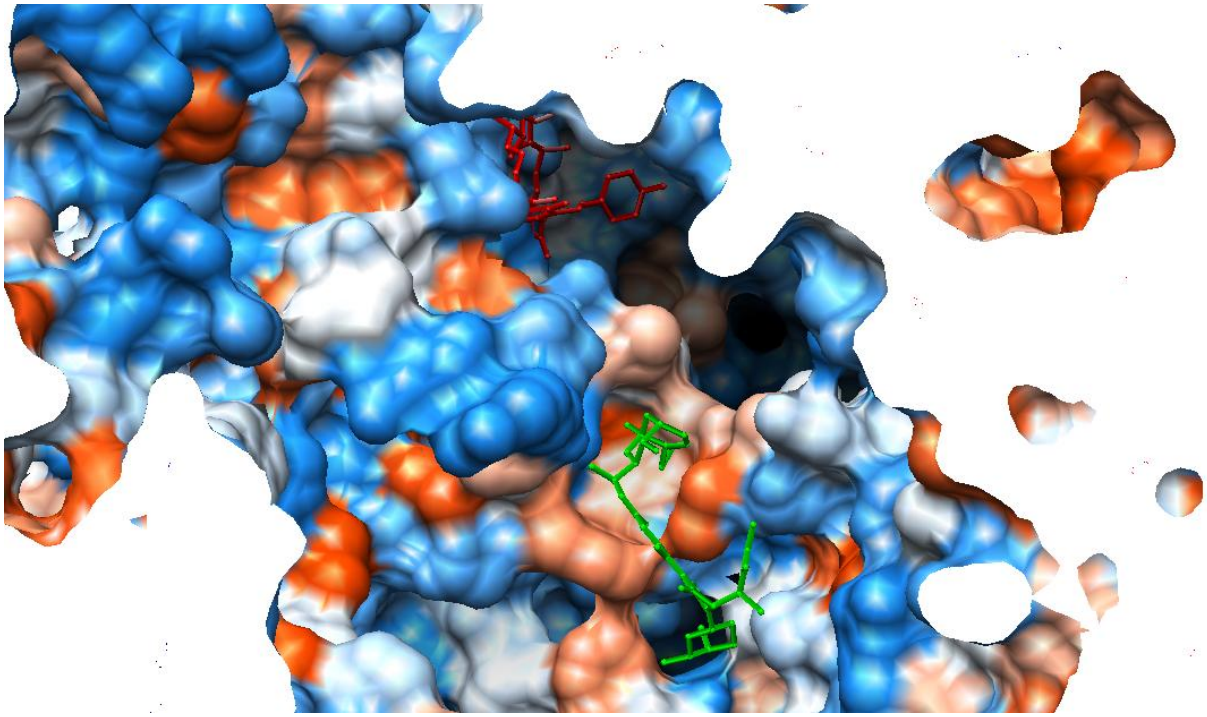


Figure 2. Active sites of Streptolydigin (in Green) and Rifampicin (in Red) on *Thermus thermophilus* RNA polymerase
In this figure, Rifampicin in PDB ID: 1YNN was superimposed to PDB ID: 1ZYR. Surface colors range from blue for the most hydrophilic to orange red for the most hydrophobic. White is used for neither hydrophilic nor hydrophobic surface.

Since RNAP needs conformation changes between straight-bridge-helix and bent-bridge-helix for its function, Streptolydigin can interfere with NTP binding and translocation of RNAP.

Previously, we developed the optimized docking scores for general usage by using various types of target proteins [6]. A total of 113 ligand-protein complexes were obtained from the Protein Data Bank (PDB). They are used as training and test sets for their developments. In the screening using the test set, the chance to discover the active ligands were significantly improved from 36.8% to 77.2% in 20-time enrichment. Since the optimized docking scores were developed using various targets not specific to certain types of proteins, the scores can be applied to virtual screening of antimicrobial agents targeting RNAP.

RNAP is essential for bacterial growth and survival [2]. However, RNAP is the least poplar target for the antibiotics. Thus, developments of antimicrobial agents for RNAP may open the door for new class of antimicrobial therapy. Also, small compound inhibitors could be useful tools for determining the interest of RNAP [3][7], or they can be a seed for new precious antimicrobial agents. Therefore, in this study, we decide to search antimicrobial agents for RNAP from a library of small fragmental compounds. From hits of the virtual screening, we finally discovered two novel agents to inhibit growth of Gram-positive bacteria *in vivo*.

2. Materials and Methods

2.1 Materials

Bacillus subtilis (NBRC 3009), *Thermus thermophilus* (NBRC 101084) and *Thermus aquaticus* (NBRC 103206) were obtained from NITE Biological Resource Center (NBRC), Japan.

Growth medium for *Bacillus subtilis* was ten grams of Polypepton (BD, NJ, USA), two grams of Yeast extract (Nakalai Tesque, Japan), and one gram of $MgSO_4.7H_2O$ (Nakalai Tesque, Japan) in one litter of distilled water. Growth media for *Thermus aquaticus* and *Thermus thermophilus* were eight grams of Polypepton, four grams of Yeast extract, and two gram of NaCl (Nakalai Tesque, Japan) in one litter of distilled water. LB medium (Nakalai Tesque, Japan) was used for *Escherichia coli*. For solid medium, 1.5% w/v agar (Nakalai Tesque, Japan) was mixed to the growth media.

Compounds 1 and 6 were purchased from OTAVA, Ukraine. Compound 2 was purchased from Princeton BioMolecular Research, USA. Compound 3 was purchased from Labotest, Germany. Compound 4 was purchased from Enamine, Ukraine. Compound 5 was purchased from Vitas-M, Russia. Compound 7 was purchased from Toronto Research Chemicals, Canada. Rifampicin was purchased from Nakarai Tesque, Japan.

2.2 Paper Disk Diffusion Assay

After bacterial culture was applied on the solid medium, six millimeters paper disks in diameter (Whatman filter papers by GE, USA) were placed. Then, two micro litters of a testing compound dissolved in DMSO was applied on each disk. The medium was incubated on 30, 37, 60 or 60 °C for *Bacillus subtilis*, *Escherichia coli*, *Thermus thermophilus* or *Thermus aquaticus*, respectively for over-night or until colonies appeared.

2.3 Screening Library

ZINC database was used for screening library in virtual screening [8]. The dataset was "clean-fragments" subset (# 12), 120727 entries from ZINC ver.10 downloaded on April 26, 2010. All the entries in this dataset satisfy XLogP [9]<=2.5, molecular weight <=250, and rotatable bonds <=5.

2.4 Target Proteins

We used two complex structures of RNAPs and antibiotics (PDB IDs: 1YNN and 1ZYR) for this study. 1ZYR is a complex of *Thermus thermophilus* RNAP and an antibiotic, Streptolydigin. 1YNN is a complex of *Thermus aquaticus* RNAP and an antibiotic, Rifampicin. Additionally, two more unrelated structures were used for the study, and they were docked to the screening library to remove “frequent hitters.” Those structures were simply selected from first and last IDs (PDB IDs: 1AAQ and 8GCH) from our previous dataset for the docking score optimizations. 1AAQ is a complex structure of human immunodeficiency virus-1 (HIV-1) protease and its inhibitor. 8GCH is structure of γ-chymotrypsin.

2.5 Binding Site Definition.

Binding sites for FRED [10] were defined as rectangular area to cover all heavy atoms of the ligands of the complexes with 5 Å cushion. Shape of binding site differs between GOLD [11] and

FRED, and it is a sphere for GOLD. Since calculation time for GOLD is much longer than FRED, we tried to assign smaller binding sites for GOLD. After trails of re-docking ligands and target proteins, the binding sites for GOLD were determined to 12, 14, 10, and 10 Å in radii for 1AAQ, 1YNN, 1ZYR, and 8GCH from the centers of the ligands, respectively.

2.6 Molecular Docking Programs and Processes.

Bissantz, et. al [12] reported the default settings performed generally well for DOCK [13], FlexX [14], and GOLD. Thus, the parameters used for this study were basically their default settings as described in the next paragraph. Molecular docking programs return several solutions for ligand-target protein complexes in the calculations, but only the best scored solution was selected as a docking result of each compound-target protein complex in each molecular docking program.

In our previous study, the combination of FRED and GOLD made a good combination for the virtual screenings [6]. Thus, we used those two docking programs for this study.

GOLD 4.0.1 is based on a genetic algorithm. GOLD uses mol2 file format for both ligands and proteins, and the mol2 files generated at the previous preparation step with SYBYL [15] were applied for the docking calculations. Calculations were processed using their default settings from a preset of '7-8 times speed-up.' Only minor modifications were made to reduce output file sizes of GOLD; 'clean_up_option save_top_n_solutions' was 1, and 'n_top_solutions' was 3. GoldScore (a force field based score) was used as a scoring function.

FRED 2.2.4 is based on a shape-based docking method. FRED requires a set of input conformers for each ligand in their in-house format. Thus, the screening library was processed with OMEGA [16] to generate a single binary file for all input ligands. FRED was simply run in command line with their default settings. After shape fittings, ligands were optimized by a knowledge-based scoring function, Chemgauss3.

All docking calculations were performed using computers running Ubuntu 4.2.4 (Linux) with Intel® Core™2 CPU 6700 at 2.66GHz and Red Hat 4.0.2 with Intel® Xeon™ 5110 at 1.60GHz.

2.7 Removals of "frequent hitters"

From docking results, we calculated average ranks of compounds for the four targets including non-related targets, and count the numbers of times compounds ranked 5000th or higher. When a compound ranked 5000th or higher in other targets, such a compound was removed as "frequent hitter." Also, a compound was removed if it has average ranks better than top 20000th.

3. Results and Discussions

3.1 Dockings and Rankings

We used two PDB entries for the target RNAPs (PDB IDs: 1YNN and 1ZYR). 1YNN is a complex of *Thermus aquaticus* RNA polymerase and antibiotics, Rifampicin. 1ZYR is a complex of *Thermus thermophilus* RNA polymerase and antibiotics, Streptolydigin. The binding sites of antibiotics differ in these two PDB entries. Rifampicin binds to β subunit of the RNA polymerase in 1YNN, and Streptolydigin binds to a boundary between β and β' subunits of the RNA polymerase in 1ZYR (Figure 2). We also prepared two other unrelated proteins for negative controls (PDB IDs: 1AAQ and 8GCH). Those were simply selected from first and last entries in the dataset previously used for the score optimization [6].

Since we found that the combination of FRED and GOLD yielded the best results in our previous optimizations [6], we used both FRED and GOLD for this study. The screening library of 120,727 entries from ZINC [8] was docked to the four PDB protein entries above by FRED and GOLD. Then, optimized scores were calculated from the docking scores obtained by those docking programs according as the equations previously developed (Eq. 1 and 2) [6].

$$Optimized\ FRED = 0.25 \times [Desolvation] - [Steric] - 1.75 \times [Metal] - 1.25 \times [Donor] - [Acc] \quad (1)$$

$$Optimized\ GOLD = [External\ Hbond] + 0.5 \times [External\ vdw] - 1.5 \times [Internal\ Hbond] + [InternalTorsion] + 0.5 \times [Internal\ vdw] \quad (2)$$

3.2 Removals of "Frequent Hitters"

After we obtained rankings of compounds in the screening library according to the optimized docking scores, we removed "frequent hitters" from the hits. "Frequent hitters" are known as compounds which frequently appeared on hits covering a wide range of targets [17]. It is not clear what frequent hitters are, and usually only several frequent hitters were determined [18]. However, we don't prefer to purchase any suspicious compounds which can be frequent hitters for *in vivo* assay. Thus, we lowered qualification limits of "frequent hitters" and removed all suspicious candidates. In this study, we removed compounds from the hits if they ranked within top 5000 compounds (Top 4.1%) more than twice, or an average rank of the compound was higher than 20000^{th} (Top 16.5%) in the four docking targets. As a result, we removed more than half of compounds from the hits (Table 1).

Table 1. Numbers of hits after removals of suspicious candidates in Top 1000 compounds

PDB ID:	1YNN	1ZYR
FRED	320	361
GOLD	69	62

More top ranked compounds in GOLD were also ranked higher in the four target proteins than those in FRED. This might be because FRED uses a shape-based docking method, and a top ranked compound in one target cannot fit to other targets due to differences in target structures, whereas GOLD focuses more on strengths of ligand-protein interactions, and the same compound can possesses high affinities for several targets.

3.3 Selection of Compounds from the Hits

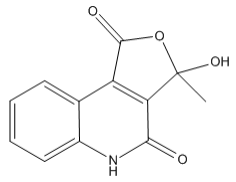
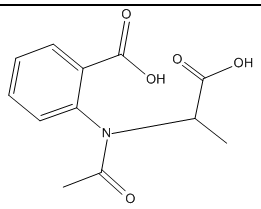
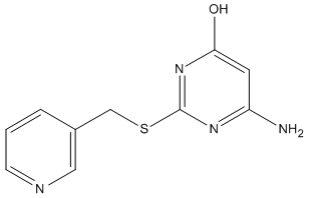
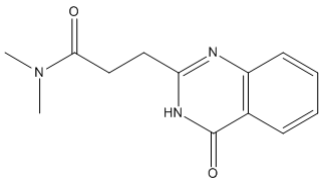
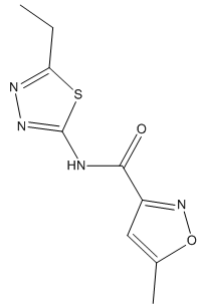
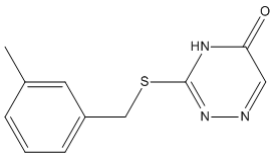
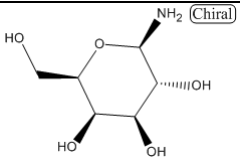
Our previous results suggested that a third part of candidates should be selected from GOLD and the rest should be from FRED. Those should be selected independently without considering ranks or scores in another docking program, since there were no substantial correlations in scores or ranks between FRED and GOLD [6].

We selected compounds from the top 1000 compounds without "frequent hitters" (Table 1). All those top compounds possessed much higher optimized scores than Rifampicin, Streptolydigin, or Sorangicin A. Thus, it is possible that we lower the standards and select more compounds as hits. However, we selected compounds only from the top 1000 as hits, since we could select enough

number of compounds for testing antimicrobial activities.

Additionally, we removed compounds with delivery times more than six weeks for our convenience. From the compounds which passed all the above criteria, we picked compounds with lower average ranks in the four targets. Finally, we selected and purchased total seven compounds; four from FRED and three from GOLD (Table 2).

Table 2. List of selected compounds

 Compound 1 (for 1ZYR by GOLD)	 Compound 2 (for 1YNN by GOLD)	 Compound 3 (for 1ZYR by FRED)
 Compound 4 (for 1ZYR by FRED)	 Compound 5 (for 1YNN by FRED)	 Compound 6 (for 1ZYR by GOLD)
 Compound 7 (for 1YNN by FRED)		

3.4 Paper Disk Agar Diffusion Assay

The seven synthetic compounds were purchased and tested for their antimicrobial activities by paper disk agar diffusion assay. The results showed *Thermus thermophilus* was susceptible to compound 3 and 6 (Figure 3).

When comparing with antibiotics, Rifampicin, zone of inhibition for 100mg/ml of the compound 6 was equivalent to zone for 0.5 mg/ml Rifampicin, and the compound 3 showed weaker inhibition. It means that antimicrobial activity of the compound 6 is approximately 1/200 of Rifampicin. Other five compounds did not show any inhibition. Similar results were observed in *Thermus aquaticus*. It is reasonable that those compounds have antimicrobial activity to *Thermus thermophilus* and *Thermus aquaticus*, because they belong to the same genus.

The compound 6 also showed an inhibition of *Bacillus subtilis*, but it was very faint. No other compounds showed any inhibition in *Bacillus subtilis*. In *E. coli*, no compounds showed any inhibition. Those susceptible to the compound 6 were *Thermus thermophilus*, *Thermus aquaticus*,

and *Bacillus subtilis*, which are all Gram-positive bacteria. *E. coli* are Gram-negative bacteria. Thus, the compounds may work only for Gram-positive bacteria.

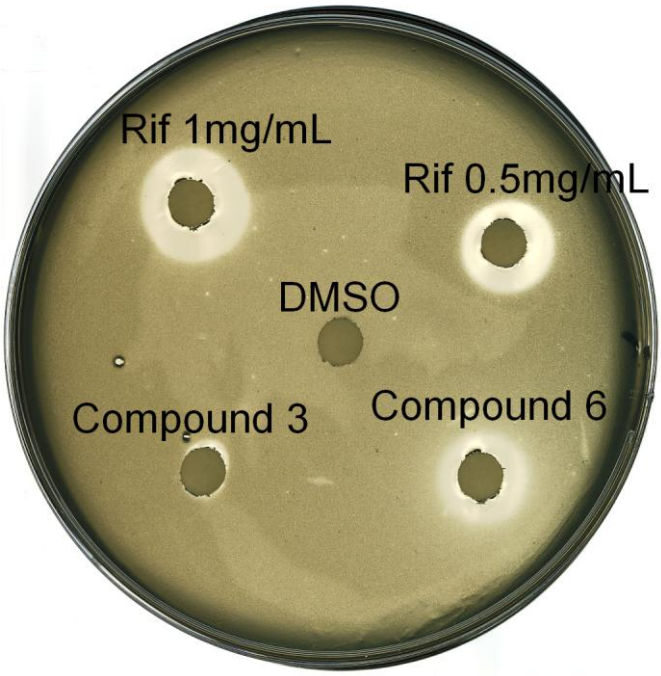


Figure 3. Results of Paper disk diffusion assay with *Thermus thermophilus*
Each disk was applied 2 μL of a compound indicated. The concentration of compounds 3 and 6 was 100 mg/mL.

3.5 Active compound 3 and 6

The docking calculations were performed in and around sites of actions of Streptolydigin with PDB ID: 1ZYR and Rifampicin with PDB ID: 1YNN, separately. From the binding site of Rifampicin, three compounds were selected and purchased, but no compounds showed antimicrobial activities.

From the biding site of Streptolydigin, four compounds were selected and purchased. Then, two compounds (Compounds 3 and 6) were found to have antimicrobial activities. By the diffusion assay, we could not conclude that those compounds can inhibit RNAP directly. However, those compounds were selected from the screening results for RNAP, and they were in a hydrophobic pocket surrounded by β and β' subunits of RNAP (Figure 4). Thus, it is considerably possible that they can inhibit RNAP activities.

We found two noble antimicrobial agents, the compounds 3 and 6, using the optimized scores. The compound 6 were actually ranked top 25th by GOLD original score, but it also ranked within top 4% twice in other targets (2778th in 8GCH and 4835th in 1YNN; Table 3). In our classifications, the compound 6 would be categorized as "frequent hitter" or its suspicious compound if the original scores were used. Therefore, we would not discover the compound 6 if we did not use the optimized scores.

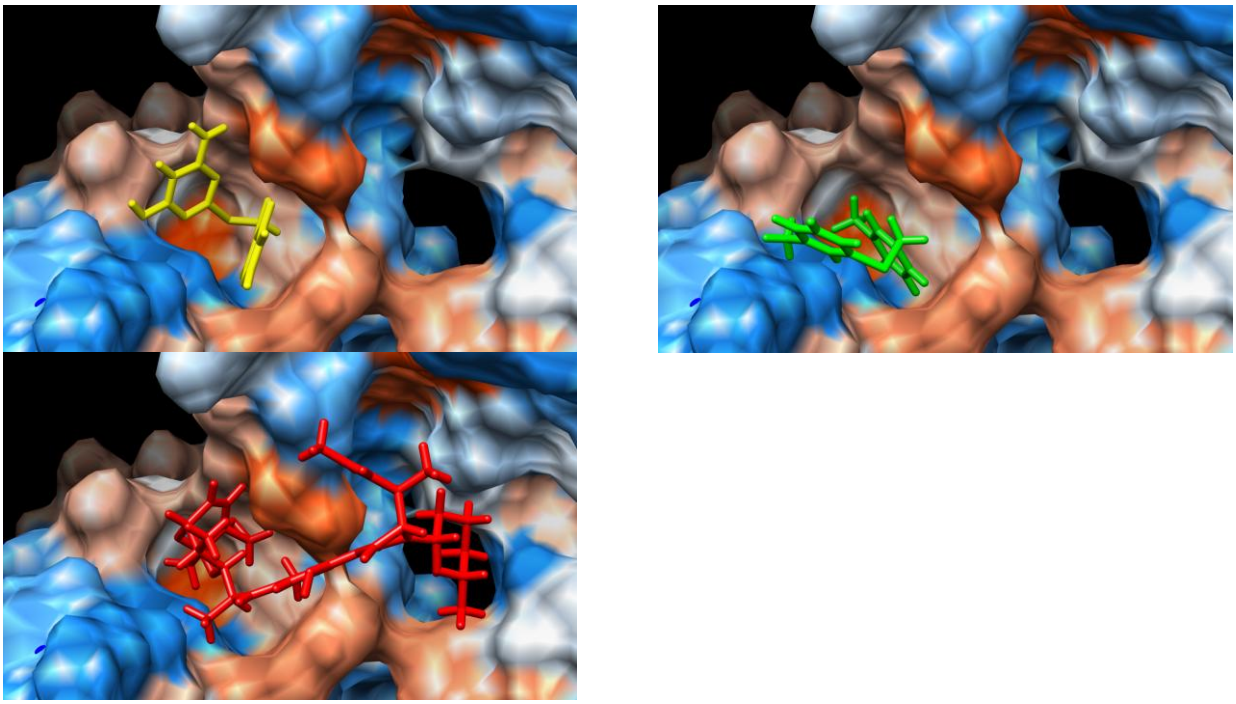


Figure 4. Docked complex structures
Either compound 3 in yellow (upper-left) or 6 in green (upper-right) was docked to *Thermus thermophilus* RNA polymerase (PDB ID: 1ZYR) by FRED or GOLD, respectively. The compound in red (lower-right) shows a binding position of Streptolydigin in 1ZYR. Surface colors range from blue for the most hydrophilic to orange red for the most hydrophobic. White is used for neither hydrophilic nor hydrophobic surface.

Table 3. Ranks by the optimized ranks and original ranks among the four targets

		Rank in 1ZYR	Average Rank	5000th or higher
Compound 3 (by FRED)	Original Score	794	46810	Once
	Optimized Score	248	42777	Once
Compound 6 (by GOLD)	Original Score	25	6325	Three times
	Optimized Score	308	28252	Once

The compound 3 were ranked higher using the optimized score, but both scores ranked within top 1000th and not categorized as "frequent hitter." Thus, the compound 3 would be selected from the hits even though the original scores were used.

In this screening, one additional active compound 6 was discovered by the introduction of the optimized score. We cannot proof the superiority of the optimized scores with just this small example. We developed the optimized scores to differentiate binders from non-binders rather than predict correct binding affinities. We can say it shows promising results that the optimized scores can be a good differentiator of binders from non-binders.

4. Conclusions

We discovered two novel antimicrobial agents for RNAP. Their structures were not similar to any known antibiotics for RNAP. They are small and less than 250 in molecular weight. They also have similar scaffold. Both have two six-membered rings and a sulfur atom between them. Since they were small compounds, and only small differences between the two agents exhibited significant difference in the antimicrobial activity, they can possibly have much higher activity as a new antimicrobial drug by modifying the compounds 3 and 6.

They showed antimicrobial activity to Gram-positive bacteria, T*hermus thermophilus* and *Thermus aquaticus* by the compound 3 and 6, and *Bacillus subtilis* by the compound 6. Gram-negative bacterium, *E. coli*, was not susceptible to those agents. Thus, the compounds may work only for Gram-positive bacteria. The antibiotic, Rifampicin, works for RNAPs of wider species including all the four bacteria used for this study. Thus, structures of RNAPs were similar at least in the binding area of Rifampicin. *E. coli* was not susceptible to the compounds, because only binding area of the agents may slightly differ in each RNA polymerase, or *E. coli* may possesses a resistance to the agents by multidrug resistance [19].

In this study, we used the optimized scores with rigorous removals of "frequent hitters." We purchased four compounds specific to the binding area of Streptolydigin (PDB ID: 1ZYR). Among the four, two exhibited antimicrobial activities. Thus, we can say that our method works. Of course, we cannot claim the superiority of our method just in this screening example. However, this result encourages introduction of our method to everyday works of virtual screenings. We hope that our screening method should be extensively tested for wide varieties of targets by many research groups.

Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081) [20].

References

[1] Bumann, D. Has nature already identified all useful antibacterial targets? *Curr. Opin. Microbiol.* **2008,** *11* (5), 387-392.

[2] Chopra, I. Bacterial RNA polymerase: a promising target for the discovery of new antimicrobial agents. *Curr. Opin. Investig. Drugs* **2007,** *8* (8), 600-607.

[3] Andre, E.; Bastide, L.; Michaux-Charachon, S.; Gouby, A.; Villain-Guillot, P.; Latouche, J.; Bouchet, A.; Gualtieri, M.; Leonetti, J. P. Novel synthetic molecules targeting the bacterial RNA polymerase assembly. *J. Antimicrob. Chemother.* **2006,** *57* (2), 245-251.

[4] Campbell, E. A.; Pavlova, O.; Zenkin, N.; Leon, F.; Irschik, H.; Jansen, R.; Severinov, K.; Darst, S. A. Structural, functional, and genetic analysis of sorangicin inhibition of bacterial RNA polymerase. *EMBO J.* **2005,** *24* (4), 674-682.

[5] Tuske, S.; Sarafianos, S. G.; Wang, X.; Hudson, B.; Sineva, E.; Mukhopadhyay, J.; Birktoft, J. J.; Leroy, O.; Ismail, S.; Clark, A. D., Jr.; Dharia, C.; Napoli, A.; Laptenko, O.; Lee, J.; Borukhov, S.; Ebright, R. H.; Arnold, E. Inhibition of bacterial RNA polymerase by streptolydigin: stabilization of a straight-bridge-helix active-center conformation. *Cell* **2005,** *122* (4), 541-552.

[6] Onodera, K.; Kamijo, S. Universal Optimizations of Scoring Functrions for Virtual Screening.

Chem-Bio Informatics Journal **2010,** *10*, 85-99.
[7] Darst, S. A. New inhibitors targeting bacterial RNA polymerase. *Trends Biochem. Sci.* **2004,** *29* (4), 159-160.
[8] Irwin, J. J.; Shoichet, B. K. ZINC--a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005,** *45* (1), 177-182.
[9] Wang, R.; Fu, Y.; Lai, L. A New Atom-Additive Method for Calculating Partition Coefficients. *J. Chem. Inf. Comput. Sci.* **1997,** *37*, 615-621.
[10] McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. Gaussian Docking Functions. *Biopolymers* **2003,** *68*, 76-90.
[11] 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* (3), 727-748.
[12] Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000,** *43* (25), 4759-4767.
[13] Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des.* **2001,** *15* (5), 411-428.
[14] Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein-ligand docking: current status and future challenges. *Proteins* **2006,** *65* (1), 15-26.
[15] *SYBYL*, Version 6.92; Tripos, Inc.: St. Louis, MO, 2004.
[16] *OMEGA*, Version 2.3.1; OpenEye Scientific Software, Inc.: Santa Fe, NM, 2008.
[17] Roche, O.; Schneider, P.; Zuegge, J.; Guba, W.; Kansy, M.; Alanine, A.; Bleicher, K.; Danel, F.; Gutknecht, E. M.; Rogers-Evans, M.; Neidhart, W.; Stalder, H.; Dillon, M.; Sjogren, E.; Fotouhi, N.; Gillespie, P.; Goodnow, R.; Harris, W.; Jones, P.; Taniguchi, M.; Tsujii, S.; von der Saal, W.; Zimmermann, G.; Schneider, G. Development of a virtual screening method for identification of "frequent hitters" in compound libraries. *J. Med. Chem.* **2002,** *45* (1), 137-142.
[18] Onodera, K.; Satou, K.; Hirota, H. Evaluations of molecular docking programs for virtual screening. *J. Chem. Inf. Model.* **2007,** *47* (4), 1609-1618.
[19] Nikaido, H. Multidrug resistance in bacteria. *Annu. Rev. Biochem.* **2009,** *78*, 119-146.
[20] Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004,** *25* (13), 1605-1612.