



REGULAR PAPER

Multiple mutations in RNA polymerase β -subunit gene (*rpoB*) in *Streptomyces incarnatus* NRRL8089 enhance production of antiviral antibiotic sinefungin: modeling rif cluster region by density functional theory

Saori Ogawa,¹ Hitomi Shimidzu,¹ Koji Fukuda,¹ Naoki Tsunekawa,² Toshiyuki Hirano,³ Fumitoshi Sato,³ Kei Yura,⁴ Tomohisa Hasunuma,⁵ Kozo Ochi,¹ Michio Yamamoto ,^{1,6} Wataru Sakamoto,¹ Kentaro Hashimoto,⁷ Hiroyuki Ogata,⁷ Tadayoshi Kanao,¹ Michiko Nemoto,¹ Kenji Inagaki,¹ and Takashi Tamura ^{1,*}

¹The Graduate School of Environmental and Life Science, Okayama University, Okayama, Japan; ²Institute for Quantitative Biosciences, The University of Tokyo, Bunkyo-ku, Japan; ³Institute of Industrial Science, The University of Tokyo, Komaba, Japan; ⁴Graduate School of Humanities and Sciences, Ochanomizu University, Tokyo, Japan; ⁵Graduate School of Science, Technology and Innovation, Kobe University, Kobe, Japan; ⁶RIKEN AIP, Tokyo, Japan; and ⁷Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto, Japan

*Correspondence: Takashi Tamura, tktamura@okayama-u.ac.jp

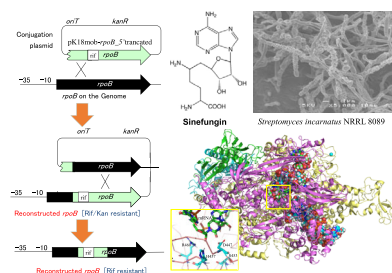
ABSTRACT

Streptomyces incarnatus NRRL8089 produces the antiviral, antifungal, antiprotozoal nucleoside antibiotic sinefungin. To enhance sinefungin production, multiple mutations were introduced to the *rpoB* gene encoding RNA polymerase (RNAP) β -subunit at the target residues, D447, S453, H457, and R460. Sparse regression analysis using elastic-net lasso-ridge penalties on previously reported H457X mutations identified a numeric parameter set, which suggested that H457R/Y/F may cause production enhancement. H457R/R460C mutation successfully enhanced the sinefungin production by 3-fold, while other groups of mutations, such as D447G/R460C or D447G/H457Y, made moderate or even negative effects. To identify why the rif cluster residues have diverse effects on sinefungin production, an RNAP/DNA/mRNA complex model was constructed by homology modeling and molecular dynamics simulation. The 4 residues were located near the mRNA strand. Density functional theory-based calculation suggested that D447, H457, and R460 are in direct contact with ribonucleotide, and partially positive charges are induced by negatively charged chain of mRNA.

Received: 27 December 2020; Accepted: 13 January 2021

© The Author(s) 2021. Published by Oxford University Press on behalf of Japan Society for Bioscience, Biotechnology, and Agrochemistry. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Graphical Abstract



Multiple mutations introduced to RNA polymerase enhanced sinefungin production. RNAP/DNA/mRNA complex model was constructed and DFT calculation estimated interaction between rif-cluster residues and mRNA.

Keywords: antiviral antibiotics, *rpoB* mutation, sparse regression, molecular dynamics simulation, density functional theory

Sinefungin is a nucleoside antibiotic that effectively suppresses the propagation of a variety of viruses, including SARS coronavirus (Bouvet et al. 2010), dengue virus (Selisko et al. 2010), vaccinia virus (Pugh, Borchardt and Stone 1978; Zheng and Shuman 2008), vesicular stomatitis virus (Li, Chorba and Whelan 2007), Epstein–Barr virus (Long et al. 1987), and feline herpes virus FHV-1 (Kuroda et al. 2019). The toxicity of sinefungin to mammals is low and is comparable to that of acyclovir, an antiherpes medicine approved for herpes zoster (Kennedy and Gershon 2018). The oral LD₅₀ of sinefungin is 1 g kg^{−1} (mouse), while that of acyclovir is >10 mg kg^{−1} (mouse) (Andrei et al. 1995; Kuroda et al. 2019). Sinefungin mainly acts via the selective inhibition of viral RNA maturation, which depends on guanine base N7-methylation of Gppp-mRNA transcripts to form m7Gppp-mRNA. This methylation is an integral step in mRNA maturation and, more significantly, protects mRNA transcripts from host 5'-exonucleases that degrade viral transcripts. Sinefungin can inhibit various methyltransferases by competing for binding with the methyl-donor substrate S-adenosyl-L-methionine (SAM). Furthermore, sinefungin inhibits pneumococcal colonization of the rat middle ear *in vivo* and biofilm growth *in vitro*. Bacterial biofilm development is reported to involve SAM-dependent methylation, which elevates the persistence of pathogen resistance to antibiotic treatment and attack by host immune systems (Yadav et al. 2014). Demands are growing for effective and highly selective drugs against viruses, parasites, and antibiotic-resistant pathogens. Exploring the potential utility of sinefungin inevitably requires an ample supply of this nucleoside antibiotic at low cost.

Nucleoside antibiotics belong to a class of microbial secondary metabolites whose biosynthesis is poorly understood but that have potential utility for development as antiviral, antitumor, and antiparasite compounds and thus are of interest as possible pharmaceuticals (Shiraishi and Kuzuyama 2019). The medicinal application of nucleoside antibiotics is hindered by their low levels of production and, sometimes, by variable production by fermentation. For nucleoside antibiotics, it has also proven difficult to identify their biosynthetic gene clusters in the genome sequence of the producer strains. Unlike major classes of antibiotics that are produced in a modular fashion through the selection and condensation of small building blocks such as isoprene, amino acids, and acyl-CoA, there are no known common modular sequences that can be used for identifying

biosynthetic gene clusters of nucleoside antibiotics. As an alternative to the molecular cloning of biosynthetic genes, which can be costly and time-consuming, several studies have enhanced antibiotic production in *Streptomyces lividans* by mutating the portion of the rif cluster in RNA polymerase (RNAP) β -subunit (*rpoB*) (Hu, Zhang and Ochi 2002). This observation was further supported by the report that antibiotic production by *Streptomyces coelicolor* and *S. lividans* was dramatically enhanced by *rpoB* mutation in the genetic background of *relA* and *relC* (Lai et al. 2002). Mutations in the *rpoB* gene of the rif region can also affect the rate of RNA synthesis, suggesting that both the timing and the extent of antibiotic production by *Streptomyces* spp. are crucially affected by the activity of the RNAP within the cell. Single mutations altering residues corresponding to *Streptomyces incarnatus* NRRL8089 RpoB D447, S453, H457, R460, and others were reported to lead to improved antibiotic production in various actinomycetes and the activation of silent genes in various bacteria (Lai et al. 2002). Strain improvement through modulating transcription machinery has an advantage of sparing the cost and effort of molecular cloning of biosynthetic genes clusters. Altering a limited number of amino acid residues in RNA polymerase β -subunit was found to activate poorly expressed gene clusters through screening for rifampicin-resistant phenotype, which accompanies the genotype of mutation in the rif cluster residues (Wehrli et al. 1968; Campbell et al. 2001). Enhanced actinorhodin production was observed for *S. coelicolor* harboring specific amino acid substitutions in the narrow region of the *rpoB* gene (Hu and Ochi 2001; Tamehiro et al. 2003; Inaoka et al. 2004).

Conceivably, altering multiple residues in the rif region in RNAP may lead to significantly higher activation of silent gene clusters than spontaneous single mutations, although this has not previously been examined for *rpoB* mutants. It has been proposed that the *rpoB* mutations can alter their secondary metabolite profiles, and consequently dormant biosynthetic gene clusters may be awakened in *rpoB* mutant strains. However, questions regarding the molecular mechanisms underlying significant alterations in antibiotic production have not been addressed. In the present study, sparse regression analysis was conducted on previously reported single *rpoB* mutations H457X in reference to the physicochemical indexes of amino acid residues, and the most relevant index selected by the machine learning was used to design multiple mutations in the rif cluster region. Then, a method was developed for introducing multiple mutations in

the *rpoB* gene of sinefungin-producing strain of *S. incarnatus* NRRL8089 by homologous recombination with gene fragment that is introduced by conjugation with *Escherichia coli* S17-1 strain and examining the effect of sinefungin production enhancement. Another theoretical approach using homology modeling, molecular dynamics simulations, and density functional theory (DFT) was taken to elucidate the roles of rif cluster residues that may have interaction to the growing strand of mRNA in the RNAP/DNA/mRNA complex.

Materials and methods

Bacterial strains and culture conditions

S. incarnatus NRRL 8089 was obtained from the Agricultural Research Service, US Department of Agriculture. Sinefungin-producing medium SF4 contained, per liter, 10 g D-fructose, 5 g glycerol, 4 g dried yeast extract, 2 g CaCO₃, and 10 g malt extract. The culture medium was adjusted to pH4 with 9 M H₂SO₄. Mycelium (50 mg wet weight) from precultures grown for 2 days at 30°C in 5 mL of SF4 medium were inoculated to 100 mL of SF4 medium and incubated at 28 °C on a rotary shaker set to 200 round/min. Sinefungin production was measured in biological triplicate, and the production levels expressed as mean ± SD.

Sparse elastic-net modeling of *rpoB* H437X mutations using AAIndex

Previously reported successful mutations in rif cluster residues are diverse in their production enhancement effects, which also seem to depend on the secondary metabolites and producer strains. Among the single *rpoB* mutations in rif cluster region, H437X mutations are most abundant in reported number, and the titer-increase effects were subject to the sparse estimation modeling using AAIndex as the explanatory function. The fold-increase and CHOC760104 were log transformed, and then, all variables including the fold-increase and explanatory variables were standardized for regression. An elastic-net regression analysis was carried out using R “glmnet” function (R version 4.0.2, package glmnet version 4.0-2) (Zou and Hastie 2005; Friedman, Hastie, Tibshirani 2010), with the penalty mixing parameter $\alpha = 0.5$. The solution path and the estimated values of the regression coefficients was selected by cross-validation (CV). For sensitivity analysis of the result, further regression analysis was carried out with alternative α values varied to 0.1 and 0.9.

The sparse modeling was not applicable to mutations in the other 3 target residues, D447, S453, and R460, due to the lack of enough mutation data published.

Modeling the *S. incarnatus* RNAP/DNA/RNA complex

A high-quality homology model for RNA polymerase in complex with template double-stranded DNA and mRNA was constructed using a pairwise sequence alignment. The amino acid sequence deduced from the genome sequence of *S. incarnatus* NRRL 8089 (Oshima et al. 2015) was aligned with the sequence of the ternary complex for *Thermus thermophilus* HB8 (PDB, 2051) (Vassilyev et al. 2007) using the ALAdeGAP threading protocol, a protein evolution approach that assumes that the insertion and deletion of residues more frequently occurs on the surfaces of secondary structures rather than their interiors (Hijikata et al. 2011). The resulting alignment was used

to build ternary folds of polypeptide chains using the SWISS-MODEL web server in alignment mode (Bordoli et al. 2009). The polypeptide scaffolds were then combined with atomic coordinates for double-stranded DNA and RNA molecules as well as 2 zinc ions from the PDB file 2051. The resulting geometry of the complex model was further optimized using the Molecular Operating Environment (MOE) platform (Chemical Computing Group, Inc., Montreal, Canada), in which the interior networks of van der Waals interactions, charge-charge interactions, hydrogen bonding and tautomer orientation were optimized using a Protonate3D matrix (Labute 2009) in the implicit Born solvent model (Labute 2008). The homology model was also bound to two Zn²⁺ ions coordinated to a set of four Cys residues, respectively. These Cys residues were conserved in *S. incarnatus* and *T. thermophilus*.

The initial homology model optimized using the MOE platform was further annealed on the molecular dynamics simulation using the AMBER16 program loaded with ZAFF prep/frcmod files for Zn²⁺ ions coordinated to a set of four Cys residues (Peters et al. 2010). The model structure was optimized with molecular dynamics (MD) simulations using the generalized Born solvation model. Simulations were conducted in the PME-MD-MPI module using the SHAKE algorithm for bonds involving hydrogen atoms. Annealing simulations with 1 ns interval were conducted by creating 500 000 consecutive frames at 2 fs increments, with 1 out of every 50 frames chosen and aligned to make a trajectory file for a 1 ns time interval. The average atomic coordinates were computed from the trajectory of 10 000 frames created by MD calculations from 10 to 11 ns.

Mutations in genomic *rpoB* genes created by recombination

E. coli JM109 was used as the host for gene cloning and sub-cloning while *E. coli* S17-1 was used as the donor for the conjugative transfer of DNA to *S. incarnatus*. *E. coli* strains were cultured in Luria-Bertani (LB) liquid or on agar plate medium at 37 °C. Kanamycin (50 mg L⁻¹) and nalidixic acid (50 mg L⁻¹) were added as necessary to maintain the plasmid. The 5'-truncated *rpoB* gene was amplified by the primers *rpoB*_frw_BglII; 5'-CGAGAGATCTGGTCCGTTCCCCGGTGTCT-3' and *rpoB*_rev_HindIII; 5'-CGAGAAGCTTCAGACCTCTTCGACGCTGC-3', then cloned into pK18mob (Schäfer et al. 1994) at the BamHI-HindIII site. Multiple mutations were introduced to *rpoB* by site-directed mutagenesis using a pair of mismatched primers and a KOD FX Neo polymerase kit (Toyobo) and confirmed by DNA sequencing (Eurofins, Tokyo). Primers are listed in Table S1 in the Supplemental Information. Conjugation was carried out by mixing 50 µL total of cell suspension of *E. coli* S17-1 harboring the mutant plasmid and spores of wild type *S. incarnatus* NRRL8089, then spotting this mixture on ISP No. 4 agar plates supplemented with 40 mM CaCl₂, which were then allowed to dry in a laminar flow cabinet. Soft agar (0.75% w/v) containing 100 mg L⁻¹ kanamycin, 100 mg L⁻¹ nalidixic acid and 400 mg L⁻¹ rifampicin was overlaid onto the plate. Conjugant colonies of *S. incarnatus* were obtained after 1 week of incubation at 28 °C, then subcultured onto an ISP No. 4 agar plate then liquid ISP-No. 4 medium. Amplification and DNA sequencing of the *rpoB* gene from the conjugant was used to confirm the presence of the desired mutation. A second homologous recombination mutant was obtained by growing the conjugant strain on ISP No. 4 medium containing 200 mg L⁻¹ rifampicin without kanamycin.

HPLC analysis of sinefungin produced by *rpoB* mutants

The effects of multiple *rpoB* mutations were assessed by measuring wet mycelial weight and sinefungin production in the supernatant of cultures of wild type and mutant *S. incarnatus* strains grown in SF4 medium. Sinefungin production was analyzed by HPLC using a CAPCELL SCX column (4.5 mm I.D. × 150 mm, Shiseido Fine Chemicals) with 0.5 M ammonium formate, pH 4.5, as the mobile phase and a flow rate of 0.5 mL min⁻¹. The eluate was monitored using a Hitachi photodiode array and concentration measured based on absorption at 260 nm. The sinefungin was identified and measured by cochromatography and calibration with authentic sinefungin (Sigma-Aldrich).

Estimation of RNA–rif cluster residue interactions by DFT calculations

To determine the molecular interactions between the ribonucleotides and nearby amino acid residues, including rif cluster residues, quantum chemistry calculations were carried out. Atomic coordinates for the QM/MM calculation were cut out on VMD by choosing atoms within a 20 Å range of H457. The resulting peptide fragments were capped with formamide (H–CO–NH–) and carboxy amides (–CO–NH₂) at their N- and C-termini, respectively, using the MOE platform. The geometry of these capping ends was optimized on MOE using the Amber99 force field in a born solvation model. For DFT calculations, the Gauss view was used to visualize the QM model structure, and calculations based on DFT was carried out at the B3LYP (d, p) level (Frisch 2009). Mulliken electronic charge was computed for each of the 1481 atoms constituting the QM model and the charge differences computed by subtraction: (peptides + RNA complex model) – (peptide model alone + RNA model alone). After performance of this subtraction, positive values indicate atoms in amino acids for which RNAP receives a partial plus charge ($\delta+$) upon interaction with an RNA strand, while negative values indicate that partial negative charges ($\delta-$) are induced by interaction with RNA strand.

Results and discussion

Sparse elastic-net modeling of *rpoB* H437X MUTATIONS using AAIndex

At present, there is no theoretical approach which can suggest candidate set of mutations in *rpoB* to improve antibiotic production. The effects of previously reported single *rpoB* mutations may be used as an empirical guide to infer mutations by taking into account physicochemical properties of amino acid residues recorded in the AAIndex database (Kawashima, Ogata and Kanehisa 1999). Sparse regression analysis was constructed to narrow down candidate sets of amino acid indices in the database that may represent the most relevant parameters for production improvement of various antibiotics. The production enhancement effect designated by the fold-increase was used as the response variable (Table S2, Supplementary Information), while the standardized values on AAIndex were incorporated as explanatory variables in terms of hydrophobicity, polarity, and interior/exterior distribution in the polypeptide chains. The estimated regression coefficients with the tuning parameter selected by cross validation (CV) were listed in the Table 1. Under the varied ratios employing ridge or lasso penalties, the elastic-net modeling conclusively suggested that CHAM820101 was the most relevant variable with the highest coefficient.

Table 1. Coefficients estimated by sparse regression analysis with elastic-net mixing parameter alpha

Amino acid index	alpha = 0.5	alpha = 0.1	alpha = 0.9
GOLD730101			
GOLD730102	0.120	0.146	
KIRIW790103	0.0101	0.142	
EISD840101	–0.158	–0.102	
COHE430101			
CIDH920105			
CIDH920101			
CIDH920102			
CHAM820101	0.289	0.173	0303
CHAM820102	0.109	–0.0984	
CHOC760104		–0.0235	
COHE430101.1			
ROSM880101		0.0268	
KUHL950101		0.0294	
MITS020101		0.00761	

AAIndex contains numeric parameter sets for 20 amino acid residues, which are subjected to sparse regression analysis against production enhancement effect via H437X point mutations previously reported. The solution paths with various alpha values are shown in Figure S1 in the Supplementary Information.

When the alpha value was changed to 0.1 (much like the ridge regression) and 0.9 (much like the lasso regression), the solution path has changed (Figure S1, Supplementary Information) but these results also supported the highest scores for CHAM820101 (Bigelow 1967). According to this physicochemical index, H437R and H437Y mutations, which very often makes productive effects, are substitution of His (0.23) by similar property values of Arg (0.291), Phe (0.290), and Tyr (0.298), while lower productivity was reported when mutation occurred to residues with lower values like Asp (0.105) and Leu (0.186) (Table S3). Accordingly, H457R and H457Y mutations were employed among our multiple mutations in *S. incarnatus rpoB*.

Effects of multiple mutations on sinefungin production in *S. incarnatus*

A library of pK18mob-*rpoB* plasmids with various mutations was constructed and transferred to *S. incarnatus* NRRL8089 via conjugation with *E. coli* S17-1, then resulting conjugant strains were selected based on rifampicin resistance. In the pK18mob-*rpoB* plasmid, 500-bp of the 5'-region of open-reading-frame of *rpoB* is truncated. When homologous recombination occurs and the rif cluster region of the genomic *rpoB* gene is altered, the *S. incarnatus* conjugant acquires resistance to 400 ppm rifampicin. Successful recombination produces a pair of *rpoB* genes, one that has complete *rpoB* gene with an altered rif cluster region, and another that has a truncated *rpoB* fragment lacking a promoter and 500-bp in the 5'-region. If the recombination occurs another way, it leaves a 5'-deleted *rpoB* with altered rif region and intact wild type *rpoB* sequence, which makes the conjugant sensitive to rifampicin.

The amino acid residues D447, S453, H457, and R460 were mutated in this study. The numbering is based on the ORF elucidated from the nucleotide sequence reported in GenBank AB516306.1 (Fukuda et al. 2010). Using the one-letter amino acid code, substituted residues was written using a capital letter while the original residues were designated using lowercase letters. For example, the D447G single mutant is designated as Gshr, and the strains bearing the double mutation D447G/R460C were designated as GshC (Table 2).

Table 2. Effect of multiple mutations in *rpoB* gene on sinefungin production

Mutant	D447	S453	H457	R460	Production ($\mu\text{g mL}^{-1}$)
dsRC	–	–	R	C	3.0 ± 0.4
GsRF	G	–	R	F	1.7 ± 0.3
GLhC	G	L	–	C	1.6 ± 0.7
GsYr	G	–	Y	–	0.9 ± 0.4
wild type	–	–	–	–	1.2 ± 0.2
GshC	G	–	–	C	0.8 ± 0.1
Gshr	G	–	–	–	0.6 ± 0.3

Altering the rif cluster region of the *rpoB* gene produced several mutant strains that enhanced sinefungin production. The time courses of changes in sinefungin production and mycelia weight of 6 *rpoB* mutants and a wild-type strain suggest that *rpoB* mutations can alter the day of production peak. Wild type strain (dsRC) and mutant strain GshC showed production peak on day 5, while production by mutant Gshr peaked after 6 days (Figure 1a). The other 4 mutant strains examined showed maximum production levels after 8 days, with relative production yields of dsRC > GsRF > GLhC > GsYr (Figure 1b). The *rpoB* genes of wild-type (dsRC) and GshC and Gshr mutant strains all have S453 and H457 while the other mutants tested have either S453L (as in GLhC), H457R (dsRC and GsRF) or H457Y (GsYr). Production enhancement effect was most prominent in the mutant strains dsRC and GsRF, which share the H457R mutation.

In contrast to the varied production of sinefungin by wild type and *rpoB* mutants, their growth measured by wet mycelia weight was within the SD of their triplicate cultures, suggesting that the multiple mutations examined in the present study did not affect the apparent growth of *S. incarnatus* (Figure 1c). The mycelia weight increases up to 25–38 mg mL^{−1} on day 6, then decreases until day 10, where the wet mycelia is 18–28 mg mL^{−1}. Sinefungin production medium is adjusted at pH4 when seed culture is inoculated. Rise of the culture pH up to 6 ± 0.4 coincides with the production peak for sinefungin for all the strains tested (Figure S2, Supplementary Information). It was also observed under confocal laser-scanning microscopy that the percentage of dead cells once drops at day period 4 to 6, and increased while sinefungin production reaches the highest amount, suggesting that secondary metabolism might impose metabolic stress while the cells are coping with acidic conditions (Supplementary Information).

Homology modeling of an *S. incarnatus* RNAP/DNA/RNA complex

Draft genome sequence of *S. incarnatus* NRRL8089 was used to predict the amino acid sequences of the α -, β -, β' -, and γ -subunits of RNA polymerase (Oshima et al. 2015). The ternary complex structure of PNAP/DNA/mRNA was modeled on its *Thermus thermophilus* HB8 counterpart, which is deposited in PDB as 2O5I (Vassilyev et al. 2007). Homology modeling essentially depends on the precise pairwise alignments and comparative modeling; however, conventional threading methods can sometimes allocate gaps in the consensus sequence of target and template sequence alignments, which can produce imprecise domain boundaries for the different secondary structural elements of the target sequence. Based on the idea that the protein interior is more conserved than loops and turns occurring on the protein surface, the ALAdeGAP algorithm assumes that the in-

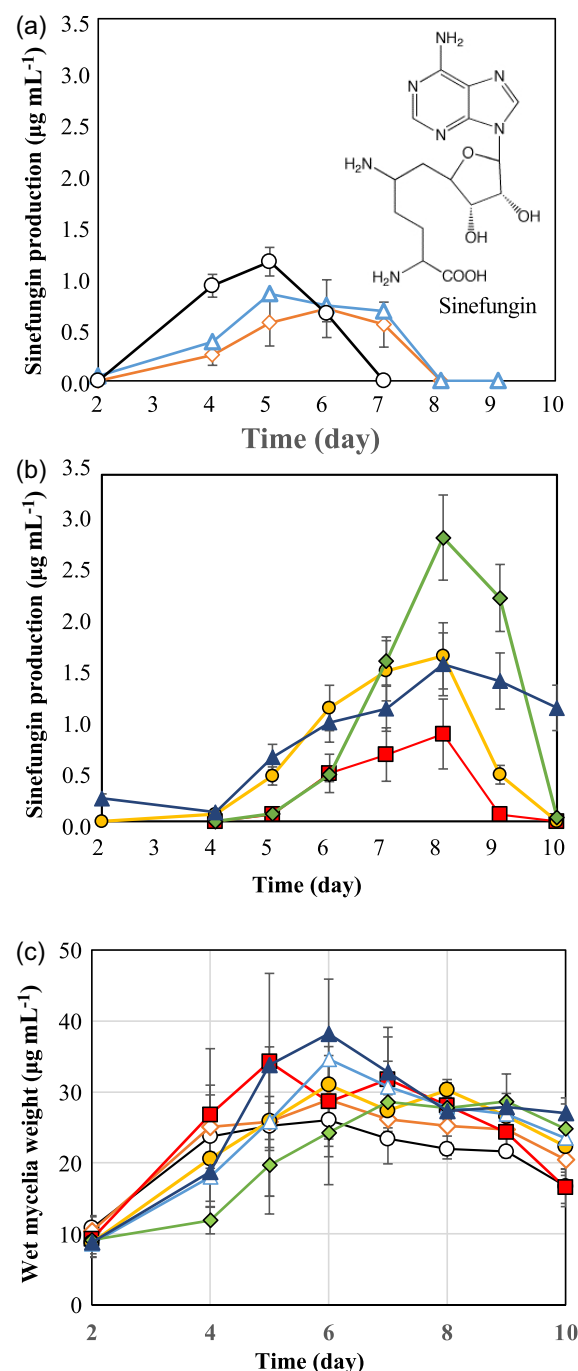


Figure 1. Effect of RNAP *rpoB* gene mutations on sinefungin production by *S. incarnatus*. (a, b) Time-dependent sinefungin production. Data are means from triplicate cultures. The supernatant solutions were analyzed on HPLC implemented with photodiode array detector. Error bars represent SDs from triplicate cultures. (c) Wet mycelial weight was measured by weighting wet pellet of 1 mL culture after centrifugation at 6000 rpm for 15 min. Deviation is also computed for biological triplicate. Symbols: ○, wild type; △, GshC; ◇, Gshr; ▲, GLhC; ◆, dsRC; ■, GsYr; and ●, GsRF.

sertion and deletion of residues occur more frequently on the surface than the interiors of secondary structures like tightly folded helices and strands (Hijikata et al. 2011). The alignment we generated was used to build the ternary folds of polypeptide chains using the SWISS MODEL web server in alignment mode (Bordoli et al. 2009). The polypeptide scaffolds were then joined

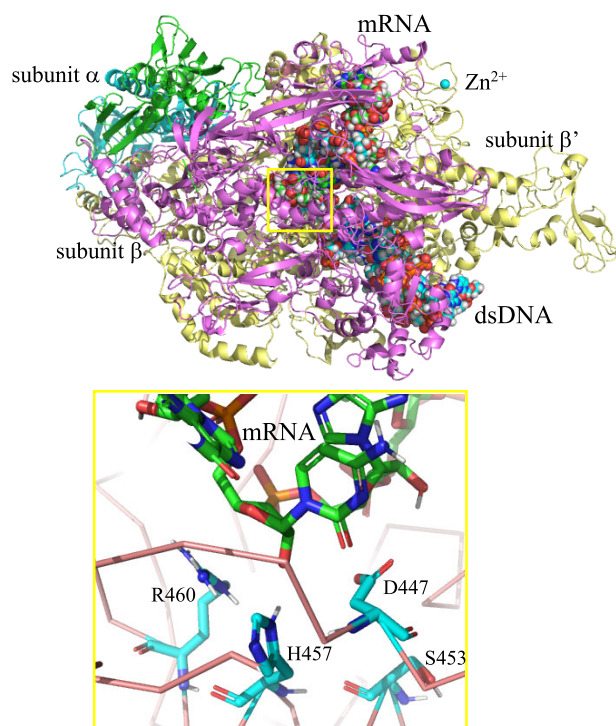


Figure 2. Modeling an *S. incarnatus* RNAP/dsDNA/mRNA complex constructed using the X-ray crystallographic coordinates of *Thermus thermophilus* HB8 RNAP (PDB, 2O5I) as the template. Double-stranded DNA and single-stranded mRNA are represented using space-filling models. Subunits are shown in cartoon form using different colors: green/cyan for two α subunits, wheat in front for the β subunit (*rpoB* product), and yellow for β' subunit. The γ subunit is located behind the β' subunit.

with atomic coordinates for two Zn ligands, each of which is coordinated by four Cys residues. The resulting geometry was further optimized using the MOE platform. During this process, the interior networks of van der Waals interactions, charge-charge interactions, hydrogen bonding, and tautomer orientation were optimized using a Protonate3D matrix (Labute 2009) in the implicit Born solvent (Labute 2008).

The initial model structure optimized with the MOE platform using the Amber99 force field was further improved by performing MD simulations using the AMBER16 program in the isotherm-isobar system. The ternary complex system was relaxed by MD simulation with a generalized Born solvation model for a few nanoseconds. Annealing simulations for 1 ns interval were conducted by creating 500 000 consecutive frames at 2 fs increments, with 1 of every 50 frames being picked up and aligned to make a trajectory file for a 1-ns time interval. We were typically able to gain an initial equilibrated conformation in the first 3–4 ns. Next, the model was allowed to reach a satisfactory equilibrated conformation by further optimization up to 11 ns. The average atomic coordinates were computed from the trajectory of 10 000 frames created by the MD calculations from 10 to 11 ns.

The resulting rif cluster region was visualized using Virtual Molecular Dynamics. The residues D447, S453, H457, and R460 were in close vicinity to the growing mRNA strand, interacting with the third and fourth ribonucleotide units from the 3'-end (Figure 2). The guanidino group of R460 was directed toward the syn-diol group of the ribose moiety of mRNA molecules. H457 also appears to have direct contact with the RNA chain but is also in close contact with R460. This positions H457 to assist in

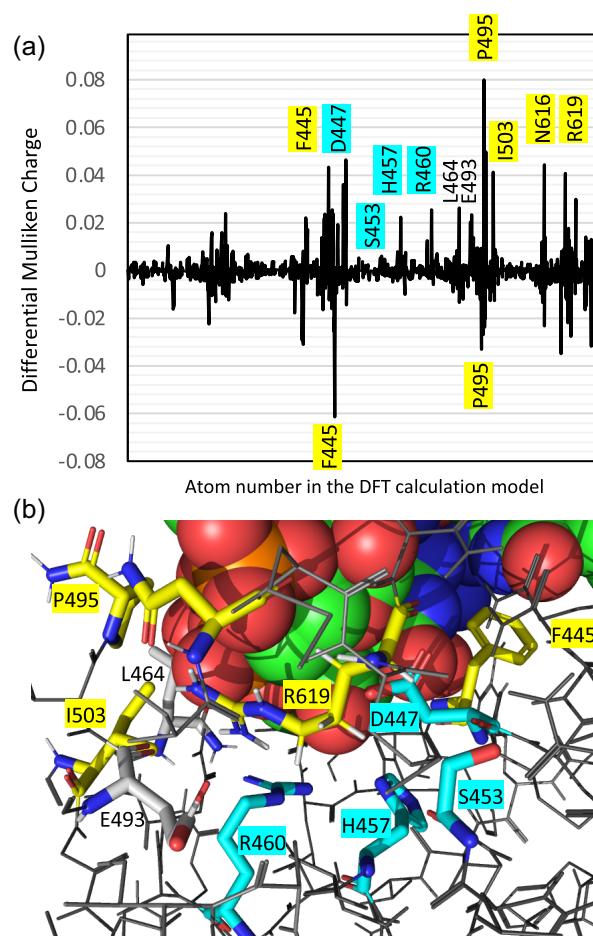


Figure 3. DFT calculations reveal that rif cluster residues interact with mRNA strands in RNAP. (a) Mulliken electronic charges with or without an mRNA moiety. (b) Rif cluster residues D447, S453, H457 and R460 (designated in cyan) are in close vicinity to the growing RNA. F445, R619, R495, and I503 (designated in yellow) are predicted to have potent interactions with the mRNA moiety.

orienting R460 such that the guanidino group is in close contact with the syn-diol group of the ribose moiety. D447 does not have direct contact with the mRNA, but is in close vicinity, while S453 interacts with D447 but has no direct contact with RNA. Those interactions are at equilibrium in molecular dynamics motions. The ribose moiety that the rif cluster residues attach to is the third ribonucleotide from the 3'-end, at which elongation is taking place. Therefore, mutations altering rif cluster residues may loosen or tighten the harness of the growing ribonucleotide on the deoxyribonucleotide template. Estimations of the levels of intermolecular interactions between RNA molecules and rif cluster residues or other residues nearby require quantum chemistry approaches limited to the vicinity of the rif cluster region.

Estimation of RNA–rif cluster residue interactions by DFT calculations

Atomic coordinates of residues within 20 Å of H457 were obtained from the protein viewer Virtual Molecular Dynamics and a QM model constructed to estimate the intermolecular charge interactions between the RNA strand and nearby amino acid residues, including rif cluster residues (Figure 3). As a conventional theoretical approach, Mulliken charges were calculated

for each atom in the QM model with or without the RNA moiety, and the difference in the 2 models calculated. Despite the subtle nature of these differences, it was obvious that electronic charges ($\delta > 0.02$) are induced in residues D447, H457, and R460 by the negatively charged mRNA strand, while the electronic state of S453 did not change with ribonucleotide presence, suggesting a less direct interaction between S453 and the RNA moiety in the RNA/DNA/RNA complex. Instead, S453 appears to have a much more direct interaction with D447. Since rif cluster residues are in close vicinity to each other, multiple mutations may have profound effects on interactions with the growing RNA strand. DFT calculations suggest that the additional nearby residues F445, R619, R495, and I503 have even more potent interactions with ribonucleotide moiety than the residues listed above (Table S4). MD models also suggests that these latter residues are also located nearby the growing RNA strand. Even though those residues have not been identified as candidate residues based on the rifampicin-resistance phenotype, they can be targets for mutations that may help drive gene expression in combination with rif cluster residues. Our genome editing technique altering rif cluster residues arbitrarily now allows any point mutation at the target residues D447, S453, H457, and R460, or at new target candidates F445, R619, R495, and I503.

Supplementary material

Supplementary material is available at [Bioscience, Biotechnology, and Biochemistry](#) online.

Acknowledgments

MD and DFT calculations were assisted by Center for Information Technology and Management, Okayama University, and HPC solutions (Tokyo, Japan). The authors are grateful to Ryobi Teien Memory Foundation for the support.

Author contribution

S.O., H.S., K.F., and M.N. contributed to genetic engineering and microbial experiments. N.T., T.H., F.S., K.Y., and T.H. performed modeling and calculation chemistry. M.Y., W.S., K.H., and H.O. contributed sparse elastic-net modeling using AAIndex. K.O., T.K., and K.I. designed multiple mutations. T.T. designed and conducted the research.

Funding

This work was supported by A-STEP under grants AS221Z03753E, AS231Z01635G, and AS242Z00693Q. The present study was also supported by PRESTO, JST.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Andrei G, Snoeck R, Reymen D et al. Comparative activity of selected antiviral compounds against clinical isolates of varicella-zoster virus. *Eur J Clin Microbiol Infect Dis* 1995;**14**:318-29.
- Bigelow CC. On the average hydrophobicity of proteins and the relation between it and protein structure. *J Theor Biol* 1967;**16**:187-211.
- Bordoli L, Kiefer F, Arnold K et al. Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc* 2009;**4**:1-13.
- Bouvet M, Debarnot C, Imbert I et al. In vitro reconstitution of SARS-coronavirus mRNA cap methylation. *PLoS Pathog* 2010;**6**:e1000863.
- Campbell EA, Korzheva N, Mustaev A et al. Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* 2001;**104**:901-12.
- Friedman JH, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;**1**:2010.
- Frisch MJ, Trucks GW, Schlegel HB et al. Gaussian09. 2009, Gaussian, Inc., Wallingford CT.
- Fukuda K, Tamura T, Ito H et al. Production improvement of antifungal, antitrypanosomal nucleoside sinefungin by rpoB mutation and optimization of resting cell system of *Streptomyces incarnatus* NRRL 8089. *J Biosci Bioeng* 2010;**109**:459-65.
- Hijikata A, Yura K, Noguti T et al. Revisiting gap locations in amino acid sequence alignments and a proposal for a method to improve them by introducing solvent accessibility. *Proteins* 2011;**79**:1868-77.
- Hu HF, Ochi K. Novel approach for improving the productivity of antibiotic-producing strains by inducing combined resistant mutations. *Appl Environ Microbiol* 2001;**67**:1885-92.
- Hu HF, Zhang Q, Ochi K. Activation of antibiotic biosynthesis by specified mutations in the rpoB gene (encoding the RNA polymerase beta subunit) of *Streptomyces lividans*. *J Bacteriol* 2002;**184**:3984-91.
- Inaoka T, Takahashi K, Yada H et al. RNA polymerase mutation activates the production of a dormant antibiotic 3,3'-neotrehalosadiamine via an autoinduction mechanism in *Bacillus subtilis*. *J Biol Chem*. 2004;**279**:3885-92.
- Kawashima S, Ogata H, Kanehisa M. AAindex: amino acid index database. *Nucleic Acids Res* 1999;**27**:368-9.
- Kennedy PGE, Gershon AA. Clinical features of varicella-zoster virus infect. *Viruses* 2018;**10**:609.
- Kuroda Y, Yamagata H, Nemoto M et al. Antiviral effect of sinefungin on in vitro growth of feline herpesvirus type 1. *J Antibi* 2019;**72**:981-5.
- Labute P. The generalized Born/volume integral implicit solvent model: estimation of the free energy of hydration using London dispersion instead of atomic surface area. *J Comput Chem* 2008;**29**:1693-8.
- Labute P. Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins* 2009;**75**:187-205.
- Lai CX, Xu J, Tozawa Y et al. Genetic and physiological characterization of rpoB mutations that activate antibiotic production in *Streptomyces lividans*. *Microbiology* 2002;**148**:3365-73.
- Li JR, Chorbha JS, Whelan SPJ. Vesicular stomatitis viruses resistant to the methylase inhibitor sinefungin upregulate RNA synthesis and reveal mutations that affect mRNA cap methylation. *J Virol* 2007;**81**:4104-15.
- Long WK, Fronko GE, Lindemeyer RG et al. Effects of S-adenosylhomocysteine and analogs on Epstein-Barr virus-induced transformation, expression of the Epstein-Barr virus capsid antigen, and methylation of Epstein-Barr virus DNA. *J Virol* 1987;**61**:221-4.
- Oshima K, Hattori M, Shimizu H et al. Draft genome sequence of *Streptomyces incarnatus* NRRL8089, which produces the nucleoside antibiotic sinefungin. *Genome Announc* 2015;**3**:e00715-15.
- Peters MB, Yang Y, Wang B et al. Structural survey of zinc-containing proteins and development of the zinc AMBER force field (ZAFF). *J Chem Theory Comput* 2010;**6**:2935-47.

- Pugh CS, Borchardt RT, Stone HO. Sinefungin, a potent inhibitor of virion mRNA(guanine-7-)-methyltransferase, mRNA(nucleoside-2'-)-methyltransferase, and viral multiplication. *J Biol Chem* 1978;**253**:4075-7.
- Schäfer A, Tauch A, Jäger W et al. Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. *Gene* 1994;**145**:69-73.
- Selisko B, Peyrane FF, Canard B et al. Biochemical characterization of the (nucleoside-2'O)-methyltransferase activity of dengue virus protein NS5 using purified capped RNA oligonucleotides (7Me)GpppAC(n) and GpppAC(n). *J Gen Virol* 2010;**91**:112-21.
- Shiraishi T, Kuzuyama T. Recent advances in the biosynthesis of nucleoside antibiotics. *J Antibiot* 2019;**72**: 913-23.
- Tamehiro N, Hosaka T, Xu J et al. Innovative approach for improvement of an antibiotic-overproducing industrial strain of *Streptomyces albus*. *Appl Environ Microbiol* 2003;**69**:6412-7.
- Vassilyev DG, Vassilyeva MN, Zhang J et al. Structural basis for substrate loading in bacterial RNA polymerase. *Nature* 2007;**448**:163-8.
- Wehrli W, Knusel F, Schmid K et al. Interaction of Rifamycin with bacterial RNA polymerase. *Proc Natl Acad Sci USA* 1968;**61**: 667-73.
- Yadav MK, Park SW, SW Chae et al. Sinefungin, a natural nucleoside analogue of S-adenosylmethionine, inhibits *Streptococcus pneumoniae* biofilm growth. *Biomed Res Int* 2014;**2014**:156987.
- Zheng S, Shuman S. Structure-function analysis of vaccinia virus mRNA cap (guanine-N7) methyltransferase. *RNA* 2008;**14**:696-705.
- Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Statist Soc B* 2005;**67**:301-20.