Title

Evolution-Guided Engineering of Non-Heme Iron Enzymes Involved in Nogalamycin Biosynthesis

Authors

Benjamin Nji Wandi,[1] Vilja Siitonen,[1] Pedro Dinis,[1] Vladimir Vukic,[2,3] Tiina A. Salminen[2] and Mikko Metsä-Ketelä[1,*]

Affiliations

[1]Department of Biochemistry, University of Turku, FI-20014 Turku, Finland

[2]Structural Bioinformatics Laboratory, Biochemistry, Faculty of Science and Engineering, Åbo Akademi University, FI-20520, Turku, Finland

[3]Faculty of Technology Novi Sad, University of Novi Sad, Novi Sad, Serbia

Corresponding Author

Mikko Metsä-Ketelä, Contact: E-mail: mianme@utu.fi, Tel. +35823336847, Fax. +35823336860

Running Title

Engineering Anthracycline Tailoring Enzymes

Keywords

protein engineering; chimeragenesis; *Streptomyces*; polyketide; computational modelling

Enzymes

EC number 1.14.11

ABSTRACT (249 / 250 words)

Microbes are competent chemists that are able to generate thousands of chemically complex natural products with potent biological activities. Key to the formation of this chemical diversity has been the rapid evolution of secondary metabolism. Many enzymes residing on these metabolic pathways have acquired atypical catalytic properties in comparison to their counterparts found in primary metabolism. The biosynthetic pathway of the anthracycline nogalamycin contains two such proteins, SnoK and SnoN, belonging to non-heme iron and 2-oxoglutarate-dependent mono-oxygenases. In spite of structural similarity, the two proteins catalyse distinct chemical reactions; SnoK is a C2–C5″ carbocyclase, whereas SnoN catalyses stereoinversion at the adjacent C4″ position. Here we have identified four structural regions involved in the functional differentiation and generated 30 chimeric enzymes to probe catalysis. Our analyses indicate that the carbocyclase SnoK is the ancestral form of the enzyme from which SnoN has evolved to catalyse stereoinversion at the neighboring carbon. The critical step in the appearance of epimerization activity has likely been the insertion of three residues near the C-terminus, which allow repositioning of the substrate in front of the iron center. The loss of the original carbocyclization activity has then occurred with changes in four amino acids near the iron center that prohibit alignment of the substrate for formation of the C2–C5″ bond. Our study provides detailed insights into the evolutionary processes that have enabled *Streptomyces* soil bacteria to become the major source of antibiotics and antiproliferative agents.

INTRODUCTION

Natural products have been an important source of new drug leads for the pharmaceutical industry. In the past 34 years, approximately two thirds of antibiotics [1] and one third of anticancer [2] agents approved for clinical use have been either natural products or their semi-synthetic derivatives. Microbial natural products [3], such as those made by soil-dwelling Gram-positive *Streptomyces* bacteria, have been a particularly rich reservoir of new chemical entities and thousands of biologically active secondary metabolites have been isolated over the years. One medically important class of natural products are the anthracyclines [4], which include compounds such as doxorubicin (**1**, Figure 1A) and aclacinomycin that are among the most effective anticancer drugs available [5]. Currently these drugs are included in approximately 500 clinical trials to explore better combinations [6]. The major contributor to the biological effects of anthracyclines is formed via intercalation to DNA and poisoning of topoisomerases [7]. but the biological response is complex and influenced by many other factors [6,8,9,10]. Nogalamycin (**2**, Figure 1A) belongs to a unique subgroup of anthracyclines, where the D-ring of the aglycone is fused to a carbohydrate unit both via a canonical O-glycosidic bond and an additional carbon–carbon bond [11]. In total, 408 bacterial-derived anthracyclines have been described to date [12], but only a limited number of metabolites of the nogalamycin-subclass e.g. arugomycin [13] (**3**, Figure 1A) have been characterized. Like all anthracyclines, nogalamycin exhibits high activity against several cancer cell lines, but was found to harbor high acute toxicity [14].

The chemical complexity of anthracyclines is reflected in the compositions of the metabolic pathways and the gene cluster responsible for nogalamycin biosynthesis codes for a total of 32 enzymes [15]. In all anthracyclines, the polyphenolic aglycone units are synthesized by canonical aromatic type II polyketide pathways, while the starting material for the

carbohydrate units is typically glucose-1-phosphate [4]. Much of the chemical diversity is generated in the tailoring steps, where the common 7,8,9,10-tetrahydro-5,12-naphtacenoquinone scaffold is modified further by an assortment of enzymes. The nogalamycin pathway contains two non-heme iron (II) and 2-oxoglutarate dependent proteins SnoK and SnoN, which function during these late-stage steps [16]. The archetypical reaction of the protein superfamily, such as the much-studied taurine dioxygenase TauD [17], is to catalyse hydroxylations at inactivated carbon centers, but many members are able to perform diverse energetically demanding chemical transformations [18]. In nogalamycin biosynthesis, SnoK is responsible for formation of the subgroup epitomizing C2–C5″ bond, while SnoN is an epimerase functioning at the adjacent C4″ carbon.

Both SnoK and SnoN have been shown to require Fe(II) for activity, and to utilize 2-oxoglutarate and molecular oxygen as co-substrates [16]. However, whereas the carbocyclase activity of SnoK has been demonstrated *in vitro* and the enzyme is able to convert **4** into **5** (Figure 1B), the C4″ epimerization activity of SnoN to form **6** has only been established *in vivo*. The enzymatic reaction of SnoN with **4** appears to lead *in vitro* to a degradation cascade, where **7** is detected as the transformation product (Fig 1B). It is intriguing to note that the oxidation state of the anthracycline does not change during the epimerization reaction. However, consumption of 2-oxoglutarate and structural analysis of the substrate complex of SnoN [16] indicates that Fe(II) is converted to the high-valent Fe(IV)=O species during the catalytic cycle. This apparent redox imbalance dictates that the overall reaction may require an unknown reductant component that is present in the cellular surroundings, but missing from the enzymatic reactions *in vitro*. This would account for the discrepancies in SnoN activities and for the putative radical chain reaction leading to the formation of **7** [16].

The abundance of microbial secondary metabolites has raised intriguing questions regarding how simple soil bacteria are able to produce such chemical complexity and diversity [19,20,21]. An important factor appears to be an evolutionary pressure promoting chemical diversity, since the appearance of a novel bioactive compound may provide significant competitive advantage to the producing organism [22]. An interesting consequence of this divergent evolution of secondary metabolism is that the functions of even highly similar proteins may greatly differ [19]. Selected examples from polyketide tailoring steps include conversion of a methyl transferase into a mono-oxygenase via insertion of a single amino acid [23], the recruitment of a polyketide cyclase as a partner in a two component mono-oxygenase system [24] and use of flavin dependent aromatic hydroxylases to catalyse dehydration reactions [25]. Here we have probed the functional differentiation of SnoK and SnoN by extensive chimeragenesis. We engineered four regions in different combinations in the two proteins, which enabled us to reconstruct the evolutionary events that have presumably led to the appearance of epimerization activity from the ancestral carbocyclization function.

RESULTS AND DISCUSSION

Genome Mining and Phylogenetic Analysis of Non-Heme Iron Enzymes Involved in Anthracycline Biosynthesis.

Analysis of public sequence databases revealed total of 11 gene clusters with an unknown function, which encoded proteins essential for formation of the anthracycline carbon frame, but also harbored genes homologous to *snoN* and *snoK*. The majority of these were found from *Streptomyces* sp., but one cluster resided in *Actinomadura* and one in *Salinispora*. Phylogenetic analysis resolved two clades, one for sequences similar to SnoK and another one similar to SnoN, when TauD from *Escherichia coli* was used as outgroup (Figure 2).

Surprisingly, all gene clusters encoded enzymes clustering together with SnoK, whereas SnoN -type enzymes could only be detected in gene clusters encoding two copies of 2-oxoglutarate and non-heme iron dependent enzymes. This distribution of genes suggests that the carbocyclization activity of SnoK is the ancestral activity from which the epimerization reaction catalysed by SnoN has evolved. The finding is further supported by the chemical characterization of arugomycin -type metabolites [13] (Figure 1A) that contain the C2–C5″ bond, but have opposite stereochemistry at C4″ in comparison to nogalamycin.

Structural Analysis of SnoK and SnoN, and Selection of Chimeragenesis Regions

The structures of SnoK and SnoN have been determined previously [16] and shown to consist of eleven β-strands forming the characteristic jelly-roll topology of the β-sandwich fold of the protein family (Figure 3A–B). The two proteins share significant structural similarity (overall r.m.s.d. 0.83 Å) despite only moderately sequence identity (37%). The mononuclear iron, which is located in the interior of the β-barrel, is accessible from the surface of the proteins through a cleft mostly lined with hydrophobic amino acids. The binding site of the co-substrate 2-oxoglutarate residing at the bottom of the substrate binding pocket is exceedingly similar in the two enzymes. In addition, one of the active site walls has two conserved residues, a tyrosine (Y68/Y74 in SnoK/SnoN) shaping the active site and a tryptophan (W57/W64 in SnoK/SnoN) that interacts with the anthracycline ring system via pi-pi stacking (Figure 3C–D). The corresponding tryptophan residue is fully conserved in all 16 proteins identified by genome mining (Figure S1). The major differences between SnoK and SnoN are found on the other wall of the active site cleft, where the loop regions between β7 and β8 and the C-termini form complementary substructures. The longer β7–β8 loop region of SnoK (Figure 3A; R2 (blue)) is compensated by the longer C-terminus in SnoN (Figure 3B; R1

(cyan)). These differences contribute to the surface accessible (SA) area, where the more open SnoK active site cavity (453 $Å^3$) is significantly larger in comparison to SnoN (277 $Å^3$)

In order to understand the functional difference between SnoK and SnoN, we superimposed the crystal structures and identified four distinct regions that might be important for the catalytic differences (Figure 3A–B; R1–R4). We selected the dissimilar C-terminal part as the first region, R1 (residues 238–248 in SnoK and residues 241–261 in SnoN; r.m.s.d. 2.69 Å; Figure 3A–B; cyan), for the chimeragenesis studies. In both proteins, the R1 segments protrude from above towards the substrates and make contacts with the B- and C-rings of the anthracycline aglycones. The shorter R1 region of SnoK hosts Q240 within hydrogen bonding distance from the carbomethoxy group at C10 (Figure 3C), while the longer R1 of SnoN contains F245 in a nearby position interacting with the ligand (Figure 3D). Neither of these residues is fully conserved in the two protein families, although all four SnoN-type enzymes are seven residues longer in the R1 region compared to the SnoK-type enzymes (Figure S1).

The situation is reversed in the β7-β8 loop, selected as region R2 (residues 164–183 in SnoK and residues 171–186 in SnoN; r.m.s.d. 3.06 Å; Figure 3A–B; blue), where six additional residues exists in the SnoK-type proteins (Figure S1). This region closes in on the substrates from below and interacts with the A- and B-rings of the anthracycline ligands. In SnoN, the substrate is sandwiched between W180 from R2 and the conserved W64 (Figure 3D). Previous studies have shown that the R2 region is mobile and binding of the anthracycline substrate triggers a conformational change [16]. In SnoK, the tryptophan residue corresponding to W180 of SnoN is replaced by F173, which is more distal to the ligand (Figure 3C) based on the modelling of the substrate to the crystal structure of the SnoK-

Fe(II)-2-oxoglutarate ternary complex. The residues equivalent for W180 and F173 appear to be fully conserved in both SnoN- and SnoK-type subfamilies, respectively (Figure S1).

The third R3 region (residues 103–106 in SnoK and residues 109–112 in SnoN; r.m.s.d. 0.14 Å; Figure 3A–B; magenta) is situated in the β4-sheet directly above the non-heme iron. It consists only of four residues, which appear to be critical for correct alignment of the L-rhodosamine unit of the substrate in front of the iron center. In SnoN, K110 forms a hydrogen bond to the C4″ hydroxyl group at the site of stereoinversion and E112 is within hydrogen bonding distance to the dimethylamine group of L-rhodosamine (Figure 3D). Both of these residues are fully conserved in the four SnoN-type enzymes (Figure S1). In contrast, the corresponding residues in SnoK, S104 and D106, respectively, are shorter in length and enable the substrate to invade deeper into the active site (Figure 3C). The R3 regions significantly contribute to the differences in the shape of the active site cavities of the two enzymes (Figure 3E–F). Interestingly, four of the 12 unknown SnoK-type proteins contain a glutamic acid akin to SnoN instead of an aspartic acid (D106 in SnoK), but the residue corresponding to S104 in SnoK is conserved in ten members of the family (see Figure S1). This suggests an important role for the serine in SnoK whereas preserving the acidic nature of the other residue seems to be adequate for the catalytic activity [16]. Previous studies provide experimental support for the suggestion as the D106A and D106N variants of SnoK abolished enzymatic activity [16]. The third residue of R3 is a fully conserved tyrosine or phenylalanine in the SnoK- and SnoN-like proteins, respectively (Y103/F109 in SnoK/SnoN, Figure S1). This residue points towards the active site, but also interacts with amino acids (residues 127–129 in SnoK and 136–138 in SnoN) within the R4 region (see below).

Finally, region R4 (residues 122–135 in SnoK and residues 131–144 in SnoN; r.m.s.d. 1.11 Å; Figure 3A–B; salmon) located in the interior of the β-barrel fold is responsible for shaping a minor portion of the binding pocked close to the D-ring of the aglycone and the C6′′ of L-rhodosamine. The first residue of R4 is a conserved aspartate, D123/D132 in SnoK/SnoN, that forms a part of the HXD-motif of mononuclear non-heme iron enzymes [26]. This residue is responsible for coordination of the mononuclear iron [16] together with the fully conserved H121/H130 and H210/213 in SnoK/SnoN (Figure S1). However, the next four amino acids of R4 after the HXD motif differ significantly in the two protein subfamilies; SnoK harbors a fully conserved E124 (Figure S1) and a partially conserved T125 followed by fully conserved G126 and L127 residues, whereas SnoN contains completely conserved F134 and A135 residues flanked by non-conserved A133 (Figure S1) and a partially conserved F136 amino acid (Figure S1, Figure 3C–D).

Enzymatic Activity Assays of the Chimeric Proteins

In order to investigate the importance of the selected regions for catalysis in SnoK and SnoN, we cloned 30 chimeric genes in a modified pBAD vector [27]. The N-terminally histidine tagged chimeras were produced in *E. coli* and purified to near homogeneity by affinity chromatography (Figure S2). The chimeras were named based on the protein scaffold, followed by identification of the exchanged region (e.g. SnoK R1 consists of the SnoK protein, where the R1 region has been replaced with the sequence from SnoN). To probe the relative activities of the chimeras, we utilized **4** (Figure 1B) as a substrate, since it can be turned over by both native enzymes, and analyzed reaction products by HPLC-UV/Vis.

The results from the activity measurements confirmed that all of the regions selected for chimeragenesis are important for catalysis (Table S1). In case of the single chimeras, SnoK

R1 gained significant activity (32%) towards the SnoN reaction (Figure 4). Analysis of an additional earlier time point revealed that substrate consumption of SnoK R1 was 107% in comparison to the wild type enzyme (Figure S3). The exchanged regions in SnoK R2 and SnoK R4 influenced catalysis similarly, but to a lesser extent with 7% and 17% gain in activity towards the SnoN reaction, respectively. Both SnoK R2 and SnoK R4 had lost significant amounts of enzymatic activity, since 85% and 48%, respectively, of the substrate was left unconsumed in the relative activity measurements (Figure 4). Surprisingly, the SnoK R3 chimera had lost all enzymatic activity. From the multiple chimeras, SnoK R12 displayed minor SnoK and SnoN activities of 10% and 12%, respectively. SnoK R14 still harbored 4% of native SnoK activity, whereas all other chimeras had lost all enzymatic activity.

The engineering efforts to modify the SnoN scaffold proved to be detrimental for the enzymatic activity and none of the chimeras gained carbocyclization activity. Of the single chimeras, SnoN R1 and SnoN R4 had lost significant enzymatic activity with 12% and 14% relative activity remaining, respectively, whereas SnoN R3 was only moderately active with 30% relative activity. Exchanging region R2 seemed to have minor effects on catalysis (Figure 4) with additional assays indicating that substrate consumption was 81% in SnoN R2 in comparison to the wild type (Figure S3). From the multiple chimeras, only SnoN R13 had retained 29% activity, while the relative activities of SnoN R12, R23 and R24 were low, but still above the detection limit at 6%, 7% and 7%, respectively. Circular Dichroism (CD) spectroscopy confirmed that the fully inactive SnoN chimeras were folded (Figure S4) and that the inactivation was likely due to the engineering efforts.

The Importance of Region R1 in the Evolution of Epimerization Activity

In order to trace the evolutionary steps that have taken place during the functional differentiation of SnoK and SnoN, we next performed molecular modelling and docking calculations to gain support for the activity assays. Region R1 was found to have a significant impact in catalysis and the SnoK R1 chimera gained substantial epimerization activity (Figure 4, Table S1). Molecular modelling of SnoK R1 revealed that the region provides several critical interactions with the ligand. The key differences in the active site of SnoK R1 in comparison to the wild type were the V239 ^{SnoK}/F245^{SnoN} and Q240^{SnoK}/L246^{SnoN} replacements (Figure 3C–D). Intriguingly, the R1 loop in SnoK R1 is stabilized similarly as in the wild-type SnoN and the position of F245^{SnoN} corresponds to that of W180 in SnoN (Figure 5A). As a consequence, **4** is stacked between F245^{SnoN} and W57 in the SnoK R1 complexes (Figure 5A). The fact that SnoK R1 gained 32% of SnoN activity (Figure 4, Table S1) indicates that formation of these stacking interactions may have been the key evolutionary event to orient **4** for the SnoN reaction (Figure 5A). Similar aromatic stacking interactions, which are further reinforced by the hydrophobic interactions of L246^{SnoN}, are not found in native SnoK (Figure 3C). It is worth noting that disrupting the polar contact between Q240 and **4** by introducing L246^{SnoN} did not totally abolish the carbocyclization reaction, since SnoK R1 retained 68% of native SnoK activity (Figure 4, Table S1).

Further investigation of the R1 region revealed that the loop containing F245^{SnoN} and L246^{SnoN} is stabilized by a structurally conserved hydrogen bond between the main chain and a conserved aspartate (D248^{SnoN} in Figure 5A; D241^{SnoK} in Figure S1). In order to validate the importance of this key area, we narrowed down the original R1 region and designed a new chimera SnoK R1.1 with an insertion of three residues (residues 238–240 of SnoK exchanged with residues 241–246 from SnoN). The SnoK R1.1 chimera gained 42% relative activity towards the SnoN reaction (Figure 4) and the predicted binding mode of **4** (Figure 5B) in the

active site was similar to that of SnoK R1 (Figure 5A). These experiments indicate that extension of the tip of the R1 loop and formation of stacking interactions may have been the original evolutionary events that have led to the appearance of epimerization activity.

The Role of Region R2 in Substrate Binding

Crystallographic evidence [16] and our results here indicate that region R2 in SnoN functions as a mobile loop, which facilitates substrate binding. Comparison of the active site cavity volumes of wild type SnoK and SnoK R2 revealed a significantly enlarged cavity in the chimera (1310 $Å^3$) in comparison to the wild type (453 $Å^3$) (Figure 5C). Even though the increased space could accommodate the substrate well, weakened stabilizing interactions could explain the lower relative activity (Figure 4, Table S1). Due to the different length of the exchanged region, the introduced aromatic $W180^{SnoN}$ of R2 is not in the same position as F173 in the wild-type SnoK and, therefore, appears to be too far away from the active site to compensate for the loss of F173 (Figure 5C). The R1 and R2 regions form complementary substructures in the proteins and the deletion of six residues in SnoN-type proteins has presumably occurred in response to the lengthening of R1 required for the epimerization activity. The significant loss of enzymatic activity in the SnoK R2 chimera reduced the likelihood that shortening of R2 would have been the initial evolutionary event in the functional diversification.

Further evidence is provided by the observation that the shortening of R2 is not essential for the epimerization function, since SnoN R2 retained full relative activity (Figure 4, Table S1). Based on the docked SnoN R2 – **5** complex (Figure 5D), the intact SnoN activity of SnoN R2 (Figure 4, Table S1) likely results from a successful reorganization of the active site. Firstly, in the absence of bulky W180 from R2, F245 of R1 and W64 stack with the ligand (Figure

5D). Consequently, C4″ of L-rhodosamine has a similar distance to the iron center as in the X-ray structure of SnoN – **5** complex. Secondly, $E175^{SnoK}$ occupies the same position as E178 in the wild-type SnoN and together with $R177^{SnoK}$ forms ionic interactions with the ligand.

The Influence of Region R3 in the Loss of Carbocyclization Activity

Region R3 appears to be responsible for fine-tuning the positioning of the carbohydrate unit of the substrate and provides a fascinating example of directional selection. The total loss of enzymatic activity in the SnoK R3 chimera (Figure 4, Table S1) may result from the reduced size of its active site cavity (384 $Å^3$) in comparison to the wild type SnoK (453 $Å^3$) . Particularly, the $S104^{SnoK}$/$K110^{SnoN}$ and $D106^{SnoK}$/$E112^{SnoN}$ replacements incorporate longer side chains in the R3 region (Figure 5E) and, thus, the substrate presumably cannot enter deep enough for the SnoK reaction in any R3 chimeras of SnoK. Furthermore, the lack of a hydroxyl group due to the $Y103^{SnoK}$/$F103^{SnoN}$ replacement may contribute to the loss of SnoK activity.

The residues in region R3 are highly efficient in quenching the ancestral carbocyclization activity, but they are not essential for the epimerization function, as shown by the activity of the SnoN R3 chimera (Figure 4, Table S1). However, contradictory to expectations the SnoN R3 chimera did not gain any SnoK activity. The docking results indicate that **4** forms a hydrogen bond with the hydroxyl group of $Y103^{SnoK}$ and binds too deep into the active site of SnoN R3 (Figure 5F). As a result, the distances of C2 of the anthracycline aglycone and C5″ of L-rhodosamine to the non-heme iron center become too long to (7.9 Å) allow the SnoK reaction to occur.

The Regulatory Role of Region R4 During the Catalytic Cycle

Even though region R4 only shapes a minor portion of the active site (Figure 6A), the engineering of this region had a significant influence in catalysis (Figure 4, Table S1). In the SnoK R4 chimera, the key interactions with Q240 and F173 are preserved like in the wild-type enzyme (Figure 3C). Despite the lack of the tight aromatic sandwich that is characteristic for substrate binding in SnoN, SnoK R4 gained 17% of *in vitro* SnoN activity by incorporating larger residues to the bottom of the active site ($T125^{SnoK}/F134^{SnoN}$, $G126^{SnoK}/A135^{SnoN}$ and $L127^{SnoK}/F136^{SnoN}$; Figure 6A-B). These replacements make the active site narrower and smaller (353 $Å^3$) particularly by $F136^{SnoN}$ near the iron center, which might contribute to the SnoN reaction by reducing the space available for the movement of the sugar ring of **4** (Figure 6A-B). Molecular dynamic simulations (Figure S5 and Figure S6) suggest that impaired ligand binding might be responsible for the slower reaction rates (Table S1), since compound **4** fluctuated in the active site cavity of SnoK R4 (3.84–6.29 Å distances between C5″ of **4** and oxygen of iron centre).

The situation differs in SnoN where the structural analysis revealed extensive interactions of R4 with R1 and R2 regions (Figure 6C). In particular, F134 of the R4 region in the wild-type SnoN seems to have an important functional role, since its movement upon substrate binding turns the disordered R2 loop into an ordered conformation, which allows W180 to stack with the ligand (Figure 6C). Additionally, during the disorder/order transition of the R2 loop, the D172-R210 salt bridge positions R2 near R4, which is linked to R1 by R239. The analysis indicates that in SnoN the R4 region mediates information on the state of the R2 loop to R1 shown to be critical for catalysis. On the contrary, the R2 loop of SnoK is constantly ordered and F173 positioned for ligand binding (Figure 6D). In light of these differences, even subtle disturbances in these interatomic interactions within SnoN are likely to have a damaging effect on its catalytic activity explaining why the majority of SnoN chimeras were inactive.

Concluding Remarks

The abundance of microbial natural products is dependent on the efficient formation of novel metabolic pathways that lead to new chemistry. Here we have probed how this has occurred after a gene duplication event in the evolution of nogalamycin-type anthracyclines from arugomycin-type molecules. We demonstrate that amino acid insertions in the C-terminal loop region (R1) have altered the alignment of the substrate in front of catalytic iron-oxo species, which has enabled novel epimerization activity to emerge. An adjacent mobile loop region (R2), which is likely to be involved in substrate entry, has subsequently shortened to accommodate these changes. The loss of carbocyclization has most likely occurred through two point mutations in the vicinity of the iron center (R3) that inhibit the original activity. Finally, these changes have been reinforced by the emergence of a regulatory region (R4) to connect ligand entry (R2) and catalysis (R1) in the newly evolved protein. More precise control over enzyme catalysis may be important for the SnoN reaction, which generates an anthracycline radical that needs to be reduced outside of the active site cavity. Detailed understanding of the natural processes how secondary metabolism biosynthetic pathways have evolved paves the way for artificial design of catalysts that can be utilized to generate novel natural products and molecular diversity by protein engineering.

Figure Legends

Figure 1. Chemical structures of selected anthracyclines and the enzymatic reactions catalysed by SnoK and SnoN. (A) Anthracyclines relevant to this study. (B) The carbocyclization reaction by SnoK converting **4** to **5** and the stereoinversion reaction by SnoN converting **4** to **7** *in vitro* and **5** to **6** *in vivo* in the nogalamycin biosynthetic pathway.

Figure 2. Phylogenetic analysis of non-heme iron and 2-oxoglutarate dependent oxygenases residing in putative anthracycline gene clusters. Eleven anthracycline-type gene clusters harboring genes homologous to non-heme iron and 2-oxoglutarate dependent oxygenases were uncovered by genome mining. The phylogenetic tree was constructed from SnoK, SnoN the 14 unknown non-heme iron enzymes, with TauD from *E. coli* used as an outgroup.

Figure 3. Comparison of SnoK and SnoN structures and their ligand binding properties. Cartoon representation of (A) SnoK crystal structure with docked compound **4** (orange) and (B) SnoN crystal structure with compound **5** (yellow). The regions (R1 – R4) selected for chimeragenesis are colored as follows: R1 (cyan), R2 (blue), R3 (magenta) and R4 (salmon). Active site of (C) SnoK with docked compound **4** and (D) SnoN with compound **5**. The co-substrate 2-oxoglutarate is shown in gray, and the iron and water molecules as orange and red spheres, respectively. The conserved tyrosine (Y68 in SnoK/Y74 in SnoN) delineating the binding site near the acidic residue (D106 in SnoK/E112 in SnoN) is not shown for clarity. The residues interacting with the ligands are shown as sticks and the key hydrophobic (< 5 Å distance) and polar interactions as dashed lines. Volume and shape of the active site of (E) SnoK and (F) SnoN.

Figure 4. The relative enzymatic activities of selected chimeras. The *in vitro* reaction products (compound **5** and **7**) of the native SnoK and SnoN are shown in green and purple, respectively. Chimeras not shown in the figure had lost all enzymatic activity (Table S1).

Figure 5. Structural analysis of 3D models for SnoK chimeras R1, R2 and R3 with docked ligands. Comparison of (A) the R1 loop (cyan) of SnoK R1 (green) and wild type SnoN (purple) and (B) close-up view of the active site of SnoK R1.1 with compound **4** (orange). Comparison of (C) the active site cavities of SnoK R2 chimera (blue/purple) and wild type SnoK (green) and (D) close-up view of the active site of SnoN R2 with compound **5** (yellow). Comparison of (E) the active site cavities of SnoK R3 chimera (magenta) and wild type SnoK (green) and (F) view of the active site of SnoK R3 with compound **4**. Residues from the chimeric regions (R1 – R4) are colored as follows: R1 (cyan), R2 (blue) and R3 (magenta). The co-substrate 2-oxoglutarate is shown in gray, while the iron and water molecules are drawn as orange and red spheres, respectively. The interactions between the ligands and the protein are shown as dashed lines.

Figure 6. Structural analysis of 3D models for R4 chimeras and conformational changes induced by the R4 region. Comparison of (A) the active site cavities of SnoK R4 chimera (salmon) and wild type SnoK (green) and (B) interactions of the SnoK R4 chimera with compound **4** (orange). (C) Disorder/order transition of SnoN during substrate binding. The wild-type SnoN without substrate (PDB ID 5EP9) is shown in gray and with **5** (PDB ID 5EQU) using the R-region color-coding. Black circles indicate the disordered area (chain break) in the wild-type SnoN without substrate. The grey arrow depicts the concerted events mediated by R4 that link the regions important for ligand entry (R2) and catalysis (R1) in SnoN. Upon the disorder/order transition, F134 (R4) moves away and allows entry of W180 (R2) to active site. F134 is stabilized by R210 and R239. Of these, R210 forms ionic interactions with D172 (preceding R2), where as R239 (preceding R1) is stabilized by the main chain hydrogen bonds to the R4 phenylalanines. F138 (R4) also forms hydrophobic interactions with the R3 residue F109. (D) Lack of disorder/order in SnoK. The SnoN structure (PDB ID 5EQU) shown in gray with violet labels for comparison. The extensive ionic interactions (black dashes) stabilize the R2 loop in SnoK to allow the R166 (R2) -R207 stacking interactions. F173 of R2 locates in the ligand binding site and interacts with V239 from R1. Furthermore, the R2 loop in SnoK is positioned by a main chain hydrogen bond (black dashes) between M170 (R2) and T125 (R4). In SnoN, ionic interactions of D172 (R2; violet dashes) with R210 stabilize R2. Residues from the chimeric regions (R1 – R4) are colored accordingly: R1 (cyan), R2 (blue), R3 (magenta) and R4 (salmon). The iron and water molecules as orange and red spheres, respectively, and 2-oxoglutarate is shown in gray. The key interactions between the ligands and the protein are shown as dashed lines.

MATERIALS AND METHODS

Multiple Sequence Alignment, Phylogenetic Trees and Structural Superimposition

For phylogenetic analysis, multiple sequence alignment was constructed using MUSCLE [28] integrated within SeaView version 4 [29] a multiplatform graphical user interface for sequence alignment and phylogenetic tree construction. Phylogenetic trees were generated using the distance-based neighbor-joining [30] method in SeaView platform using 1000 rounds of bootstrap analysis and the sequence of TauD from *E. coli* to root the tree. The phylogenetic tree was draw with the program FigTree. For structure based sequence alignment, SnoN (Protein Data Bank (PDB) Identification Code (ID): 5EQU:A) and SnoK (PDB ID: 5EPA:A) structures were first superimposed with VERTAA [31] implemented in Bodil [32] and then *E. coli* TauD (PDB ID: 1GY9:A) was superimposed on them using the conserved ion binding residues (Figure S1. H121, D123 and H210 in SnoK) and the conserved 2-oxoglutarate binding arginine (Figure S1. R221 in SnoK). Finally, the sequences of related proteins were aligned to the prealigned structure-based sequence alignment using MALIGN implemented in Bodil [32]. Root-mean square deviations (r.m.s.d.) of the whole enzyme structures as well as the chimeric region pairs were calculated using the algorithm "align" in PyMol (The PyMOL Molecular Graphics System).

Bacterial Strains and Culture Conditions

Escherichia coli TOP10 (Invitrogen) was used as the cloning and production host. The *E. coli* strain was cultivated either in Luria–Bertoli or 2×yeast extract/tryptone medium (2xTY) at 23 - 37℃. For the production of the substrates and standards, *Streptomyces albus* [33] was used and it was cultivated either in NoS-soyE1 [22], tryptone soya broth (TSB) (Oxoid), and mannitol soya flour medium (MS). *S. albus* were cultivated for 7 days in 50 ml batches in 250 ml Erlenmeyer flasks in NoS-soyE1 at 30 °C. The antibiotics used for selection were ampicillin (50 μg/ml; Sigma-Aldrich) and apramycin (50 μg/ml; Sigma-Aldrich) for *E.coli* and *S. albus*, respectively.

Production and Purification of Metabolites.

The metabolites used in the study were produced from *S. albus* and purify with preparative HPLC. Compounds **4** was isolated from the strain *S. albus/pSnoΔK* whereas compounds **5** and **6** were isolated from the strain *S. albus/pSnogaori+N* [16]. The *S. albus* strains were cultivated for 7 days and the cells were discarded. XAD7-resin (20 $g \cdot L^{-1}$; Rohm and Haas) was used to extract the metabolites from the supernatant after which the resin was washed several times with water and the bounded metabolites were extracted from the resin with methanol. Majority of the impurities in the extracted crude extract were removed by running the crude extract through an open LH20 column (GE Healthcare) in methanol and collected in fractions. The fractions containing the desired metabolites were determined by SCL-10Avp HPLC with a SPD-M10Avp diode array detector (Shimadzu) using a Kinetex column (2.6 µm C18 100 Å, LC Column 100 x 4.6 mm, Ea, Phenomenex). The fractions were concentrated with rotavapor RII (BUCHI) and subjected to preparative HPLC (LC-20AP, model; CBM-20A, SHIMADZU) with a reverse phase column (SunFire Prep C18, 5 mm, 10 × 250 mm; Waters). For all metabolites, a mobile phase gradient from 10% acetonitrile to 70% acetonitrile including 18 mM ammonium acetate, pH 3.6, was used.

Cloning Strategies

The chimeric gene constructs were amplified with Phusion DNA polymerase (ThermoFister) using pBADN and pBADK [16] as templates. The four-primer overhang extension method [34] was used for generating the chimeric constructs. In general, for the construction of each chimeric enzyme, a first round of PCR with a complementary oligodeoxyribo-nucleotide primers (forward and reverse, used to introduce the chimeric fragment) paired respectively with the reverse and forward primers of the native constructs (see Table S2) was performed to generate two DNA fragments possessing overlapping ends. In the second round of PCR, these

two fragments serve as the template where the overlapping ends anneal and the primers for native constructs were used to further amplify and generate the desired chimeric constructs. The *snoK R1.1* chimera was ordered as a synthetic DNA fragment (ThermoFisher). All PCR products and plasmids DNA were extracted from agarose (BioNordika) gel using GeneJet gel extraction kit (ThermoFisher) and GeneJet plasmid miniprep kit (ThermoFisher) respectively. The amplified genes were cloned into pBAD/HisBΔ-plasmid as PstI/HindIII fragments, and verified by sequencing (Eurofins MWGOperon) before protein production.

Production and Purification of Proteins

The native proteins and their chimeras with additional AHHHHHHHRSAD sequences were produced in *E.coli* TOP10 cells as N-terminally His-tagged recombinant proteins. The cells were cultured at 30–37℃ until the OD600 reached 0.5–0.7 after which the cells were induced with l-arabinose (0.02 % [w/v]) and allowed to express the proteins for 15–18 h at 23℃. After production, the cells were pelleted and resuspended in buffer A (50 mM sodium phosphate, 50 mM NaCl, 5 mM imidazole, 10% glycerol [v/v], pH 7.5). The cells were lysed by sonication and the cell debris was removed by centrifugation and the crude lysate was mixed with TALON Superflow (GE Healthcare) to bind the recombinant proteins. The impurities were washed with buffer A and the target proteins were eluted from the column with B-buffer (A-buffer with 250 mM imidazole). The PD-10 column (GE Healthcare) was used for desalting and the proteins were stored in C-buffer (50 mM sodium phosphate, 50 mM NaCl, 10% glycerol, pH 7.5) with glycerol added to 40% (v/v). The purity for all purified proteins was evaluated by SDS-PAGE and their concentrations were estimated photometrically at 280 nm (NanoDrop2000, Thermo Scientific) and stored at - 20℃ before enzymatic reactions.

Enzyme Assays

The enzymatic reactions were performed in triplicates using 4.5 μM of every enzymes, 150 μM of **4**, 90 μM 2-oxoglutarate, 100 μM Fe(II)SO_4, 200 μM L-ascorbate in 100 mM Na-phosphate, pH 7.5, in a 200 μL reaction volume for 45 min at 293 K. One negative and two positive control reactions were performed. No enzyme was added to the negative control reaction whereas the two positive control reactions contained the native enzymes SnoK and SnoN. To obtain relative kinetic data for earlier time points prior to depletion of the substrate, the same conditions were the same except that 1.5 μM of SnoK and SnoK R1, and 1 μM of SnoN and SnoN R2 were used with 50 μM of **4**. The SnoK/SnoK R1 and the SnoN/SnoN R2 pair reactions were incubated for 15 min and 3 min respectively. The reactions were extracted twice under basic conditions with chloroform, dried with concentrator plus (Eppendorf) and dissolved with methanol. The reaction products were analysed by a SCL-10Avp HPLC with a SPD-M10Avp diode array detector (Shimadzu) using a Kinetex column (2.6 μm C18 100 Å, LC Column 100 x 4.6 mm, Ea, Phenomenex) with a gradient from 10% acetonitrile containing 18 mM ammonium acetate pH 3.6 to 70% containing 18 mM ammonium acetate pH 3.6. Product yields were calculated based on integration of peak areas at 265 nm. Circular dichroism (CD) spectroscopy was used to measure the secondary structure of the non-active SnoN chimeras at a concentration of 0.05 mg/ml in 20 mM Sodium phosphate buffer pH 7.5 at 25 °C. This was measured with Chirascan CD spectrometer (Applied Photophysics) equipped with a 2 mm path-length cell over the range 200–260 nm.

Computational Modelling

3D Modeling of the chimeras

Three-dimensional coordinates for the molecular structures and sequences of SnoN (PDB ID: 5EQU:A) and SnoK (PDB ID: 5EPA:A) were retrieved from the PDB [16]. The sequences for SnoN and SnoK R1-R4 chimeras (Figure S1) were aligned with the wild type SnoN and SnoK

sequences. Five homology models for each SnoN and SnoK chimera were generated using the alignment and the X-ray structures of SnoN and SnoK as the templates, respectively. The best models were selected by considering the smallest value of the normalized discrete optimized molecule energy (DOPE) [35]. Final 3D models were verified by Ramachandran plot and using the Structural Analysis and Verification Server [36,37].

Molecular docking simulations

All simulations were performed using OPLS3 force field [38]. Prior to the protein preparation, the ligand was extracted from the SnoN structure. The protein structures were prepared for simulations using Maestro Protein Preparation Wizard: the hydrogen atoms were added, the protonation types were solved using Epik (pH: 7.0±2.0), the bond orders were assigned and the water molecules were removed. The molecular docking simulations were performed using the Glide program [39]. The extra precision mode with flexible ligand and the Epik state penalties to docking score was used. The MM-GBSA method was performed to calculate ligand binding affinities using VSGB 2.0 solvation model [40]. The residues within a 8.0 Å distance from the ligand were assigned flexible. Using the described protocol, the SnoN substrate was successfully re-docked into the X-ray structure of SnoN protein with a root mean square deviation (RMSD) of 1.22 Å, suggesting that the methods used in the present study are appropriate.

Molecular dynamics simulations

To evaluate obtained docking poses, the MD simulations were performed using the Desmond software [41]. Water molecules and ions were added to the each system consisting of a ligand, protein, 2-oxoglutarate, Fe(IV)=O, docked ligand, TIP3P water molecules, Na+ ions, and Cl− ions. Na+ and Cl− ions were added by replacing water molecules to ensure the overall charge

neutrality. 50 ns MD trajectories were conducted using OPLS3 force field under a constant NpT condition (T=298 K and p=1 atm). The Nose–Hoover chain thermostat method, with a relaxation time of 1 ps, and the isotropic Martyna-Tobias-Klein barostat method, with a relaxation time of 2 ps, were used. The cutoff radius of the Lennard-Jones (LJ) and electrostatic interactions were set to 9.0 Å. The MD time step was set to 1 fs. The trajectories were analysed using the standard Simulation Interaction Diagram included in the Desmond program. The main results of the MD simulations are uploaded as supplementary material (SM). Obtained results were visualized using the program PyMol. The active site cavities were identified in the SnoK and SnoN proteins using MetaPocket 2.0 [42]. The volumes of the active site cavities were calculated using CASTp online program, with probe size of 1.4 Å [43].

Accession IDs

The wild type proteins SnoK (Uniprot, Q9RN60; PDB, 5EPA) and SnoN (Uniprot, Q9EYI0; PDB, 5EQU) utilized in this work have previously been deposited to appropriate databases.

ACKNOWLEDGMENTS

This study was supported by the Academy of Finland (grant no. 285971 to MM-K) and Sigrid Juselius Foundation and Tor, Joe, and Pentti Borg's Foundation to TAS. Instruct-FI and Biocenter Finland are acknowledged for the infrastructure support in bioinformatics (J.V. Lehtonen), translational activities and structural biology and CSC IT Center for Science for laboratory and computational infrastructure support at the Structural Bioinformatics Laboratory, Åbo Akademi University. The assistance of P. Rosenqvist in Circular Dichroism measurements in acknowledged.

ASSOCIATED CONTENT

Supporting Information

Seven figures and two tables are included in the supplemental information.

AUTHOR CONTRIBUTIONS

BNW, VS, PD, VV, TAS and MM-K planned experiments; BNW, VS, VV, TAS performed experiments; BNW, VS, PD, VV, TAS and MM-K analysed data; BNW, VV, TAS, and MM-K wrote the paper; TAS, and MM-K provided funding.

REFERENCES

1. Newman DJ & Cragg GM (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* 75, 311–335.
2. Newman DJ & Cragg GM (2016) Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* 79, 629–661.
3. Katz L & Baltz RH (2016) Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* 43, 155–176.
4. Metsä-Ketelä M, Niemi J, Mäntsälä P & Schneider G (2008) Anthracycline biosynthesis: Genes, enzymes and mechanisms. In *Anthracycline Chemistry and Biology I: Biological Occurence and Biosynthesis, Synthesis and Chemistry,* (Krohn, K., Ed.), pp 101−140. Springer-Verlag, Berlin/Heidelberg.
5. Weiss RB (1992) The anthracyclines: will we ever find a better doxorubicin? *Semin. Oncol.* 19, 670–86.
6. Pang B, Qiao X, Janssen L, Velds A, Groothuis T, Kerkhoven R, Nieuwland M, Ovaa H, Rottenberg S, Van Tellingen O, Janssen J, Huijgens P, Zwart W & Neefjes J (2013)

Drug-induced histone eviction from open chromatin contributes to the chemotherapeutic effects of doxorubicin. *Nat. Commun.* 4, 1908–1913.

7. Nitiss J (2009) Targeting DNA topoisomerase II in cancer chemotherapy. *Nat. Rev. Cancer* 9, 338–350.

8. Vejpongsa P & Yeh ETH (2014) Topoisomerase 2β: A promising molecular target for primary prevention of anthracycline-induced cardiotoxicity. *Clin. Pharmacol. Ther.* 95, 45–52.

9. Tacar O, Sriamornsak P & Dass CR (2013) Doxorubicin: An update on anticancer molecular action, toxicity and novel drug delivery systems. *J. Pharm. Pharmacol.* 65, 157–170.

10. Gewirtz DA (1999) A critical evaluation of the mechanisms of action proposed for the antitumor effects of the anthracycline antibiotics adriamycin and daunorubicin. *Biochem. Pharmacol.* 57, 727–741.

11. Arora SK (1983) Molecular Structure, Absolute Stereochemistry, and Interactions of Nogalamycin, a DNA-Binding Anthracycline Antitumor Antibiotic. *J. Am. Chem. Soc.* 105, 1328–1332.

12. Elshahawi SI, Shaaban KA, Kharel MK & Thorson JS (2015) A comprehensive review of glycosylated bacterial natural products. *Chem. Soc. Rev.* 44, 7591–7697.

13. Kawai H, Hayakawa Y, Nakagawa M, Furihata K, Seto H & Otake N (1984) Studies on arugomycin, a new anthracycline antibiotic part II. structural elucidation of arugomycin. *Tetrahedron Lett.* 25, 1941–1944.

14. Moore D. J, Brown TD, LeBlanc M, Dahlberg S, Miller TP, McClure S & Fisher RI (1999) Phase II trial of menogaril in non-Hodgkin's lymphomas: A southwest oncology group trial. *Invest. New Drugs* 17, 169–172.

15. Siitonen V, Claesson M, Patrikainen P, Aromaa M, Mäntsälä P, Schneider G & Metsä-

Ketelä M (2012) Identification of Late-Stage Glycosylation Steps in the Biosynthetic Pathway of the Anthracycline Nogalamycin. *ChemBioChem* 13, 120–128.

16. Siitonen V, Selvaraj B, Niiranen L, Lindqvist Y, Schneider G & Metsä-Ketelä M (2016) Divergent non-heme iron enzymes in the nogalamycin biosynthetic pathway. *Proc. Natl. Acad. Sci.* 113, 5251–5256.
17. Proshlyakov DA & Hausinger RP (2011) Transient Iron Species in the Catalytic Mechanism of the Archetypal α-Ketoglutarate-Dependent Dioxygenase, TauD. In *Iron-Containing Enzymes* (de Visser SP & Kumar D, eds.), pp. 67–87. Royal Society of Chemistry, Cambridge.
18. Islam MS, Leissing TM, Chowdhury R, Hopkinson RJ & Schofield CJ (2018) 2-Oxoglutarate-Dependent Oxygenases. *Annu. Rev. Biochem.* 87, 585–620.
19. Metsä-Ketelä M (2017) Evolution inspired engineering of antibiotic biosynthesis enzymes. *Org. Biomol. Chem.* 15, 4036–4041.
20. Fewer DP & Metsä-Ketelä M (2019) A Pharmaceutical Model for the Molecular Evolution of Microbial Natural Products. *FEBS J.*, in press. doi: 10.1111/febs.15129
21. Fischbach MA, Walsh CT & Clardy J (2009) The evolution of gene collectives: How natural selection drives chemical innovation. *Proc. Natl. Acad. Sci.* 105, 4601–4608.
22. Firn RD & Jones CG (2000) The evolution of secondary metabolism - a unifying model. *Mol. Microbiol. Subscr.* 37, 989–994.
23. Grocholski T, Dinis P, Niiranen L, Niemi J & Metsä-Ketelä M (2015) Divergent evolution of an atypical *S* -adenosyl-l-methionine–dependent monooxygenase involved in anthracycline biosynthesis. *Proc. Natl. Acad. Sci.* 112, 9866–9871.
24. Siitonen V, Blauenburg B, Kallio P, Mäntsälä P & Metsä-Ketelä M (2012) Discovery of a two-component monooxygenase SnoaW/SnoaL2 involved in nogalamycin biosynthesis. *Chem. Biol.* 19, 638–646.

25. Kallio P, Patrikainen P, Belogurov GA, Mäntsälä P, Yang K, Niemi J & Metsä-Ketelä M (2013) Tracing the evolution of angucyclinone monooxygenases: Structural determinants for C-12b hydroxylation and substrate inhibition in PgaE. *Biochemistry* 52, 4507–4516.

26. Hegg EL (1997) The 2-His-1-Carboxylate Facial Triad—An Emerging Structural Motif in Mononuclear Non-Heme Iron (II) Enzymes. *Eur. J. Biochem.* 250, 625–629.

27. Kallio P, Sultana A, Niemi J, Mäntsälä P & Schneider G (2006) Crystal structure of the polyketide cyclase AknH with bound substrate and product analogue: Implications for catalytic mechanism and product stereoselectivity. *J. Mol. Biol.* 357, 210–220.

28. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.

29. Gouy M, Guindon S & Gascuel O (2010) Sea view version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224.

30. Saitou N & Nei M (1987) The neighbour-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, 406–425.

31. Johnson MS, Lehtonen JV (2000) Comparison of protein three- dimensional structures. In *Bioinformatics, sequence, structure and databanks*, (Higgins D, Taylor W eds), pp 15–50. Oxford University Press, Oxford.

32. Lehtonen JV, Still DJ, Rantanen VV, Ekholm J, Björklund D, Iftikhar Z, Huhtala M, Repo S, Jussila A, Jaakkola J, Pentikäinen O, Nyrönen T, Salminen T, Gyllenberg M & Johnson MS (2004) BODIL: a molecular modelling environment for structure- function analysis and drug design. *J. Comput. Aided. Mol. Des.* 18, 401–419.

33. Chater KF & Wilde LC (1980) Streptomyces albus G mutants defective in the SalGI restriction-modification system. *J. Gen. Microbiol.* 116, 323–334.

34. Ho SN, Hunt HD, Horton RM, Pullen JK & Pease LR (1989) Site-directed mutagenesis

by overlap extension using the polymerase chain reaction. *Gene* 77, 51–59.

35. Sali A & Sali A (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci.*, 2507–2524.

36. Lüthy R, Bowie JU & Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. *Nature* 356, 83–85.

37. Wiederstein M & Sippl MJ (2007) ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35, 407–410.

38. Harder E, Damm W, Maple J, Wu C, Reboul M, Xiang JY, Wang L, Lupyan D, Dahlgren MK, Knight JL, Kaus JW, Cerutti DS, Krilov G, Jorgensen WL, Abel R & Friesner RA (2016) OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* 12, 281–296.

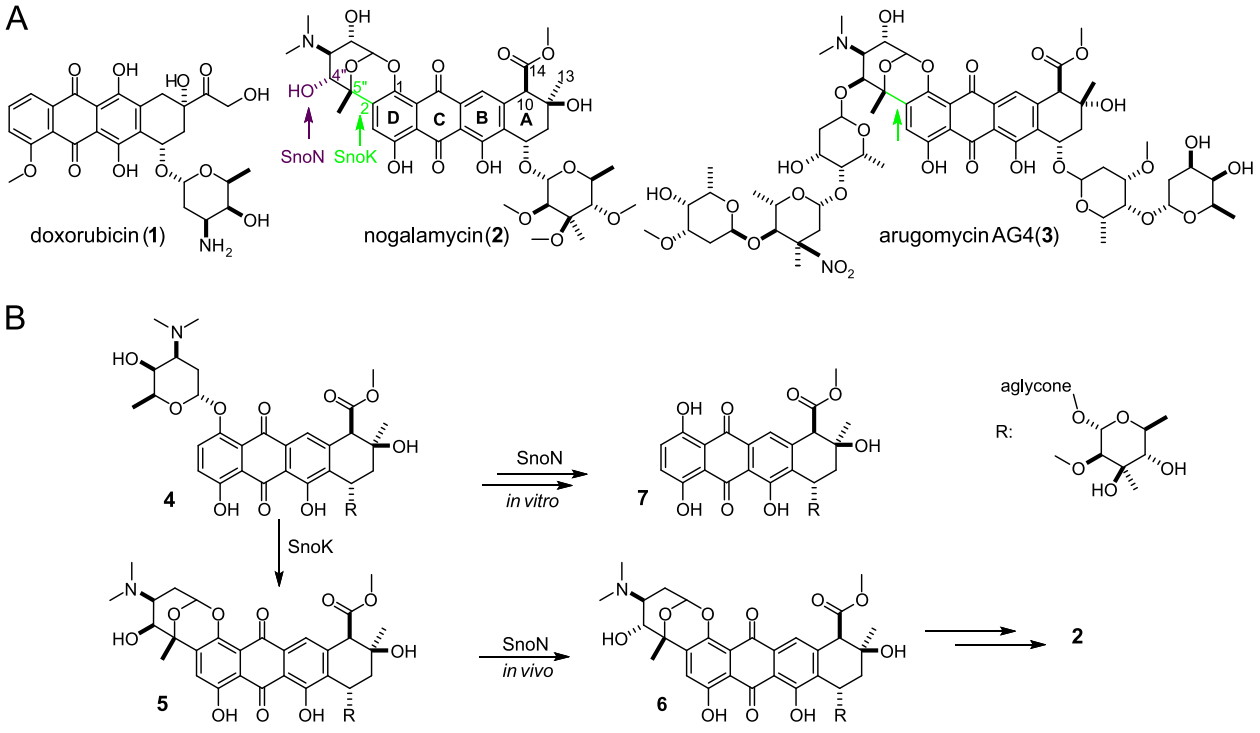
39. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC & Mainz DT (2006) Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* 49, 6177–6196.

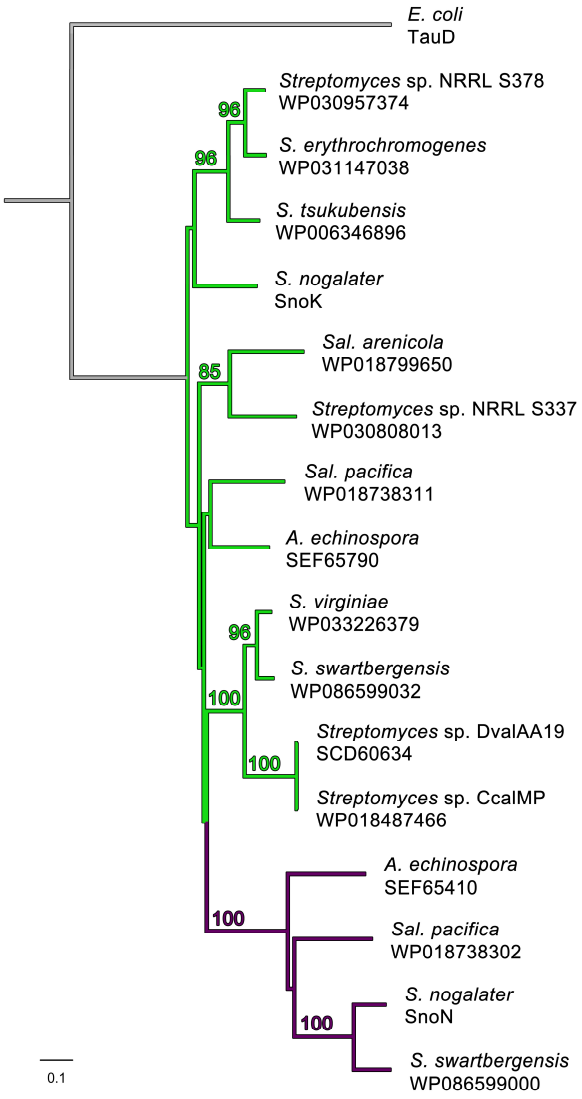
40. Hou T, Wang J, Li Y, Wang W, Houa T, Wangb J, Lia Y & Wang W (2011) Assessing the performance of the MM/PBSA and MM/GBSA methods: I. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Comput. Sci.* 51, 69–82.

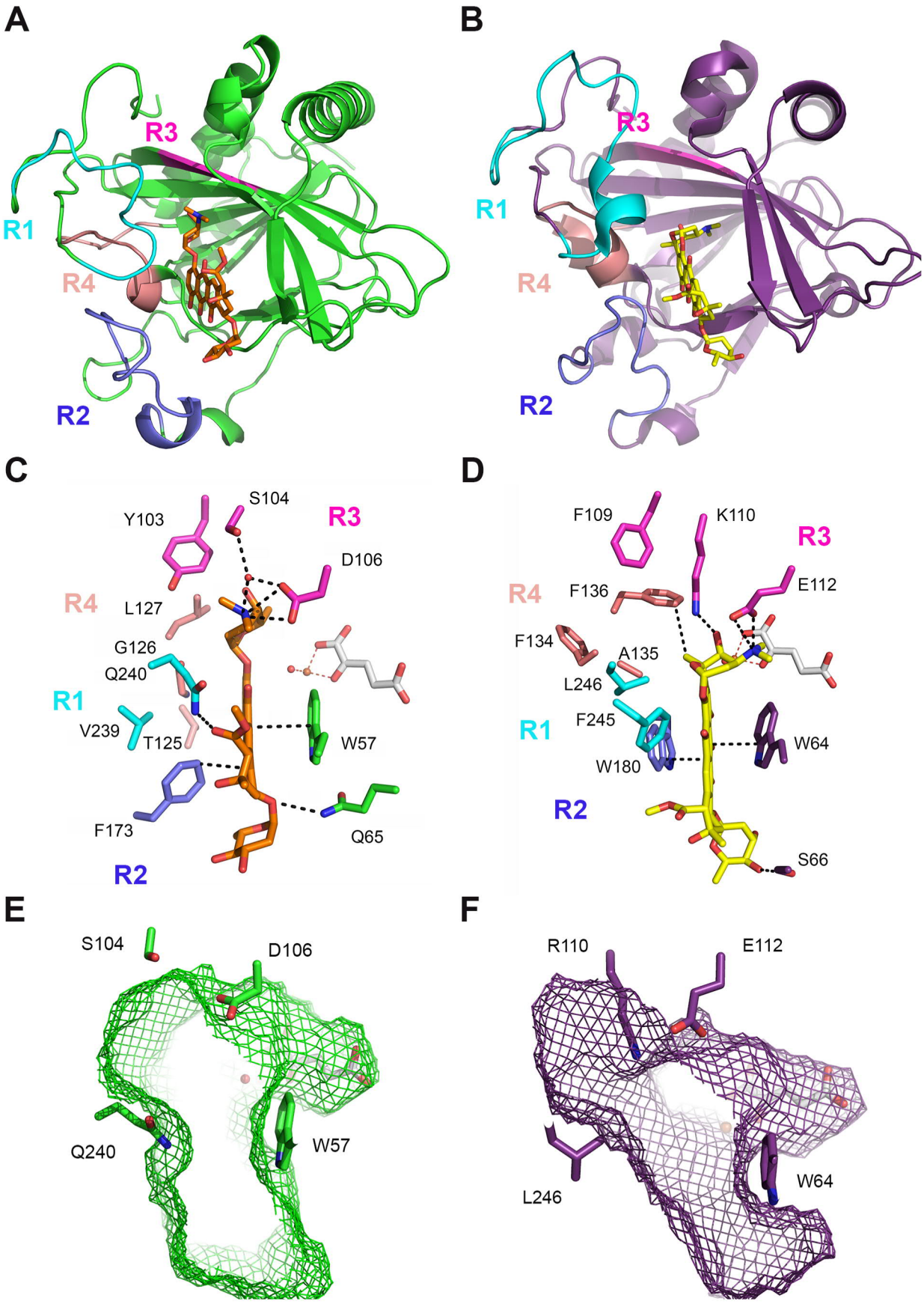
41. Bowers KJ, Sacerdoti FD, Salmon JK, Shan Y, Shaw DE, Chow E, Xu H, Dror RO, Eastwood MP, Gregersen BA, Klepeis JL, Kolossvary I & Moraes MA (2006) Scalable algorithms for molecular dynamics simulations on commodity clusters. *Proc. 2006 ACM/IEEE Conf. Supercomput. - SC '06*, 84.

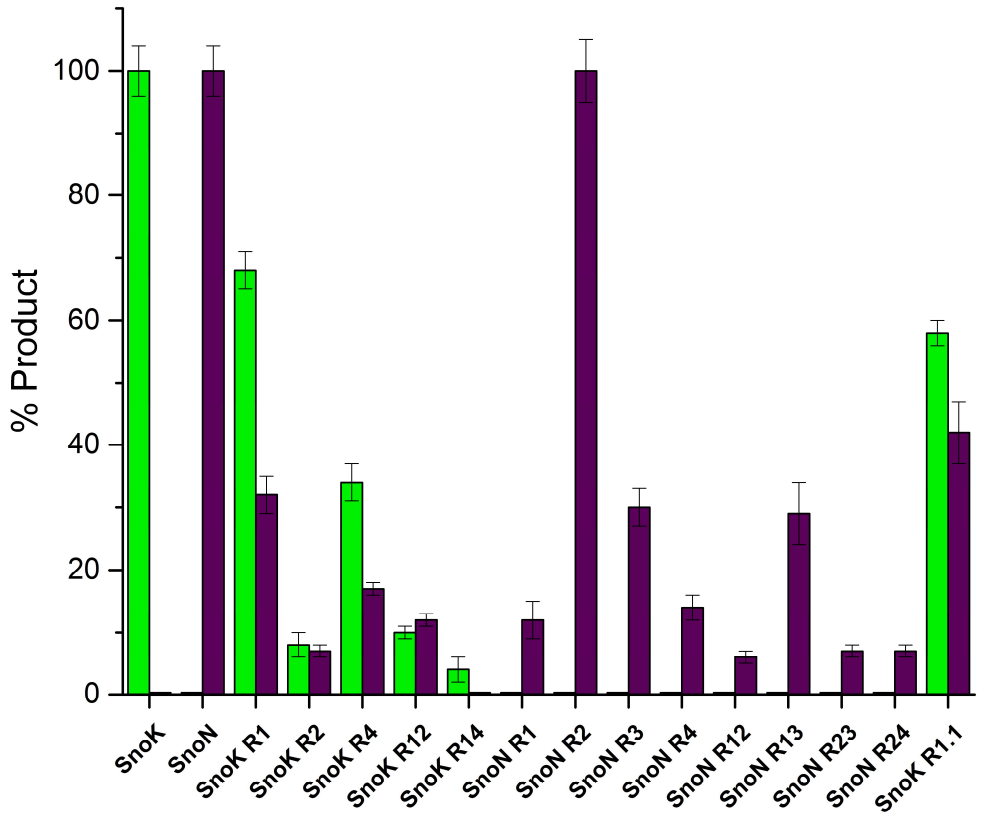
42. Huang B (2009) MetaPocket : A Meta Approach to Improve Protein Ligand Binding Site Prediction. *OMICS* 13, 325–330.

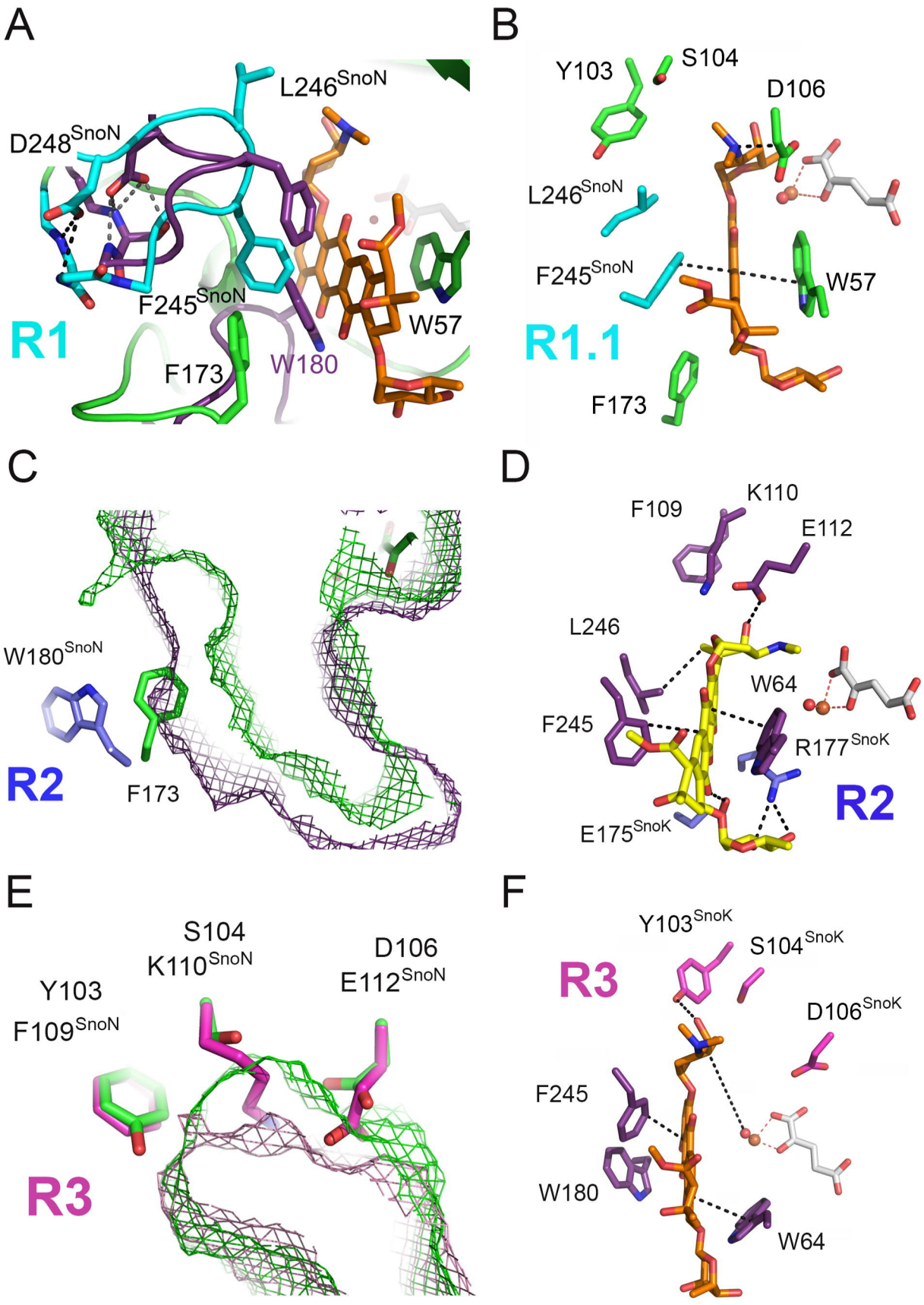
43. Tian W, Chen C, Lei X, Zhao J & Liang J (2018) CASTp 3.0: Computed atlas of surface topography of proteins. *Nucleic Acids Res.* 46, W363–W367.

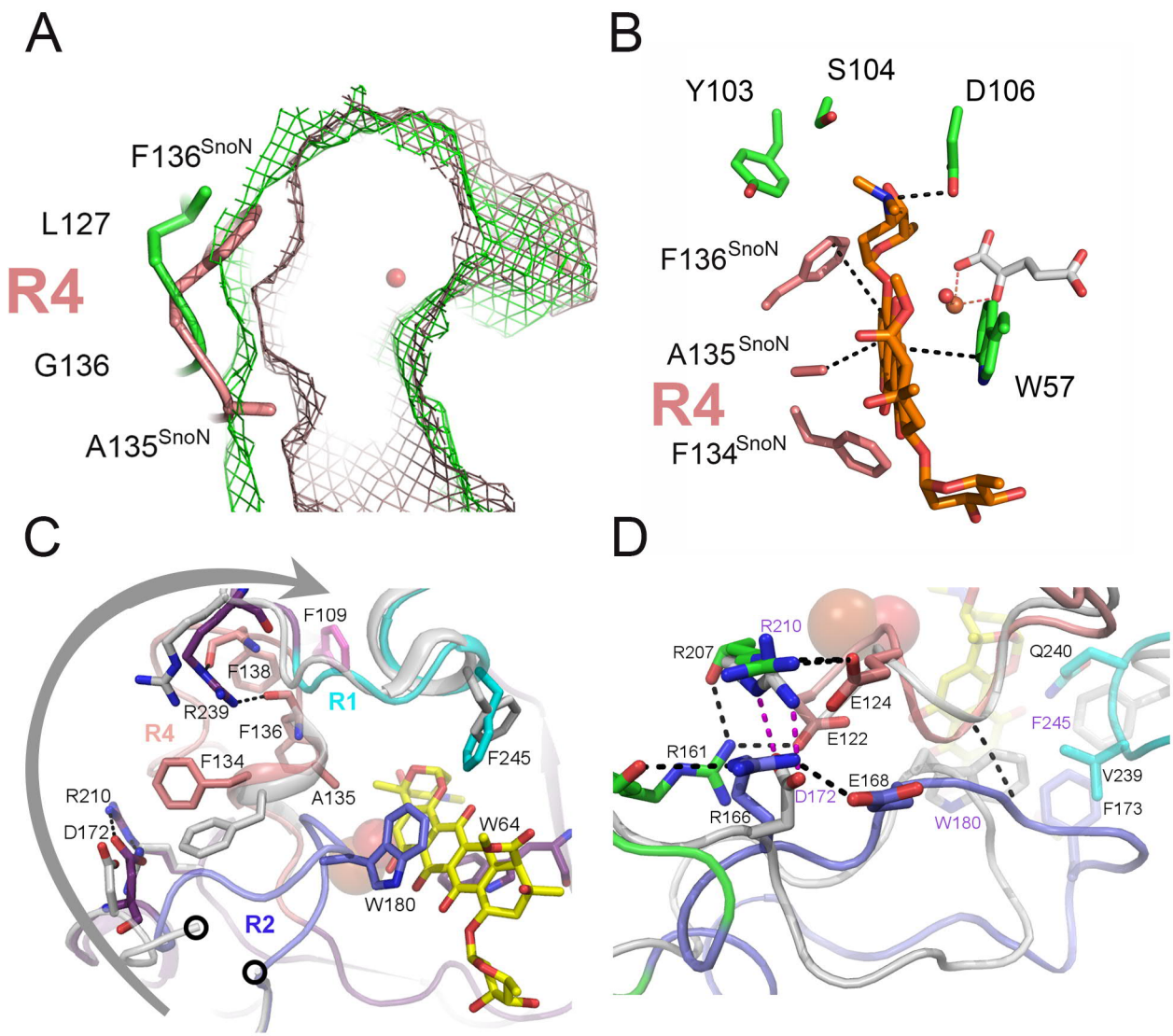

A
doxorubicin (1)
nogalamycin (2)
arugomycin AG4(3)
SnoN
SnoK
B
4
5
7
6
2
SnoN
in vitro
SnoK
SnoN
in vivo
aglycone
R:


E. coli
TauD
Streptomyces sp. NRRL S378
WP030957374
96
S. erythrochromogenes
WP031147038
96
S. tsukubensis
WP006346896
S. nogalater
SnoK
Sal. arenicola
WP018799650
85
Streptomyces sp. NRRL S337
WP030808013
Sal. pacifica
WP018738311
A. echinospora
SEF65790
S. virginiae
WP033226379
96
S. swartbergensis
WP086599032
100
Streptomyces sp. DvalAA19
SCD60634
100
Streptomyces sp. CcalMP
WP018487466
A. echinospora
SEF65410
100
Sal. pacifica
WP018738302
S. nogalater
SnoN
100
S. swartbergensis
WP086599000
0.1


A
R3
R1
R4
R2
B
R3
R1
R4
R2
C
Y103
S104
R3
D106
R4
L127
G126
Q240
R1
V239
T125
W57
F173
Q65
R2
D
F109
K110
R3
R4
F136
E112
F134
A135
L246
R1
F245
W64
W180
R2
S66
E
S104
D106
Q240
W57
F
R110
E112
L246
W64


% Product
0
20
40
60
80
100
SnoK
SnoN
SnoK R1
SnoK R2
SnoK R4
SnoK R12
SnoK R14
SnoN R1
SnoN R2
SnoN R3
SnoN R4
SnoN R12
SnoN R13
SnoN R23
SnoN R24
SnoK R1.1


A
L246SnoN
D248SnoN
F245SnoN
W57
R1
F173
W180
B
Y103
S104
D106
L246SnoN
F245SnoN
W57
R1.1
F173
C
W180SnoN
R2
F173
D
F109
K110
E112
L246
W64
F245
R177SnoK
R2
E175SnoK
E
S104
K110SnoN
D106
E112SnoN
Y103
F109SnoN
R3
F
Y103SnoK
S104SnoK
R3
D106SnoK
F245
W180
W64


A
F136SnoN
L127
R4
G136
A135SnoN
B
Y103
S104
D106
F136SnoN
A135SnoN
R4
F134SnoN
W57
C
F109
F138
R239
R1
R4
F136
F134
F245
A135
R210
D172
W64
W180
R2
D
R210
R207
Q240
E124
F245
E122
R161
V239
E168
D172
F173
R166
W180